IC3K 2024

16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management

PROCEEDINGS

Volume 3: KMIS

Porto, Portugal

17 - 19 November, 2024

EDITORS

Le Gruenwald
Elio Masciari
Jorge Bernardino

https://ic3k.scitevents.org/

SPONSORED BY

INSTICC

PAPERS AVAILABLE AT

SCITEPRESS DIGITAL LIBRARY

IC3K 2024

Proceedings of the
16th International Joint Conference on
Knowledge Discovery, Knowledge Engineering and
Knowledge Management

Volume 3: KMIS

Porto - Portugal

November 17 - 19, 2024

Sponsored by
INSTICC - Institute for Systems and Technologies of Information, Control and Communication

ACM In Cooperation
SIGAI - ACM Special Interest Group on Artificial Intelligence

In Cooperation with
AAAI - Association for the Advancement of Artificial Intelligence
APPIA - Portuguese Association for Artificial Intelligence

Edited by Le Gruenwald, Elio Masciari and Jorge Bernardino

Printed in Portugal
ISSN: 2184-3228
ISBN: 978-989-758-716-0
DOI: 10.5220/0000193400003838
Depósito Legal: 536889/24

https://kmis.scitevents.org
kmis.secretariat@insticc.org

Brief Contents

INVITED SPEAKERS

Carlo Sansone
University of Naples Federico II
Italy

Nirmalie Wiratunga
Robert Gordon University, Aberdeen
United Kingdom

João Gama
University of Porto
Portugal

ORGANIZING COMMITTEES

CONFERENCE CHAIR

Jorge Bernardino, Polytechnic of Coimbra - ISEC, Portugal

PROGRAM CO-CHAIRS

Le Gruenwald, University of Oklahoma, School of Computer Science, United States
Elio Masciari, University of Napoli Federico II, Italy

SECRETARIAT

Ana Rita Paciência, INSTICC, Portugal

GRAPHICS PRODUCTION AND WEBDESIGNER

Inês Teles, INSTICC, Portugal

WEBMASTER

João Francisco, INSTICC, Portugal
Carolina Ribeiro, INSTICC, Portugal

PROGRAM COMMITTEE

Leon Abdillah, Bina Darma University, Indonesia

Nekane Aramburu, University of Deusto, Spain

Michael Arias, Universidad de Costa Rica, Costa Rica

Benjamin Aziz, Buckinghamshire New University, United Kingdom

Giuseppe Berio, University of South Brittany, France

Claudio Biancalana, Roma Tre University, Italy

Kelly Braghetto, University of São Paulo, Brazil

Adriana Burlea-Schiopoiu, University of Craiova, Romania

Marcirio Chaves, Pontifical Catholic University of Rio Grande do Sul, Brazil

Eric Cheng, The Education University of Hong Kong, Hong Kong

Vincent Cheutet, Université de Lyon, INSA Lyon, DISP (EA4570), France

Chin Wei Chong, Multimedia University, Malaysia

Ritesh Chugh, Central Queensland University, Australia

Mikael Collan, Lappeenranta University of Technology, Finland

Laurent d'Orazio, University of Rennes, France

James Denford, Royal Military College of Canada, Canada

Mihaela Dinsoreanu, Technical University of Cluj-Napoca, Romania

Michael Dzandu, University of Westminster, United Kingdom

Kurt Engemann, Iona College, United States

Rik Eshuis, Eindhoven University of Technology, Netherlands

Benoit Eynard, Sorbonne University, University of Technology of Compiegne, France

Joan-Francesc Fondevila-Gascón, CECABLE (Centre d'Estudis sobre el Cable), UPF, URL, UdG (EU Mediterrani) and UOC, Spain

Matteo Gaeta, University of Salerno, Italy

Francisco García-Sánchez, University of Murcia, Spain

Severin Grabski, Michigan State University, United States

Gabriel Guerrero-Contreras, University of Cádiz, Spain

Keith Harman, Liberty University, United States

Anca Daniela Ionita, University Politehnica of Bucharest, Romania

Dimka Karastoyanova, University of Groningen, Netherlands

Marite Kirikova, Riga Technical University, Latvia

Erhun Kundakcioglu, Ozyegin University, Turkey

Dominique Laurent, ETIS Laboratory CNRS UMR 8051 - Cergy-Pontoise University - ENSEA, France

Fernando Laurindo, University of Sao Paulo, Brazil

Michael Leyer, University of Rostock, Germany

Kecheng Liu, The University of Reading, United Kingdom

Carlos Malcher Bastos, Universidade Federal Fluminense, Brazil

Ra'ed Masa'deh, The University of Jordan, Jordan

Nada Matta, University of Technology of Troyes, France

Sally McClean, University of Ulster, United Kingdom

Brahami Menaouer, National Polytechnic School of Oran (ENPOran), Algeria

Birendra Mishra, Anderson Graduate School of Management, United States

Michele Missikoff, ISTC-CNR, Italy

Owen Molloy, University of Galway, Ireland

Oswaldo Moscoso-Zea, Universidad Tecnológica Equinoccial, Ecuador

Rohana Ngah, Universiti Teknologi MARA, Malaysia

Pham Ngoc, National Economics University, Vietnam

Alex Norta, Tallinn University of Technology, Estonia

Mazni Omar, Universiti Utara Malaysia, Malaysia

Faizuniah Pangil, Universiti Utara Malaysia, Malaysia

Iraklis Paraskakis, South East European Research Centre, Greece

Wilma Penzo, University of Bologna, Italy

Quoc Trung Pham, HCMC University of Technology, Vietnam

Dara Pir, City University of New York, United States

Pedro Luiz Pizzigatti Correa, Universidade de Sao Paulo, Brazil

Nicolas Prat, Essec Business School Paris, France

Arkalgud Ramaprasad, University of Illinois at Chicago, India

Nataša Rupcic, University of Rijeka, Croatia

Higor Santos, Pernambuco University, Brazil

Stanko Škec, University of Zagreb, Croatia

Alessandro Stefanini, University of Pisa, Italy

Costas Vassilakis, University, Peloponnese, Greece

Anthony Wensley, University of Toronto, Canada

Wei Zhang, UMASS BOSTON, United States

AUXILIARY REVIEWERS

Sowelu Avanzo, Italy, University of Turin, Italy

Aihua Han, University of Reading, United Kingdom

Alexandr Kormiltsyn, Tallinn University of Technology, Estonia

Hadi Novandish, TLU.ee, Estonia

SELECTED PAPERS BOOK

A number of selected papers presented at KMIS 2024 will be published by Springer in a CCIS Series book. This selection will be done by the Conference Chair and Program Co-chairs, among the papers actually presented at the conference, based on a rigorous review by the KMIS 2024 Program Committee members.

FOREWORD

This book contains the proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. This year, IC3K was held in Porto, Portugal, from 17 to 19 November, 2024. It was sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC). IC3K 2024 was also organized in collaboration with the ACM Special Interest Group on Artificial Intelligence, the Association for the Advancement of Artificial Intelligence and the Portuguese Association for Artificial Intelligence.

The purpose of the IC3K series is to bring together researchers, engineers, and practitioners in the areas of Knowledge Discovery and Information Retrieval (KDIR), Knowledge Engineering and Ontology Development (KEOD) and Knowledge Management and Information Systems (KMIS).

IC3K this year, as in previous years, was composed of 3 subconferences, each specializing in one of the aforementioned knowledge areas, namely KDIR, KEOD, and KMIS.

IC3K 2024 received 175 paper submissions from 47 countries of which 21.14% were accepted and published as full papers. A double-blind paper review was performed for each submission by at least 2 but usually 3 or more members of the International Program Committee, consisting of established researchers and domain experts.

The high quality of the IC3K 2024 program is enhanced by the keynote lectures delivered by distinguished speakers who are renowned experts in their fields: Carlo Sansone (University of Naples Federico II, Italy), Nirmalie Wiratunga (The Abderdeen Robert Gordon University, United Kingdom) and João Gama (University of Porto, Portugal).

The conference is complemented by a Special Session on Ontologies for Digital Twin modelling, chaired by Fatma Chamekh. In addition, some tutorials on relevant topics have been offered to the conference audience.

All papers presented will be available in the SCITEPRESS Digital Library and will be submitted for evaluation for indexing by SCOPUS, Google Scholar, The DBLP Computer Science Bibliography, Semantic Scholar, Engineering Index and Web of Science / Conference Proceedings Citation Index.

In recognition of the best papers, several awards will be presented at the closing session of the conference, based on the combined scores of the paper reviewers, as assessed by the Programme Committee, and the quality of the presentation, as assessed by the session chairs at the conference venue.

Authors of selected papers will be invited to submit extended versions for inclusion in a forthcoming book of IC3K Selected Papers to be published by Springer, as part of the CCIS Series. Some papers will also be selected for publication of extended and revised versions in the special issue of the Springer Nature Computer Science Journal.

The program for this conference required the dedicated efforts of many people. Firstly, we must thank the authors, whose research efforts are herewith recorded. Next, we thank the members of the Program Committee and the auxiliary reviewers for their diligent and professional reviewing. We would also like to express our sincere gratitude to the invited speakers for their invaluable contributions and for taking the time to prepare their presentations. Finally, a word of appreciation for the hard work of the INSTICC team; organizing a conference of this level is a task that can only be achieved by the collaborative effort of a dedicated and highly competent team.

We wish you all an exciting and inspiring conference. We hope to have contributed to the development of our research community, and we look forward to presenting more research at the next edition of IC3K, details of which can be found at https://ic3k.scitevents.org.

Le Gruenwald
University of Oklahoma, School of Computer Science, United States

Elio Masciari
University of Napoli Federico II, Italy

Jorge Bernardino
Polytechnic of Coimbra - ISEC, Portugal

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Multimodal Deep Learning in Medical Imaging

Carlo Sansone
Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, University of Naples Federico II, Naples, Italy

Abstract: In this talk, we will consider how Deep Learning (DL) approaches can profitably exploit the presence of multiple data sources in the medical domain. First, the need to be able to use information from multimodal data sources is addressed. Starting from an analysis of different multimodal data fusion techniques, an innovative approach will be proposed that allows the different modalities to influence each other. However, in medical applications it is often very difficult to obtain high quality and balanced labelled datasets due to privacy and sharing policy issues. Therefore, several applications have leveraged DL approaches in data augmentation techniques, proposing models that can create new realistic and synthetic samples. Consequently, a new data source can be identified, namely a synthetic data source. In this context, a data augmentation method based on deep learning, specifically designed for the medical domain, will be presented. It exploits the biological characteristics of images by implementing a physiologically-aware synthetic image generation process.

BRIEF BIOGRAPHY

Carlo Sansone is currently Full Professor of Computer Engineering at the Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione of the University of Naples Federico II. His basic interests cover the areas of image analysis, pattern recognition and machine and deep learning. From an applicative point of view, his main contributions were in the fields of biomedical image analysis, biometrics, intrusion detection in computer networks and image forensics. He coordinated several projects in the areas of artificial intelligence, biomedical images interpretation and network intrusion detection. Prof. Sansone is a member of the IEEE and of the International Association for Pattern Recognition (IAPR). In 2012 he was elected Vice-President of the GIRPR (the Italian Association affiliated to the IAPR) for two terms (four years).

Sansone, C.
Multimodal Deep Learning in Medical Imaging.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, page 5
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Intelligent Reuse of Explanation Experiences: The Role of Case-Based Reasoning in Promoting Best Practice in Explainable AI

Nirmalie Wiratunga
RGU's School of Computing, United Kingdom

Abstract: The EU now requires that machine learning models provide an explanation of their decisions. Different stakeholders may have different backgrounds, competencies, and goals, which may require different types of explanations. Interpreting and explaining machine learning (ML) models can be done in various ways, and there are many options available. However, it's difficult to know which method or combination of methods to use for different AI models and different deployment situations. The iSee project is trying to tackle this question. In this talk we will discuss why Case-Based Reasoning (CBR) is well placed to promote best practices in Explainable AI (XAI). We will also explore how CBR can be used to reason about end-users' XAI experiences and enable the sharing and reusing of such experiences through the iSee platform (https://isee4xai.com/). The talk will present the key components that facilitate reasoning in iSee – an ontology to model experiences, cases to capture experiences, a retrieval engine to identify best practice, and an interactive interface to engage with end-users

BRIEF BIOGRAPHY

Nirmalie Wiratunga is a Professor in Intelligent Systems at RGU's School of Computing, and the Associate Dean for Research in the school, with over two decades of experience in computer science and AI research. She has held positions such as postdoctoral researcher on EPSRC funded projects, and was appointed Readership in 2009, and Professorship in 2016. Nirmalie lead's the Artificial Intelligence & Reasoning Research Group (AIR) in the School of Computing. She has been involved in numerous funded AIR projects, including the development of platforms for reusable explainable AI experiences and initiatives in healthcare.

Wiratunga, N.
Intelligent Reuse of Explanation Experiences: The Role of Case-Based Reasoning in Promoting Best Practice in Explainable AI.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, page 7
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Recent Advances in Learning from Data Streams

João Gama
University of Porto, Portugal

Abstract: Learning from data streams is a hot topic in machine learning and data mining. This talk presents two problems and discusses streaming techniques to solve them. The first problem is the application of data stream techniques to predictive maintenance. We propose a two-layer neuro-symbolic approach to explain black-box models. The explanations are oriented toward equipment failures. For the second problem, we present a streaming algorithm for online hyperparameter tuning. The Self hyper-parameter Tuning (SPT) algorithm is an optimisation algorithm for online hyper-parameter tuning from non-stationary data streams. SPT is a wrapper over any streaming algorithm and can be used for classification, regression, and recommendation.

BRIEF BIOGRAPHY

João Gama is a Full Professor at the Faculty of Economy, University of Porto. He is a researcher and vice-director of LIAAD, a group belonging to INESC TEC. He got the PhD degree from the University of Porto, in 2000. He is a Senior member of IEEE.He has worked on several National and European projects on Incremental and Adaptive learning systems, Ubiquitous Knowledge Discovery, Learning from Massive, and Structured Data, etc. He served as Co-Program chair of ECML 2005, DS 2009, ADMA 2009, IDA 2011, and ECML/PKDD 2015. He served as track chair on Data Streams with ACM SAC from 2007 till 2016. He organized a series of Workshops on Knowledge Discovery from Data Streams with ECML/PKDD, and Knowledge Discovery from Sensor Data with ACM SIGKDD. He is the author of several books on Data Mining (in Portuguese) and authored a monograph on Knowledge Discovery from Data Streams. He authored more than 250 peer-reviewed papers in areas related to machine learning, data mining, and data streams. He is a member of the editorial board of international journals ML, DMKD, TKDE, IDA, NGC, and KAIS. He (co-)supervised more than 12 PhD students and 50 MSc students.

Gama, J.
Recent Advances in Learning from Data Streams.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, page 9
ISBN: 978-989-758-716-0; ISSN: 2184-3228

PAPERS

FULL PAPERS

Scientific Claim Verification with Fine-Tuned NLI Models

Miloš Košprdić[1][a], Adela Ljajić[1][b], Darija Medvecki[1][c], Bojana Bašaragin[1][d]
and Nikola Milošević[1,2][e]

[1]*The Institute for Artificial Intelligence Research and Development of Serbia, Fruškogorska 1, Novi Sad, Republic of Serbia*
[2]*Bayer A.G. Reaserch and Development, Mullerstrasse 173, Berlin, Germany*
{*milos.kosprdic, adela.ljajic, darija.medvecki, bojana.basaragin, nikola.milosevic*}*@ivi.ac.rs*

Keywords: Claim Verification, Deep Learning Models, Natural Language Inference, PubMed, SciFact Dataset.

Abstract: This paper introduces the foundation for the third component of a pioneering open-source scientific question-answering system. The system is designed to provide referenced, automatically vetted, and verifiable answers in the scientific domain where hallucinations and misinformation are intolerable. This Verification Engine is based on models fine-tuned for the Natural Language Inference task using an additionally processed SciFact dataset. Our experiments, involving eight fine-tuned models based on RoBERTa Large, XLM RoBERTa Large, DeBERTa, and DeBERTa SQuAD, show promising results. Notably, the DeBERTa model fine-tuned on our dataset achieved the highest F1 score of 88%. Furthermore, evaluating our best model on the HealthVer dataset resulted in an F1 score of 48%, outperforming other models by more than 12%. Additionally, our model demonstrated superior performance with a 7% absolute increase in F1 score compared to the best-performing GPT-4 model on the same test set in a zero-shot regime. These findings suggest that our system can significantly enhance scientists' productivity while fostering trust in the use of generative language models in scientific environments.

1 INTRODUCTION

In the scientific field, the veracity of information is paramount, especially as large language models (LLMs) become increasingly integrated into research methodologies. The paper (Košprdić et al., 2024) outlines a novel open-source initiative designed to mitigate the risks of inaccuracies, or "hallucinations", in answers generated by LLMs in the biomedical domain. Central to this initiative is a sophisticated architecture comprising three core components: an information retrieval system that utilizes both semantic and lexical search combination techniques to retrieve scientific papers from PubMed[1]; a Retrieval Augmented Generation (RAG) module with a generative model fine-tuned to produce referenced answers based on the retrieved scientific papers; and a verification engine tasked with cross-checking these generated answers against scientific papers to ensure accuracy and to identify potential hallucinations. This system aims to enhance the productivity of researchers by providing reliable information and to instill trust in the use of generative language models within scientific domains where misinformation can have serious repercussions.

According to (Guo et al., 2022), claim verification is a process performed after a claim is generated and evidence is retrieved. The authors also segment the claim verification process into two components: *verdict prediction*, where claims are assigned truthfulness labels, and *justification production*, where explanations for verdicts must be generated (Guo et al., 2022). Our system generates claims using RAG, and we use the retrieved documents from the PubMed repository as evidence. In this paper, we are going to perform the verdict prediction task by transforming it into the natural language inference (NLI) task to predict one of the labels: support, contradict, and no_evidence. The development of a model specifically tailored for textual entailment or NLI represents a crucial element of the verification engine in (Košprdić et al., 2024). In this paper, we are going to present our effort on enhancing and fine-tuning different models to achieve accuracy in the task of scientific

[a] https://orcid.org/0000-0002-4180-0050
[b] https://orcid.org/0000-0001-7326-059X
[c] https://orcid.org/0000-0002-4180-0050
[d] https://orcid.org/0000-0002-7679-1676
[e] https://orcid.org/0000-0003-2706-9676
[1] https://pubmed.ncbi.nlm.nih.gov/

Košprdić, M., Ljajić, A., Medvecki, D., Bašaragin, B. and Milošević, N.
Scientific Claim Verification with Fine-Tuned NLI Models.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 15-25
ISBN: 978-989-758-716-0; ISSN: 2184-3228

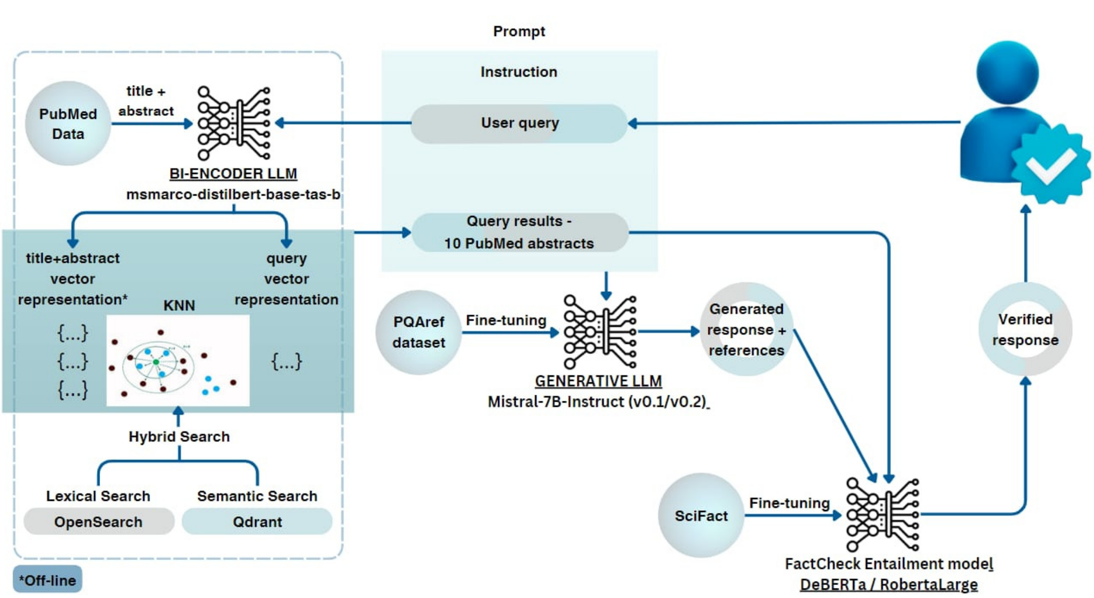


Figure 1: Architecture of our verification system.

claims verification.

By fine-tuning state-of-the-art models for the NLI task in the scientific field, we plan to allow our model to detect the subtle differences in the statements between claims supported by evidence, those that contradict the evidence, and those for which evidence is lacking. This will not only build upon the foundational work presented in (Košprdić et al., 2024; Ljajić et al., 2024) for claim verification but also introduce a novel methodology and models designed to provide claim verification by fine-tuning models for textual entailment tasks in the biomedical domain. The model fine-tuned for textual entailment will enhance the automated scientific claim verification process in (Košprdić et al., 2024), complementing retrieval and generation steps in (Bašaragin et al., 2024), by verifying generated text.

Our work's contributions are twofold: firstly, we fine-tuned different deep learning models on the SciFact dataset to significantly improve textual entailment predictions in the biomedical domain, achieving state-of-the-art results. Secondly, we provide comparative results of these models fine-tuned with the SciFact dataset.

The paper is organized as follows: Section 2 provides an overview of our comprehensive system, detailing the RAG module and the verification of claims produced. Section 3 presents current datasets and methods for the claim verification task. The transformation of the chosen dataset, as well as the models used for this task and their parameters, are presented in Section 4. Results are given in Section 5, with an accompanying error analysis in Section 6. Finally, Section 7 presents the conclusions drawn from our study and outlines directions for future research.

2 OVERALL SYSTEM DESIGN

Our verification system encompasses two main processes: Retrieval Augmented Generation (RAG) and the Verification Engine. The RAG process, consisting of two components, is based on a fine-tuned large language model (LLM) for referenced question-answering. In this setup, retrieved relevant abstracts from PubMed are provided to the LLM as input through a prompt. The output is an answer based on these PubMed abstracts, with each statement appropriately referenced, facilitating subsequent verification by the Verification Engine.

In this section, we describe the three main components of our system (see Figure 1), to offer clarity to the entire process which contains claim verification as its second and final verification step.

The Information Retrieval Component (IR) utilizes data from the PubMed database[2], which contains citations and biomedical literature from various sources. The IR system incorporates both sparse vectors (lexical index) and dense vectors (semantic index), facilitating lexical, semantic, and hybrid searches.

For lexical retrieval, based on BM25, we use OpenSearch[3] to create an index of PubMed articles

[2]https://pubmed.ncbi.nlm.nih.gov/download/
[3]https://opensearch.org/

by concatenating the titles and abstracts into a single indexed field. For semantic retrieval, based on dense vectors, we employ the Qdrant[4] vector database. To generate vector embeddings, we utilize a bi-encoder sentence transformer model pre-trained on the MS-Marco dataset[5], which had the highest performance on the Passage Retrieval Task at the time of indexing[6].

Hybrid search in our system combines the lexical and semantic IR components. To implement a hybrid search, we normalized the scores from these two IR methods to a scale ranging from 0 to 1. The scores from each search method are then multiplied by the respective importance weights. This approach allows for the identification of both direct matches and enhances the discovery of semantically related phrases and text segments, even when exact textual matches are absent. The selected documents from the IR Component are then passed to the Generative Component, which is responsible for generating the appropriate response.

The Generative Component receives the user query and 10 retrieved documents as its input. It consists of a generative model, currently Mistral-7B-Instruct-v0.2[7], which we additionally fine-tuned for the task of question-answering with references using the QLoRA methodology (Dettmers et al., 2023) and a dataset of randomly selected questions from the PubMedQA dataset (Jin et al., 2019). The output of the model is an answer to the user query which contains a reference for each of the claims generated based on the relevant articles.

The Verification Component is designed to verify the claims created by the RAG module. The specific type of verification and the models used for this process are described in detail in Section 4. Since the answer of the generative model can contain multiple claims/claim parts, our goal is to create a model for verifying individual claim parts, and how we combine them will be decided in the future.

3 RELATED WORK

The task of determining a claim's veracity is dubbed differently in the literature: from *verdict prediction* (Tan et al., 2023) or *veracity prediction* (Vladika et al., 2024) to *claim verification* (Wadden et al., 2020), to name only a few, and usually forms a part of a multi-component pipeline aiming to produce as factual results as possible (Tan et al., 2023). The task typically consists of assessing a claim and its corresponding evidence and categorizing them into one of three distinct labels: support/evidence, contradiction/refute, and no_evidence/not_enough_info. The process of verifying scientific claims can be conceptualized within the framework of natural language inference (NLI), treating the claim verification task as a multi-class classification task, a perspective that aligns with prior research (Thorne et al., 2018; Wadden et al., 2020).

Different techniques have been utilized and developed to improve claim verification, and recently, transformer models have achieved state-of-the-art performance in fact-checking in general, both in general and scientific domains (Tan et al., 2023). Input for these models usually consists of concatenated claim and evidence pairs from which they create representations for classifying relationships between the two (Tan et al., 2023). Including the entire context of the evidence (e.g. entire document) ensures minimal loss of information and produces better results during inference (Wadden et al., 2022).

Interestingly, both general-purpose and domain-specific large language models are used for the task of scientific claim verification (Vladika and Matthes, 2023). The successful use of general-purpose models for this task is supported by the work of the creators of the SciFact dataset, who released the VeriSci model (Wadden et al., 2020). This model uses RoBERTa-large (Liu et al., 2019) model pre-trained on the FEVER dataset (Thorne et al., 2018) and fine-tuned on the SciFact dataset as a component for label prediction, since it demonstrated the strongest performance compared to other tested scenarios. This model also showed better accuracy compared to SciBERT (Beltagy et al., 2019), BioMedRoBERTa (Gururangan et al., 2020), and RoBERTa-base when all were trained only on the SciFact dataset.

While general-domain datasets for the task of claim verification have existed since 2014 (Vlachos and Riedel, 2014), the first dataset for scientific claim verification, SciFact, appeared in 2020 (Wadden et al., 2020). The number of different scientific claim verification datasets continued to grow ever since, from those that collect claims from social media posts (Mohr et al., 2022), Wikipedia and Internet in general (Diggelmann et al., 2021; Sarrouti et al., 2021), different web portals (Kotonya and Toni, 2020; Vladika et al., 2024), science exam questions (Tan et al., 2023) or publications (Malaviya et al., 2023).

[4] https://qdrant.tech/

[5] https://huggingface.co/sentence-transformers/msmarco-distilbert-base-tas-b

[6] https://www.sbert.net/docs/pretrained-models/msmarco-v3.html

[7] https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2

However, SciFact is still one of the rare datasets containing claims from research papers and is the most used dataset for building scientific claim verification systems to date (Vladika and Matthes, 2023).

The work of (Sarrouti et al., 2021) shows that the choice of the in-domain dataset for fine-tuning makes a significant difference. The authors conducted experiments on several baseline models: BERT (Devlin et al., 2019), SciBERT, BioBERT(Lee et al., 2019), and T5 (Raffel et al., 2020), trained and evaluated on the HealthVer dataset (Sarrouti et al., 2021), with T5 demonstrating superior performance over all other models. They also tested the BERT-base model fine-tuned on FEVER (Thorne et al., 2018), SciFact, PubHealth, and HealthVer datasets, and assessed its performance on the HealthVer test set. Their findings suggest that despite the FEVER dataset's size advantage over SciFact and HealthVer, the model achieved superior F1 scores when trained on SciFact and HealthVer datasets. Since FEVER is based on Wikipedia sentences, this supports the notion that training on in-domain claims yields more substantial benefits for domain-specific claim verification tasks.

The positive effect of in-domain datasets is further confirmed by the work of (Tan et al., 2023), who performed in-domain fine-tuning first using Med-Fact and Gsci-Fact datasets (Tan et al., 2023) and then using SciFact, HealthVer, and CLIMATE-FEVER (Diggelmann et al., 2021) on BERT, DeBERTa (He et al., 2021), SciBERT, Longformer, and BioBERT. In comparison with only fine-tuning the models using SciFact, HealthVer, and CLIMATE-FEVER (Diggelmann et al., 2021), they were able to achieve an improved performance for most models, with DeBERTA performing best in almost all scenarios.

4 OUR MODEL FOR CLAIM VERIFICATION

Similarly to the prompt-generated answers to certain questions in generating long-form responses that contain multiple claims (Wei et al., 2024), the responses generated by our RAG module contain multiple claims supported by references to PubMed abstracts as evidence. This structured approach ensures that each claim is supported by relevant scientific literature, enhancing the credibility and reliability of the generated responses.

The model's performance is evaluated through various metrics, including macro and weighted precision, recall, and F1-score, and also accuracy, which collectively provide a comprehensive understanding of its effectiveness in the task.

In our recent exploration of advanced natural language processing techniques, we opted for a fine-tuning process for 3 cutting-edge models: RoBERTa-large (Liu et al., 2019), XLM-RoBERTa Large (Conneau et al., 2019a), and DeBERTa Large (He et al., 2021), alongside DeBERTa SQuAD - a DeBERTa model fine-tuned using the SQuAD dataset. Our primary objective was to enhance their performance on the task of textual entailment.

The selection of DeBERTa was motivated by its outstanding performance in claim verification experiments (Tan et al., 2023). RoBERTa was chosen for its modified attention mechanisms and its strongest performance on the label prediction task (Wadden et al., 2020). XLM Roberta exhibited better performance on English data compared to RoBERTa (Conneau et al., 2019b). DeBERTa SQuAD was selected due to its training on question-answer pairs, specifically tailored for question answering tasks.

Fine-tuning was conducted using the SciFact (Wadden et al., 2020) dataset, which is specifically designed to support the development and evaluation of automated fact-checking systems focusing on scientific claims. Our objective in fine-tuning these models with the SciFact dataset was twofold: to improve their performance in discerning textual entailment within scientific texts and to compare their effectiveness in this specialized domain.

4.1 Dataset Transformation

The SciFact dataset comprises pairs of claims and evidence within the biomedical domain — a domain noted for its classification challenges even among human experts. We anticipate that it will serve as an optimal dataset for the fine-tuning of models aimed at claim verification within our system.

The SciFact dataset is structured into two separate files: *corpuses* and *claims*. The *corpuses* file contains the titles and abstracts of scientific articles, providing a rich source of evidence. The *claims* file includes various claims linked to these scientific articles via an identification number, facilitating the process of verifying the claims against the provided evidence. Initially, the claims file is divided into three parts: training, validation, and test. However, the publicly available version of the dataset does not include labels for the test partition. As a result, the 300 examples in the test partition were excluded from our analysis. To maximize the data available for our experiments, we combined the training and validation subsets, yielding a total of 1,711 examples. This combined dataset serves as the foundation for training, validating, and evaluating our claim verification models.

Initially, the titles and abstracts extracted from the corpus file underwent individual cleaning procedures. This process involved eliminating redundant spaces, excluding special characters present at the ends of the texts, removing surplus parentheses and brackets, and filtering out superfluous information contained within the abstracts, such as *(ABSTRACT TRUNCATED AT 250 WORDS)*. Recognizing the informative value inherent in article titles, we made a deliberate choice to concatenate them with their corresponding abstracts to construct a comprehensive response to the claims. Throughout this concatenation process, special attention was paid to titles lacking terminal punctuation. To maintain coherence and facilitate comprehension, we ensured that a concluding punctuation mark was appended to such titles. This step was essential to mitigate potential ambiguity and prevent the model from misinterpreting the concatenated text in instances where unpunctuated titles were directly joined with the initial sentence of the abstract, creating potentially unrelated sentences.

Subsequently, using the identification number, we integrated information from both files, ensuring that each claim was matched with its corresponding concatenation of the title and abstract, along with one of three labels: *no evidence*, *support*, or *contradict*. The SciFact dataset was initially designed so that the same labels for an abstract were repeated if they were found in multiple sentences of that abstract. However, since we opted to consider the entire concatenated title and abstract as a single response to the claim, rather than treating each sentence separately, we performed a deduplication of these instances. This step ensured that redundant combinations were removed. Ultimately, this process yielded 1,213 unique claim+[title+abstract] combinations, which form the final dataset[8] for our experiments. This comprehensive and deduplicated dataset provides a robust foundation for training and evaluating our claim verification models.

Descriptive statistics of the combined dataset formed in this manner revealed that approximately 36% of claim+[title+abstract] combinations are labeled as *no evidence*, about 42% are labeled as *support*, and around 22% are labeled as *contradict*. We divided the dataset into training, validation, and test subsets in a ratio of 80:10:10, ensuring that the proportion of each label is preserved across all subsets. This stratified division helps maintain the consistency of label distribution, which is crucial for accurately training and evaluating the model's performance.

[8]https://huggingface.co/datasets/MilosKosRad/SciFact_VerifAI

4.2 Experiments and Training Parameters

Our study aims to perform a comparative analysis of various transformer models, which have been fine-tuned and evaluated using the additionally processed SciFact dataset.

Determining one of the three relationships (*no evidence*, *support*, *contradict*) in which a claim and an evidence can stand, we decided to consider as a Textual Entailment task. This task can be viewed as the (2) Sequence Classification task into 3 classes by appropriately structuring the input data. Therefore, the task of claim verification could be conceptualized as a multi-class classification task.

To address this task, transformer models were fine-tuned utilizing the concatenation of a claim (*c*) and a corresponding PubMed title+abstract concatenation, serving as evidence (*e*). Since the 4 models we fine-tuned with the formatted SciFact dataset for the Textual Entailment task are based on either BERT or RoBERTa architecture, the inputs for them are:

$$[CLS]c[SEP]e[SEP]$$

for models with the basic architecture of BERT, and

$$<s> c </s></s> e </s>$$

for models with the basic architecture of RoBERTa.

The objective of our model is to accurately assign one of the following labels (*l*): *no_evidence*, *support*, or *contradict*. Based on the received input, the model's prediction is formalized as:

$$l\{c,e\} \in \{no\_evidence, support, contradict\}$$

This formulation allows for a systematic evaluation of the model's ability to discern and categorize the relationship between biomedical claims and their corresponding evidential support.

All trainings used the ADAM optimizer (Kingma and Ba, 2014) with a learning rate of 1e-5, a weight decay of 0.01, and were conducted on a single DGX NVIDIA A100-40GB GPU using the PyTorch framework and Hugging Face Transformer library. The number of epochs for all models was initialized at 15, with an early stopping strategy implemented on the validation subset based on the F1 metric to determine the optimal model checkpoint. We applied two distinct values for the early stopping hyperparameter for each of the 4 models (3 and 4), resulting in a total of 8 fine-tuned models for evaluation. Model evaluation was conducted by exact prediction-label matching, employing standard performance metrics of F1 score, accuracy, precision, and recall.

Table 1: The results of eight fine-tuned models 80% of SciFact data, validated on 10% of SciFact data, and tested on remaining 10% of data.

		RoBERTa L_{SF}				XLM RoBERTa L_{SF}				DeBERTa$_{SF}$				DeBERTa SQuAD$_{SF}$			
		NE*	S	C	wa	NE	S	C	wa	NE	S	C	wa	NE	S	C	wa
3	P	0.71	0.55	0.00	0.48	0.83	0.69	0.54	0.71	0.83	0.86	0.85	0.84	0.86	0.90	0.82	0.87
	R	0.73	0.82	0.00	0.61	0.89	0.67	0.52	0.71	0.86	0.84	0.81	0.84	0.86	0.88	0.85	0.87
	F1	0.72	0.66	0.00	0.53	0.86	0.68	0.53	0.71	0.84	0.85	0.83	0.84	0.86	0.89	0.84	**0.87**
	Acc	0.61				0.71				0.84				0.87			
4	P	0.85	0.75	0.67	0.77	0.75	0.76	0.71	0.74	0.88	0.90	0.88	0.89	0.82	0.91	0.88	0.87
	R	0.89	0.76	0.59	0.77	0.91	0.67	0.63	0.75	0.95	0.88	0.78	0.89	0.93	0.84	0.81	0.87
	F1	0.87	0.76	0.63	0.77	0.82	0.71	0.67	0.74	0.91	0.89	0.82	**0.88**	0.87	0.88	0.85	0.87
	Acc	0.77				0.75				0.89				0.87			

* NE: no_evidence, S: support, C: contradict, wa: weighted average, P: precision, R: recall, F1: F1 score
Acc: accuracy

Table 2: Results of the DeBERTa model fine-tuned on the 80% and 90% of the SciFact dataset end evaluated on the HealthVer test set.

	DeBERTa$_{SF-80}$				DeBERTa$_{SF-90}$			
	NE	S	C	wa	NE	S	C	wa
P	0.46	0.70	0.66	0.60	0.47	0.67	0.69	0.59
R	0.94	0.25	0.15	0.50	0.88	0.29	0.27	0.52
F1	0.62	0.37	0.24	**0.44**	0.61	0.40	0.39	**0.48**
Acc	0.50				0.52			

5 RESULTS

Our models underwent a three-stage evaluation process. First, we assessed the models on a test subset of the transformed SciFact dataset (refer to Section 5.1) to evaluate their performance and identify the best model. Next, we tested the optimal model on an external dataset (Section 5.2). Finally, we compared the performance of our top model against GPT-4 models in Section 5.3.

5.1 In-Domain Evaluation

As indicated in Table 1, the best-performing model with an early stopping patience of 3 was DeBERTa SQuAD, achieving an F1-score of 0.87. In comparison, the DeBERTa model with an early stopping patience of 4 achieved the highest F1-score of 0.88.

The results also reveal that the *CONTRADICT* class poses the most significant challenge for the models, which is anticipated given that this class constitutes only 22% of the dataset, resulting in fewer examples for training and evaluation. This imbalance likely contributes to the models' difficulties in accurately predicting this class, highlighting the need for more targeted strategies to improve performance in underrepresented categories.

5.2 Out-of-Domain Evaluation

To evaluate our best-performing model[9], DeBERTa fine-tuned with an early stopping patience parameter set to 4, on a dataset distinct from the one used for training and in-domain evaluation, we chose the HealthVer dataset. This dataset is designed for evidence-based fact-checking of health-related claims, allowing researchers to assess the validity of real-world claims by evaluating their truthfulness against scientific articles.

As can be seen in Table 2 (DeBERTa$_{SF-80}$), we obtained a weighted average F1 score of 0.44 and an accuracy of 0.50. Comparing these results to those

[9]https://huggingface.co/MilosKosRad/TextualEntailment_DeBERTa_preprocessedSciFACT

Table 3: Comparison of the DeBERTa$_{SF}$ model with GPT-4 models.

	DeBERTa$_{SF}$				GPT-4				GPT-4 Turbo				GPT-4o			
	NE	S	C	wa	NE	S	C	wa	NE	S	C	wa	NE	S	C	wa
P	0.88	0.90	0.88	0.89	0.85	0.77	0.84	0.82	0.93	0.81	0.65	0.82	0.72	0.91	0.74	0.80
R	0.95	0.88	0.78	0.89	0.80	0.94	0.59	0.81	0.64	0.92	0.81	0.80	0.89	0.80	0.63	0.80
F1	0.91	0.89	0.82	**0.88**	0.82	0.85	0.70	**0.81**	0.76	0.86	0.72	0.79	0.80	0.85	0.68	0.79
Acc	0.89				0.81				0.80				0.80			

reported by the authors in (Sarrouti et al., 2021), who fine-tuned a BERT-base model with SciFact and evaluated it on the HealthVer test set, we observe that they achieved an F1 score of 0.36 and an accuracy of 0.39. This demonstrates that our DeBERTa model fine-tuned on the transformed SciFact dataset improved upon these results.

In Section 5.1, we identified DeBERTa fine-tuned with 80% of the transformed SciFact dataset as the optimal model for Textual Entailment. Subsequently, we evaluated this model out-of-domain on the HealthVer dataset, observing superior performance compared to previous state-of-the-art (SOTA) models, with an absolute increase of 8% in F1 score. Given that the test subset, constituting 10% of the transformed SciFact dataset, had already been utilized for in-domain evaluation, we incorporated it into the training set. Subsequently, we retrained the DeBERTa model on 90% of the data from the transformed SciFact dataset. Upon evaluating the new model on the HealthVer dataset, we observed a further absolute improvement of 4% in the F1 metric (refer to Table 2, DeBERTa$_{SF-90}$ model). Furthermore, the exploration of augmenting the training dataset underscores the adaptability and robustness of our methodology.

5.3 Comparison with GPT-4 Models

We utilized the same test set as for our in-domain evaluation – 10% of our transformed SciFact, comprising 122 examples and encompassing three classes, to assess the performance of GPT-4, GPT-4 Turbo, and GPT-4o in zero-shot mode. The specific prompt employed for this testing was as follows:

> Critically asses whether the statement is supported, contradicted or there is no evidence for the statement in the given abstract. Output SUPPORT if the statement is supported by the abstract. Output CONTRADICT if statement is in contradiction with the abstract and output NO_EVIDENCE if there is no evidence for the statement in the abstract.

For all models, the temperature parameter was set to 0 to minimize randomness and generate the most deterministic outputs, while the max_tokens parameter was set to 350 to allow sufficient context generation. This approach enabled us to directly compare the performance of our fine-tuned transformer-based model with that of the GPT-4 series models in zero-shot regime under identical conditions.

In our experiments, we observed that our transformer-based model for claim verification outperformed GPT-4, GPT-4 Turbo, and GPT-4o (refer to Table 3). Specifically, our model demonstrated superior performance across various evaluation metrics, including both accuracy and F1-score. These results highlight the efficacy of our fine-tuning approach and the robustness of our model architecture in handling the complexities of claim verification tasks. The consistent outperformance of our model over the aforementioned state-of-the-art models underscores its potential for real-world applications and further establishes its credibility within the domain of automated fact-checking.

Additionally, our model is open-source, providing transparency and flexibility that are crucial for industries such as pharmaceuticals and biomedicine, where stringent process control is required. Unlike closed models, our open-source solution allows for comprehensive customization and verification, ensuring that the claim verification process adheres to the rigorous standards necessary in these fields.

6 ERROR ANALYSIS

An error analysis was undertaken to scrutinize misclassified claims within the in-domain evaluation subset of the transformed SciFact dataset, leveraging our top-performing model, DeBERTa$_{SF}$. As depicted in Figure 2, a total of 14 claim-abstract pairs were inaccurately classified, revealing the distribution of errors across different classes. While the model exhibits commendable performance in the NO_EVIDENCE and SUPPORT classes, it demonstrates a relatively higher misclassification rate in the CONTRADICT class, suggesting a focal point for potential enhancement.

The frequent misclassification of the SUPPORT class as the NO_EVIDENCE class primarily stems from the inclusion of numerical data in the evidence, encompassing various measures and variations. Additionally, the model's inability to recognize abbreviations of specific terms as equivalent contributes to this misclassification. Moreover, the presence of intricate details such as biological processes and chemical reactions may confound the model, especially when it lacks explicit fine-tuning to handle such nuanced information.

On the other hand, instances where genuine SUPPORT claims are incorrectly classified as CONTRADICT pose significant challenges within our dataset. These misclassifications are often attributed to the semantic complexity inherent in clinical trial data, characterized by complex immunology terminology, detailed descriptions, and intricate comparisons. Furthermore, the model encounters difficulties in discerning the alignment between specific time frames provided in the evidence and categorizing the claim as contradictory, even when the evidence unequivocally supports the claim.

Notably, the CONTRADICT class exhibits the highest proportion of errors, with instances erroneously classified as SUPPORT in 4 examples and as NO_EVIDENCE in 2 examples out of a total of 27 CONTRADICT examples in the test set. In light of these observations, we will delve into the most problematic cases within this class.

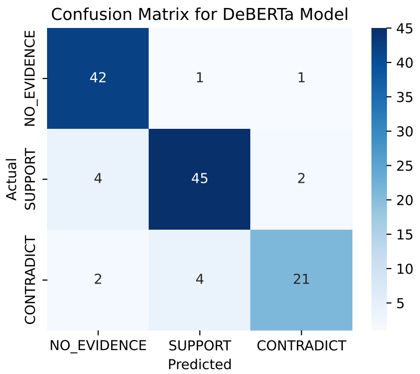


Figure 2: Confusion matrix for DeBERTa$_{SF}$ model.

Misclassifying claims from CONTRADICT to SUPPORT poses significant challenges, particularly in practical applications. These instances often exhibit surface-level similarities, marked by the presence of common keywords and phrases, without sufficiently interpreting the underlying semantic relationships. Consequently, the model may overlook clear contradictions in the evidence, relying instead on generalizations stemming from the presence of related terms. Furthermore, this misclassification is exacerbated by the semantic complexity inherent in the content of the evidence, along with the intricate analytical nuances embedded within scientific concepts. Additionally, challenges arise in comprehending the context surrounding complex biological processes, further complicating the accurate classification of claims.

For example, considering the claim ”The genomic aberrations found in matasteses are very similar to those found in the primary tumor”, several reasons contribute to the model misclassifying it as supporting the claim:

- The word ”metastases” is incorrectly written as ”matasteses” in the claim, but correctly spelled in the evidence, which could confuse the model.
- The claim implies a straightforward similarity between the genomic makeups of metastases and primary tumors. However, the evidence delves into the complex evolutionary process and genetic diversity of metastases, discussing timing, routes of dissemination, and genetic signatures of metastatic processes.
- The evidence describes metastases evolving in parallel, while the claim suggests a linear evolution, a critical distinction the model fails to recognize.

Misclassifying claims from CONTRADICT to NO_EVIDENCE may stem from inadequate representation of nuanced contradictions in scientific contexts within the training data. The absence of sufficient examples illustrating subtle contradictions may limit the model's ability to accurately discern and classify such instances. Moreover, the presence of semantic complexity, numerical data in experiments, variations in experiment timelines, and implicit contradictions further exacerbate the model's challenges in inference. Consequently, the model may struggle to make accurate classifications in scenarios where these factors interact, leading to instances of misclassification within our dataset.

For instance, considering the claim ”The most prevalent adverse events to Semaglutide are cardiovascular.”, several factors contribute to the misclassification:

- The evidence includes various numerical data points, such as medication dosages (e.g., 2.5 mg, 5 mg), percentages of change in hemoglobin A1c levels, body weight changes, and percentages of patients experiencing adverse events, which poses a challenge for the model.

- The contradiction is not explicitly stated, as the claim focuses specifically on cardiovascular adverse events as the most prevalent for Semaglutide. However, the evidence primarily discusses Semaglutide's efficacy in glycemic control, mentioning adverse events more broadly, with an emphasis on gastrointestinal events.
- The evidence contains extensive detail about the clinical trial's setup, participant demographics, dosage specifics, efficacy outcomes, and overall adverse events, overwhelming the NLI model, which needs to filter out irrelevant information to focus on the claim.

7 CONCLUSION AND FUTURE WORK

This paper outlines the development and evaluation of a Verification Engine as part of an open-source scientific question-answering system. By fine-tuning models for the Natural Language Inference task on a processed SciFact dataset, we aimed to provide referenced, automatically vetted, and verifiable answers.

Our in-domain evaluation identified the DeBERTa model, fine-tuned with 80% of the transformed SciFact dataset, as the optimal performer with an F1 score of 0.88. Testing this model on the HealthVer dataset, we achieved an F1 score of 0.44 and accuracy of 0.50, surpassing previous benchmarks and demonstrating significant improvements. Further experiments revealed that augmenting the training dataset to 90% led to an additional 4% increase in the F1 metric, emphasizing the robustness of our approach. Comparisons with GPT-4 models in a zero-shot regime showed our model's superior performance with a 7% absolute increase in F1 score to the best-performing GPT-4 plain model, highlighting its effectiveness in claim verification tasks.

Our model's open-source nature offers significant benefits for domains like pharmaceuticals and biomedicine, where rigorous process control is essential. Unlike closed models, our solution allows for transparency and customization, ensuring adherence to strict industry standards. Overall, our verification system enhances scientific productivity and establishes a reliable framework for automated fact-checking, crucial for maintaining the accuracy and integrity of scientific information.

Given, that the SciFact dataset contains challenging examples, we believe that the performance of the model tested on this dataset may be also underestimated compared to the real-world claims generated by a large language model. Nevertheless, we aim to further improve our methodology for claim verification in the future. We are planning to conduct research in the following directions:

- Generate a new dataset for claim verification - while SciFact is a decent dataset for biomedical claim verification using literature, we have noticed challenges in the dataset. Some claims may be short, and unclear without further context and this context is missing from the labels. We aim to collaborate with the industry and create a cleaner dataset, that overcomes these challenges.
- We understand that there are limitations of NLI, given even the most powerful neural architectures. Therefore, we aim to create a more comprehensive method, based on the combination of textual entailment task, text similarity, and chains of thoughts in fine-tuned large language models.

Enhancement in the dataset and a more comprehensive methodology will even further push the state-of-the-art in this domain and may contribute to the establishment of trust in generative search engines. Also, the area of claim verification is important as it may contribute to the adaptation of generative AI in the scientific domain, where the adaptation is currently limited due to the phenomenon of hallucinations, which makes verification of generated texts time-consuming and generated text unusable.

REFERENCES

Bašaragin, B., Ljajić, A., Medvecki, D., Cassano, L., Košprdić, M., and Milošević, N. (2024). How do you know that? teaching generative language models to reference answers to biomedical questions. In Demner-Fushman, D., Ananiadou, S., Miwa, M., Roberts, K., and Tsujii, J., editors, *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 536–547, Bangkok, Thailand. Association for Computational Linguistics.

Beltagy, I., Lo, K., and Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China. Association for Computational Linguistics.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2019a). Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2019b). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Dettmers, T., Pagnoni, A., Holtzman, A., and Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. *arXiv preprint arXiv:2305.14314*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Diggelmann, T., Boyd-Graber, J., Bulian, J., Ciaramita, M., and Leippold, M. (2021). CLIMATE-FEVER: A Dataset for Verification of Real-World Climate Claims. arXiv:2012.00614 [cs].

Guo, Z., Schlichtkrull, M., and Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., and Smith, N. A. (2020). Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

He, P., Liu, X., Gao, J., and Chen, W. (2021). Deberta: Decoding-enhanced bert with disentangled attention. In *International Conference on Learning Representations*.

Jin, Q., Dhingra, B., Liu, Z., Cohen, W. W., and Lu, X. (2019). Pubmedqa: A dataset for biomedical research question answering. *arXiv preprint arXiv:1909.06146*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kotonya, N. and Toni, F. (2020). Explainable Automated Fact-Checking for Public Health Claims. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7740–7754, Online. Association for Computational Linguistics.

Košprdić, M., Ljajić, A., Bašaragin, B., Medvecki, D., and Milošević, N. (2024). Verif.ai: Towards an open-source scientific generative question-answering system with referenced and verifiable answers.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Ljajić, A., Košprdić, M., Bašaragin, B., Medvecki, D., Cassano, L., and Milošević, N. (2024). Scientific qa system with verifiable answers.

Malaviya, C., Lee, S., Chen, S., Sieber, E., Yatskar, M., and Roth, D. (2023). Expertqa: Expert-curated questions and attributed answers. *arXiv preprint arXiv:2309.07852*.

Mohr, I., Wührl, A., and Klinger, R. (2022). CoVERT: A Corpus of Fact-checked Biomedical COVID-19 Tweets. In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Odijk, J., and Piperidis, S., editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 244–257, Marseille, France. European Language Resources Association.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21(1).

Sarrouti, M., Ben Abacha, A., Mrabet, Y., and Demner-Fushman, D. (2021). Evidence-based Fact-Checking of Health-related Claims. In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3499–3512, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tan, N., Nguyen, T., Bensemann, J., Peng, A., Bao, Q., Chen, Y., Gahegan, M., and Witbrock, M. (2023). Multi2Claim: Generating Scientific Claims from Multi-Choice Questions for Scientific Fact-Checking. In Vlachos, A. and Augenstein, I., editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2652–2664, Dubrovnik, Croatia. Association for Computational Linguistics.

Thorne, J., Vlachos, A., Christodoulopoulos, C., and Mittal, A. (2018). FEVER: a large-scale dataset for fact extraction and VERification. In Walker, M., Ji, H., and Stent, A., editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics.

Vlachos, A. and Riedel, S. (2014). Fact Checking: Task definition and dataset construction. In Danescu-Niculescu-Mizil, C., Eisenstein, J., McKeown, K., and Smith, N. A., editors, *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA. Association for Computational Linguistics.

Vladika, J. and Matthes, F. (2023). Scientific Fact-Checking: A Survey of Resources and Approaches.

In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6215–6230, Toronto, Canada. Association for Computational Linguistics.

Vladika, J., Schneider, P., and Matthes, F. (2024). HealthFC: Verifying Health Claims with Evidence-Based Medical Fact-Checking.

Wadden, D., Lin, S., Lo, K., Wang, L. L., van Zuylen, M., Cohan, A., and Hajishirzi, H. (2020). Fact or Fiction: Verifying Scientific Claims. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7534–7550, Online. Association for Computational Linguistics.

Wadden, D., Lo, K., Wang, L. L., Cohan, A., Beltagy, I., and Hajishirzi, H. (2022). MULTIVERS: Improving scientific claim verification with weak supervision and full-document context. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 61–76. Association for Computational Linguistics.

Wei, J., Yang, C., Song, X., Lu, Y., Hu, N., Tran, D., Peng, D., Liu, R., Huang, D., Du, C., and Le, Q. V. (2024). Long-form factuality in large language models.

Semantic Support Points for on the Fly Knowledge Encoding in Heterogenous Systems

Daniel Spieldenner[1][a], André Antakli[1][b], Torsten Spieldenner[2][c] and Harkiran Sahota[2][d]

[1]*Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Campus D3 2, Saarbrücken, 66123, Saarland, Germany*

[2]*Lappeenranta-Lahti University of Technology LUT, Yliopistonkatu 34, Lappeenranta, 53850, Finland*

{daniel.spieldenner, andre.antakli}@dfki.de, tor.spiel@gmail.com, fourth_author@student.lut.fi

Keywords: Intelligent Systems, Semantic Web, Interoperability, Multi Agent Systems.

Abstract: Connecting participants in heterogenous environments and allowing for seamless interaction is a challenging task that requires profound knowledge of the system envolved and data exchanged. The knowledge on how to integrate and use certain systems is usually given in the form of manuals, code documentation or needs to be derived by the developer by understanding and interpreting source code or data sets. With this work, we propose an approach to provide this knowledge about data via so called semantic support points, making use of Semantic Web technologies and appropriate ontologies, in a resource efficient fashion by creating semantic representations only when and where needed. Instead of just providing semantic meta data to describe actors in the environment, we make it possible to embed actual data values from data sources like databases and payloads exchanged between systems into a virtual semantic cloud, allowing for interactions purely on the semantic data representations and propagating computation results back to the system layer automatically.

1 INTRODUCTION

Every day our world is becoming more and more complex, systems are growing in size, data is produced in unprecedented amounts and exchanged in increasingly smart environments, from smart homes equipped with devices ready for the internet of things (Khanna and Kaur, 2020; Nižetić et al., 2020; Lee and Lee, 2015) to smart cities (Wang et al., 2020; Jafari et al., 2023; Khan et al., 2020) to the vision of national or even international data spaces. (Halevy et al., 2006; Franklin et al., 2005; Curry, 2020; Solmaz et al., 2022; Zillner et al., 2021)

Maintaining these systems, consisting of often very heterogenous devices, services and data sources, some of them decades old from times long before concepts like interoperability or the *Semantic Web*(Lassila et al., 2001) were are thing, poses an enormous challenge, making it necessary to understand data, transform data to be understood by the recipient and making data discoverable and available in general.

This is especially true for complex intelligent systems, which consist of a range of actors working together autonomously. For instance, in a smart home, sensors and control units collaborate to maintain the resident's preferred room temperature and lighting. Similarly, an entire shop floor with machines forming an intricate production line operates in a coordinated manner. However, every participant in even the most complex system was designed by someone with a certain intention about the task the participant is expected to perform, under which conditions, with certain prerequisites and expected outcomes. Usually this knowledge is ”encoded” in manuals, software documentation, code comments or what we would consider ”common knowledge”, for example the fact that a thermometer is supposed to measure temperature.

We aim to bring this implict knowledge contained by design within the actors themselves to the environment itself, allowing the participants to emanate this ”knowledge” when and where needed as requested and to weave somewhat of a volatile, connected fabric of semantics above the actual system layer, as suggested by (Spieldenner, 2023), expressing and encoding expert knowledge concerning the available systems, the intended interpretation of data values, capabilities of actors and the possible connection be-

[a] https://orcid.org/0000-0002-8262-7733
[b] https://orcid.org/0000-0002-4066-7649
[c] https://orcid.org/0000-0003-3034-9345
[d] https://orcid.org/0000-0003-1160-1904

Spieldenner, D., Antakli, A., Spieldenner, T. and Sahota, H.
Semantic Support Points for on the Fly Knowledge Encoding in Heterogenous Systems.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 26-37
ISBN: 978-989-758-716-0; ISSN: 2184-3228

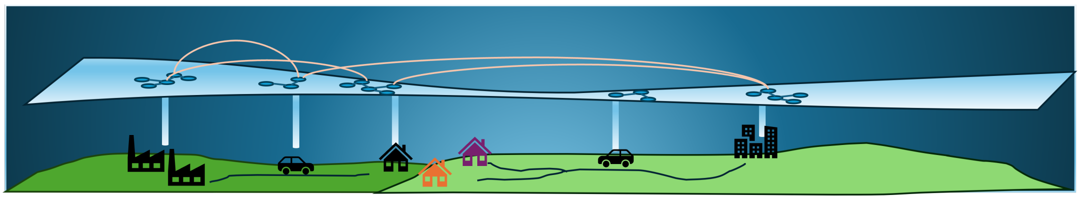


Figure 1: Our vision of semantic support points, spanning a layer of semantics abstracting from the world below.

tween participants with the help of ontologies (Breitman et al., 2007) and principles of the *Semantic Web*. (Lassila et al., 2001)

With such a representation at hand, we would allow any system to benefit from concepts like semantic communication (Lan et al., 2021; Qin et al., 2021), semantic interoperability (Heiler, 1995) as well as reasoning and semantic queries.

In this paper we present a method to bridge the gap between a world consisting of legacy data sources, devices and services without any semantic information attached and the aforementioned semantic data fabric, using *semantic support points*: federated, light weight, runtime configurable and non-intrusive micro services tailored to each individual participant.

Each of these micro services have access to individual distributed data transformation definition files, creating a semantic representation of individual or aggregated data values at runtime on request as well as data selection and transformation on the semantic level for transformation back to non-semantic data representations like JSON.

Pairs of support points allow for bidirectional communication between what we consider the *system layer*, i.e. the collection of participants without any semantic information attached, and arbitrary actors on the *semantic layer*, interacting directly and only with knowledge graphs provided by the semantic support points. The knowledge graphs will be derived from non-semantic data emitted from non-semantic interfaces or data sources, and enriched with human written expert knowledge by providing transform rules between non-semantic data formats to RDF, and selection and transform rules to extract data from knowledge graphs and create precisely defined JSON objects accordingly.

This approach enables concepts like semantic interoperability, semantic queries and reasoning on knowledge graphs in any environment, even if it was intially designed without any attempt to provide semantic capabilities.

As a proof of concept, we demonstrate the proposed approach by applying it to a real life example of an automotive production line, using semantic support points to extract information from heterogenous data sources, transforming it to semantic data and feeding it to a multi agent system, using the semantic nature of the mapped data values for deriving optimized production plans by using a reasoning engine. The results of the agent frameworks' plan optimization are then propagated back to a range of systems via non-semantic input tailored specifically to each participant's interface and data model.

The paper is structured as follows:

An overview over the relevant background and related approaches aiming to realize a concept of a semantic data integration is given in section 2. In section 3, we introduce our own concept of *semantic support points*, minimal service implementations allowing to access transform rules based on semantic concepts about systems and data sources to generate knowledge graphs including knowledge about the system interfaces themselves as well as the data exchanged. Section 4 demonstrates how to put the concept into practical use by using semantic support points to enhance a system representing an automotive production line by adding a semantic multi agent system, using the semantic representation of actors on the system level and the data they provide to optimize production plans with the help of semantic reasoning. The results of the exemplary realization and their relevance as proof of concept are summarized in 5 before we finally conclude in section 6, discuss open questions and provide an outlook on possible future work.

2 RELATED WORK

The idea of creating a semantic abstraction layer on top of an existing architecture involves a wide range of research domains and technologies. We will make use of semantic web concepts to describe data values and system interfaces, take some inspiration from ambient intelligent systems and employ multi agent systems to interact autonomously with the environment.

In the following, we provide background information to these topics and conclude with a brief overview over existing semantic layer approaches.

```
@prefix ex: <http://www.example.org/> .
@prefix rdfs: <http://www.w3.org
    /2000/01/rdf-schema#> .

<ex:placeholder> a <rdfs:Resource>;
                 <rdfs:label> "Example" .
```

Figure 2: Example for a very simple RDF graph in turtle syntax.

2.1 RDF, OWL and the Semantic Web

In order to include expert knowledge directly into our semantic layer, we will make use of *semantic web* (Lassila et al., 2001) technologies, namely *RDF graphs* (Manola et al., 2004) as data format, making use of *OWL ontologies* (Breitman et al., 2007; McGuinness et al., 2004) to define the concepts and relations relevant in a given environment, both well established W3C standards.

RDF allows us to make statements about things in the world in terms of triples, consisting of a subject, predicate and object. The place of the subject position in such a triple is taken by a *resource* with a *unique resource identifier* (URI). This resource does not only represent an entity in the abstract, digital world, but can directly relate to a physical thing in the real world, providing an address where to access information about this entity via the URI.

A RDF triple on its own does not necessarily express any knowledge yet and can be not much more than arbitrary strings, providing no additional valuable information aside from existing links between resources. Using appropriate ontologies by including their respective namespaces and asserting resources in our RDF graph to be members of classes defined there, we can include these semantics, modelled by domain experts, into our knowledge graph, and with that to interfaces and data values of a participant in our system. In order to define these ontologies, we make use of the *Web Ontology Language* (OWL), allowing us to define concepts like sub class relationships, class unions, exclusions, properties and so on.

2.2 Applications of the Semantic Web Idea

Thinking about systems consisting of actors with each actor provided with the capability to express knowledge about itself, making it in some sense intelligent, brings the idea of *Ambient intelligence* to mind, as Weber et al described it as *"the vision of a future filled with smart and interacting everyday objects offers a whole range of fascinating possibilities"* (Weber et al., 2005; Chung et al., 2020), or Dunne et al. put it in a recent survey concerning the current state of ambient intelligence research: *"AmI can be synonymous with terms such as smart homes, smart environments, and intelligent environments. It is an umbrella term for a set of technologies that are embedded into the physical surroundings — seamlessly — to create an invisible user interface augmented with AI"* (Dunne et al., 2021).

Furthermore, encoding actual expert knowledge within the ambient intelligent environment by putting into effect principles of the semantic web is a widely regarded as worthwhile approach (Santofimia et al., 2011; Razzak et al., 2013; Ngankam et al., 2022; Padilla-Cuevas et al., 2021).

This gives rise to the need for flexibly maintainable, quick to set up and extend, systems, allowing for a seamless integration of both knowledge as such as well as actors to work on that knowledge.

While the semantic, ontology side of view is examined in detail, providing rather a theoretical concept of knowledge encoding, the practical application of these concepts to existing systems is often lacking.

Concepts like the *Semantic Web of Things* (Ruta et al., 2012) aim to bridge the gap between semantically annotated data on the one hand and a real system world on the other hand by introducing the idea of including knowledge about every participant via per device knowledge base generation, united in what Rute et al. refer to as *"ubiquitous knowledge base"*. The most practical approach we are aware of to actually putting the idea of the semantic web of things into action was proposed by Antoniazzi et al. (Antoniazzi and Viola, 2019)

Building on the W3C vision[1] of the Web of Things (Zeng et al., 2011), Charpenay investigated methods to semantically describe web things (Charpenay et al., 2016) and actual semantic data integration. (Charpenay et al., 2018). Semantic data integration is realized by an automatic transform step to JSON-LD, based on the keys of the original JSON object and the ontology to be used for annotation.

While this again shows the interest in and the need for semantic annotation not only on a device interface meta data level, but actual semantic data integration, for our vision this approach is still too limited in two regards: first, the focus on W3C WoT poses a restriction when it comes to system representation as the assumption that participants can be expressed in terms of sensors, actions and so on does not necessarily hold in every case. Furthermore, the key based JSON to JSON-LD approach does not provide the level of freedom we would like to realize when it comes to including actual human expert knowledge into the semantic

[1] https://www.w3.org/WoT/

representation generated.

Vdovjak et al presented their vision of a *semantic layer* in (Vdovjak and Houben, 2002), however in our vision of enriching a system with semantic information, we aim for giving the user, in our case system maintainers and developers, more control over expressing their intentions, making the construction of the layer as such a more interactive process.

Lu and Ashgar suggested a semantic communication layer for cyber physical systems (Lu and Asghar, 2020), proposing an architecture consisting of a *"physical layer"* representing the actual physical things, a *"cyber layer"* sensing the environment of a thing and planning future behavior, propagating messages through a *semantic layer* to enrich it wiht semantic information and finally transmit it through a *communication layer* to other cyber physical systems. While this work emphasizes the importance of considering semantic data layers in future applications, the approach focusses more on a centralized communication platform design and secure recommendation, rather than providing an actual permanently present knowledge based representation of a world possibly consisting of more than just cyber physical systems.

Entirely abstracting from any specific data formats, ontologies, middlewares or other implementations, Spieldenner described the idea of a semantic *medium*, a virtual, semantic reflection of a collection of systems, enhancing the options of multi agent systems to act within the system by exploiting the semantic information available, and making actual changes in the "real world" by directly linking outcomes in the semantic world to actual actions in the real world. (Spieldenner, 2023)

However, to the best of our knowledge, until today there still is a lack of actual implementations allowing actors to interact on the semantic level, exploit knowledge derivation and reasoning capabilities or allow for actual data exchange down to the system level or dynamic embedding of actual data values into the ubiquitous knowledge graph.

3 THE SEMANTIC SUPPORT POINT CONCEPT

We want the semantic support points to enable semantic integration and data exchange with the following properties:

Distributed. Rather than one monolithic middleware handling all data transform and transmission on one platform, we aim to provide *semantic support points* on top of the system layer, acting like a source of semantic data to be discover- and queriable for participants acting on the semantic level.

Decoupled. Two separate semantic support points should always be decoupled from each other, in the sense that losing access to one support point does not influence the sanity of the other. Each support point should always be able to perform on its own, containing all necessary data to add its specific semantic information to the abstraction layer.

Non-Invasive. Adding semantic information to an existing system via an abstraction layer should be seen as an optional data offer for actors aiming for exploiting available semantic information, e.g. for reasoning to generate new knowledge, achieving interoperability by working on semantic interface abstractions or to have access of semantic data querying capabilities. To this end, semantic access points should be realized in such a way that they provide interfaces that existing systems are able to use (e.g. HTTP endpoints, event streams, message queues etc ...)

Non-Persisting. No state and no data should be stored in a specific support point. All data transformation and access should happen on demand, while necessary data transform rules are provided as necessary via *transform rule injection routes* into each semantic support point.

Domain Independent. The semantic support point implementation itself should be minimal and independent. All (semantic) information needed to describe the abstracted interface or data point should be provided via external sources. Domain independence along with the non-invasive, non-persisting and non-invasive properties ensure that the concept of semantic support points can be applied to any environment, in any domain, to an arbitrary extent, without the necessity of hosting or configuring a pre-implemented middleware beyond providing appropriate semantic transform rules to the semantic support points.

Any semantic support point implementation should be kept as minimal as possible, providing only the most basic, same features per support point. First of all, connection components that allow for data flow into the component as well as out of it. This dataflow is routed through a transform processing step, with transform rules provided via an external injection point. Receiving transform rules from outside sources ensures distributed and non-persisting properties of the support point, as no data is stored within any support point at any time and can be routed as needed. Further, it ensures the domain independence, as the

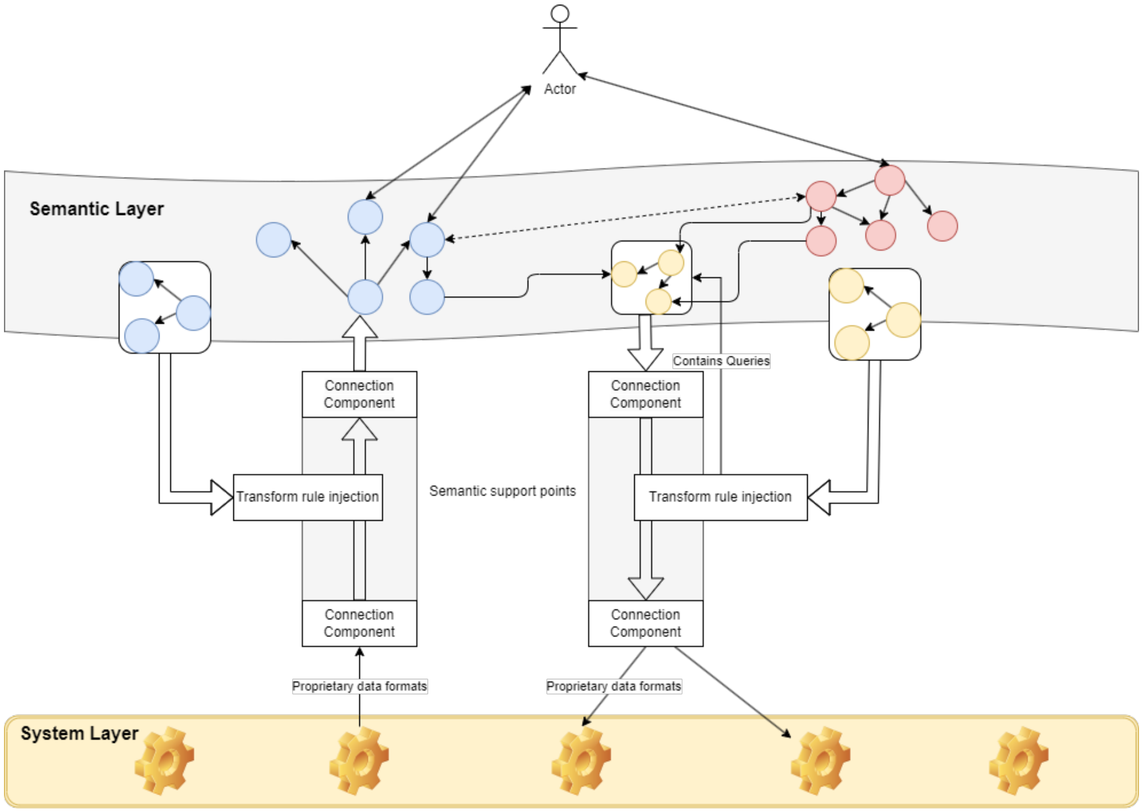


Figure 3: High level overview of the semantic support point concept.

domain solely depends on the transformation route, not on the support point implementation.

The connection point can implement any communication protocol necessary to connect the systems on the system layer it provides semantic support for as needed, be it HTTP, event driven systems, message buses or whatever else might exist out there.

This is true for both transformation directions. In case of a support point generating RDF from non-RDF sources, the outgoing connection component must provide interfaces for any actors that want to interact with it on a semantic manner. The semantic layer as such is non-persisting and provides data only as needed.

See figure 3 for a schematic overview of the semantic support point concept. Data is emitted in semantic format, interpretable as knowledge graph, to be provided to an actor intending to process semantically annotated data. This data processing may include operations like reasoning, combining information from several knowledge graphs, or linking concepts from a knowledge graph emitted by a semantic support point to a larger ontology (red knowledge graph in the diagram).

In order to propagate results of computations on the semantic level back to the system layer, again a transform rule is used that may contain additional information on how to query the dataset provided as result of the actor's processes. The sub knowledge graph resulting from this querying step is then routed back through a support point and transformed to a data object according to the transformation rules provided via the injection route.

Notice that transformation rules, as they are written in RDF, can be understood as part of the virtual semantic layer and as such can provide additional information for actors acting on the semantic level, e.g. for service discovery and composition by exploiting information about the handling of certain data values contained in the transform file knowledge graph.

This concept of semantic support points fulfills the intended properties listed above: it is *distributed* in the sense that the minimal support point implementations can be hosted independent from each other where needed and where possible, *decoupled* in the sense that one support point crashing does not affect the runtime of any other support point, or any available transformation rules, *non-invasive* in the sense that existing interfaces do not need to be changed but can connect to support points to send and receive data as-is, *non-persisting* as no knowledge graphs that are results of semantic transformation are stored, and as actors that interact with the semantic layer don't work on a persisted knowledge graph directly, but via the connection components of semantic support points and *domain independent* in the sense that domain is defined by the separately provided transformation rules, not by the support point implementations themselves.

4 USING SEMANTIC SUPPORT POINTS: AN EXAMPLE

In order to demonstrate how to combine the capabilities micro transformation service mapping data from and to RDF to a full abstraction layer we give an example making use of a semantic multi agent system.

Each agent maintains an individual view onto the world in its own knowledge base. This has multiple benefits compared to trying to render the knowledge of the entire world continuously and storing it in one central spot.

First of all it ensures encapsulation between multiple agents acting potentially simultaneously.

Second the we do not need to store a large amount of data representing the entire world at once, which would require not only a large data storage to do so, but also long computation times to process the queries to find relevant data. Instead, each agent can decide when and which data it needs. Imagine it as asking someone who has knowledge you need to explain it to you, or as only looking at the pages of a documentation that cover the points of interest for you.

4.1 Mapping Between Structured Data an RDF

In the following we provide some inspiration on how to realize the concept of a semantic support point in a real world application.

For mapping non-semantic data into a RDF knowledge graph, we entirely rely on freely available 3rd party libraries, while for the mapping back from a semantic representation to a explicitly specified interface, we propose our own solution.

4.1.1 From Non-Semantic Data Formats RDF

In order to generate RDF graphs from different data formats, several approaches exist (Hildebrand et al., 2019; Méndez et al., 2020; Sahoo et al., 2009), however due to its wide acceptance and the benefit of being able to manage mapping files in a non-central, flexible distributed way, we aim for using *RML* (Dimou et al., 2014).

Mapping files written in RDF allow, in our opinion, the best trade off between complexity and the flexibility for system maintainers or service developers to freely express their intentions and their interpretation of the role of their system in terms of the generated RDF graph. Moreover, the possibility to store mapping files in a federated manner, making it easy to add mappings for new concepts, updated existing mappings to reflect changes in the environment and remove knowledge that is not longer needed at runtime support the idea of a seamless integration of new concepts, weaving our linked data fabric quilt patch by patch without having to interfere with the underlying physical world.

In order to provide RML mapping functionality to our system, we wrap *CARML* [2] into a micro service offering either an HTTP endpoint for direct calls or capabilities to connect to an event stream. Via what we call an *mapping file injection route* mapping files to be used for mapping the incoming data are received from a configurable remote endpoint, the resulting RDF graph can either be forwarded via HTTP to a pre configured external endpoint or written to an event based message bus.

4.1.2 From RDF to Explicitly Specified Non-Semantic Data Formats

In order to access the semantic abstraction layer and propagating knowledge from there back to specific systems, we aim to use a similar approach as RML provides, but in the opposite direction.

While JSON-LD (Sporny et al., 2020) as out of the box available serialization format seems like an obvious choice at first glance, it does not provide the functionality we need. First of all, we are not interested in just generating a JSON representation from an entire RDF graph, which would be the case if we just take the JSON-LD representation of some given RDF, but we want to be able to select the information we are interested in. This requires at least enabling SPARQL queries as a preprocessing step. Also, we need more control over the structure of the generated JSON object, the key labels, value types and possible addition of literals not included in the RDF dataset.

Say we want to generate a simple JSON object like given in 4 as an example with the goal to propagate a measured room temperature to some fictional smart thermostate in order to decide whether the heater should switch off or continue heating. The measured room temperature was published by the temperature sensors to our semantic abstraction layer, however the information that this value now should be forwarded

[2]https://github.com/carml/carml

```
{
    "id": "wekbnfi35r",
    "targetDevice": "thermostate",
    "temperature": 21
}
```


Figure 4: JSON example including a ”type” not derivable from the RDF dataset.

to a thermostate is knowledge of the recipient system's developer, not something a third party would have published as semantic data.

This step of adding additional information in the process of making the semantically available data useful for the actual physical systems is something that can not be accomplished by just serializing existing RDF data to JSON-LD.

JSON schema (Pezoa et al., 2016) is not a feasible approach as it is cumbersome to describe larger objects and it lacks proper integration of semantic data. (Pezoa et al., 2016). Working on and with the semantic abstraction layer is supposed to happen in terms of expressing interpretations and intentions, instead of being a purely technical task. Therefore, mapping rules are provided and applied on a semantic level, analogously to RML mapping files.

While approaches on how to derive non-semantic data objects from RDF knowledge graphs are investigated (Allocca and Gougousis, 2015; Grassi et al., 2023), for our intentions they lack the possibility to enforce restrictions on the structure of the generated JSON or the exact definition of key names to use.

For this reason, we developed *POSER*[3] (Spieldenner, 2022) to provide data mapping from RDF graphs into a JSON object, following the structure given by transformation rules formulated in terms of a minimal JSON ontology. Briefly summarized, the intended outcome is described in terms of a nested hierarchy of resources representing JSON objects and datatypes, along with a datatype header defining queries and subgraph matches to extract them from a given knowledge graph and put them in relation to the JSON structure defined in terms of the JSON ontology.

These transformation rules can be injected into a semantic support point, analogously to RML mapping file in the inverse direction.

RDF data is pushed by or read from an actor on the semantic level side by the incoming connector component, filtered or queried according to the datatype header, written into a JSON object constructed according to the rules in the transformation file and then provided to systems on the system layer level via the outgoing connection component.

[3] https://github.com/spidan/poser

4.2 Integration of a Semantic MAS

As an application example to serve as proof of concept, we show an integration of a semantic multi agent system into an automotive production system, aggregating data from heterogenous data sources, exploiting the semantic representation by performing reasoning operations and deriving optimized production plans that are again provided to different elements on the system level.

Examples for agent systems capable of working on semantic data are the semantic web-based MAS platform SEAGENT (Dikenelli et al., 2005)introduced in 2005, the system of Sabbatini et al. (Sabbatini et al., 2022), in which OWL and RDF knowledge graphs are used to train machine learning systems to realize learning agents in the Semantic Web or (Ciortea et al., 2018) in which a web-based MAS for manufacturing is introduced, interacting with Linked-Data and Web-of-Things environments to develop new behaviors from the semantic environment. However, as these systems are often designed with specific applications in mind tailored for specific applications and require adaptation to work fully exploit the benefits of semantic data sources, such as deduction of new facts via reasoning or logic operations on the semantic data set, we decided to use our own semantic Multi Agent System *AJAN* (Antakli et al., 2023) .

For additional information on (MAS) and its capabilities, we'd like to refer you to previous works where it has been used in various Semantic Web and non-Semantic Web based environments. For example, in an smart living environment in which agents use the W3C Web of Things (WoT) architecture (see (Alberternst et al., 2021)), to optimize the production of a virtual factory floor using an AJAN-based MAS (see (Spieldenner and Antakli, 2022)), to coordinate language courses (see (Antakli et al., 2023)), or to control simulated human-robot collaboration scenarios (see (Antakli et al., 2019)).

4.3 Application Scenario

The EU-funded AIToC[4] project developed an assistance system (see Figure 5) for on-site workers and production planners in the automotive sector. A key component is the Operation Reasoner (see Figure 6), an agent-based planner using the MAS-framework AJAN[5] [6] to generate assembly plans for visualization

[4] EU-Project AIToC: aitoc.eu/

[5] AJAN on GitHub: github.com/aantakli/AJAN-service

[6] Operation Reasoner AJAN-model (ZIP-File) including RML and POSER mapping descriptions: github.com/aantakli/AJAN-

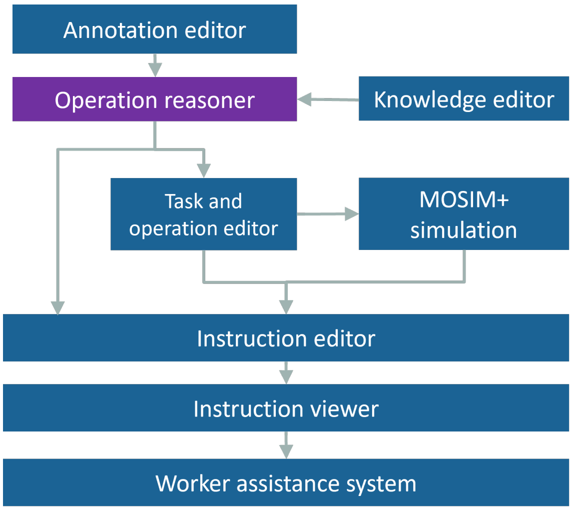


Figure 5: Assistance Pipeline.

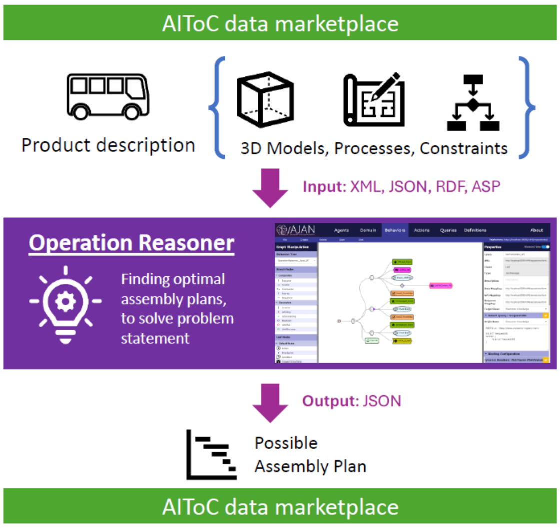


Figure 6: Operation Reasoner.

and verification.

The Operation Reasoner is a Web Service combining general process rules and specific product annotations to derive required assembly plans using the Answer Set Programming (ASP)(Vladimir, 2008) solver clingo[7]. Therefore, the Operation Reasoner uses previously defined product descriptions as an input for clingo and derives possible assembly plans from resulting stable models. The product descriptions are stored in the project data marketplace, the first point of exchange for the various services of the different project partners to exchange data with each other. These descriptions including annotated 3D models of assembly parts (originated from the Annotation editor) and process definitions and constraints (originated from the Knowledge editor), are defined in JSON respectively AutomationML (XML), RDF and ASP. For the use within the Operation Reasoner, product descriptions are translated into RDF, the data format internally used by our MAS, using RML.

The fact that the required aggregated and homogenized planning problem is contained in the agent's knowledge enables the AJAN-based ASP-reasoning(Antakli et al., 2021) of instructions in the next step. In order to make the generated assembly plans available to other components and services within the project architecture, they must be translated back into JSON, generally used by services from the project service market place, and stored in the data marketplace. The Operation Reasoner uses the approach presented in 4.1.2 to translate the RDF-based data into the target data format. Based on the resulting possible plans, other services such as a motion synhtesis based Simulation can intuitively display such assembly plans in a 3D simulated factory or via the Instruction Viewer to a real worker on site for assistance.

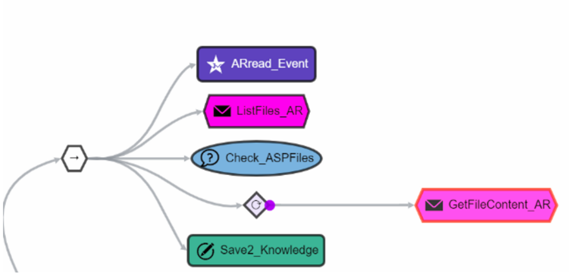
As an example of picking up data from the system layer, the processing of the semantic layer and the forwarding of the results to the system layer, we delve a bit deeper in showing the first step in the planning process of an Operation Reasoner agent: the reading of heterogeneous product description information.

In Figure 7a, the BT to fulfill this task of collecting data from the real world of an Operation Reasoner agent in charge of, is presented. First, the agent receives an instruction (purple colored BT node) to generate optimal assembly plans, by pointing it to multiple data points from the system layer, which consists the needed information to perform the reasoning and planning task. The querying of this information is done by the pink colored BT nodes, where the *transform rule injection routes*, as defined in 4.1, are given via the node property fields: *Data Mapping*, *Endpoint Mapping* and *Response Mapping*. This is done in form of endpoints to read the respective mapping data from, as shown in Figure 7b. Via a *SPARQL Construct query*, the information relevant for an API call on system level, is collected from the agent's knowledge base, transformed to a proper JSON object matching the recipients API and performed as an HTTP POST request.

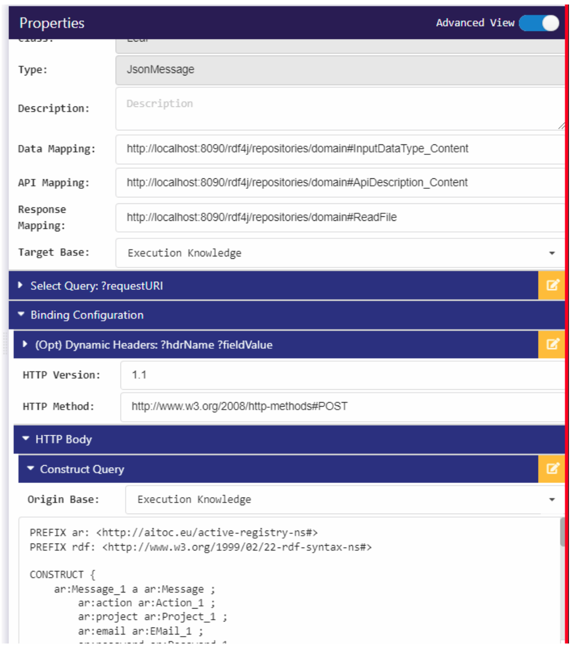
In order to ensure the correct data is provided to the JSON API, during the transformation step we provide a definition of datatypes we are interested in (see figure 8). An example of the JSON object to be generated is given in figure 9. For the sake of brevity, we

packages/blob/main/packages/Operation_Reasoner.ajan

[7]ASP Solver clingo: https://potassco.org/clingo/



(a) Behavior Tree executing collecting data for plan generation and data forwarding.



(b) Properties for the Behavior Tree given above.

Figure 7: Behavior tree and its execution properties.

did not include the full API transform definition used by the transform engine. The response of this request is again received by the agent and mapped back to RDF via the RML service, allowing the actual assembly planning task in the next step.

5 RESULTS

By realizing the concept of semantic support points, we were able to efficiently abstract from data provided by heterogenous systems, providing a semantic representation and relations inbetween data to be consumed by and worked on a semantic multi agent system. For example, worker instructions were generated by collecting data from different sources, like properties of tools and parts from 3D models, process descriptions in ASP and general properties of

```
:InputDataType_Content {
  json:EntryPoint a ar:Message;
    ar:action ar:Action;
    ar:project ar:Project;
    ar:file ar:File;
    ar:email ar:EMail;
    ar:password ar:Password .

  ar:Action ar:value iots:Number .
  ar:Project ar:value iots:Number .
  ar:File ar:value iots:Number .
  ar:EMail ar:value iots:Number .
  ar:Password ar:value iots:Number .
}
```

Figure 8: Data types to be used to determine the specific values in a generated JSON object.

```
{
    "action": "getFileAsJSON",
    "project": 17,
    "file": "context_information.lp",
    "email": "X.X@example.com",
    "password": "asfkalnknakfsnfjb55"
}
```

Figure 9: Example of JSON object to be generated.

the environment given in JSON, transformed to RDF, aggregated, used in a planning and reasoning engine and finally the results being returned to different systems like animation synthesis or plan generation UI in JSON (see figure 10 for an UI example).

Notice how the interplay between RML and JSON mappings, along with the concept and importance of the on-demand semantic layer, becomes especially apparent here. We cannot rely on the information gathered in the first step being available in one specific data format. Instead, it is provided by a plethora of heterogeneous sources: some in JSON or XML format, some consisting of ASP rules, and others directly providing RDF data. This diversity makes a certain level of interoperability necessary.

This is, in this case, achieved by exploiting the semantic interoperability capabilities of the semantic layer, making the data available in an aligned format as soon as the agent requests it.

Further more, in order to perform the reasoning tasks for plan generation, we are required to have data available in RDF format, with the results being provided in RDF again. This constitutes processes on a purely semantic level with no direct relation to interfaces in the physical world, making a concise, configurable and query enriched transform step from RDF to JSON necessary.

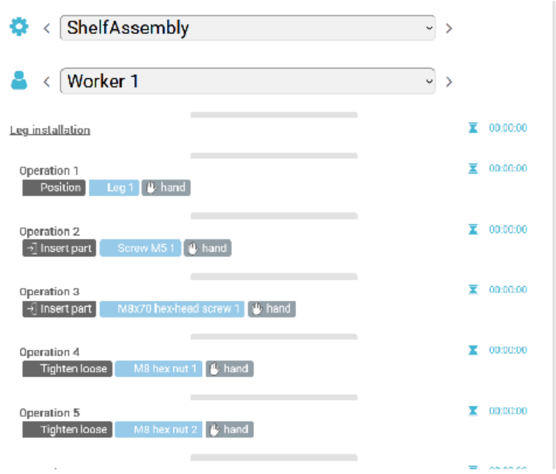


Figure 10: UI to model instructions for workers, pre-configured by results of reasoning on agent level converted back to JSON.

This demonstrates the idea behind and the power of the semantic layer by moving beyond simple interoperable data exchange in an IoT environment, which is already a complex task. By combining raw data with expert knowledge about data interpretation (the data transform step), we fully exploit the possibilities of this approach. Additionally, we incorporate actual intention.

6 CONCLUSION AND DISCUSSION

We have introduced the concept of federated semantic support points to abstract from interfaces in a heterogenous, non-semantic environment, allowing for a unified view on the underlying system world. Minimal stateless semantic transformation services that represent these support points are used to generate a RDF graph from non-RDF data, using established mapping technologies, and again propagating the results of actions performed on the data fabric back to the system layer by querying and transforming data according to semantic descriptions of both structure and data types of the consuming interfaces.

The data fabric created by the semantic support points is non-persistent, generated on demand by actors wishing to consuming the semantic data representation on request. Data is encoded on the fly from data emitted by the system layer and results are immediately propagated back to the system layer to be consumed by recipients. Transformation rules are stored and managed seperately form actual semantic support point implementations, making our approach highly flexible and domain independent.

In contrast to semantic interoperability approaches common in the dataspace or semantic web of things world, not only the participants and entire data objects are semantically described in terms of static meta data, but each data value itself can be embedded into the semantic fabric at runtime.

As a proof of concept we demonstrated on-demand data aggregation and using a multi agent system capable of working on semantic linked data to derive instruction suggestions for workers in an automotive factory. Based on data like product descriptions, 3D models, pre-defined processes and constraints we generated a unified semantic view enriched with expert knowledge to derive optimized production plans that could not only be immediately consumed and virtually be tested by a semantic multi agent system, but also rendered to a human readable representation for actual factory workers.

For this proof of concept, we decided to follow a plug in like approach, extending the agent system itself with the capabilities to generate a semantic view on the world and propagate changes made by the agents back to the system layer. However, extracting the actual support point functionality to fully cover the concept presented in section 3 would only require minor code changes and hosting of the support points as separate services. In our case it was more efficient to realize a plugin approach, as the MAS was the only actor working on the semantic layer. We consider this as another proof of flexibility of the approach when it comes to actual implementation details.

We are highly interested in exploring the possibilities provided by mapping files that are already written in RDF and with that contain semantic information and expert knowledge when it comes to service discovery and composition. In addition with, possibly distributed, service registries containing meta data based information about available services, these semantic instructions given in the mapping files, seem like a promising way to generate another semantic meta data layer on top of the actual linked data fabric, allowing for knowledge based service discovery and composition, strengthening the idea of an intention driven data exchange even more, moving towards a system providing a certain level of *pragmatic interoperability* (Neiva et al., 2016)

Further work is necessary in terms of runtime behavior and resources needed for time or resource critical applications. To our anecdotal findings, the mapping instructions provided are the main impact on resources needed in terms of triple stores while resources needed to handle the semantic representations generated at runtime are proportional to the size of the non-semantic data payloads emitted and consumed on system level. As this dynamic data is discarded as soon as a certain process ends, the data fabric itself only creates negligible overhead.

However, for our use cases these considerations were not relevant at this point and will be subject to future work.

ACKNOWLEDGEMENTS

This work has been supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) in the projects *ForeSightNEXT* (01MK23001G) and *TwinMap* (13IK028J).

REFERENCES

Alberternst, S., Anisimov, A., Antakli, A., Duppe, B., Hoffmann, H., Meiser, M., Muaz, M., Spieldenner, D., and Zinnikus, I. (2021). From things into clouds–and back. In *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, pages 668–675. IEEE.

Allocca, C. and Gougousis, A. (2015). A preliminary investigation of reversing rml: From an rdf dataset to its column-based data source. *Biodiversity data journal*, (3).

Antakli, A., Kazimov, A., Spieldenner, D., Rojas, G. E. J., Zinnikus, I., and Klusch, M. (2023). Ajan: An engineering framework for semantic web-enabled agents and multi-agent systems. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 15–27. Springer.

Antakli, A., Spieldenner, T., Rubinstein, D., Spieldenner, D., Herrmann, E., Sprenger, J., and Zinnikus, I. (2019). Agent-based web supported simulation of human-robot collaboration. *In Proc. of the 15th Int. Conf. on Web Information Systems and Technologies (WEBIST)*, pages 88–99.

Antakli, A., Vozniak, I., Lipp, N., Klusch, M., and Müller, C. (2021). Hail: Modular agent-based pedestrian imitation learning. In *Proc. 19th Int. Conf. on Practical Applications of Agents and Multi-Agent Systems (PAAMS)*. Springer.

Antoniazzi, F. and Viola, F. (2019). Building the semantic web of things through a dynamic ontology. *IEEE Internet of Things Journal*, 6(6):10560–10579.

Breitman, K. K., Casanova, M. A., and Truszkowski, W. (2007). Ontology in computer science. *Semantic Web: Concepts, Technologies and Applications*, pages 17–34.

Charpenay, V., Käbisch, S., and Kosch, H. (2016). Introducing thing descriptions and interactions: An ontology for the web of things. In *SR+ SWIT@ ISWC*, pages 55–66.

Charpenay, V., Käbisch, S., and Kosch, H. (2018). Semantic data integration on the web of things. In *Proceedings of the 8th International Conference on the Internet of Things*, pages 1–8.

Chung, E., Lefrançois, M., and Boissier, O. (2020). Increasing interoperability in the web of things using autonomous agents (position paper). In *18e Rencontres des Jeunes Chercheurs en Intelligence Artificielle 2020 (RJCIA)*.

Ciortea, A., Mayer, S., and Michahelles, F. (2018). Repurposing manufacturing lines on the fly with multi-agent systems for the web of things. In *AAMAS*, pages 813–822.

Curry, E. (2020). *Real-time linked dataspaces: Enabling data ecosystems for intelligent systems*. Springer Nature.

Dikenelli, O., Erdur, R. C., and Gumus, O. (2005). Seagent: a platform for developing semantic web based multi agent systems. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1271–1272.

Dimou, A., Vander Sande, M., Slepicka, J., Szekely, P., Mannens, E., Knoblock, C., and Van de Walle, R. (2014). Mapping hierarchical sources into rdf using the rml mapping language. In *2014 IEEE International Conference on Semantic Computing*, pages 151–158. IEEE.

Dunne, R., Morris, T., and Harper, S. (2021). A survey of ambient intelligence. *ACM Computing Surveys (CSUR)*, 54(4):1–27.

Franklin, M., Halevy, A., and Maier, D. (2005). From databases to dataspaces: a new abstraction for information management. *ACM Sigmod Record*, 34(4):27–33.

Grassi, M., Scrocca, M., Carenini, A., Comerio, M., and Celino, I. (2023). Composable semantic data transformation pipelines with chimera.

Halevy, A., Franklin, M., and Maier, D. (2006). Principles of dataspace systems. In *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 1–9.

Heiler, S. (1995). Semantic interoperability. *ACM Computing Surveys (CSUR)*, 27(2):271–273.

Hildebrand, M., Tourkogiorgis, I., Psarommatis, F., Arena, D., and Kiritsis, D. (2019). A method for converting current data to rdf in the era of industry 4.0. In *Advances in Production Management Systems. Production Management for the Factory of the Future: IFIP WG 5.7 International Conference, APMS 2019, Austin, TX, USA, September 1–5, 2019, Proceedings, Part I*, pages 307–314. Springer.

Jafari, M., Kavousi-Fard, A., Chen, T., and Karimi, M. (2023). A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future. *IEEE Access*, 11:17471–17484.

Khan, H. H., Malik, M. N., Zafar, R., Goni, F. A., Chofreh, A. G., Klemeš, J. J., and Alotaibi, Y. (2020). Challenges for sustainable smart city development: A conceptual framework. *Sustainable Development*, 28(5):1507–1518.

Khanna, A. and Kaur, S. (2020). Internet of things (iot), applications and challenges: a comprehensive review. *Wireless Personal Communications*, 114:1687–1762.

Lan, Q., Wen, D., Zhang, Z., Zeng, Q., Chen, X., Popovski, P., and Huang, K. (2021). What is semantic communication? a view on conveying meaning in the era of machine intelligence. *Journal of Communications and Information Networks*, 6(4):336–371.

Lassila, O., Hendler, J., and Berners-Lee, T. (2001). The semantic web. *Scientific American*, 284(5):34–43.

Lee, I. and Lee, K. (2015). The internet of things (iot): Applications, investments, and challenges for enterprises. *Business horizons*, 58(4):431–440.

Lu, Y. and Asghar, M. R. (2020). Semantic communications between distributed cyber-physical systems towards collaborative automation for smart manufacturing. *Journal of manufacturing systems*, 55:348–359.

Manola, F., Miller, E., McBride, B., et al. (2004). Rdf primer. *W3C recommendation*, 10(1-107):6.

McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(10):2004.

Méndez, S. J. R., Haller, A., Omran, P. G., Wright, J., and Taylor, K. (2020). J2rm: An ontology-based json-to-rdf mapping tool. In *ISWC (Demos/Industry)*, pages 368–373.

Neiva, F. W., David, J. M. N., Braga, R., and Campos, F. (2016). Towards pragmatic interoperability to support collaboration: A systematic review and mapping of the literature. *Information and Software Technology*, 72:137–150.

Ngankam, H. K., Pigot, H., and Giroux, S. (2022). Ontodomus: a semantic model for ambient assisted living system based on smart homes. *Electronics*, 11(7):1143.

Nižetić, S., Šolić, P., Gonzalez-De, D. L.-d.-I., Patrono, L., et al. (2020). Internet of things (iot): Opportunities, issues and challenges towards a smart and sustainable future. *Journal of cleaner production*, 274:122877.

Padilla-Cuevas, J., Reyes-Ortiz, J. A., and Bravo, M. (2021). Ontology-based context event representation, reasoning, and enhancing in academic environments. *Future Internet*, 13(6):151.

Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., and Vrgoč, D. (2016). Foundations of json schema. In *Proceedings of the 25th international conference on World Wide Web*, pages 263–273.

Qin, Z., Tao, X., Lu, J., Tong, W., and Li, G. Y. (2021). Semantic communications: Principles and challenges. *arXiv preprint arXiv:2201.01389*.

Razzak, F. et al. (2013). The role of semantic web technologies in smart environments.

Ruta, M., Scioscia, F., and Di Sciascio, E. (2012). Enabling the semantic web of things: framework and architecture. In *2012 IEEE Sixth International Conference on Semantic Computing*, pages 345–347. IEEE.

Sabbatini, F., Ciatto, G., and Omicini, A. (2022). Semantic web-based interoperability for intelligent agents with psyke. In *Explainable and Transparent AI and Multi-Agent Systems: 4th Inter. Workshop (EXTRAAMAS 2022)*, pages 124–142. Springer.

Sahoo, S. S., Halb, W., Hellmann, S., Idehen, K., Thibodeau Jr, T., Auer, S., Sequeda, J., and Ezzat, A. (2009). A survey of current approaches for mapping of relational databases to rdf. *W3C RDB2RDF Incubator Group Report*, 1:113–130.

Santofimia, M. J., Fahlman, S. E., Del Toro, X., Moya, F., and Lopez, J. C. (2011). A semantic model for actions and events in ambient intelligence. *Engineering Applications of Artificial Intelligence*, 24(8):1432–1445.

Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., and Sánchez, L. (2022). Enabling data spaces: Existing developments and challenges. In *Proceedings of the 1st International Workshop on Data Economy*, pages 42–48.

Spieldenner, D. (2022). Poser: A semantic payload lowering service. In *WEBIST*, pages 249–256.

Spieldenner, T. (2023). Linked data as medium for distributed multi-agent systems.

Spieldenner, T. and Antakli, A. (2022). Behavior trees as executable representation of milner calculus notations. In *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE.

Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., and Lindström, N. (2020). Json-ld 1.1. *W3C Recommendation, Jul*.

Vdovjak, R. and Houben, G.-J. (2002). Providing the semantic layer for wis design. In *Advanced Information Systems Engineering: 14th International Conference, CAiSE 2002 Toronto, Canada, May 27–31, 2002 Proceedings 14*, pages 584–599. Springer.

Vladimir, L. (2008). What is answer set programming. In *AAAI*, volume 8.

Wang, J., Yang, Y., Wang, T., Sherratt, R. S., and Zhang, J. (2020). Big data service architecture: a survey. *Journal of Internet Technology*, 21(2):393–405.

Weber, W., Rabaey, J., and Aarts, E. H. (2005). *Ambient intelligence*. Springer Science & Business Media.

Zeng, D., Guo, S., and Cheng, Z. (2011). The web of things: A survey. *J. Commun.*, 6(6):424–438.

Zillner, S., Gomez, J. A., García Robles, A., Hahn, T., Le Bars, L., Petkovic, M., and Curry, E. (2021). Data economy 2.0: From big data value to ai value and a european data space. In *The elements of big data value: Foundations of the research and innovation ecosystem*, pages 379–399. Springer International Publishing Cham.

Values and Enablers of Lessons Learned Practices: Investigating Construction Industry Context

Jeffrey Boon Hui Yap[a]
Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman (UTAR), Kajang, Selangor, Malaysia
bhyap@utar.edu.my

Keywords: Knowledge Management, Lessons Learned, Values, Enablers, Construction Industry, Project Management.

Abstract: In the realm of construction project management, the value of "lessons learned (LL)" cannot be overstated. LL, as an important approach for effective project management and continuous improvement, is analysed in this study, with the aim to advance the impact of LL by determining the values of LL practices and examining the enablers that positively influence LL practices in the construction industry. A detailed literature review has revealed nine (9) values and seven (7) enablers of LL practices relevant to the construction industry's context. Using a questionnaire survey involving 129 Malaysian construction professionals selected based on non-probability techniques, the significance of the values and enablers is prioritised based on mean scores. Findings reveal that LL practices help to avoid making similar past mistakes, optimize project performance and engender collaborative learning in the project team. Individual-related enablers are perceived to be more influential than organisational-related enablers in implementing LL in construction projects. Collective and conscious efforts in fostering a learning culture are crucial to encourage the construction industry to embrace LL practices and help individuals and organisations thrive.

1 INTRODUCTION

The construction industry acts as a catalyst for economic growth in a developing country such as Malaysia - increasing the country's income, work opportunities, and infrastructure. However, the industry is under ever-increasing pressure to deliver projects faster, with better quality and with lower costs. Good management practices are crucial in achieving these demands. As Disterer (2002, p. 519) advocates, "success of projects depends heavily on the right combination of knowledge and experience". Correspondingly, Meredith et al. (2017, p.302) accentuate, "past knowledge...should be built into estimates of future project performance". In advocating knowledge representation for efficient re-use of project memory, (Bekhti et al., 2011) underscore the need for designers to learn from past project experiences to deal with new design problems. Construction companies are project-based organisations since much of their knowledge is generated on-site, from projects they carry out. As such, the development of knowledge and expertise from project learning practices is critical in construction.

Knowledge is critical for construction companies to succeed and maximization of value through enhancing competencies, confidence, effectiveness, competitiveness, and sustainability. Knowledge management (KM) processes can prevent re-invention of the wheel, facilitate innovation; and lead to increased agility, efficiency, flexibility, quality, learning, better decision making, better teamwork and supply chain integration, improved project performance, higher client satisfaction, and organisational growth (Eken et al., 2015; KPMG Consulting, 2000; Yap & Lock, 2017). A recent Malaysian study in the construction industry reveals the most important benefits of KM are to improve quality, enhance decision-making, raise quality, circumvent the repeat of past mistakes and enable knowledge exchange (Yap et al., 2022). Likewise in Portugal, the practitioners acknowledged the most significant aspects of KM in the management of construction projects are associated with the exchange of experiences between project team

[a] https://orcid.org/0000-0003-4332-0031

Yap, J.
Values and Enablers of Lessons Learned Practices: Investigating Construction Industry Context.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 38-47
ISBN: 978-989-758-716-0; ISSN: 2184-3228

members, the sharing of information among stakeholders and continuous improvements (Marinho & Couto, 2022).KM practices can positively enhance the effectiveness, efficiency and efficacy of project personnel. The construction industry is project-based but very much knowledge-intensive. Multi-disciplinary teams (i.e. architect, engineer, quantity surveyor and contractor) are involved and project delivery relies heavily on previous experience/heuristics. Thus, lessons learned (LL) in construction projects should be captured and reused in future projects.

Effective management of LL is vital for the generation of project knowledge and supports continuous learning in project-based industries such as construction. In this vein, decision-making processes are further enhanced by gaining insights from the "know-what", "know-how" and "know-why". The value-addedness of learning is directly linked to project performance. This being the case, it is necessary to determine how LL can add value to construction project delivery and examine the enablers that positively influence LL practices in the construction industry. The research questions in the present study are:

Q1: Why do we need to capture LL in construction projects?

Q2: What are the enablers that positively influence LL practices in construction?

2 LESSONS LEARNED (LL) PRACTICES AND THE CONSTRUCTION INDUSTRY

LL is a critical variable for success (Kerzner, 2017) - providing a platform for reflection, growth, and development by extracting knowledge from past experiences. The four dimensions of LL are: When? What about? How know? and What is included? In the construction industry's context, LL is the intellectual asset used to create value based on previous projects and contribute to the organisation's learning agenda (Carrillo et al., 2013). Likewise, the Project Management Institute's PMBOK Guide (Project Management Institute, 2017) underscores the need for using existing knowledge and creating new knowledge to achieve the project's objectives and contribute to organisational learning. Considering this, the positive and negative aspects of projects are needed to learn from past experiences, particularly in avoiding the repetition of costly mistakes that can jeopardise project performance and damage a company's reputation as well as increasing the likelihood of attaining success that proved to be effective and profitable. Thus, LL is very beneficial for similar future work and improves the company's competitiveness, such as improved decision-making, problem-solving and innovation. LL is particularly vital for improving future performance (Love et al., 2016) and for organisations to realise a competitive edge if used properly (Hlupic et al., 2002).

Extensive knowledge is generated throughout the construction project delivery from start to finish. Most professionals acquire knowledge mostly through meetings with more experienced personnel as well as lessons learned from completed projects (Marinho & Couto, 2022). Knowledge gained from past projects can be leveraged to improve the capability and productivity of construction companies (Dave & Koskela, 2009). For example, knowledge reuse can significantly contribute to better expert judgment and improved time-cost performance (Yap & Skitmore, 2020). In a Spanish study, Forcada et al. (2013) observed the top KM benefits being: employee experience exchange, group work improvement and efficiency improvement. They further explained that effective management of project knowledge is vital in enhancing continuous improvements from LL. For example, the project team can better excel in project management via sharing LL and advanced practices, which can be transferred within and between projects (Terzieva, 2014). However, knowledge dissemination remains a challenge and the value of LL has yet to be fully capitalised (Debs & Hubbard, 2023).

To develop the competency of project personnel, Yap & Shavarebi (2022) proposed sharing past project experiences which lead to the expansion of cognitive ability, expert judgement and better-informed decision-making; ultimately resulting in better project results. Tacit knowledge is developed from experience and is hard to formalise but it is considered to be more important than explicit knowledge (Forcada et al., 2013; Teerajetgul & Chareonngam, 2008). Tacit knowledge can be captured by talking to experts and reflecting on the LL from others. For example, using storytelling learning to communicate LL (Duffield & Whitty, 2016). However, some construction companies fail to recognise the value of LL and perceive LL to be project-specific (Carrillo et al., 2013). Some construction professionals, on the other hand, do not want to share their problems or are not willing to learn from other people's mistakes (Carrillo et al., 2013). Knowledge sharing behaviour among construction project members are influenced by two driving

Table 1: Summary of enablers of LL practices.

Ref	Enablers	Authors																Total
		(Kululanga & Mccaffer, 2001)	(Levin & Cross, 2002)	(Tsai, 2002)	(MacNeil, 2003)	(Carrillo et al., 2004)	(Van Den Hooff & Ridder, 2004)	(Rego et al., 2009)	(Theriou et al., 2011)	(Javernick-Will, 2012)	(Tan et al., 2012)	(Carrillo et al., 2013)	(Duffield & Whitty, 2016)	(Longwe et al., 2015)	(Dang & Le-Hoai, 2019)	(Dang et al., 2019)	(Yang et al., 2019)	
Individual																		
B1	Sharing culture			√	√			√	√	√				√				6
B2	Honouring of commitment	√			√					√								3
B3	Peer recognition				√	√				√	√							4
B4	Reciprocity and trust		√					√		√						√		4
Organisational																		
B5	Perceived value						√			√		√						3
B6	Financial/ social motivation	√			√	√				√	√					√		6
B7	Workplace culture	√						√	√		√		√		√		√	7
B5	Perceived value						√			√		√						3

modes, namely trust-driven and incentive-driven (Cheng & Yin, 2024). According to the Construction Industry Institute (CII) (2012), best practice is "a process or method that, when executed effectively, leads to enhanced project performance". In the construction project management context, best practices or rather proven practices can be defined as something that works well on a repetitive basis that leads to a competitive advantage (Kerzner, 2017). Some of the learning in projects can evolve into best practices that can be standardized.

Table 1 presents a list of the most frequently cited enablers of LL practices from previous literature. There enablers are divided into individual- and organisational-related.

3 RESEARCH METHODOLOGY

A positivist paradigm employing the deductive approach is adopted to objectively examine the practice of capturing LL in the construction industry. A quantitative research design with a cross-sectional field survey was employed, as it provides an efficient and economical means to gather feedback from a large number of professionals currently working in the construction industry for statistical analyses. The methodological flowchart for the study is presented in Figure 1.

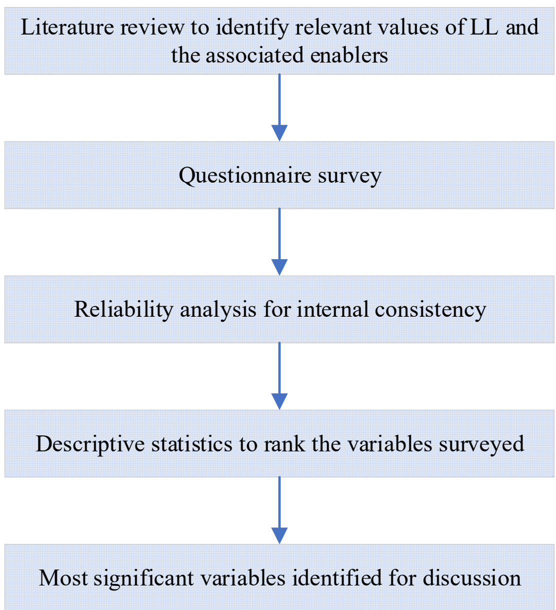


Figure 1: Methodological flowchart for the study.

The Statistical Package for Social Sciences (SPSS) version 23 was used to analyse the data collected. The analyses were done to prioritise the value of LL and the associated enablers/inhibitors according to their descriptive statistics (mean scores and standard deviations).

3.1 Questionnaire Design

The questionnaire was designed based on the literature review and consultation with industry subject matter experts. The questions were drafted clearly and concisely to create easy-to-understand materials and limited to a 15-minute completion time to prevent survey fatigue. The questionnaire contains three parts. Part I deals with the respondents' demographic information, in terms of their educational background, years of industry experience and the type of projects involved. Part II contains the question; Do you agree with the following value of lessons learned in construction projects? on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Part III provided a list of enablers identified through the detailed literature review (Table 1). For each enabler, the respondents were requested to indicate their level of agreement on a similar five-point Likert scale as in Part II.

3.2 Survey Respondents and Demographics

The sampling frame consisted of professionals from the three key parties in construction, namely clients, consultants and contractors in Malaysia. Non-probability techniques of purposive and convenience with snowball sampling are used to select respondents to yield reasonable responses. In this study, the unit of analysis is construction professionals, as they are the actors directly involved in project delivery. The reason for engaging a variety of professions (i.e. clients, consultants and contractors) was to ensure different perspectives pertaining to LL practices in construction are represented.

The questionnaire pilot involved 30 targeted construction professionals to ensure clarity and unambiguity. Following a successful pilot test, the questionnaire remained unaltered for the main survey whereby another 170 questionnaires were electronically distributed. Overall, 129 valid responses were collected after follow-up reminders, attaining a response rate of 64.50%. The sample size (> 100) is adequate for meaningful statistical analyses (Roscoe, 1975; Yap & Skitmore, 2018). Additionally, the Yamane sampling approach led to the determination of 100 samples at a 90% confidence level for a population size over 100,000 (Israel, 1992; Yap et al., 2022).

Table 3 indicates the demographic profile of the respondents, with 90 questionnaires (70%) from respondents with at least a bachelor's degree. Nearly 50% had more than 10 years of working experience in construction. 57.4% of respondents are involved in building projects. These are considered sufficient to obtain sound judgment from qualified respondents for this perception-based study.

Table 2: Demographic profile of respondents.

Profile	Description	Frequency	Percentage (%)
Academic qualification	Master's degree	33	25.6
	Bachelor's degree	57	44.2
	Diploma	37	28.7
	Certificate	2	1.6
Working experience	0 to 5 years	41	31.8
	6 to 10 years	26	20.2
	11 to 15 years	23	17.8
	16 years and above	39	30.2
Type of project	Building	74	57.4
	Infrastructure	55	42.6

4 RESULTS AND DISCUSSIONS

4.1 Questionnaire Reliability

Table 3 summarises the α values for the two categories of variables, viz. values and enablers of LL, which is greater than which is higher than the threshold of 0.70 needed to establish the internal reliability of the scale used (Yap, Lim, et al., 2021). This denotes that good overall reliability was obtained on the research instrument used.

Table 3: Measurement of internal consistency.

Category	Number of items	Cronbach's alpha, α
Values of LL	9	0.867
Enablers of LL	7	0.759

4.2 Mean Scores and Ranking LL Values

Table 4 presents the mean scores and standard deviations (SDs) of each value surveyed. A close examination of Table 5 reveals that all 8 values have mean scores higher than 4.0, which is regarded as very significant in the rating scale. This implied that the majority of the respondents either agreed or strongly agreed with the evaluated values. The five most significant values of capturing LL in construction projects are:

1. A2: Avoiding the same mistakes from happening in upcoming projects (mean = 4.519, SD = 0.574);
1. A5: Better performance or procedure by adopting lessons learned from other projects (mean = 4.519, SD = 0.574);
3. A9: Promote a collective environment to attain the project team's shared goals through the sharing of personal experiences (mean = 4.519, SD = 0.651);
4. A1: Ensuring good practices in previous projects that are successful are being re-used in upcoming projects (mean = 4512, SD = 0.626); and
5. A3: Developing new ideas or methods through lessons learned (mean = 4.496, SD = 0.697).

The data indicates that there is a consistent emphasis on the value of integrating lessons learned from previous projects across several dimensions, with very similar mean scores suggesting a high level of agreement among respondents. The most valuable aspect of capturing LL for construction projects is to avoid the recurrence of similar mistakes. The interview participants from Yap & Skitmore's (2020) study specifically emphasized that "past experiences will tell you what you can do and enrich one's expert judgment" and "individual needs to learn from his/her mistakes and not repeat the same mistake twice". Given that project mistakes are the major contributing factor to rework and time-cost overruns, capturing and sharing critical LL can help construction professionals avoid repeating the same mistakes and reinventing the wheel in future projects. The other highly perceived importance of LL is to enhance productivity, efficiency and smarter working. LL is needed to build absorptive capacity and drive towards performance improvement in the construction industry (Love et al., 2016).

Third, LL practices are a collaborative technique to encourage project team members to share their personal experiences, which will then contribute to a collective environment in attaining shared goals. Sharing knowledge between team members is crucial to achieving organisational learning and collective competence (Yap, Shavarebi and Skitmore, 2021). It is worth noting that trust and collaboration are significant knowledge factors for construction projects (Teerajetgul & Charoenngam, 2006). The fourth value of LL is related to the reuse of some best practices from other successful projects. LL is handy project knowledge that can be reused and employed as best practices to increase the likelihood of repeating project delivery success (Yap & Shavarebi, 2022). The fifth value is making LL the base to foster innovation and developing new ideas/methods/solutions from long 'trial and error' ending with successes and failures in the construction projects. According to Kolb & Kolb (2009), people learn best in situations such as brainstorming sessions that call for the generation of ideas. The recent developments in information and communications technology (ICT) tools have further advanced the way people share knowledge and ideas – for improvement and innovation (Carrillo, 2005; Yap et al., 2022).

Table 4: Ranking the values of LL.

The values of capturing LL in the construction industry context	Overall (N=129)		
	Mean	SD	Rank
A2: Avoiding the same mistakes from happening in upcoming projects.	4.519	0.574	1
A5: Better performance or procedure by adopting lessons learned from other projects.	4.519	0.574	1
A9: Promote a collective environment to attain the project team's shared goals through the sharing of personal experiences.	4.519	0.651	3
A1: Ensuring good practices in previous projects that are successful are being re-used in upcoming projects.	4.512	0.626	4
A3: Developing new ideas or methods through lessons learned.	4.496	0.697	5
A4: Transforming individual knowledge to organisational knowledge by sharing lessons learned.	4.481	0.663	6
A6: Facilitate project planning (forecasting ability) using lessons learned from previous projects.	4.450	0.637	7
A7: Improvise project monitoring and control processes using lessons learned from previous projects.	4.326	0.709	8
A8: The quality and quantity of lessons learned in the construction industry are influenced by the size and difficulty of the project.	4.326	0.752	9

4.3 Mean Scores and Ranking LL Enablers

Table 5 presents the enablers of LL in the construction industry according to their significance. All the enablers have a mean value above 4.00 and are therefore considered relevant and very significant. The topmost five enablers are:

1. B3: Peer recognition (mean = 4.450, SD = 0.661);
2. B1: Sharing culture (mean = 4.434, SD = 0.705);
3. B2: Honouring of commitment (mean = 4.411, SD = 0.645);
4. B4: Reciprocity and trust (mean = 4.403, SD = 0.724); and
5. B7: Workplace culture (mean = 4.364, SD = 0.706);

Four of the five enablers are related to individual aspects.

Table 5: Ranking of enablers.

Enablers	Overall (N=129)		
	Mean	SD	Rank
B3	4.450	0.661	1
B1	4.434	0.705	2
B2	4.411	0.645	3
B4	4.403	0.724	4
B7	4.364	0.706	5
B6	4.333	0.654	6
B5	4.248	0.729	7

4.3.1 Peer Recognition (Individual)

A construction project team involve various experts from different skills, knowledge, experience and professional background. All the parties work as a team to complete a project, although there is a hierarchical structure. Every stakeholder is allowed to share their perception or knowledge while carrying out a knowledge-sharing (KS) session. People like the feeling of being recognised and thankful when they share their knowledge, and information and contribute to the project team, especially agreement from seniors (MacNeil, 2003). Some people just need a "thank you" to get affirmation from colleagues, which in turn, helps to improve the workplace culture (Javernick-Will, 2012).

In addition, peer recognition from colleagues, employees or seniors encourages a person to be more self-confident and willing to share their knowledge with others (Rahman et al., 2018). It also encourages self-development as well as engenders innovative and new knowledge or ideas because they have self-confidence and allows them to feel and look like an expert. (Carrillo et al., 2004) believe that peer recognition is more significant than financial incentives because it only provides tiny opportunities for success. According to Tan et al. (2012), peer recognition also assists others in finding the solutions to problems, as a result of self-confidence in sharing knowledge with others.

4.3.2 Sharing Culture (Individual)

When individuals interact with each other in a team, it creates a learning environment and sharing culture in the organisation that brings benefits to the organisation (Longwe et al., 2015). People are actually learning by sharing tacit knowledge or their own experience with others and hence become explicit knowledge (Rego et al., 2009). Nonetheless, knowledge gained from LL is difficult to transform from tacit to explicit knowledge and be shared with others in a team. Communication is key to sharing knowledge. For example, breakfast or lunch gatherings are useful platforms for exchanging previous experiences (Fong, 2005). However, if an individual is capable of gathering, recreating, utilising and sharing knowledge, will bring advantages to an organisation (MacNeil, 2003). Moreover, knowledge sharing (KS) with competitors by an individual is a type of coopetition. Coopetition creates common interests between individuals and competitors. The knowledge gained from competitors allows individuals to benefit themselves and also benefits an organisation. In this circumstance, an organisation allows the development of new ideas, skills, information, knowledge and technology from others (Tsai, 2002).

People who contribute and share the tacit knowledge that is stored in their brains and minds create a sharing and learning environment (Chugh et al., 2015; Yap et al., 2022). It encourages other members of the organisation to share their knowledge because everyone knows that "knowledge is power" (Theriou et al., 2011). A workplace culture that encourages knowledge sharing and learning allows individuals to improve, which in turn, improves productivity and increases the competitive advantage of an organisation (Javernick-Will, 2012).

4.3.3 Honouring of Commitment (Individual)

A construction project involves a lot of professionals from different backgrounds/departments such as architecture, engineering, cost consulting and project management. During the management and delivery of construction projects, the project team members want to appear consistent with the project objectives and have made their intentions to share their knowledge explicit – they will want to live up to these intentions and honour their commitment (Leal et al., 2017). Once team members are involved in a problem or issue, they would like to remain involved in it to give advice, information, knowledge or solutions until the problem or issue is eventually solved (Javernick-Will, 2012). This is because people like to show self-worth and be respected by others. Another way to explain is that people want to be compatible with others. After their purpose of sharing knowledge is made clear, the individuals want to stay up with these promises and respect their pledge or even to be a leader. In investigating knowledge exchange behaviours among virtual communities in China, Luo et al. (2021)observed that affective and normative commitment can significantly influence the knowledge contributors' sharing intention.

Leaders play an important role in an organisation, as a leader can inspire the team members to commit to the project (Kululanga & Mccaffer, 2001). An individual who wants to build a group should draft a sanction and attend a series of meetings on preparedness judgement or evaluation, to show that they are well-connected, leadership and management support. People ensure that they keep up to date and remain active in the society. All of the above is to ensure leaders of the teams or organisation merit their commitment and ensure they perform their own best. (MacNeil, 2003b).

4.3.4 Reciprocity and Trust (Individual)

The environment and relationships within a group of people are very important, as they also influence the success or failure of a project. To facilitate LL practices in the construction industry, people must learn to reciprocate (Dang et al., 2019). Some people are more willing to share knowledge with those people who helped and supported them before when those people faced some issues or problems. People will think that it is the way to pay back as they helped them before. It is a mutual benefit relationship (Javernick-Will, 2012). This can be understood by the adage that "people treat you like the way you treat them". It is a two-way relationship.

The norm of reciprocity also indirectly creates trust relationships among people. People are also more willing to share knowledge when trust exists. Trust is also a two-way relationship same as reciprocity, to tighten the relationship within a team (Rego et al., 2009). Thus, knowledge exchange is better and faster if people in the organisation trust each other. When trust exists, people provide and share useful knowledge willingly. Therefore, people are also likely to hear, consume and learn the knowledge shared by other people (Levin & Cross, 2002). It reduces conflicts between the people in the organisation by the existence of reciprocity and trust.

4.3.5 Workplace Culture (Organisational)

In a successful KM system, organisational culture is the most crucial facilitating factor. An organisation should share their vision and mission with all the employees or team members (Yang et al., 2019). "Work as a team is better than one", because teamwork increases collaboration and allows brainstorming to develop or create more ideas and thus improve productivity (Theriou et al., 2011). When every party have the same vision and the same target as the organisation, they are more likely to contribute and complete the project efficiently and effectively (Kululanga & Mccaffer, 2001). The culture of the workplace highly affects a person's behaviour and attitude, therefore affecting the performance of an organisation (Rego et al., 2009b). A person who works in a positive workplace culture will be influenced by the environment of the organisation and participate in any activities actively. When working in a negative workplace culture, the person will have the same feelings and will not want to contribute to the organisation (Tan et al., 2012). For example, a student would perform better in a good class, because they are studying under positive influence, although the student does not have a good basic.

Furthermore, practitioner shares their visions, committed leadership and reward creativity and innovation depending on the culture of the workplace (Dang et al., 2019). Therefore, a workplace culture influences the success of LL practices and also affects the success of an organisation (Duffield & Whitty, 2016).

5 CONCLUDING REMARKS

From a detailed literature review nine (9) values and seven (7) enablers of LL practices in the construction industry were identified. The opinions of construction professionals currently working in Malaysia were obtained through a cross-sectional self-administered questionnaire survey. The underlying aim of ranking the values and enablers is towards recognizing and embracing the importance of LL practices in the complex construction environment to increase the chances of project success as well as cultivate a culture of learning and improvement that can benefit construction organisations in all aspects of their operations. Findings reveal that avoiding the same mistakes from happening in upcoming projects, better performance or procedure by adopting lessons learned from other projects and promoting a collective environment to attain the project team's shared goals through the sharing of personal experiences are the leading values of performing LL. Construction organisations that prioritise LL practices not only can take advantage of lessons from previous successes and failures but also enhance project outcomes with improved ability to plan, schedule and estimate their future projects. The most influential enablers are peer recognition, sharing culture and honouring of commitment. Collective and conscious efforts in fostering a learning culture are crucial to encourage the construction industry to embrace LL practices – help individuals and organisations thrive.

While the study makes several contributions to LL practices in construction project management, it is limited by the single data collection method using field survey possibly causing mono-method bias. Nevertheless, this is substantiated by triangulating the findings by cross-referencing the research literature for theoretical validation. Although the use of a self-completion questionnaire form is widely used to gather quantitative data from a large and diverse sample for statistical analyses, it does not allow researchers to probe or clarify participants' responses. An interpretative approach using in-depth interviews and/or case studies could be further employed to collect rich real-world project experiences from construction professionals, as well as to validate the statistical results. The rating of the values and enablers of LL practices on a five-point Likert scale may not be completely reliable as different respondents may perceive the scale differently when they attach their interpretation of the different scale points. It is worth noting that the Likert scale is commonly used to measure people's opinions, perceptions and attitudes in behavioural sciences and construction project management studies. Further studies would benefit by investigating some of the formal and informal best practices for capturing LL at various phases of construction project delivery, particularly on how emerging digital technologies have revolutionized KM practices in the construction context.

REFERENCES

Bekhti, S., Matta, N., & Djaiz, C. (2011). Knowledge representation for an efficient re-use of project memory. *Applied Computing and Informatics*, *9*(2), 119–135. https://doi.org/10.1016/j.aci.2011.05.004

Carrillo, P. (2005). Lessons learned practices in the engineering, procurement and construction sector. *Engineering, Construction and Architectural Management*, *12*(3), 236–250.

Carrillo, P., Robinson, H., Al-Ghassani, A., & Anumba, C. (2004). Knowledge management in UK construction: Strategies, resources and barriers. *Project Management Journal*, *35*(2), 46–56.

Carrillo, P., Ruikar, K., & Fuller, P. (2013). When will we learn? Improving lessons learned practice in construction. *International Journal of Project Management*, *31*(4), 567–578.

Cheng, F., & Yin, Y. (2024). Organizational antecedents and multiple paths of knowledge-sharing behavior of construction project members: evidence from Chinese construction enterprises. *Engineering, Construction and Architectural Management*, *31*(3), 957–975. https://doi.org/10.1108/ECAM-07-2022-0614

Chugh, R., Wibowo, S., & Grandhi, S. (2015). Mandating the transfer of tacit knowledge in Australian universities. *Journal of Organizational Knowledge Management*, *2015*, 1–10. https://doi.org/10.5171/2015

Construction Industry Institute. (2012). CII best practices guide: Improving project performance. In *Construction Industry Institute*.

Dang, C. N., & Le-Hoai, L. (2019). Relating knowledge creation factors to construction organizations' effectiveness: Empirical study. *Journal of Engineering, Design and Technology*, *17*(3), 515–536.

Dang, C. N., Le-Hoai, L., & Peansupap, V. (2019). Linking knowledge enabling factors to organizational performance: Empirical study of project-based firms. *International Journal of Construction Management*. https://doi.org/10.1080/15623599.2019.1637097

Dave, B., & Koskela, L. (2009). Collaborative knowledge management: A construction case study. *Automation in Construction*, *18*(7), 894–902.

Debs, L., & Hubbard, B. (2023). Gathering and disseminating lessons learned in construction companies to support knowledge management. *Construction Economics and Building*, *23*(1/2), 56–76.

Disterer, G. (2002). Management of project knowledge and experiences. *Journal of Knowledge Management*, *6*(5), 512–520.

Duffield, S. M., & Whitty, S. J. (2016). Developing a systemic lessons learned knowledge model for organisational learning through projects. *International Journal of Project Management*, *33*(2), 1280–1293.

Duffield, S., & Whitty, S. J. (2016). How to apply the systemic lessons learned knowledge model to wire an organisation for the capability of storytelling. *International Journal of Project Management*, *34*(3), 429–443.

Eken, G., Bilgin, G., Dikmen, I., & Birgonul, M. T. (2015). A lessons learned database structure for construction companies. *Procedia Engineering*, *123*, 135–144.

Fong, P. S. (2005). Building a knowledge sharing culture in construction project teams. In C. J. Anumba, C. Egbu, & P. M. Carrillo (Eds.), *Knowledge Management in Construction* (pp. 195–212). Wiley-Blackwell. https://www.researchgate.net/publication/283256545

Forcada, N., Fuertes, A., Gangolells, M., Casals, M., & MacArulla, M. (2013). Knowledge management perceptions in construction and design companies. *Automation in Construction*, *29*(January), 83–91.

Hlupic, V., Pouloudi, A., & Rzevski, G. (2002). Towards an integrated approach to knowledge management: “Hard”, “soft” and “abstract” issues. *Knowledge and Process Management*, *9*(2), 90–102.

Israel, G. D. (1992). Determining sample size. *University of Florida, IFAS Extension*, *PE0D6*(April 2009), 1–5. https://edis.ifas.ufl.edu/pdffiles/PD/PD00600.pdf

Javernick-Will, A. (2012). Motivating knowledge sharing in engineering and construction organizations: Power of social motivations. *Journal of Management in Engineering*, *28*(2), 93–202.

Kerzner, H. R. (2017). *Project management: A systems approach to planning, scheduling, and controlling.* (12th ed.). John Wiley & Sons.

Kolb, A. Y., & Kolb, D. A. (2009). Experiential learning theory: A Dynamic, holistic approach to management learning, education and development. In S. Armstrong & C. Fukami (Eds.), *The Sage handbook of management learning, education and development*.

KPMG Consulting. (2000). *Knowledge management: Research report 2000*.

Kululanga, G. K., & Mccaffer, R. (2001). Measuring knowledge management for construction organizations. *Engineering, Construction and Architectural Management*, *8*(5–6), 346–354.

Leal, C., Cunha, S., & Couto, I. (2017). Knowledge sharing at the construction sector - Facilitators and inhibitors. *Procedia Computer Science*, *121*, 998–1005. https://doi.org/10.1016/j.procs.2017.11.129

Levin, D. Z., & Cross, R. (2002). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, *50*(11), 1477–1490.

Longwe, T., Lord, W., & Carrillo, P. (2015). The impact of employee experience in uptake of company collaborative tool. *Proceedings of the 31st Annual ARCOM Conference Lincoln, UK*, 621–629.

Love, P. E. D., Teo, P., Davidson, M., Cumming, S., & Morrison, J. (2016). Building absorptive capacity in an alliance: Process improvement through lessons learned. *International Journal of Project Management*, *34*(7), 1123–1137.

Luo, C., Lan, Y., (Robert) Luo, X., & Li, H. (2021). The effect of commitment on knowledge sharing: An empirical study of virtual communities. *Technological Forecasting and Social Change*, *163*. https://doi.org/10.1016/j.techfore.2020.120438

MacNeil, C. M. (2003). Line managers: Facilitators of knowledge sharing in teams. *Employee Relations*, *25*(3), 294–307.

Marinho, A. J. C., & Couto, J. (2022). Contribution to improvement of knowledge management in the construction industry - Stakeholders’ perspective on implementation in the largest construction companies. *Cogent Engineering*, *9*(1). https://doi.org/10.1080/23311916.2022.2132652

Meredith, J. R., Shafer, S. M., & Mantel, S. J. (2017). *Project management: A managerial approach* (10th ed.). John Wiley & Sons.

Project Management Institute. (2017). *A guide to the project management body of knowledge (PMBOK Guide)* (6th ed.). Project Management Institute, Inc.

Rahman, M. S., Mannan, M., Hossain, M. A., Zaman, M. H., & Hassan, H. (2018). Tacit knowledge-sharing behavior among the academic staff: Trust, self-efficacy, motivation and Big Five personality traits embedded model. *International Journal of Educational Management*, *32*(5), 761–782. https://doi.org/10.1108/IJEM-08-2017-0193

Rego, A., Pinho, I., Pedrosa, J., & Pina E. Cunha, M. (2009). Barriers and facilitators to knowledge management in university research centers. *Journal of the Iberoamerican Academy of Management*, *7*(1), 33–47.

Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences* (2nd ed.). Holt, Rinehart and Winston.

Tan, H. C., Carrillo, P. M., & Anumba, C. J. (2012). Case study of knowledge management implementation in a medium-sized construction sector firm. *Journal of Management in Engineering*, *28*(3), 338–347. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000109

Teerajetgul, W., & Chareonngam, C. (2008). Tacit knowledge utilization in Thai construction projects. *Journal of Knowledge Management*, *12*(1), 164–174.

Teerajetgul, W., & Charoenngam, C. (2006). Factors inducing knowledge creation: Empirical evidence from Thai construction projects. *Engineering, Construction and Architectural Management*, *13*(6), 584–599.

Terzieva, M. (2014). Project knowledge management: How organizations learn from experience. *Procedia Technology*, *16*, 1086–1095.

Theriou, N., Maditinos, D., & Theriou, G. (2011). Knowledge management enabler factors and firm performance: An empirical research of the Greek medium and large firms. *European Research Studies Journal*, *14*(2), 97–134.

Tsai, W. (2002a). Social structure of “coopetition” within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organization Science*, *13*(2), 179–190.

Tsai, W. (2002b). Social structure of “coopetition” within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organization Science*, *13*(2), 179–190.

Van Den Hooff, B., & Ridder, J. A. (2004). Knowledge sharing in context: The influence of organizational commitment, communication climate and CMC use on knowledge sharing. *Journal of Knowledge Management*, *8*(6), 117–130.

Yang, Y., Brosch, G., Yang, B., & Cadden, T. (2019). Dissemination and communication of lessons learned for a project-based business with the application of information technology: a case study with Siemens. *Production Planning & Control*. https://doi.org/10.1080/09537287.2019.1630682

Yap, J. B. H., Lim, B. L., & Skitmore, M. (2022). Capitalising knowledge management (KM) for improving project delivery in construction. *Ain Shams Engineering Journal*, *13*(6), 101790. https://doi.org/10.1016/j.asej.2022.101790

Yap, J. B. H., Lim, B. L., Skitmore, M., & Gray, J. (2021). Criticality of project knowledge and experience in the delivery of construction projects. *Journal of Engineering, Design and Technology*. https://doi.org/10.1108/JEDT-10-2020-0413

Yap, J. B. H., & Lock, A. (2017). Analysing the benefits, techniques, tools and challenges of knowledge management practices in the Malaysian construction SMEs. *Journal of Engineering, Design and Technology*, *15*(6), 803–825.

Yap, J. B. H., & Shavarebi, K. (2022). Enhancing project delivery performances in construction through experiential learning and personal constructs: competency development. *International Journal of Construction Management*, *22*(3), 436–452. https://doi.org/10.1080/15623599.2019.1629864

Yap, J. B. H., Shavarebi, K., & Skitmore, M. (2021). Capturing and reusing knowledge: analysing the what , how and why for construction planning and control. *Production Planning & Control*, *32*(11), 875–888.

Yap, J. B. H., & Skitmore, M. (2018). Investigating design changes in Malaysian building projects. *Architectural Engineering and Design Management*, *14*(3), 218–238.

Yap, J. B. H., & Skitmore, M. (2020). Ameliorating time and cost control with project learning and communication management Leveraging on reusable knowledge assets. *International Journal of Managing Projects in Business*, *13*(4), 767–792.

Evaluation of the Contribution of Knowledge Management to Efficiency in the Manufacturing Industry Through Machine Learning

Juan Ibujés-Villacís[a]
Facultad de Ciencias Administrativas, Escuela Politécnica Nacional, Quito, Ecuador
Juan.ibujes@epn.edu.ec

Keywords: Business Management, Efficiency, Knowledge Management, Manufacturing Industry, Machine Learning.

Abstract: Knowledge management (KM) has been instrumental for organizations to improve their efficiency. The objective of this research is to determine the contribution of knowledge management (KM) to manufacturing industry efficiency, using machine learning models to predict the relevant KM factors that should be taken into account to improve efficiency. Given the quantitative nature of the research, in the first phase, data on variables associated with KM factors and efficiency were collected and processed. In the second phase, four supervised machine learning models were developed to predict which manufacturing companies are efficient in their production process based on a set of KM factors. The study was based on information from 142 manufacturing companies in the province of Pichincha, Ecuador. The results show that the relevant KM factors that contribute to business efficiency are policies and strategies, organizational structure, technology, incentive systems and organizational culture. This pioneering study in Ecuador allows predicting the relevant KM factors that impact the efficiency of manufacturing firms. This article contributes to the field of knowledge management and provides information on the KM factors that manufacturing firms should focus on to achieve greater efficiency.

1 INTRODUCTION

Enterprises are currently leveraging machine learning (ML) technology to optimize various areas of business management, such as analyzing purchase history, personalizing product recommendations, and predicting customer behaviors (Akerkar, 2019; Hemachandran & Rodriguez, 2024). However, the potential of ML is not limited to these applications; it can also play a crucial role in strategic decision making and improving operational efficiency.

Many companies in different economic sectors have implemented artificial intelligence to increase efficiency, improve their operations, and predict future needs and behaviors in real time, allowing them to offer better experiences to their customers (Anshari et al., 2023; Pagani & Champion, 2024). These technologies help companies optimize resources and capabilities, contributing significantly to their strategic objectives.

From a knowledge management (KM) perspective, many companies develop strategies such as knowledge exploitation, acquisition, sharing and exploration to improve knowledge management companies (Bolisani & Bratianu, 2018). However, these strategies do not always translate into efficiency gains, probably due to the lack of data for informed decision making.

The purpose of this research is to design and develop machine learning models that have an impact on predictive analysis, identifying which manufacturing companies are operationally efficient based on practices associated with KM. This research is pioneering in the Ecuadorian context, since there are no studies in which machine learning is used to predict business management results.

Methodologically, it has a quantitative approach and a survey was used as a research technique, taking 142 manufacturing companies in Pichincha, Ecuador, as a random sample. The survey collected data on factors related to KM and efficiency based on previous studies (Ibujés-Villacís & Franco-Crespo, 2022). With these data, several supervised machine learning models were developed, including multiple

[a] https://orcid.org/0000-0001-8439-3048

Ibujés-Villacís, J.
Evaluation of the Contribution of Knowledge Management to Efficiency in the Manufacturing Industry Through Machine Learning.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 48-59
ISBN: 978-989-758-716-0; ISSN: 2184-3228

linear regression, where KM factors were considered as independent variables and efficiency factors as dependent variables.

Knowledge management brings numerous benefits to companies such as the optimization of efforts and the improvement of operational efficiency. It allows identifying and leveraging best practices, as well as avoiding errors and rework (Pagani & Champion, 2024). This study, by training algorithms with data from medium-sized manufacturing companies, contributes to identify the factors of KM that are relevant to determine efficiency in the Ecuadorian manufacturing industry. Its results will enable companies to develop strategies to optimize resources and capabilities in achieving business objectives.

The paper begins with an overview of knowledge management, organizational efficiency and machine learning. Then, four multiple linear regression models are presented to predict variables associated with business efficiency from KM-related variables. Through machine learning, algorithms are developed to identify significant KM variables that impact efficiency. Finally, results are discussed, conclusions, limitations and possible directions for future studies are presented.

2 THEORETICAL ELEMENTS

2.1 Knowledge Management

Knowledge can be treated both as an object with attributes and properties, and as a process involving a set of cognitive activities performed by individuals or organizations with the objective of creating or adding value (Davenport & Prusak, 1998; Saulais & Ermine, 2019). In the organizational context, this value manifests itself in various forms, such as the creation of new business models, increased profitability, improved organizational efficiency, innovations in products and processes, and increased customer satisfaction (Andreini & Bettinelli, 2017).

Knowledge management (KM) in organizations is one of the most important collective capabilities, as it is the key to professional growth and profitability strength in the 21st century (Manning & Manning, 2020). In addition, it is fundamental to improve efficiency and promote innovations in products and processes (Newell, 2015).

According to North & Kumta (2018), KM is oriented in two main directions, as shown in Figure 1. The first, focused on the operational management of symbols until knowledge becomes a competitive advantage. The second, focused on strategic knowledge management, which consists of determining what type of knowledge, data or symbols the organization needs to realize its strategies.

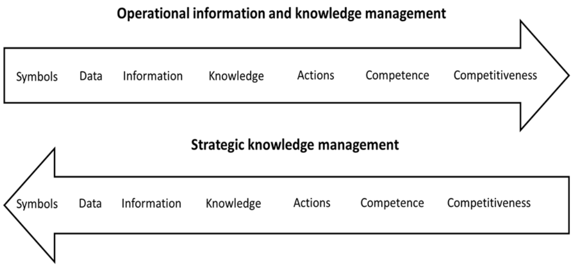


Figure 1: Knowledge management and competitiveness. *Note*: Image adapted from Knowledge Management. Value Creation Through Organizational Learning (p. 35), by Klaus North and Gita Kumta, 2018, Springer.

Knowledge management is multidimensional. In the static dimension, the organization focuses on maintaining, replicating and exploiting available knowledge as an internal capability of the organization, leveraging internal human talent and existing technological infrastructure (Endres, 2018; Kaur, 2019). In the dynamic dimension, the organization performs activities to acquire, convert and apply externally generated knowledge.

In recent years, due to the vast amount of data available and the development of computer science, KM has gained renewed importance in organizations. This resurgence has been driven by advances in machine learning and artificial intelligence (Bhupathi et al., 2023; Uden et al., 2014; Weber, 2023).

2.2 Efficiency in the Industry

Efficiency is a key indicator that reflects a company's ability to operate economically. The key indicators of efficiency focus on physical-technical performance and costs (Zanda, 2018). Efficiency assesses whether resources are being utilized to their maximum productive capacity, i.e., whether productive factors are being utilized at one hundred percent or whether there is idle capacity (Cachanosky, 2012).

In the context of Ecuadorian industry, efficiency has also been studied as an indicator of innovation and its relationship with sustainable development objectives (Ibujés-Villacís & Franco-Crespo, 2019, 2023a, 2023b). These studies highlight the importance of efficiency not only from an economic perspective, but also from a sustainability and innovation approach.

Several corporate performance factors are specifically related to efficiency, and the application of these factors depends on the context and careful management of each one (Albornoz, 2009). In this study, relevant factors were selected for medium-sized manufacturing companies in Pichincha, based on previous studies conducted in these companies (Ibujés-Villacís & Franco-Crespo, 2022).

This research focuses on the impact of knowledge management (KM) on the efficiency of manufacturing companies. For this purpose, a set of factors were considered associated with both knowledge management and efficiency. The objective is to determine how certain KM factors can predict efficiency in these companies. By understanding the relationship between KM and efficiency, organizations can develop more effective strategies to optimize their operations and improve their overall performance.

2.3 Machine Learning

Machine learning, predictive modeling and artificial intelligence are closely related terms (Shmueli et al., 2023). This field of study endows computers with the ability to learn without the need to be explicitly programmed. In machine learning, a computer program learns from experience with respect to a set of tasks, progressively improving its performance as it accumulates experience (Akerkar, 2019).

Machine learning generally begins with the simplified representation of reality using a model (Burger, 2018). Models are mathematical tools that describe systems and capture relationships in the data provided (Kuhn & Silge, 2022). Unlike dashboards, which provide a static picture of the data, models allow understanding and predicting future trends (Burger, 2018).

There are several machine learning models, such as regression, clustering and neural networks, all based on algorithms. The three main types of models are: regression models, classification models and mixed models combining both approaches.

To meet the objective of this research, a supervised learning algorithm will be used to model the relationships between KM input variables and efficiency output variables. Machine learning is currently a fundamental tool for decision making in business (Pagani & Champion, 2024; Weber, 2023). In particular, this research will employ a multiple linear regression model to determine the relationship between a set of corporate efficiency variables (dependent variables) and another set of knowledge management variables (independent variables).

Machine learning requires training a model with a data set, which represents a percentage of the total available data. The training results are evaluated to determine if the errors decrease and if the model fits correctly. If errors persist, the model needs to be modified and refined (Burger, 2018).

Training data are crucial for fitting machine learning models and, in many cases, are used to perform cross-validation during the training phase of the model. This validation consists of splitting the data into two subsets, one for training and one for testing, which allows further refinement of the model (Burger, 2018; Hastie et al., 2023).

This research is based on supervised machine learning, since it is required to make predictions about the efficiency of companies based on a data set that relates two defined categories: KM and corporate efficiency. These data were obtained through surveys of manufacturing companies in Pichincha, Ecuador.

2.4 Multiple Linear Regression

Multiple linear regression (MLR) is a statistical technique used to model the relationship between a dependent variable and two or more independent variables. MLR seeks to find the best line (or hyperplane in higher dimensions) that fits the data optimally. This involves determining the coefficients that minimize the difference between the values predicted by the model and the actual values observed in the data set.

Mathematically, the multiple linear regression model is expressed as equation 1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (1)$$

Where

Y is the dependent variable.

$X_1, X_2, \ldots\ldots, X_n$: independent variables.

$\beta_0, \beta_1, \beta_2, \ldots., \beta_n$: coefficients representing the slope of each independent variable.

ϵ: is the error term, which captures the variation not explained by the model.

MLR is especially useful for understanding how multiple independent factors contribute to a particular outcome. In this study, MLR is used to analyze and predict the relationship between dependent variables related to company efficiency and set of independent variables related to knowledge management.

In the scope of this research, which focuses on medium-sized manufacturing companies in

Pichincha, Ecuador, the dependent variables are related to corporate efficiency, as shown in Table 2. The independent variables, on the other hand, are related to knowledge management, as shown in Table 1. The use of the MLR allows us to identify which factors of knowledge management have a significant impact on the efficiency of these companies.

3 METHODOLOGY

Figure 2 shows the complete process to achieve the research objective, starting with the determination of the sample and ending with the results obtained after the application of machine learning.

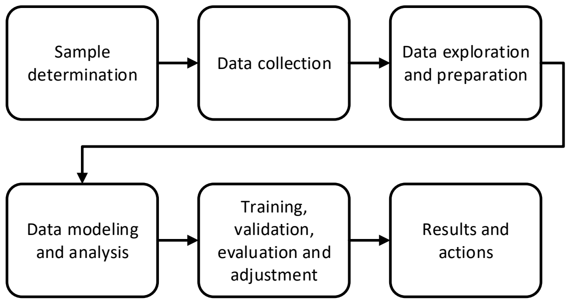


Figure 2: Process for data analysis.

3.1 Sample Determination

The scope of the study is companies in the manufacturing sector in the province of Pichincha, where Quito, the capital of Ecuador, is located. This economic sector was chosen because of its significant contribution to the country's economy, contributing 14.2% to Ecuador's total production (MIPRO, 2021).

The study population includes medium-sized manufacturing companies that are active and have been operating for at least five years. These companies have between 50 and 199 employees, annual revenues between US$1 million and US$5 million, and an asset value of less than US$4 million (SUPERCIAS, 2021). As of November 2020, medium-sized manufacturing companies in Pichincha that had submitted their economic and financial reports for 2019 totaled 338 (SUPERCIAS, 2020).

To determine the sample size, proportional sampling was used for a finite population. The sampling was probabilistic and with equal probabilities. The selection of companies was done by simple random sampling, without replacement, to ensure the greatest representativeness of the sample (Latpate et al., 2021; Lohr, 2019).

To obtain a representative (n) and adequate sample of the population, equation 2 (Lohr, 2019; Ott & Longnecker, 2016) was applied.

$$n = \frac{Z^2Npq}{E^2(N-1)+Z^2pq} \tag{2}$$

The parameters used to calculate the sample were: N = 338 (study population), E = 10 % (sampling error percentage), Z = 1.96 (95 % confidence level), p = 0.5 (probability of success) and q = 0.5 (probability of failure). With these parameters it was determined that n = 75 companies. The study was applied to 142 companies, exceeding the required sample size, which reduced the sampling error to 6 % and maintained the confidence level at 95 %.

3.2 Data Collection

Data collection was carried out by means of a survey addressed to the top managers of the companies included in the study sample. A closed-ended questionnaire was used to evaluate 85 items distributed in two main sections. The KM is represented by 35 variables grouped into seven factors, while the efficiency of the companies is represented by four variables, as detailed in Tables 1 and 2.

This questionnaire was subjected to content validation by experts, considering four categories: coherence, relevance, clarity and sufficiency of the questions. To ensure these qualities, a pilot test was conducted with the participation of ten experts from academia and industry. Based on the validation and the comments received, the suggested improvements were incorporated and the final version of the questionnaire was prepared.

To respond to the questionnaire, company managers were asked to rate each of the items using the psychometric instrument called Likert scale (Bertram, 2018). A 10-point scale was used, with 1 representing very low agreement and 10 representing very high agreement with the argument presented in each item.

The surveys were conducted using a Google form, applied electronically from June to September 2021. A total of 250 questionnaires were sent by e-mail to the companies that were the subject of the study. Each survey complied with ethical research standards: informed consent, voluntary participation, confidentiality and absence of physical or psychological risk to participants.

Table 1: Knowledge management factors and variables.

Knowledge management variables	Notation
Policies and strategies (PS)	
Policies for the acquisition and generation of organizational knowledge.	PS1
Policies for the storage, sharing and use of knowledge organizational.	PS2
Implementation of properly documented processes, procedures and routines	PS3
Establishment of alliances with public and private organizations.	PS4
Development of dynamic plans to overcome internal and external barriers.	PS5
Permanent focus on continuous improvement.	PS6
Systematic combination of existing and new knowledge.	PS7
Organizational structure (OS)	
Internal organizational structures dedicated to research and development.	OS1
Regulations established for the access and use of knowledge.	OS2
Agility in the processes to access organizational knowledge.	OS3
Facilities for the horizontal flow of knowledge within the organization.	OS4
Facilities for the vertical flow of knowledge within the organization.	OS5
Technology (TG)	
Use of technology for the methodical storage of knowledge.	TG1
Use of information systems for accessing, sharing and utilizing the organizational knowledge.	TG2
Application of ICT for access, exchange and use of knowledge.	TG3
Utilization of corporate social networks for collaboration and knowledge of the environment.	TG4
Persons (PP)	
Years of employee experience.	PP1
Employees' level of education.	PP2
Age of employees.	PP3
Foreign language proficiency of employees.	PP4
Gender diversity among employees.	PP5
Incentive systems (IS)	
Economic incentives for generating, sharing and using knowledge.	IS1
Training offered as an incentive for generating, sharing and using the knowledge.	IS2
Days off granted as an incentive for generating, sharing, and using the knowledge.	IS3
Public recognition as an incentive for generating, sharing and utilizing the knowledge.	IS4
Organizational culture (OC)	
Importance of personal values.	OC1
Positive attitude towards work.	OC2
Respect for the company's principles and regulations.	OC3
Application of best practices.	OC4
Staff empowerment for decision making.	OC5
Creation of a collaborative and synergistic work environment.	OC6
Communication (CM)	
Formal communication in the work environment.	CM1
Informal communication in the work environment.	CM2
Effective communication with all hierarchical levels.	CM3
Fluent communication in physical and virtual spaces.	CM4

Note: ICT: Information and communication technologies.

Table 2: Efficiency variables.

Efficiency variables	Notation
Reduced production and marketing costs.	CS1
Application of best practices.	CS2
Reduced product delivery time.	CS3
Increased benefit/cost ratio.	CS4

3.3 Data Exploration and Preparation

Exploratory data analysis is a crucial phase in the modeling process in machine learning, as it provides valuable information about the nature and quality of the data (Costa-Climent et al., 2023). This phase is essential because its results can influence the decisions made during the modeling process and improve the effectiveness and interpretation of the resulting models. In this research, the variables used in supervised learning correspond to KM factors and efficiency factors. In all cases, the variables are quantitative.

The algorithm chosen to relate the KM variables (inputs) to the efficiency variables (output) was multiple linear regression. Since the responses to each question range from 1 to 10, no outliers were found. Therefore, no histograms, boxplots or scatter plots were performed to visualize the distribution of the data and detect possible outliers.

The relationships between each of the variables that make up the seven KM factors were explored to detect multicollinearity of the independent variables. Multicollinearity occurs when two or more independent variables in a model are highly correlated with each other (Lantz, 2023). The presence of multicollinearity can cause several problems in regression analysis, including instability in coefficient estimation, increased coefficient variance, and unreliable coefficients.

The correlation between the independent variables made it possible to eliminate those with a correlation coefficient greater than 0.7. These ten variables were: PS1, PS6, OS3, TG1, TG2, OC2, OC3, OC4, OC6, CM4; thus leaving 25 variables corresponding to the KM for the analysis.

3.4 Data Modeling and Analysis

The approach chosen for the model in this research is supervised machine learning. Supervised models are those in which a machine learning model is trained and fit with labeled data, i.e., known quantities (Burger, 2018).

To evaluate the impact of KM on manufacturing efficiency, a multiple linear regression model was chosen. This model was selected for several reasons. First, due to the nature of the data, since all variables are quantitative. Second, the amount of data available facilitates the application of the proposed model. The model is represented by equation 3.

$$Y = f(X) + \varepsilon \; = \beta_0 + \beta_1 X_1 \; + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (3)$$

The impact of the KM factors on four variables related to efficiency was evaluated. For this reason, four multiple linear regression models were developed and are described in Table 3.

Table 3: Multiple linear regression models.

Model	Y	X
1	Y1= CS1	PS2, PS3, PS4, PS5, PS7 OS1, OS2, OS4, OS5 TG3, TG4 PP1, PP2, PP3, PP4, PP5 IS1, IS2, IS3, IS4 OC1, OC5 CM1, CM2, CM3
2	Y2= CS2	
3	Y3= CS3	
4	Y4= CS4	

3.5 Training, Validation, Evaluation and Adjustment

The database used contains 142 records and 31 variables, of which 25 are associated with KM and four with efficiency. All variables are quantitative. To evaluate the performance of the predictive model, the data were divided into two subsets: training data (80 %) and test data (20 %).

Cross-validation is a technique used in machine learning and statistics to evaluate the performance of a predictive model. It consists of dividing the data set into multiple training and test subsets, training and evaluating the model on different combinations of these subsets (Boehmke & Greenwell, 2020). In this study, the K-fold technique with ten divisions (folds) was used. This subdivision allowed obtaining more stable estimates of the model performance, providing a more robust evaluation by averaging the results across the different data splits.

A recipe was used to define a set of preprocessing steps that were applied to the data sets prior to modeling. This recipe served as a template for data preprocessing. Next, a workflow was created to model the MLR, integrating the MLR model and the preprocessing steps defined in the recipe, allowing to train and evaluate the model in an integrated and consistent way.

Model validation was performed using the root mean squared error (RMSE) value, which measures the level of dispersion of the residual values and calculates the square root of the mean value of the squared difference between the actual and predicted value for all data points. The RMSE is calculated as the square root of the mean of the squared errors between the model predictions and the actual values in the test set (Kuhn & Silge, 2022).

The RMSE formula is given in equation 4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

Where n is the number of observations in the test set, yi are the actual values of the dependent variable and $\hat{yi}$ are the model predictions for the dependent variable. A model performs well the lower the RMSE value and the closer this value resembles the value obtained between the training and test data (Kuhn & Silge, 2022). Both modeling and data analysis were performed using the RStudio programming language.

4 RESULTS

4.1 Relationship Between KM and Reduction of Production and Marketing Costs

The relationship between KM and cost reduction was evaluated using a multiple linear regression model CS1 = f (X)+Ɛ. Table 4 shows that three KM variables belonging to the factors of organizational structure, incentive system and communication are significant and have a direct relationship with cost reduction. These results indicate that the model is viable.

Table 4: KM variables that impact cost reduction.

KM variable	Coefficient	Pr(>|t|)
OS4	0.281	0.022
IS1	0.241	0.026
CM2	0.261	0.008

R^2 = 0.575, F = 4.61, p-value model = 6.1e-08

Notes:
OS4: Facilities for the horizontal flow of knowledge within the organization, IS1: Economic incentives for generating, sharing and using knowledge, CM2: Informal communication in the work environment. Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship assessment statistic, p: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the following function: CS1= 0.28 OS4 + 0.24 IS1 + 0.26 CM2.

The RMSE of the best model with the training data is 2.79, a value similar to that obtained with the test data, which is a positive sign that the model is robust and has good generalizability. Table 5 reviews the statistical assumptions of the model, while Figure 3 shows these results graphically.

Table 5: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.681	Ok.
Heteroscedasticity	p = 0.243	Ok.
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

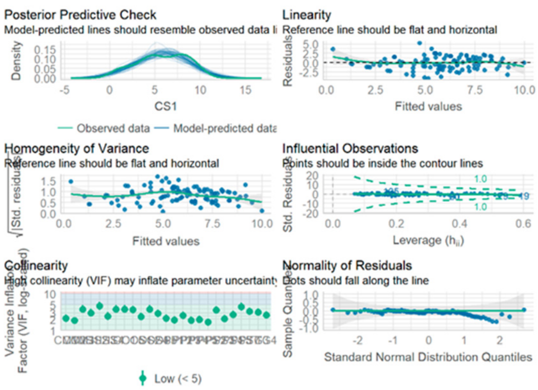


Figure 3: Graphs of statistical assumptions. *Note*: Image obtained from RSudio.

4.2 Relationship Between KM and the application of best practices

The relationship between KM and the application of best practices is evaluated using the multiple regression model CS2 = f (X)+Ɛ. Table 6 shows that six knowledge management variables belonging to factors such as technology, incentive system, organizational culture and communication are significant, and have a direct relationship with the application of best practices. These results indicate that the model is viable.

Table 6: KM variables impacting the application of best practices.

KM variable	Coefficient	Pr(>|t|)
Intercept	-1.643	0.039
TG3	0.284	0.017
TG4	0.183	0.025
IS1	0.214	0.011
OC5	0.309	0.006
CM2	0.244	0.001
CM3	0.393	0.000

R^2 = 0.728, F = 9.3, p-value model = 1.75e-15

Notes:
TG3: Application of ICT for access, sharing and use of knowledge, TG4: Use of corporate social networks for collaboration and leveraging knowledge of the environment, IS1: Economic incentives for generating, sharing and using knowledge, OC5: Empowerment of staff for decision making, CM2: Informal communication in the work environment, CM3: Effective communication with all hierarchical levels, Pr(>|t|): Significance statistic of the variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, p: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS2 = -1.64306 + 0.28 TG3 + 0.18 TG4 + 0.21 IS1 + 0.31 OC5 + 0.24 CM2 + 0.39 CM3.

The RMSE of the best model with the training data is 2.25, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 7 shows the statistical assumptions of the model, while Figure 4 shows these results graphically.

Table 7: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.433	Ok
Heteroscedasticity	p = 0.405	Ok
Autocorrelated residuals	p = 0.002	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

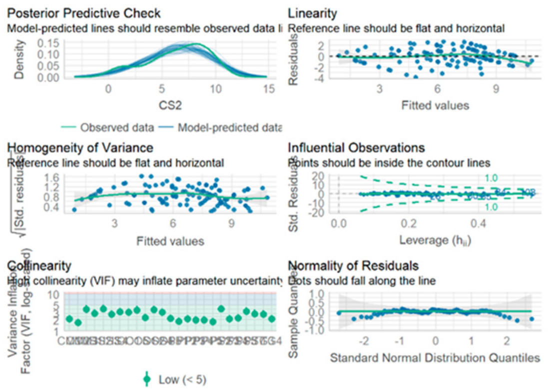


Figure 4: Graphs of statistical assumptions. *Note*: Image obtained from RSudio.

4.3 Relationship Between KM and Reduction of Product Lead Time

The relationship between KM and product lead time reduction is evaluated using the multiple regression model CS3 = f (X)+Ɛ. Table 8 shows that four KM variables belonging to the factors: organizational structure, organizational culture and communication are significant and have a direct relationship with the reduction of product lead time. Additionally, one variable belonging to the policies and strategies category is shown to have a significant and indirect relationship. The results indicate that the model is viable.

Table 8: KM variables impacting product lead time reduction.

KM variable	Coefficient	Pr(>|t|)
PS2	-0.378	0.003
OS4	0.223	0.046
OC1	0.234	0.049
CM1	0.314	0.006
CM2	0.309	0.000

R^2 = 0.677, F = 7.11, p-value model = 3.46e-12

Notes:
PS2: Policies for the storage, sharing and use of organizational knowledge, OS4: Facilities for the horizontal flow of knowledge within the organization, OC1: Importance of personal values, CM1: Formal communication in the work environment, CM2: Informal communication in the work environment, Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, *p*: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS3 = -0.38 PS2 + 0.22 OS4 + 0.23 OC1 + 0.31 CM1 + 0.31 CM2.

The RMSE of the best model with the training data is 2.93, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 9 reviews the statistical assumptions of the model, while Figure 5 shows these results graphically.

Table 9: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.386	Ok
Heteroscedasticity	p = 0.134	Ok
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

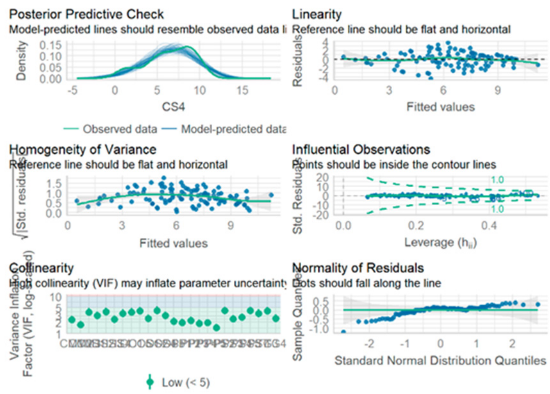


Figure 5: Analysis of statistical assumptions. *Note*: Image obtained from RSudio.

4.4 Relationship Between KM and Increase in Benefit/Cost Ratio

The relationship between KM and the increase in the benefit/cost ratio is evaluated using the multiple regression model CS4 = f (X)+Ɛ. Table 10 shows that two KM variables belonging to the factors: persons and communication are significant and have a significant and direct relationship with the increase in the benefit/cost ratio. Additionally, it is shown that one variable belonging to the policies and strategies category has a significant and indirect relationship. These results indicate that the model is viable.

Table 10: KM variables that have an impact on the increase in benefit/cost ratio.

KM variable	Coefficient	Pr(>|t|)
PS2	-0.305	0.038
PP5	0.171	0.040
CM2	0.229	0.009

R^2 = 0.548, F = 4.18, p-value model = 3.76e-12

Notes:
PS2: Policies for the storage, sharing and use of organizational knowledge, PP5: Development of dynamic plans to overcome internal and external barriers, CM2: Informal communication in the work environment, Pr(>|t|): Significance statistic of variable X, R^2: Coefficient of determination, F: Model relationship evaluation statistic, *p*: Significance statistic of the results.

The statistical results of the model indicate that it is significant and viable as a whole. The model is represented by the function: CS4 = -0.31 PS2 + 0.17 PP5 + 0.23 CM2. The RMSE of the best model with the training data is 2.49, a value similar to that obtained with the test data. This coincidence is a positive sign that the model is robust and has good generalizability. Table 11 reviews the statistical assumptions of the model, while Figure 6 shows these results graphically.

Table 11: Statistical assumptions.

Supposed	Value obtained	Evaluation
Normality of waste	p = 0.001	Warning
Heteroscedasticity	p = 0.943	Ok
Autocorrelated residuals	p = 0.001	Warning
Multicollinearity	All variables <5	Low Correlation
Outliers	None	OK

Note: Statistics obtained from RSudio.

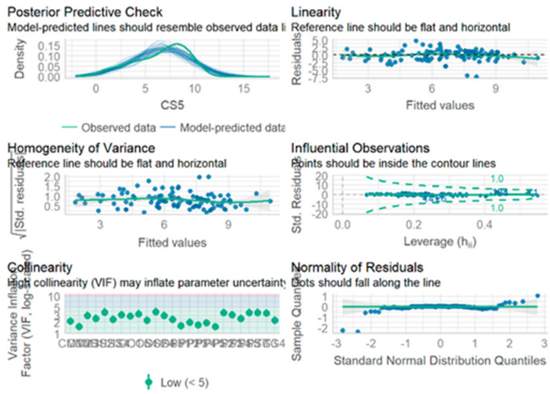


Figure 6: Analysis of statistical assumptions. *Note*: Image obtained from RSudio.

5 DISCUSSION

Through the machine learning developed in this research, it has been shown that the presence of certain KM variables in business organizations can predict efficiency in operational management. Multiple linear regression was used to describe the relationship between a target variable and a set of explanatory characteristics, and to use this relationship to predict the value of the target variable.

Evaluating the relationship between KM and the reduction of production and marketing costs, the MLR model showed that facilities for the horizontal flow of knowledge within the organization, economic incentives for generating, sharing and using knowledge, and informal communication in the work environment have a positive and significant impact on cost reduction in manufacturing companies.

Regarding the relationship between KM and the use of best practices, the results show that the application of ICT, the use of social networks, economic incentives to personnel, the empowerment of personnel in decision making, and effective and informal communication at all hierarchical levels have a positive impact on the use of good practices in the industrial sector.

Evaluating the relationship of KM with product lead time reduction, it was shown that the level of employee education, facilities for horizontal knowledge flow within the organization, the importance of personal values, and formal and informal communication have an impact on the optimization of product lead time.

Regarding the relationship between KM and the increase in the benefit/cost ratio, it was shown that the level of employee education, the development of dynamic plans to overcome internal and external barriers, and informal communication in the work environment have a direct impact on this relationship in manufacturing companies.

Table 12 shows that of the 35 KM variables distributed in seven factors, 11 variables have a significant impact on the efficiency of manufacturing companies. In addition, the number of times these variables appear in the models is shown. The KM factor that contributes most to efficiency is communication, followed by policies and strategies, organizational structure, technology, incentive systems, and organizational culture. The factor that does not yet contribute substantially to KM is persons.

The results obtained with each of the models are consistent with the assertion that KM directly leads to a reduction in operating costs (Piening & Salge, 2015) and contributes to the development of innovations (OECD & Eurostat, 2018). In addition, KM provide more efficient and effective management of companies, allowing informed decisions to be made to meet customer needs by analyzing large data sets (Hemachandran & Rodriguez, 2024).

Table 12: KM factors impacting efficiency.

Factors	Significant variables	n
Policies and strategies	PS2	2
Organizational structure	OS4	2
Technology	TG3	1
	TG4	1
Persons	PP5	1
Incentive system	IS1	2
Organizational culture	OC1	1
	OC5	1
Communication	CM1	1
	CM2	3
	CM3	1

Notes: n: Number of times the variables are present in the models studied.

5.1 Theoretical Implications

Among the theoretical implications of this research, it was determined that there are key factors related to KM that impact the efficiency of companies. Of the 35 initial variables, 11 were found to be the most influential on the efficiency of manufacturing companies. This shows that the efficiency of companies depends on a set of variables related to the broad concept of knowledge management.

In the manufacturing industry it has been concluded that all factors associated with KM should be taken into account.

However, there are factors such as communication, policies and strategies, organizational structure, technology, personnel incentives and organizational culture that are relevant in predicting the efficiency of companies.

5.2 Practical Implications

The main practical contribution of this research lies in the identification of the relevant factors of KM that impact the efficiency of manufacturing companies. This allows strategic decisions focused on cost optimization, the application of best practices, the reduction of product delivery time, and the benefit/cost ratio.

By identifying these factors, companies can make informed decisions in real time to focus on efficient industrial processes by intervening in specific KM variables. Learning from existing data will enable companies to design solutions based on solid information, aimed at solving efficiency problems.

In addition, these informed decisions will enable companies:

- Design effective policies and strategies.
- Invest in appropriate technology.
- To optimally manage its human talent.
- Create motivating incentive policies.
- Establish beneficial strategic alliances.
- Modify its organizational structure to improve efficiency.

These substantial components of KM, discussed in this study, provide a practical framework for manufacturing companies to improve their operational efficiency and competitiveness in the marketplace.

6 CONCLUSIONS

The purpose of this study was to design and develop a series of machine learning models for predictive analysis of the identification of operationally efficient industries from the application of practices associated with knowledge management. Multiple linear regression models were used to demonstrate the impact of KM in predicting company efficiency.

In each model the independent variables represented the KM, and the dependent variables represented the operating efficiency of the companies. After eliminating correlated variables, 25 variables associated with KM factors were used: policies and strategies, organizational structure, technology, persons, incentive system, organizational culture and communication. The variables related to efficiency included cost reduction, application of best practices, reduction of delivery time, and increase in the benefit/cost ratio.

Four models were developed and 11 KM variables were found to significantly impact the efficiency of manufacturing companies. The KM factors that contribute most to efficiency are policies and strategies, organizational structure, technology, incentive systems, and organizational culture. Consequently, it has been shown that the application of certain KM factors in organizations can predict their efficiency and improve organizational performance. These findings underscore the importance of KM as a strategic tool for improving operational efficiency in manufacturing companies, providing a practical framework for informed decision making and the implementation of effective business practices.

6.1 Limitations and Future Studies

One of the limitations of this study is that knowledge management is a relatively new topic for the management of Ecuadorian business organizations. To mitigate this limitation, the surveys included sufficient introductory information to facilitate respondents understanding and response to the questionnaire.

The results of this research highlight the relevance of KM in various aspects of business management and provide a solid foundation for future research. It is recommended that further studies explore the impact of KM in areas such as the use of new technologies, innovation, resilience, and business sustainability, among others. These studies could delve deeper into how KM can contribute more comprehensively to improving the efficiency and performance of manufacturing firms in Ecuador.

REFERENCES

Akerkar, R. (2019). *Artificial Intelligence for Business*. Springer. https://doi.org/10.1007/978-3-319-97436-1_3

Albornoz, M. (2009). Indicadores de innovación: las dificultades de un concepto en evolución. *CTS: Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, *5*(13), 9–25. https://www.redalyc.org/articulo.oa?id=92415269002

Andreini, D., & Bettinelli, C. (2017). Business Model Innovation. From Systematic Literature Review to Future Research Directions. In *International Series in Advanced Management Studies*. Springer. https://doi.org/10.1007/978-3-319-53351-3

Anshari, M., Syafrudin, M., Tan, A., Fitriyani, N., & Alas, Y. (2023). Optimisation of Knowledge Management (KM) with Machine Learning (ML) Enabled. *Information (Switzerland)*, *14*(1), 1–15. https://doi.org/10.3390/info14010035

Bertram, D. (2018). *Likert Scales*. http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf

Bhupathi, P., Prabu, S., & Goh, A. P. I. (2023). Artificial Intelligence-Enabled Knowledge Management Using a Multidimensional Analytical Framework of Visualizations. *International Journal of Cognitive Computing in Engineering*, *4*(January), 240–247. https://doi.org/10.1016/j.ijcce.2023.06.003

Boehmke, B., & Greenwell, B. (2020). *Hands-On Machine Learning with R*. Taylor & Francis.

Bolisani, E., & Bratianu, C. (2018). Generic Knowledge Strategies. In *Emergent Knowledge Strategies* (4th ed., Vol. 4, Issue July, pp. 147–174). Springer. https://doi.org/10.1007/978-3-319-60657-6

Burger, S. (2018). *Introduction to machine learning with R : rigorous mathematical analysis*.

Cachanosky, I. (2012). Eficiencia técnica, eficiencia económica y eficiencia dinámica. *Procesos de Mercado: Revista Europea de Economía Política*, *IX*(2), 51–80.

Costa-Climent, R., Haftor, D. M., & Staniewski, M. W. (2023). Using machine learning to create and capture value in the business models of small and medium-sized enterprises. *International Journal of Information Management*, *73*(January), 102637. https://doi.org/10.1016/j.ijinfomgt.2023.102637

Davenport, T. H., & Prusak, L. (1998). Knowledge Generation. In *Working Knowledge: How Organizations Manage What They Know* (p. 226). Harvard Business School Press.

Endres, H. (2018). Frameworks and Theories around Dynamic Capabilities. In *Adaptability Through Dynamic Capabilities: How Management Can Recognize Opportunities and Threats* (pp. 13–28). Springer Gabler. https://doi.org/10.1007/978-3-658-20157-9

Hastie, T., Tibshirani, R., James, G., & Witten, D. (2023). An Introduction to Statistical Learning. In *Springer Texts* (2nd ed., Vol. 102).

Hemachandran, K., & Rodriguez, R. (2024). Artificial intelligence for business: An implementation guide containing practical and industry-specific case studies. In *Artificial Intelligence for Business: An Implementation Guide Containing Practical and Industry-Specific Case Studies* (1st ed.). Routledge. https://doi.org/10.4324/9781003358411

Ibujés-Villacís, J., & Franco-Crespo, A. (2022). Determinant factors of innovation management in the manufacturing industry of Pichincha , Ecuador. *Journal of Technology Management & Innovation*, *17*(1), 50–70. https://doi.org/10.4067/S0718-27242022000100050

Ibujés-Villacís, J., & Franco-Crespo, A. (2023a). La eficiencia como indicador de innovación. El caso de las compañías de manufactura en Pichincha, Ecuador. *Revista Innovar*, *33*. https://doi.org/10.15446/innovar.v33n89.107039

Ibujés-Villacís, J., & Franco-Crespo, A. (2023b). Relationship between Productivity and Efficiency with Sustainable Development Goals: The Case of the Manufacturing Industry in Pichincha, Ecuador. *Revista de Métodos Cuantitativos Para La Economía y Empresas*, *35*, 34–56. https://doi.org/10.46661/revmetodoscuanteconempresa.5475

Ibujés-Villacís, J., & Franco-Crespo, A. A. (2019). Use of ICT and its relationship with the Objectives of Sustainable Development in Ecuador. *RETOS. Revista de Ciencias de La Administración y Economía*, *9*(17), 37–53. https://doi.org/10.17163/ret.n17.2019.03

Kaur, V. (2019). Review of Literature. In *Knowledge Based Dynamic Capabilities. The Road Ahead in Gaining Organizational Competitiveness* (pp. 21–78). Springer. https://doi.org/10.1007/978-3-030-21649-8

Kuhn, M., & Silge, J. (2022). *Tidy Modelling with R* (1st ed.). O'Reilly Media, Inc. https://www.tmwr.org/

Lantz, B. (2023). *Machine Learning with R* (4th ed., Vol. 4, Issue 1). Packt Publishing Pvt Ltd. www.packt.com

Latpate, R., Kshirsagar, J., Kumar Gupta, V., & Chandra, G. (2021). Simple Random Sampling. In *Advanced Sampling Methods* (pp. 11–36). Springer. https://doi.org/10.1007/978-981-16-0622-9

Lohr, S. L. (2019). Simple Probability Samples. In *Sampling. Design and Analysis* (2nd ed., pp. 25–72). CRC Press.

Manning, M. J., & Manning, M. S. (2020). Knowledge Assets Management. In *Total Innovative Management Excellence (TIME). The Future of Innovation* (pp. 354–398). CRC Press.

MIPRO. (2021). Cifras de industrias. In *Gobierno del Ecuador*. https://www.produccion.gob.ec/wp-content/uploads/2021/06/Presentación-Industria-Junio-2021.pdf

Newell, S. (2015). Managing knowledge and managing knowledge work: What we know and what the future holds. *Journal of Information Technology*, *30*(1), 1–17. https://doi.org/10.1057/jit.2014.12

North, K., & Kumta, G. (2018). Knowledge in Organisations. In *Knowledge Management. Value Creation Through Organizational Learning* (2nd ed., pp. 33–66). Springer. http://www.springer.com/series/10099

OECD, & Eurostat. (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation* (4th ed., Issue October). OECD. https://doi.org/10.1787/9789264304604-en

Ott, R. L., & Longnecker, M. (2016). Inferences About Population Central Values. In *An Introduction to Statistical Methods & Data Analysis* (Seventh, pp. 232–299). Cengage Learning.

Pagani, M., & Champion, R. (2024). *Artificial Intelligence for Business Creativity*. Routledge. https://doi.org/10.4324/9781003287582

Piening, E. P., & Salge, T. O. (2015). Understanding the antecedents, contingencies, and performance implications of process innovation: A dynamic capabilities perspective. *Journal of Product Innovation Management*, *32*(1), 80–97. https://doi.org/10.1111/jpim.12225

Saulais, P., & Ermine, J.-L. (2019). *Knowledge Management in Innovative Companies*. Wiley. https://www.ptonline.com/articles/how-to-get-better-mfi-results

Shmueli, G., Bruce, P., DeoKar, A., & Patel, N. (2023). *Machine Learning for Business Analytics* (Vol. 5, Issue 1). Wiley.

SUPERCIAS. (2020). La eficiencia de las empresas manufactureras en el Ecuador. In *Investigación y estudios sectoriales*. https://investigacionyestudios.supercias.gob.ec/wp-content/uploads/2020/01/eficienciamanufactura_FINAL.pdf

SUPERCIAS. (2021). *Ranking de compañías*. Superintendencia de Compañías, Valores y Seguros. https://appscvs.supercias.gob.ec/rankingCias/

Uden, L., Wang, L., Corchado, J., Yang, H.-C., & Ting, I.-H. (2014). *The 8th International Conference on Knowledge Management in Organizations: Social and Big Data Computing for Knowledge Management*. Springer. https://doi.org/10.1007/978-94-007-7287-8

Weber, F. (2023). Artificial Intelligence for Business Analytics: Algorithms, Platforms and Application Scenarios. In *Artificial Intelligence for Business Analytics: Algorithms, Platforms and Application Scenarios*. Springer. https://doi.org/10.1007/978-3-658-37599-7

Zanda, S. (2018). The Compatibility of Effectiveness and Efficiency: The Pillars of Barnard's Theory of Cooperation. In *Building Efficient Management and Leadership Practices, Innovation, Technology, and Knowledge Management* (pp. 109–128). Springer. https://doi.org/10.1007/978-3-319-60068-0_8 109

Charting the Transformation of Enterprise Information Management: AI Explainability and Transparency in EIM Practice

Lufan Zhang[a] and Paul Scifleet[b]
Swinburne University of Technology, Melbourne, Victoria, Australia
{lufanzhang, pscifleet}@swin.edu,au

Keywords: Enterprise Information Management, Explainable AI, AI Transparency.

Abstract: Today's data-intensive environment poses significant challenges for enterprises in managing their vital information assets that often exceed manual capabilities. Despite a promising potential to assist, there's mistrust and misunderstanding of the values AI presents to Enterprise Information Management. This paper investigates the current state of AI-led changes to EIM practices and proposes an approach to improve understanding of AI's transformative role and impact on EIM. By charting AI use in EIM platforms across five areas - AI development, AI techniques, AI-integrated EIM capabilities, AI applications, and AI impacts – along with practice-based criteria for evaluating AI-integrated EIM solutions, this paper lays the foundation for explainable and transparent AI in EIM.

1 INTRODUCTION

In the current massive data environment enterprises face substantial challenges in managing their most vital information resources. The last decade has generated more data, documents and records than any previous decade of human activity, however most information resources remain predominantly unstructured and poorly controlled (Kolandaisamy et al., 2024), making them less reliable, retrievable, and accessible than ever before (Jaillant, 2022). The overwhelming volume of information is exceeding in-house expertise and the manual or semi-automated approaches that most enterprises usually take to implement architectures for information control. Consequently, it is not surprising to see increasing attention being paid to applying Artificial Intelligence (AI) and Machine Learning (ML) based solutions in Enterprise Information Management (EIM) (Baviskar et al., 2021) including for, the classification of digital assets (Huddart, 2022), authoritative records control, taxonomy and metadata management (Duranti et al., 2022), and screening for sensitive and confidential information communicated via email (Schneider et al., 2019).

Despite the promising potential of AI-based approaches to enterprises' IM needs, current evidence indicates that the inherent complexity and opacity of AI cause mistrust and misunderstanding among EIM practitioners about the transformational opportunities and value AI presents (Adadi & Berrada, 2018). While explainable AI (XAI) research initiatives aim to improve the transparency and understandability of complex AI solutions for end users, these approaches are primarily algorithm-centric and highly technical, often falling short in adequately addressing the needs of non-expert users (Barredo Arrieta et al., 2020). In contrast to the AI experts, programmers, and data analysts, who typically interact with AI at algorithm and model design levels (Bunn, 2020; Langer et al., 2021), EIM practitioners do not need to understand AI algorithm functions and the reasons behind the generation of specific outcomes. Their first experience of AI is often through interface interactions. This might involve experimenting with publicly available tools like ChatGPT or investigating the use of AI product integration in other workplace tools. EIM professionals are more concerned with the practical applications and utility of AI across the information management lifecycle (Haresamudram et al., 2023). Solutions based on an *explainable AI in the context of EIM* are required to meet the practical

[a] https://orcid.org/0009-0004-5148-564X
[b] https://orcid.org/0000-0003-2776-1742

Zhang, L. and Scifleet, P.
Charting the Transformation of Enterprise Information Management: AI Explainability and Transparency in EIM Practice.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 60-73
ISBN: 978-989-758-716-0; ISSN: 2184-3228

needs and interests of IM practitioners. To address this need, this study investigates the current state of AI use in EIM systems and the changes this is bringing to information management practice, this paper presents the findings from an environmental scan of AI integration into EIM platforms, focusing on how understandable new AI-based solutions are for practitioners and, on the characteristics required for explainable AI in EIM.

The research findings present current issues and challenges in EIM as prioritised by 20 leading EIM platform providers between August 2022 and November 2023. Research outcomes include the categorisation of AI use by platforms into five areas required to support the explainability and understandability of AI for practitioners seeking to adopt new approaches. These include describing 1) how *AI development* is taking shape, 2) what underlying *AI techniques are being used*, 3) how *AI integration* maps to *EIM capabilities*, 4) what *AI applications* are available for use, and 5) how *AI impacts* on EIM practices. Moreover, to evaluate the extent to which information provided by the 20 AI-integrated EIM platforms is clear and transparent for practitioners, an outcome of this research is a model for evaluating AI transparency in EIM based on six practical criteria: *1) Provision of AI development details 2) Provision of AI function details 3) Provision of AI impacts (benefits & risks) 4) Provision of real-world use cases 5) User experience design for AI-integrated interface 6) Human-AI interaction.*

The contribution of the research is twofold. Firstly, it adds to knowledge in EIM by uncovering the role and impact of AI in EIM practices, with five practice-based categories to improve the description and understanding of AI integration in EIM recommended. This offers a practical contribution for EIM practitioners seeking to leverage AI in their work, and is supported by a further six criteria for AI transparency, developed as an outcome of this research, that both vendors and practitioners can work towards achieving. Both contribute to the field of Explainable AI by addressing the needs of non-experts seeking to work with AI and, through this, promoting human agency in explainable AI (XAI).

The following sections of the paper discuss key topics in related research, followed by an elaboration of the research design and findings. The paper then concludes with a discussion of the current state of AI-led changes in EIM practices and reflects on how AI-integrated EIM practices can be facilitated.

2 RELATED WORK

2.1 EIM Issues and Challenges

Enterprise Information Management is an overarching concept that encompasses a range of related information systems and information management work practices including Enterprise Content Management (ECM), Electronic Records Management (ERM), Document Management (DM), and Knowledge Management (KM) (AIIM, 2024). Notably, these terms are often used interchangeably in industry (Scifleet et al., 2023). EIM can be broadly defined as the integrated, enterprise-wide, strategic management of all types (physical, digital, differing sources and formats) of enterprise information assets over their entire lifecycle of business use (Jaakonmäki et al., 2018; Williams et al., 2014). The term *information asset* encompasses all data, information, documents, and records required for the everyday work practices of business (Scifleet et al. 2023). Hausmann et al.'s research on enterprise information readiness further summarises EIM as a comprehensive initiative for managing information assets throughout the entire lifecycle to unlock value, with a focus on ensuring regulatory compliance. A key goal is to eliminate information silos across business departments and areas of work, ensuring the availability of well-structured information when needed (Hausmann et al., 2014).

Notably, many of the EIM issues identified for business in an industry survey by Hausmann and Williams et al. in 2014 remain relevant (Hausmann et al., 2014; Williams et al., 2014), if not more pronounced following a survey conducted a decade later (Scifleet et al. 2023). Hausmann et al., (2014) identified the challenges for practitioners in managing an increased volume and variety of business information and prioritised compliance and assurance as central with new technologies and new types of data, such as social media impacting businesses. Both the 2014 and 2023 survey results revealed that while enterprises self-rated highly in achieving conformance goals, they continued to struggle when working with information to achieve performance objectives requiring timely access to critical information, sharing information (both internally and externally), managing the information lifecycle, deriving value from, and delivering actionable business intelligence (Hausmann et al., 2014; Scifleet et al., 2023; Williams et al., 2014). Additionally, the 2023 survey highlights the compounding and negative effect of an overwhelming growth in employee-created data through an increasing number of applications.

In this research, we have built from the many well-known challenges presented by Hausmann et al (2014) and others to revisit priorities current in practice today: with a focus on the role that platform providers can play by providing advanced, transparent, understandable and usable AI-based solutions for acquiring, organizing, storing, retrieving, and sharing information assets within an organization. EIM systems are beginning to integrate AI across areas of information management that are traditionally labour-intensive and hard to achieve, e.g. taxonomy development, classification, process automation, search and findability (Duranti et al., 2022). Despite the promising features that AI can bring to the EIM space, we still lack a holistic understanding of how AI is positioned in EIM practice. Understanding the steps that are being undertaken to achieve AI integration and transform EIM practice is a critical area for research.

2.2 AI and ML in the EIM Context

Within the EIM field, we have found the term *Artificial Intelligence (AI*) serving as an overarching concept encapsulating many aspects of cognitive computing and advanced programming aimed at performing tasks that typically require human intelligence, though in most cases it is being used to describe narrow, task-specific uses of AI (Meske et al., 2022). As a subset of AI, *Machine Learning (ML)* specifically focuses on developing algorithms and models that enable computers to learn from data, allowing them to undertake information processing to deliver outputs and make predictions, or decisions, without the need for additional explicit human programming or human intervention (Barredo Arrieta et al., 2020). Martens and Provost (2014) have demonstrated how ML can be applied to document classification tasks by selecting a minimal set of words to successfully classify document features for management purposes. In a study by Ragano et al. (2022), semi-supervised ML models were used to evaluate audio quality in digital sound archives, demonstrating potential benefits for other multimedia, such as video calls and streaming services. Sambetbayeva et al. (2022) highlight the document management challenges in the context of the data deluge and propose the use of ML techniques to enhance document retrieval. The potential of AI tools for managing digital records has gained widespread recognition in records management with Duranti et al' (2022) and others highlighting that the requirements for collecting and indexing digital records in a reproducible manner far exceed manual capabilities (Duranti et al., 2022; Schneider et al., 2019). AI technologies, such as Recurrent Neural Networks (RNN) (Shabou et al., 2020), Handwritten Text Recognition (HTR) (Goudarouli et al., 2019), and Chatbots (Gupta & Kapoor, 2020) are all being proposed as means for reducing labour intensive work, increasing efficiency and effectiveness with examples of their role in facilitating record classifications, access to paper-based archival information, and establishing new knowledge available. What is at issue, is just how understandable and useful these technologies are for IM practitioners in everyday work, where they are tasked with explaining application, use, and value to the enterprise?

2.3 Explainable AI (XAI) and AI Transparency

The practical implementation of AI and ML in real-world business settings is often met with scepticism by industry professionals (Modiba, 2023). Predominately this scepticism stems from perceptions of AI as an uninterpretable "black box", with significant concerns about transparency, trustworthiness, and a need to improve understanding of how AI systems produce their outcomes (Adadi & Berrada, 2018). Consequently, there is a growing prioritisation for *Explainable AI (XAI)*, with an overarching goal of improving the accessibility of intelligent systems (Meske et al., 2022). The concept of XAI demands a better explanation about how AI-generated outcomes are achieved. This has resulted in substantial progress within the AI community, where XAI is seen as a sub-field of AI (Adadi & Berrada, 2018; Barredo Arrieta et al., 2020) that aims to provide users with the ability to see inside the black-box. This is typically facilitated through the lens of another algorithm that has the role of describing the logic and decision-making processes of the AI and generating a report that confirms its operations. However, in turning to computational solutions to read an algorithm and and report on the veracity of black-boxed behaviours so that the AI can be trusted, for example in a legal claim, we are at risk of using one opaque method to describe another with little gain for practitioners (Adadi & Berrada, 2018). The Human-Computer Interaction (HCI) community has approached this differently, by bringing a human-centred perspective to XAI that emphasises the significant role of humans in the explanatory process, arguing the importance of being able to engage different user-stakeholder groups and their needs in AI development and use (Meske et al., 2022).

HCI research has focussed on understanding users' perceptions of XAI methods, with research evaluating different HCI-XAI methods for their explanatory capabilities: having the right method in place to explain AI to a user can, in turn, contribute to better more understandable AI systems design (Wang & Yin, 2021; Wanner et al., 2022). Even so, many approaches to XAI still remain technical and often struggle to be translated into practical implementations when it comes to assisting non-expert users in making decisions about AI in work-based contexts (De et al., 2020). The need to improve explainability for non-experts from a practice perspective remains (Brennen, 2020; Bunn, 2020).

EIM practitioners' engagement with AI will be through those tools in EIM platforms that they use to meet their daily information management needs. AI will be assessed on the practical benefits gained. Instead of delving into detailed micro-level explanations of AI models, there is a need to link the *understandability of AI* to EIM practices, with an emphasis on being able to identify and explain system-level applications in AI-integrated EIM transparently, in as open and clear way as possible for practitioners.

AI transparency has been identified as a key requirement for AI technologies by the AI HLEG (the European Commission's High-Level Expert Group on Artificial Intelligence), and the transparency of data processing in AI applications is mandated by the GDPR (General Data Protection Regulation) (Felzmann et al., 2019). Broadly referring to the principle of making AI systems understandable, explainable, and accountable, AI transparency concerns disclosing information about an AI system, typically to support judgments regarding fairness, trustworthiness, safety, efficacy, accountability, and compliance with regulatory and legislative frameworks (Andrada et al., 2023). The problem with AI's "black box" is a clear lack of transparency; however, the concept of AI transparency itself is often opaque (Kiseleva et al., 2022). AI transparency research primarily targets algorithmic transparency, aiming to provide visibility into the underlying algorithms and neural networks to help rationalize the outcomes produced by complex programming (Andrada et al., 2023). However, algorithmic transparency alone does not address the needs of different AI stakeholder groups and thus fails to make AI systems more understandable to non-experts (Felzmann et al., 2019; Haresamudram et al., 2023). To clarify different types of AI transparency and what greater transparency might entail, Andrada et al. (2022) offer a taxonomy that includes two main types of AI transparency: reflective transparency and transparency-in-use. Reflective transparency encompasses information transparency, material transparency, and transformation transparency. Transparency-in-use focuses on ensuring the interface itself is intuitive, allowing users to understand and navigate systems to complete their tasks. Similarly, Haresamudram et al. (2023) have proposed three levels of transparency relevant to diverse stakeholder groups and contexts: 1) algorithmic transparency, 2) interaction transparency, and 3) social transparency, however the categories remain broad. Despite a better understanding of the different aspects of AI transparency for various stakeholder groups, research on how AI transparency translates into applied settings is limited (Haresamudram et al., 2023). This study addresses the gap by operationalizing AI transparency with the EIM context in mind, providing a set of six practical evaluation criteria for AI transparency.

3 RESEARCH METHODOLOGY AND DESIGN

3.1 Research Objectives

The integration of AI into EIM platforms will transform information management practice not simply as a by-product of smarter off-the-shelf automation, but because EIM practitioners will adapt to the use of AI by altering the conventions and practical knowledge of everyday work in EIM, and by contributing to the design of AI solutions specific to EIM (De Certeau & Mayol, 1998) (Fensham et al., 2020). This research then, contributes to an improved understanding of the transformative role of AI and its impact by investigating the current state of AI-led changes to practice and presents the foundations for a practice-based descriptive framework for explainable and transparent AI in EIM. The approach taken is based on a sociotechnical and practice theory perspective (Orlikowski & Scott, 2016), holding the viewpoint that both people (the practitioners) and technologies (the platforms) have the agency to influence and shape each other, and addresses the following research objectives (RO) and questions (RQ):

RO1 – To understand current EIM challenges faced by practitioners in the massive data environment.

- RQ1(a) What current EIM issues are faced by practitioners?
- RQ1(b) How are the issues in EIM practice being presented by platform vendors?

RO2 – To describe how AI is being brought into EIM practice by EIM platforms.

- RQ2(a) How is AI for EIM being developed and integrated into platforms?
- RQ2(b) What are the impacts, benefits, and advantages that AI is having on the delivery of EIM services?

RO3 – To devise a working definition of AI explainability and transparency for AI-integrated EIM practices and evaluate the understandability of AI-led changes in EIM platforms.

- RQ3(a) What does AI explainability and transparency mean in applied EIM contexts?
- RQ3(b) How transparent is the information provided by the platforms regarding AI applications?

3.2 Research Design

This study's approach is qualitative and based on the environmental scanning (ES) of publicly available Web resources to establish awareness of products, services and strategies constituting the relatively new and emergent delivery of AI in EIM platforms. ES, which has its roots in business analysis, is a method applied by researchers to gather and analyse information concerning the domain of interest from publicly available resources to establish situational awareness of the environment and plan actions accordingly (Auster & Choo, 1994; Zhang et al., 2011). While initially applied by businesses for strategic purposes there has been a shift in the use of ES in recent years to academic research, where the identification and analysis of current, publicly available information resources is critical to research domain awareness (Lau et al., 2020; Yin et al., 2021). As the AI-based changes that are taking place in EIM are arriving from vendors integrating AI into their platforms as products for clients, scanning publicly available information from vendors' websites provides this study with a method appropriate for establishing domain awareness and a starting point for understanding the changes that AI integration in EIM brings.

Data collection for the environmental scan was undertaken in two stages between August 2022 and October 2023, following a series of steps outlined for qualitative media analysis (Altheide & Schneider, 2013), depicted in Figure 1.

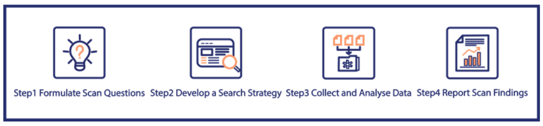


Figure 1: Steps in Environmental Scanning.

Stage one resulted in data collection from 28 leading EIM platforms between August - September 2022, with stage two following in August- October 2023 concentrating on a subset of 20 EIM platforms that were identified from the first stage, because of their more focused discussion of AI integration. While not initially planned for, this allowed the research to map a significant industry change corresponding to the public release of ChatGPT and other OpenAI initiatives from November 2022 (OpenAI, 2022), with a burst of discussion about AI and AI impacts taking place in EIM following ChatGPT hype.

Steps 1 & 2 Scan Questions and Search Strategy

Starting an environmental scan involves initiating a search strategy that is based on the research questions. To locate relevant vendors, platforms and service providers, we applied a search strategy comprising broad terms, main terms and related terms, relevant to the focus. Undertaking a Google search with the broadest concepts first, including information management, information management services and information management service providers. The same approach was taken for closely related service areas of, content management, document management, records management and knowledge management.

Following initial search results, information was collected from 140 pages, with the scan identifying more than 70 EM companies including 42 service providers and 28 platform providers. We consider EIM platform providers to be companies that design, develop, and provide integrated technology platform solutions (all software, database, network and cloud components) e.g., OpenText, Oracle NetSuite, Objective, Hyland, Microsoft M365, and EIM service providers as companies who focus on the provision of information management consultancy services, e.g. Access, Astral, TIMG, Cube Records Management, Document Logistix are examples. While EIM service providers also undertake software development to further customise major platforms for clients, they are not platform developers. For the purposes of this study, our starting point to examining changes in practice is the arrival of AI in EIM platforms.

We next focussed on applying inclusion and exclusion criteria to refine the list of EIM platform providers to a list of platforms with Web resources describing explicitly, the incorporation of AI into their offerings. As a result, a pool of 20 AI-integrated EIM vendor platforms was selected for further data collection and analysis

Steps 3 & 4 Data Collection, Analysis and Reporting

The unit of analysis for the study comprised three main components of information typically available at each of EIM platforms' websites: 1) platform overviews, 2) AI feature descriptions, and 3) real-world AI-based case studies. The presence of real-world AI-based case studies showcased by the vendors serves as a significant indicator of AI transparency, by describing real EIM use cases for AI.

A "three-level" data collection strategy was implemented for collecting data from the vendors. Navigating "three-levels deep" from the platform overview page ensured that data collection was independent from the platform's website homepage and architecture and helped to locate the same type of information consistently. The "three-levels" are defined as: Level 1, providing an overview of the EIM platform; Level 2, offering a general insight into AI features; and Level 3, providing specific details about AI features. A data collection template was used to ensure the same details were collected for each platform including, platform provider name, platform name, collection date, URLs, platform overview, AI feature descriptions, and the presence of real-world AI-based case studies. The raw data collected for each vendor was initially saved to a Word document using the template and then imported into NVivo for further analysis. Excel has been used for supporting analysis and to assist in the visual presentation of the findings.

Thematic content analysis was used to analyse the collected data, following an inductive, ground-up approach in NVivo while acknowledging that the starting questions and topics have framed the analysis (Williamson & Johanson, 2018). The coding process commenced without preconceived categories or themes about AI integration in EIM; instead, the AI-specific categories presented emerged from the research findings during coding. This was supported by sense-making and fact-checking aligned with the current state of cognitive computing, AI, and ML. To ensure reliability, the research team regularly discussed and refined the coding, with inter-coder reliability checks conducted to reach consensus on themes, topics, and categories, leading to a rigorous process of topic reduction and confirmation.

4 RESULTS

4.1 EIM Issues and Challenges

The thematic content analysis identified eight EIM issues for practitioners as described by the 20 EIM platforms: immature digitalisation, information security, privacy, and compliance (ISPC) risk, information silos, poor information findability, information overload, poor information sharing, poor information utilisation, and information quality concerns (Figure 2).

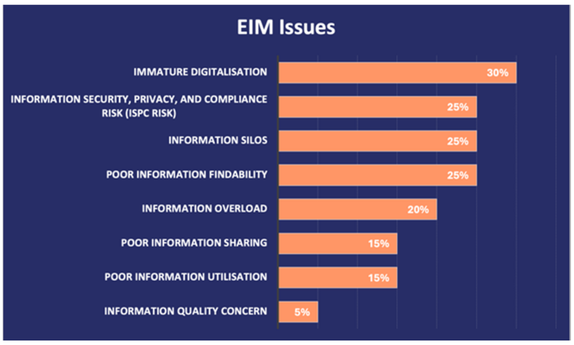


Figure 2: EIM Issues as prioritised by EIM Platforms.

Immature digitalisation (N=6, 30%) refers to lacking sufficient digitalising capabilities to leverage digital technologies for the automation of manual 'paper-driven' processes. Although OCR techniques have been widely utilised for converting documents into digital forms, full automation remains a critical challenge for many enterprises in managing their information assets.

Information security, privacy, and compliance (ISCP) risk (N=5, 25%) refers to the challenges faced by enterprises in meeting information security, privacy and regulatory compliance over their information assets. While flexible and improved collaboration enables better productivity, unauthorised access to information and lack of proper control continue to cause data breaches. Enterprises also raise concerns about identifying and protecting their customers' sensitive and personally identifiable information (PII).

Information silos (N=5, 25%) and poor information findability (N=5, 25%) are closely linked issues that significantly impact information sharing (N=3, 15%). Information silos occur when data is stored in multiple or geographically dispersed locations, with fragmentation resulting from information being spread across different systems, tools, siloed repositories, and disconnected end-line-of-business applications. Siloed information leads to poor information findability, making it difficult for employees to access information when needed. This impacts productivity and hinders both internal and external information sharing, resulting in lower productivity and higher information risks.

Information overload is another critical issue identified (N=4, 20%), that arises from the overwhelming scale of information coming from different sources and channels, and is closely linked, in platform discussion, with poor information utilisation (N=3, 15%). This includes challenges in knowledge discovery and obtaining data-driven business insights.

While not prioritised by EIM platforms (N=1, 5%), information quality concerns must be seen as remaining critical. Enterprises raise concerns about the quality of their information and the trustworthiness of decisions made based on this (Scifleet, 2023). The accuracy and reliability of analytical insights always depend heavily on data quality, and the success of AI projects in enterprises will rely heavily on the quality of the datasets that AI models are trained with.

4.2 AI Explainability in EIM Platforms

Table 1 categorises this study's findings across five key areas for explainability (XAI) that practitioners can consider when seeking to understand and evaluate AI use in EIM platforms: AI Development, AI Techniques, AI-integrated EIM Capabilities, AI Applications in EIM, and AI Impacts. These areas support the explainability and understandability of AI applications for EIM practitioners.

Table 1: Five categories of AI use in EIM.

Category	Definition
AI Development	Refers to *the way AI is developed* inhouse or adopted by an EIM platform, to develop specific AI solutions and AI model training.
AI Techniques	Refers to *types of approaches* (e.g. computer vision, generative AI, deep learning) employed in EIM platforms' AI offerings.
AI-integrated EIM Capabilities	Refers to the *EIM capabilities* for managing enterprise information assets through AI integration, e.g. AI-powered information capture.
AI Applications in EIM	Refers to the *underlying AI applications* that support EIM capabilities e.g. Automated data classification, Automated workflows.
AI Impacts	Refers to *benefits and advantages* identified for integrating AI into EIM practices across the EIM lifecycle.

4.2.1 AI Development

We consider two sub-categories important as part of AI Development: AI solutions and AI training and deployment. AI solutions refer to the integration of an AI capability directly into an EIM platform either as native AI solutions (developed in-house) or by including third-party AI solutions. Among the 20 platforms analysed, 12 are explicit about how they are developing AI, while 8 are not. That 40% do not share this level of detail remains a significant explainability concern for practitioners. As shown in Figure 3, the majority of the 12 platforms take a Native-AI approach (N=9, 75%) to development. Others adopt a third-party AI solution (N=3, 25%), working with well-known AI service providers, including Clarifai, Microsoft Cognitive Services and OpenAI, to offer AI capabilities to their clients.

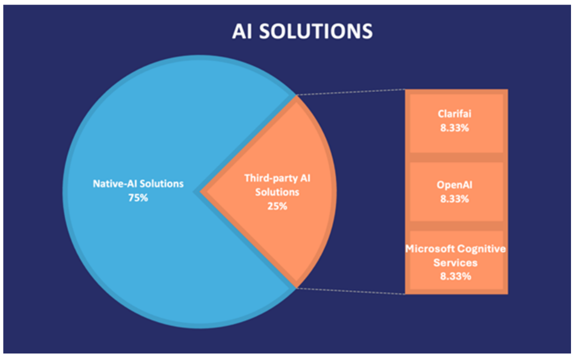


Figure 3: AI Development – AI Solutions.

Training and deployment encompass key aspects of AI model development and use, including data use with AI, human-AI interaction, pre-built AI models, customisation capability, model performance improvement, explainable features, and AI limitations (Figure 4).

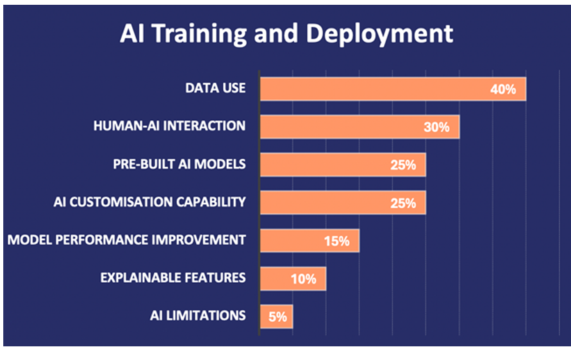


Figure 4: AI Development - AI Training and Deployment.

Data use with AI (N=8, 40%) involves datasets for training AI models and data use by deployed AI products. *Understandability*, about how AI works with data is crucial for addressing concerns regarding the security and sensitivity of proprietary datasets within enterprises. While some platforms clarify that datasets are used with enterprises' consent for model training purposes, concerns remain about the security and sensitivity of data used in deployed AI products.

Human-AI interaction (N=6, 30%) represents a key feature of AI development in the EIM platforms, where the interaction between humans and AI can take various forms. For example, humans can validate AI-produced results and provide corrections or feedback to help AI systems improve continuously.

Pre-built AI models (N=5, 25%) include pre-built or out-of-the-box AI models that are shipped with the platform. Without the burden of undertaking any further development, practitioners can interact with these pre-built AI models at the application level. However, this does not lower the burden for explainability as some platforms offering pre-built AI models also provide AI customisation capabilities (N=5, 25%), allowing integrated AI applications, including AI models and generative AI prompts, to be customised to specific use cases and business needs. The options for customising AI models vary widely, ranging from allowing practitioners to build an AI model from scratch to simply tuning model parameters or selecting from various models or model versions to achieve optimised results. Model performance improvement (N=3, 15%) includes enabling continuous learning capabilities and incorporating human-in-the-loop verification and feedback based on proper business context.

Explainable features and limitations (N=3, 15%) represent features available on EIM platforms that indicate the accuracy and reliability of AI-generated results. For example, platforms provide accuracy rankings for suggested filing locations, or use different colours to indicate certainty levels in indexing. Despite its importance for improving users' trust in AI-generated outputs, only a few platforms offer this. Additionally, only one platform in our analysis acknowledges the limitations of AI applications, noting that AI performance relies heavily on the quality of the training data used.

4.2.2 AI Techniques

Seven significant sub-categories of AI techniques emerged from the content analysis: Advanced Character Recognition (ACR), Generative AI, AI-integrated Robotic Process Automation (RPA), Computer Vision, Natural Language Processing (NLP), Deep Learning, and Generic AI and ML (Figure 5).

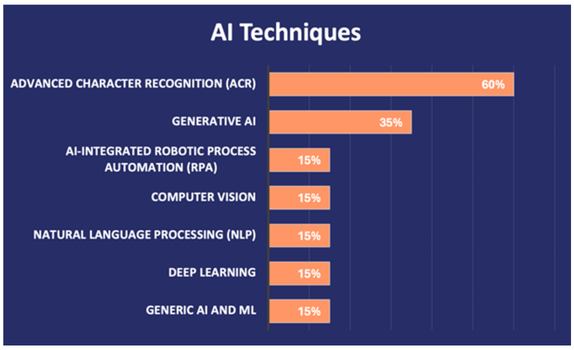


Figure 5: AI Techniques.

Among the seven main categories emerged from AI techniques, it is not surprising that Advanced Character Recognition (ACR) (N=12, 60%) appears to be the most employed AI technique in the solutions offered by EIM platforms. With a long history of using OCR for information capture in EIM, ACR is familiar to practitioners. ACR applies AI in various character recognition technologies including Optical Character Recognition (OCR), Intelligent Character Recognition (ICR), and Zonal OCR. These technologies identify and extract text from images, scanned documents, or other visual sources, converting it into editable, searchable text with high accuracy and efficiency.

The second-ranked AI technique making its way in EIM is Generative AI (N=7, 35%), including large language models (LLMs) and generative AI-based chatbots. Since OpenAI released ChatGPT in November 2022, generative AI has significantly changed the way AI tools are thought about for work tasks and are now serving multiple purposes such as generating text, images, and audio and videos. The integration of generative AI chatbots in EIM platforms is transforming EIM practices, including search and retrieval, knowledge-based reporting and digital asset management.

Other underlying AI techniques that are being brought to EIM include AI-integrated Robotic Process Automation (RPA), Computer Vision, Natural Language Processing (NLP), and Deep Learning. While disclosing information about these AI techniques across the platforms provides some transparency for practitioners, details can be dense in terms of applying specific AI techniques across the EIM Lifecycle to improve understandability. Adding to this problem, we found that some platforms are using generic AI and ML terms (N=3, 15%) without explanation, providing no useful information to help practitioners determine the use of these tools for specific IM needs.

4.2.3 AI-Integrated EIM Capabilities

Six sub-categories of EIM capabilities that result from AI inclusions were identified from the analysis, we refer to these as AI *powering* of the capability: AI-powered Business Process Automation (BPA), AI-powered Information Capture, AI-powered Information Search and Retrieval, AI-powered Information Security, Privacy and Compliance (ISPC), AI-powered Business Intelligence, and AI-powered eDiscovery (Figure 6).

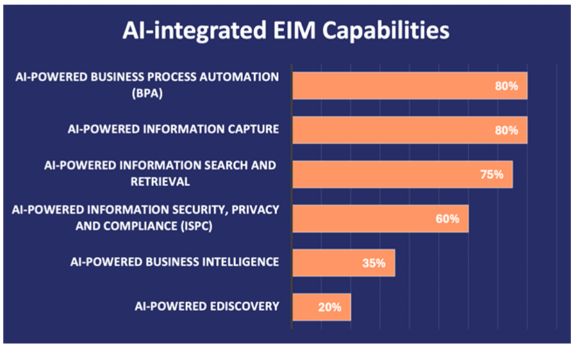


Figure 6: AI-integrated EIM Capabilities.

AI-powered Business Process Automation (BPA) identified across almost all platforms (N=16, 80%), refers to using AI to automate and streamline business processes, e.g. automating repetitive tasks and workflows, and aligns well with the most common EIM issue of immature digitalization, particularly for automating manual processes.

AI-powered Information Capture (N=16, 80%) refers to the application of AI to automate the collection, conversion, organisation and filing of data from source materials. For instance, ACR is often used in AI-powered information capture to process large volumes of scanned paper documents, making them readily available for downstream processing.

AI-powered Information Search and Retrieval (N=15, 75%) involves the application of AI in information search and retrieval processes, including natural language and semantic search across textual, visual, and multimedia data (text, image, video, and audio files). The application of computer vision for image searching is increasing and generative AI-based chatbots are providing new interfaces for employee search queries.

AI-powered Information Security, Privacy, and Compliance (ISPC) (N=12, 60%) involves applying AI to information security, compliance and governance in EIM. This includes applying AI to information security tasks, e.g. detecting threats, and anomalies for data protection. Additionally, AI is utilized for identifying and ranking sensitive and confidential information, including Personally Identifiable Information (PII) and proprietary information. Importantly, AI is being applied in information governance and compliance by automating metadata management and retention and disposal schedules.

AI-powered Business Intelligence (BI) (N=7, 35%) refers to utilising AI for various data analysis and reporting tasks, including relationship analysis, sentiment and behavioural analysis. AI techniques like NLP are used to analyse large volumes of text, identifying relationships across documents and records that are not otherwise apparent.

AI-powered eDiscovery (N=4, 20%) refers to the integration of AI technologies into the process of identifying, preserving, collecting, reviewing, and producing electronically stored information (ESI) for use in legal contexts, e.g. court proceedings, investigations, and other compliance matters. AI can be applied to automate and enhance tasks that would traditionally be time-consuming and labour-intensive, such as identifying required documents based on keywords, concepts, or patterns that occur in a document or even predicting the relevance of documents to a particular court case.

4.2.4 AI Applications in EIM

Closely related to AI-integrated EIM capabilities, AI applications in EIM represent the underlying applications of specific AI techniques that support the high-level AI-integrated EIM capabilities. As shown in Figure 7, eight sub-categories emerged from the content analysis: Automated Workflows, Automated Data Classification, Automated Content Creation, Automated Information Recognition, AI Analytics, Help, Assistance and Recommendation Services, Automated Translation and Automated Security Monitoring. Notably, a single AI-integrated EIM capability can be supported by multiple AI applications. For instance, AI-powered information capture is supported by automated information recognition, automated data classification and automated workflows simultaneously, highlighting the complex nature of AI-integrated EIM practices.

Automated workflows (N=18, 90%) are the most common AI applications, where AI is leveraged to automate various workflows, including document workflow and document control workflow. These applications help automate repetitive, manual tasks, allowing employees to focus on higher-priority activities. This aligns with the most common AI-integrated EIM capability: AI-powered Business Process Automation (BPA).

Automated data classification (N=16, 80%) utilizes AI to classify data based on its content, context, or other attributes. This includes tasks such as metadata generation, auto-tagging, auto-indexing, and document classification.

Automated content creation (N=15, 75%) uses AI to generate content, including converting paper-based documents into fully searchable digital documents, automatic form creation and completion, document summarization for reporting and knowledge creation, transcript generation from speech to text, and alt text generation for images.

Automated Information Recognition (N=13, 65%) uses AI to identify information or patterns, such as passport numbers, phone numbers, or driver's licenses from documents. This includes tasks like data extraction and data validation.

AI Analytics (N=8, 40%) use AI to derive insights from data and processes, including content analytics (text analytics, image analytics, audio and video analytics), sentiment analytics (intention analysis and behaviour analysis), and process analytics.

Help, Assistance, and Recommendation Services (N=7, 35%) use AI to provide users with various recommendations. This includes suggesting similar assets, recommending file storage options, providing visualization suggestions and offering transformation suggestions.

Automated Translation (N=3, 15%) uses AI to translate text or speech from one language to another. This capability enhances global business reach and information sharing across different languages.

Automated Security Monitoring (N=2, 10%) uses AI to identify and detect anomalies and threats in content, generating timely alerts to users to protect against data loss. This includes tasks such as anomaly and threat detection, security alert generation, and containing data leakage.

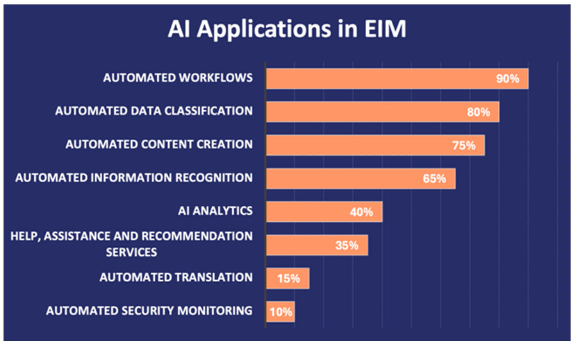


Figure 7: AI Applications in EIM.

4.2.5 AI Impacts

Six key categories identifying how AI integration impacts EIM practices were found in the analysis: AI-workplace benefits, AI-enterprise strategic, financial and reputational benefits, AI-user experience benefits, AI-information security, privacy and compliance (ISPC) benefits, AI-information quality improvements, AI-collaboration improvements, AI-customer gains, AI-business sustainability and continuity (Figure 8).

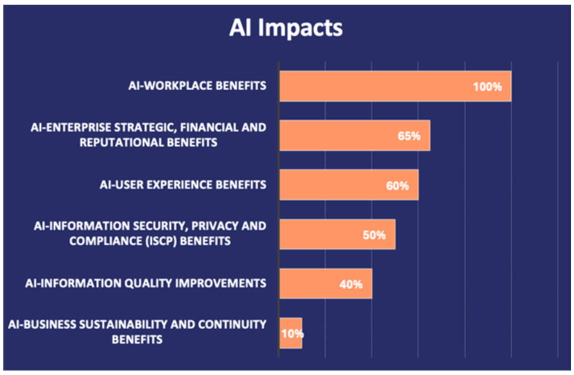


Figure 8: AI Impacts on EIM Practices.

Among these, AI workplace benefits (N=20, 100%) stand out prominently as advantages described across all EIM platforms, including operational benefits, collaboration improvements and employee gains. This aligns with a commonly listed value proposition for AI, that it enhances business operations and employee performance. Associated with this, AI-Enterprise strategic, financial and reputational benefits (N=13, 65%) feature highly, including strategic perspective (greater competitive advantage, more informed decision-making, and richer business insights for uncovering new business opportunities), financial perspective (costsaving and increased ROI), and reputational perspective (brand consistency, cohesive and unified brand experience and customer gains).

AI-user experience benefits (N=12, 60%) centre around a user-friendly including natural language interfaces, no-code environments, reduced dependency on technical expertise, user autonomy, and self-sufficiency, collectively enhancing the ease of AI adoption for users. However, the connection between both workplace and user benefits to AI technology is rarely clearly presented.

AI-Information security, privacy and compliance benefits (N=10, 50%) include enhanced information security, data loss protection, privacy, compliance and regulation adherence, and governance. AI-Information Quality Improvements (N=8, 40%) refers to the improvements regarding to all aspects of Information quality, including integrity, accuracy of data, completeness of data, reduced data errors, and more.

4.3 AI Transparency in EIM

Based on the findings and literature, this study proposes that a contextual working definition of *AI transparency* in EIM that enables the evaluation of transparency in AI-integrated EIM offerings is needed. Critically we find that AI Transparency in EIM solutions must encompass *information transparency* (disclosing information relevant to EIM practitioners) and *transparency-in-use* (intuitive user interface and human in the loop) to enable a better explainability: for understanding, trust and adoption of AI applications for EIM practitioners, and note the interdependency in these concepts.

4.3.1 AI Transparency Evaluation

To apply AI transparency in EIM, this study has developed six practice-based evaluation criteria that can be used by practitioners: *information transparency* – 1) Provision of AI development details (AI development and AI techniques), 2) Provision of AI function details (AI-integrated EIM Capabilities and AI applications), 3) Provision of AI impacts (Benefits & Risks), 4) Provision of real-world use cases, and *transparency-in-use* – 5) User experience design for AI-integrated interface, 6) Human-AI interaction.

Table 2 evaluates the transparency of current AI-integrated EIM solutions across the 20 platforms in this study, based on these criteria[1]. In terms of information transparency, all platforms offer insights into how AI can support EIM capabilities and the benefits it brings to EIM practices. However, information regarding AI development and the underlying AI techniques used is not fully disclosed, with only a few platforms providing details regarding data use with AI, including how datasets are utilised in AI model training or operational deployments. This lack of transparency on critical technical aspects such as AI techniques and data handling, can raise significant concerns for practitioners, particularly around security, trust and AI adoption in EIM.

Notably, AI risk and proven AI success use cases in the real world are rarely provided by platforms, which can hinder the trust and adoption of AI-integrated EIM offerings among practitioners. Additionally, while most platforms highlight benefits like intuitive, user-friendly interface designs for practitioners to work with their AI products, there is a lack of clarity regarding how humans can align AI with their work practice.

[1] Platform vendors are not identified by name in the Table 2, details can be provided by the authors on application.

5 CONCLUSION

5.1 Discussion of Findings

Through an environmental scan of 20 leading EIM vendor platforms, this study identified eight interrelated EIM workplace challenges prioritised by the platforms. Compared to the EIM challenges faced by enterprises a decade ago, these issues remain unresolved, if not more problematic. With the emergence and widespread use of generative AI, there EIM systems are turning to AI to address these challenges.

To understand how AI is transforming EIM practices, this study starts by examining what is available to practitioners, charting the role and impact of AI across five areas: AI development, AI techniques, AI-integrated EIM capabilities, AI applications, and AI impacts. EIM platforms are adopting a native-AI approach and developing AI capabilities internally and this is likely to result in a good fit-for-purpose. However, there is a clear lack of information available to support the understandability of AI's role across the information management lifecycle and this needs to be addressed.

Regarding AI-EIM capabilities and integration into practice; AI-powered information capture, search, and retrieval, supporting consistent information filing and organization, enhancing the discovery, sharing, and use of information, and AI-powered business process automation are all extremely promising. AI integration aims to facilitate automated security monitoring and help address compliance risks. In the face of information overload and information silos, generative AI is valuable for extracting relevant information and curating knowledge tailored to practitioners' needs. However, the risks associated with AI use are rarely mentioned.

5.2 AI Explainability and Transparency in EIM

The study's findings address the need for contextualized AI explainability and transparency in EIM, and, in addition to developing evaluative categories for understanding AI, we have presented a working definition of AI transparency for EIM practitioners with six criteria covering *information transparency* and *transparency-in-use* available for consideration. By evaluating the transparency of AI-integrated EIM solutions offered by vendor platforms

(Table 2), we find that AI transparency can be improved to facilitate explainability and understanding, trust, and adoption of AI by practitioners. When working with AI, practitioners might not be interested in all the details but require proven success in real-world scenarios that address similar EIM needs.

Regarding ease of use and confidence in AI-produced outcomes, more transparency is needed about how practitioners can interact with AI. This includes understanding how to determine the accuracy or confidence in the results produced by AI systems and how practitioners can be included in the loop to improve this process.

5.3 Limitations and Outlook

There are two limitations to this work. Firstly, the analysis is based on scanning the publicly available information provided on the websites of 20 leading EIM platforms. Both the size of the sample and the information sourced are limited. Future work may incorporate more public discourse on AI-integrated EIM offerings, such as industry reports and platform blogs, to achieve a more comprehensive analysis. Secondly, while this study explains the need for contextualized AI transparency *for EIM practitioners* and proposes a working definition of AI explainability and transparency with evaluation criteria, more work is needed to verify these criteria *with EIM practitioners*. That constitutes the second stage of this study. Practitioner interviews have been completed and will be reported at a later stage. This research takes the first step towards making AI applications more understandable, explainable, and transparent in EIM. This work also contributes to the field of Explainable AI by addressing the needs of non-experts in applying and working with AI, thereby promoting human agency in explainable AI (XAI) initiatives.

Table 2: Evaluation of AI transparency in EIM.

Transparency Evaluation criteria								
	Information Transparency						Transparency-in-use	
# Platform	1) Provision of AI development details		2) Provision of AI function details	3) Provision of AI impacts		4) Provision of real-world use cases	5) User experience design for AI-integrated interface	6) Human-AI interaction
	AI Development	AI Techniques		AI Benefits	AI Risk			
#1	√	√	√	√			√	
#2		√	√	√				
#3	√		√	√			√	√
#4		√	√	√			√	
#5		√	√	√			√	
#6			√	√				
#7	√	√	√	√		√	√	√
#8	√	√	√	√		√	√	
#9	√		√	√		√		
#10	√	√	√	√				
#11	√	√	√	√	√		√	√
#12		√	√	√			√	
#13	√	√	√	√				
#14	√	√	√	√		√	√	
#15	√	√	√	√		√		
#16	√	√	√	√			√	√
#17	√	√	√	√			√	
#18	√	√	√	√				
#19	√	√	√	√		√	√	√
#20	√	√	√	√				√

ACKNOWLEDGEMENTS

This research was funded by the ARC Industrial Transformation Training Centre for Information Resilience (CIRES). The authors gratefully acknowledge the support provided by CIRES, which has been instrumental in conducting this work.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

AIIM. (2024, April 5). *Intelligent Information Management Glossary.* https://www.aiim.org/what-is-information-management

Altheide, D. L., & Schneider, C. J. (2013). Qualitative media analysis (2nd edition ed.). Sage.

Andrada, G., Clowes, R. W., & Smart, P. R. (2023). Varieties of transparency: exploring agency within AI systems. AI & SOCIETY, 38(4), 1321-1331. https://doi.org/10.1007/s00146-021-01326-6

Auster, E., & Choo, C. W. (1994). How senior managers acquire and use information in environmental scanning. Information Processing & Management, 30(5), 607-618.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI [Article]. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.20 19.12.012

Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. V. (2021). Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. IEEE Access, 9, 72894-72936.

Brennen, A. (2020). What do people really want when they say they want "explainable AI?" we asked 60 stakeholders. Conference on Human Factors in Computing Systems - Proceedings,

Bunn, J. (2020). Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). Records Management Journal, 30(2), 143-153.

De Certeau, M., & Mayol, P. (1998). The Practice of Everyday Life: Living and Cooking. Volume 2 (Vol. 2). U of Minnesota Press.

De, T., Giri, P., Mevawala, A., Nemani, R., & Deo, A. (2020). Explainable AI: A Hybrid Approach to Generate Human-Interpretable Explanation for Deep Learning Prediction. Procedia Computer Science, 168, 40-48. https://doi.org/https://doi.org/10.1016/j.procs.2020.02.255

Duranti, L., Abdul-Mageed, M., Hofman, D., & Sullivan, P. (2022). I Trust AI, the latest InterPARES research project. Anuario Escuela de Archivología(13), 36-55.

Felzmann, H., Fosch Villaronga, E., Lutz, C., & Tamò Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. 6, 1-14. https://doi.org/10.1177/2053951719860542

Fensham, R., Threadgold, T., Webb, J., Schirato, T., & Danaher, G. (2020). Understanding Bourdieu. Routledge.

Goudarouli, E., Sexton, A., & Sheridan, J. (2019). The challenge of the digital and the future archive: through the lens of the national archives uk. Philosophy & Technology, 32, 173-183.

Gupta, A., & Kapoor, N. (2020). Comprehensiveness of archives: A modern AI-enabled approach to build comprehensive shared cultural heritage. arXiv preprint arXiv:2008.04541.

Haresamudram, K., Larsson, S., & Heintz, F. (2023). Three Levels of AI Transparency. Computer, 56(2), 93-100. https://doi.org/10.1109/MC.2022.3213181

Hausmann, V., Williams, S. P., Hardy, C. A., & Schubert, P. (2014). Enterprise Information Management Readiness: A Survey of Current Issues, Challenges and Strategy. Procedia technology, 16, 42-51. https://doi.org/10.1016/j.protcy.2014.10.066

Huddart, K. (2022). Artificial intelligence powered digital asset management: Current state and future potential. Journal of Digital Media Management, 11(1), 6-17.

Jaakonmäki, R., Simons, A., Müller, O., & vom Brocke, J. (2018). ECM implementations in practice: objectives, processes, and technologies. Journal of Enterprise Information Management, 31(5), 704-723.

Jaillant, L. (2022). How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. Archival Science, 22(3), 417-436.

Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations [Review]. Frontiers in Artificial Intelligence, 5. https://doi.org/10.3389/frai.2022.879603

Kolandaisamy, R., Rajagopal, H., Kolandaisamy, I., & Sinnappan, G. S. (2024). The Smart Document Processing with Artificial Intelligence.

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research [Article]. Artificial Intelligence, 296, Article 103473. https://doi.org/10.1016/j.artint.2021.103473

Lau, F., Antonio, M., Davison, K., Queen, R., & Bryski, K. (2020). An environmental scan of sex and gender in electronic health records: analysis of public information

sources. Journal of Medical Internet Research, 22(11), e20050.
Martens, D., & Provost, F. (2014). Explaining Data-Driven Document Classifications. MIS quarterly, 38(1), 73-100. https://doi.org/10.25300/MISQ/2014/38.1.04
Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. Information systems management, 39(1), 53-63. https://doi.org/10.1080/10580530.2020.1849465
Modiba, M. (2023). User perception on the utilisation of artificial intelligence for the management of records at the council for scientific and industrial research. Collection and Curation, 42(3), 81-87.
OpenAI. (2022, November 30). *Introducing ChatGPT.* https://openai.com/index/chatgpt/
Orlikowski, W. J., & Scott, S. V. (2016). Digital work: A research agenda. A research agenda for management and organization studies, 88-95.
Ragano, A., Benetos, E., & Hines, A. (2022). Automatic Quality Assessment of Digitized and Restored Sound Archives. Journal of the Audio Engineering Society.
Sambetbayeva, M., Kuspanova, I., Yerimbetova, A., Serikbayeva, S., & Bauyrzhanova, S. (2022). Development of Intelligent Electronic Document Management System Model Based on Machine Learning Methods. Eastern-European Journal of Enterprise Technologies, 1(2), 115.
Schneider, J., Adams, C., DeBauche, S., Echols, R., Mckean, C., Moran, J., & Waugh, D. (2019). Appraising, processing, and providing access to email in contemporary literary archives. Archives and manuscripts, 47(3), 305-326.
Scifleet, P., Felsbourg, M., Dang, C., & et al. (2023). Information governance and information management as a service: survey report 2023 [Report]. Swinburne University of Technology. https://apo.org.au/node/323899
Shabou, B. M., Tièche, J., Knafou, J., & Gaudinat, A. (2020). Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data. Records management journal, 30(2), 175-200.
Wang, X., & Yin, M. (2021). Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. International Conference on Intelligent User Interfaces, Proceedings IUI,
Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. Electronic Markets, 32. https://doi.org/10.1007/s12525-022-00593-5
Williams, S. P., Hausmann, V., Hardy, C. A., & Schubert, P. (2014). Managing enterprise information: Meeting performance and conformance objectives in a changing information environment. International journal of information systems and project management, 2(4), 5-36. https://doi.org/10.12821/ijispm020401
Williamson, K., & Johanson, G. (2018). Research methods: information, systems and contexts (Second edition. ed.). Chandos Publishing.
Yin, X., Liu, H., Webster, J., Trieu, K., Huffman, M. D., Miranda, J. J., Marklund, M., Wu, J. H., Cobb, L. K., & Li, K. C. (2021). Availability, formulation, labeling, and price of low-sodium salt worldwide: environmental scan. JMIR public health and surveillance, 7(7), e27423.
Zhang, X., Majid, S., & Foo, S. (2011). The Contribution of Environmental Scanning to Organization Performance. Singapore Journal of Library & Information Management, 40.

Uncertainty Analysis in Population-Based Dynamic Microsimulation Models: A Review of Literature

Miia Rissanen[1,2] and Jyrki Savolainen[1,3]
[1]Lappeenranta-Lahti University of Technology, Yliopistonkatu 34, Lappeenranta, Finland
[2]Keva, Unioninkatu 43, Helsinki, Finland
[3]CSC – IT Center for Science, Tehdaskatu 15, Kajaani, Finland
Miia.rissanen@keva.fi, jyrki.savolainen@lut.fi

Keywords: Dynamic Microsimulation, Population-Based Modeling, Uncertainty Analysis, Monte Carlo Simulation.

Abstract: This paper reviews population-based dynamic microsimulation (DMs) models used in policy analysis and decision support of social systems and demographics. The application of uncertainty analysis (UA) methods is examined focusing on how probabilistic Monte Carlo (MC) simulation technique is being used and reported. Secondly, inspired by the expanding possibilities of data, this analysis examines the models' capability to uncover finer temporal variations beyond traditional yearly intervals and the use of near real-time data in the reported studies. The analysis of the 44 studies included in this preliminary literature review reveals a lack in the rigorous application of UA and transparent communication of results, particularly in the social sciences. Despite the advances of data availability and modeling, no research attempts were found that would indicate a shift of paradigm from historical data-driven models to real-time data. It is suggested that DM studies in this context could benefit from some mutually agreed standardized reporting guidelines for UA. This literature review serves as a preliminary exploration of the topic, highlighting the need for a more comprehensive and systematic survey to thoroughly assess the current state of research.

1 INTRODUCTION

Dynamic microsimulation (DM) models are analytical tools to simulate the behavior of individual units over time and predict recurring events based on historical data. These models integrate data analysis, computational methods, and computer experiments to support ex-ante policy analysis, government planning and decision making. (Brown & Harding, 2002; Harding, 2007; O'Donoghue, 2014; O'Donoghue & Sologon, 2023; Sauerbier, 2002; Spielauer and Duplirez, 2019). Throughout the simulation, each micro-unit, representing diverse population characteristics (e.g., age, employment, health status), evolves independently through stochastic processes, with their states updated over time according to current conditions and attributes—a phenomenon referred to as "dynamic aging" (see e.g., Burgard et al., 2020; Dekkers, 2015).

Many popular DMs (see in detail e.g., Harding, 2007; O'Donoghue, 2001) were initially developed to address concerns about population aging and to assess affordability of the future social protection system. Over the last decade, their applications in health and labour market studies have been growing (O'Donoghue & Dekkers, 2018). Unlike population-aggregating macroscopic approaches, DMs consider individuals separately, which is crucial for understanding the complex interconnections between factors such as demographics, education, employment, and health that influence future economic and health outcomes. For a general introduction to DMs and their applications, the reader is advised to refer to, e.g., O'Donoghue (2001, 2014), O'Donoghue and Dekkers (2018), Klevmarken (2008), and Zaidi and Rake (2001).

Times of uncertainty, such as the Ukraine war, COVID-19 and past financial crises, have created new demands for real-time simulation and "nowcasting" (O'Donoghue & Sologon, 2023; see also Navicke et al., 2014) to facilitate timely decision-making in rapidly evolving economic landscape. Digital trace data from web browsing and mobile applications provide unprecedented regional and temporal data granularity, enabling close-to-real time modeling of social phenomena, such as predicting disease spread (Burgard et al., 2021; Kashyap & Zagheni, 2023; Li et al., 2024; O'Donoghue &

Rissanen, M. and Savolainen, J.
Uncertainty Analysis in Population-Based Dynamic Microsimulation Models: A Review of Literature.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 74-84
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Dekkers, 2018). With more real-time data, simulation models could better capture short-term fluctuations instead of producing predictions only on an annual level, thus hiding seasonal variations and timely insights, e.g., related to healthcare demands or labour force participation. However, it seems common that administrative data used in many popular DMs targeted to public policy analysis (see again e.g., Harding, 2007) typically has a time lag (O'Donoghue & Sologon, 2023), even if such data is generated constantly as by-products of administrative transactions.

Despite the data revolution enhancing simulation capabilities (Crato, 2023; Margetts & Dorobatu, 2023; O'Donoghue & Sologon, 2023), the proper accounting of modeling uncertainty in DMs remains challenging. To address the inherent stochasticity when simulating individual behaviour and demographic and economic changes is complex, particularly given the (too) high expectations for perfect modeling accuracy (Burgard and Schmaus, 2019; Gilbert et al., 2018; O'Donoghue, 2014; O'Donoghue & Dekkers, 2018; Sharif et al., 2012, 2017). In modeling studies, this often shifts the focus from probabilistic thinking back to traditional, deterministic scenario analysis with single-point estimates, although it is well-known (see e.g., Burgard and Schmaus, 2019; Sharif et al., 2012) that for DMs to be useful, they must thoroughly analyze potential impacts on populations under various scenarios. It's crucial to examine not just the outcomes but also the processes leading to them, incorporating comprehensive uncertainty analysis (UA) of various sources of variation. The authors discussing uncertainty and stochasticity in demographic modeling include Alho and Lassila (2023), Xue et al. (2021), Sabelhaus and Topoleski (2007), and Lee and Tuljapurkar (1994).

Monte Carlo (MC) simulation is a key method for handling uncertainty in DMs, offering a robust approach to systematically explore how variations in inputs affect model outputs. This numerical method involves random sampling from distributions and repeated simulations using the sampled values. The Markov Chain Monte Carlo (MCMC) method, in turn, draws mutually dependent samples to generate random sequences of state transitions based on probabilistic rules (e.g., from logit models build on historical data). This process is repeated hundreds or thousands of times to simulate the expected behavior of the object of interest over time, with calibration performed at each step using newly generated parameters. As such, the MC simulation mitigates misinterpretations from single simulations by examining a broad spectrum of possible outcomes, thereby capturing the inherent variability in simulated population dynamics (Burgard and Schmaus, 2019; Marois & Aktas, 2021; Rutter et al., 2011). Confidence intervals (CIs) communicate variability in outcomes, with larger sample sizes and higher number of simulation iterations leading to narrower CIs and more precise estimates (Burgard et al., 2020; Smithson, 2003; Spielauer & Dupriez, 2019).

Previous literature reviews and surveys on DMs, such as O'Donoghue (2014), Li and O'Donoghue (2013), and O'Donoghue and Dekkers (2018), provide a comprehensive overview of DMs developed over decades (see also Spielauer, 2007; Zaidi & Rake, 2001). In past reviews, the lack of standardization in reporting practices and incomplete validation of models stay as ongoing topic (see also Burgard & Schmaus, 2019; Lee et al., 2024). However, past studies have not delved deeper into the use of probabilistic methods, specifically MC approach and related reporting in demography DM studies, although best practices of UA have been proposed by e.g., Burgard and Schmaus (2019), Lee et al. (2024) and Caro (2012). Another gap pertains to the scarcity of literature examining whether enhanced data accessibility in terms of granularity and timeliness have spurred advancements in models capable of delivering more accurate and timely forecasts, compared to "traditional" DMs those run simulations in yearly intervals and are initialized using historical data with a time lag of several years (see O'Donoghue & Sologon 2023).

This paper addresses these mentioned gaps by conducting a preliminary literature review using the Scopus Database, targeting publications from 2000 onwards with "Dynamic Microsimulation" and "Population" or "Demography" in the title, abstract, or keywords. The search was limited to peer-reviewed journals, conference proceedings, books, and reviews in English, yielding 158 results. After content analysis, based on the title and abstract, 44 documents focused on dynamic microsimulation modeling works targeted mainly to model demographic dynamics were selected. In this initial review, the focus is on addressing aspects of uncertainty analysis rather than technical details. Thus, technical/model introduction reports about the model construction (e.g., Andreassen et al., 2020; Münnich et al., 2021.) were not reviewed since these do not focus primarily on conducting simulations, but rather introduce e.g., the modules and data requirements. Also, as the focus is primarily on DMs, some publications utilizing combined micro-macro simulations are not included and, duplicates were also excluded.

In the following section, preliminary findings of the review and discussion with the objective to inspect the scale and scope of the MC applications are provided with aggregated knowledge on the conventions such as number of simulations run and the use of CIs. Additionally, the modelers' decisions regarding the number of simulations in MC, and other possible discussion of uncertainty aspects together its mitigation methods are emphasized. Secondly, the review reveals the time span of the forecasts (e.g., annual) and the possible specification of being spatial or agent-based model (ABM). These reflects (from one perspective) to the data aspects in terms of timeliness and granularity. The paper also identifies studies that aim to utilize near real-time information or continuously updating models. Considering future research, other findings related to emerging technologies, mainly ML-oriented works, are acknowledged although this research mainly omits the technical details about the models.

The paper concludes with suggestions for future research. Conclusions are drawn from available publication details, and while the literature review is not comprehensive, it lays the groundwork for a more in-depth study on these schemes.

2 RESULTS AND DISCUSSION

In the following results section and related discussion, the reader may find it helpful to refer to Table 1, which presents basic information of the modeling works (author, year), the brief summary of main modeling purpose and the findings related to the MC simulation and data aspects, as detailed in the previous section. We do not specify whether the MC is used only in some model parts. Also, if the use of MC method is not reported, but repeated simulations are applied, it is categorized under the MC. If other methods are clearly reported, such as bootstrapping, they are marked.

2.1 Results

In most of the reviewed studies (30 out of 44) MC/repeated simulations is applied (see, Table 1). In the set of these 30 studies reporting practices vary: seven works did not directly report on using the MC method, but it was shown that the simulation had been indeed run repeatedly. In 13 entries the number of simulations run was not reported and notably, 16 studies (out of 30) did not report CIs. Yet only two

Table 1: Reviewed studies of DMs. Legend: [MC]=Monte Carlo method used (Yes/No or "-" if unclear and additional NR=not reported, if repeated simulation applied without reporting the method or "B" if bootstrapping is applied instead of MC), [Simrun]=number of simulations run (NR=not reported and "-" if MC not applied), [CI]=confidence intervals used (if MC used, otherwise "-") [Simstep]=Forecast period (A=annual, M=monthly, D=daily) + Detail (spatial (S)/ agent-based (AB). *In progress = not yet available since study ongoing but reported to be applied.

Auth. & Year	Study purpose	MC/Simrun/ CI/Simstep + Detail
Aransiola et al. 2024	To assess if expanding Social Assistance could reduce infant and child mortality in Brazil.	Yes/10000/Yes/A
Archer et al. 2021	To project the prevalence of chronic diseases and their economic impacts using the Future Elderly Model (FEM) in the UK.	Yes/100/Yes/A
Atella et al. 2021	To project future individual health status across OECD countries by applying several FEM models.	Yes/NR/Yes/A
Baldini et al. 2008	To assess the characteristics of the long-term disabled in Italy and the evolution of public expenditure for long-term care.	Yes/NR/No/A
Ballas et al. 2005	To simulate the basic components of population change in Ireland using spatial SMILE model.	Yes/NR/No/A + S
Ballas et al. 2005	To simulate urban and regional populations in UK.	No/-/-/A + S
Becker et al. 2024	To assess the efficiency of COVID-19 mitigation strategies with the CEACOV model in U.S.	-/-/-/D
Bonin et al. 2015	To model monetary value of family policy measures with ZEW model in Germany.	No/-/-/A
Böheim et al. 2023	To model the impact of health and education on labor force participation in US and Germany.	Yes/12/No/A
Brouwers et al. 2016	To study the effects of an ageing population on inpatient and elderly care with SESIM-LEV model in Sweden.	No/-/-/A
Chen et al. 2019	To model fiscal sustainability of healthcare by projecting the health of future elders using FEM model for Singapore.	No/-/-/A
Craig et al. 2022	To simulate the long-term health impacts in UK.	Yes/10000/Yes, In progress*/A

Table 1: Reviewed studies of DMs. Legend: [MC]=Monte Carlo method used (Yes/No or "-" if unclear and additional NR=not reported, if repeated simulation applied without reporting the method or "B" if bootstrapping is applied instead of MC), [Simrun]=number of simulations run (NR=not reported and "-" if MC not applied), [CI]=confidence intervals used (if MC used, otherwise "-") [Simstep]=Forecast period (A=annual, M=monthly, D=daily) + Detail (spatial (S)/ agent-based (AB). *In progress = not yet available since study ongoing but reported to be applied. (cont.).

Ernst et al. 2023	To analyse migration impacts on demographics in Germany.	Yes/NR/No/ A + S
Flannery & O'Donoghue 2011	To study the fiscal and redistributive impacts of different higher education finance structures using the LIAM model in Ireland.	No/-/-/A
Fukawa 2011	To project health/long-term care expenditures with the INAHSIM-II model in Japan.	Yes/NR/No/ A
Head et al. 2024	To model time individuals spent in different health states in UK.	Yes/100/Yes/ A
Horvath et al. 2023	To project healthcare costs over the lifecycle using microWELT model in Austria.	No/-/-/A
Ben Jelloul et al. 2023	To forecast morbidity of population aged +60 and identify causing factors in France.	No/-/-/A
Jiang & Li 2024	To project the population size and share of late middle-aged/older people with difficulties/dependence on activities of daily living (DL) and instrumental activities of DL with the CHARISMA model in China.	Yes, NR/1000/No/ A
Keegan 2021	To simulate the distributional impact of pension policy scenarios on superannuation savings using the APPSIM model in Australia.	No/-/-/A
Khalil et al. 2024	To predict demographic dynamics in Canada with STELARS model.	No/-/-/A + AB
Kingston et al. 2018	To predict the survival and (risk/disease) characteristics and related health expectancies in UK using PACSim model.	Yes, NR/10/- /A
Kirn &Dekkers 2023	To simulate with MIDAS_CH model the distribution of pension income and its underlying processes in Switzerland.	No/-/-/A
Knoef et al. 2013	To analyse the income distribution of the Dutch elderly.	Yes/NR/No/ A
Kopasker et al. 2024	To project changes in psychological distress given predicted economic outcomes. from a tax-benefit UKMOD model with SimPaths model in UK.	Yes, NR/1000/Yes /A
Lawson 2016	To model how demographic change is likely to affect household spending patterns in the UK.	Yes/5/Yes/A
Li et al. 2024	To model the spread of COVID-1 in China.	Yes, NR/10/Yes/D + S, AB
Maitino et al. 2020	To study the future socio-demographic structure and the effects of social security programmes in Italy.	Yes/NR/No/ A + S
Marois & Aktas 2021	To project the health of cohorts for selected EU countries to study the effects of risk factors and education on future health trajectories using ATHLOS-Mic model.	Yes/NR/No/ A
May et al. 2022	To project the health and service use among elderly in Ireland using TILDA model.	Yes/25/No/A
Milne et al. 2016	To model child development from birth to age 13 with MELC model and studying e.g., changes in family circumstances and early education in New Zealand.	Yes, NR/10/Yes/A
Nadeau et al. 2013	To model physical activity to inform population health policies using POHEM-PA model in Canada.	Yes, B/40/Yes/An nual
Patxot et al. 2018	To model the impact of retirement decision and demographics on pension sustainability in Spain.	-/-/-/A
Rasella et al. 2021	To analyse the prospective effects of fiscal policies on childhood health in the EU countries and in Italy.	Yes/1000/Ye s/A
Rephann & Holm 2004	To model economic-demographic effects of immigration in Sweden using the SVERIGE model.	Yes/NR/No/ A + S
Spielauer & Dupriez 2019	To model a variety of demographic and health characteristics with DYNAMIS-POP model in Canada.	Yes/NR/No/ A
Spooner et al. 2021	To model epidemics with spatial SPENSER model in UK.	Yes, NR/1000/Yes /D + S
Tamborini et al. 2022	To analyze socioeconomic gaps in retirement benefits using the MINT model in U.S.	No/-/-/M, A
Tikanmäki et al. 2015	To analyse impacts of the pension reform on working lives using ELSI model in Finland.	Yes/NR/No/ A
van Sonsbeek & Gradus 2005	To simulate the budgetary impact of the 2006 regime change in the Dutch disability scheme.	Yes/NR/No/ A
Walker 2004	To model the likelihood that more Australians aged 65-70 will work +15 hours per week in a changing employment environment.	Yes/NR/No/ A
Wu et al. 2011	To model several demographic processes under various scenarios with a Moses in UK.	Yes/NR/No/ A + S, AB
Zhang & Miller 2024	To predict the location of new housing supply in U.S.	No/-/-/- + S, AB
Zhang et al. 2023	To model the processes of developing depression and care-seeking behaviors among U.S children and adolescents.	Yes, NR/20/Yes/ M

studies reported using 10 000 and four studies 1000 simulation runs. In the remainder, the number of simulations vary from five to forty.

Additionally, there is a notable variability in depth across publications about the discussion of the sources and mitigation of uncertainty. Many studies indirectly or directly, but briefly, address uncertainty when discussing issues like data availability and sample size (see e.g., Becker et al., 2024; Kirn and Dekkers, 2023; Rephann & Holm, 2004) or mention it broadly as parameter/statistical/MC uncertainty. Also, authors commonly discuss of "model error" or “model-based bias”, which intersects with the uncertainty concept (see e.g., Atella et al., 2021; Jiang et al., 2021; Knoef et al., 2013; Kopasker et al., 2024; Lawson, 2016; Marois & Aktas, 2021; Spielauer & Dupriez, 2019).

Focusing on data aspects in terms of data timeliness, only three pandemic-related models and two exceptions from other disciplines seems to offer sub-annual observation periods in models, reaching daily or monthly level accuracy in simulation results. To mention, in their multi-morbidity modeling study, although simulation results are presented in yearly interval, Kingston et al. (2018) updated individual’s characteristics monthly over the simulation time period “to achieve a more realistic evolution for characteristics which jointly influence each other”, similarly than Böheim et al. (2023) regarding labour force status.

Also, to the best of our understanding, only epidemiology models by Becker et al. (2024) and Spooner et al. (2024) target to produce forecasts using near-real-time data with updates. In rest of the models, it seems common to use administrative statistics with a time lag of at least 2-3 years in model initialization.

Lastly, nine studies reviewed are by their nature spatial, including epidemiology studies. Four studies combined the ABM method with DMs, three of them being also spatial (see again Table 1).

2.2 Results Analysis

2.2.1 Uncertainty Analysis

The literature covered indicates varying practices in the application of the MC method and related reporting, highlighting a need for common standards and/or strategies to improve the transparency and comparability of demographic models in the research field. This will not only improve the accuracy of individual studies but also facilitate more robust analyses and comparisons across different research efforts in demographic modeling. We justify this claim by the often inadequate depth of the discussion (and missing information) of MC related details, such as the number of simulation rounds (and reasons that led to the number) and lack of CIs. The lack in reporting CIs aligns also with Smithson (2003), who noticed that different disciplines vary considerably how frequently they report CIs in published research (see also Lappo, 2015; O’Donoghue 2014, 332; O’Donoghue & Dekkers, 2018). Kingston et al. (2018) notes the lack of CIs as one of their study limitations, although the authors also highlight that running the simulation iteratively reveals a small range of prevalence for multi-morbidity (less than 1 %), even when the error in transition rates is disregarded. Knoef et al. (2013) reported not using CIs due the “computational reasons”. Lappo (2015), however, states that the omission of reporting CIs may be since many microsimulation users are not statisticians, perhaps so be in the case of social sciences. However, this indicates a prevalent lack of established practices in employing methods to convey information on result variability across study disciplines (see e.g., Li & O’Donoghue, 2013).

To further explore practices related to the MC method, some authors provide the basis (or tests made) for selecting the number of simulations such as Rasella et al. (2021), who state that a thousand simulation runs was chosen after ensuring that the estimates were stable and additional runs did not alter the point estimates (see also Van Sonsbeek & Gradus, 2005). Spielauer and Dupriez (2019) claimed that 24 iterations make MC variation neglectable, whereas Aransiola et al. (2024) performed 10k rounds to ensure the variation of the parameter values. Overall, the selection of the number of MC simulation runs has received only limited attention even though it is a crucial factor for generating meaningful predictions (see e.g., Byrne 2013; Kennedy 2019, Kennedy et al. 2000). There is a position to analyze more comprehensively the specific factors contributing to the large variation in the number of simulations, especially within studies investigating the same phenomena and “sharing” the same uncertainty elements. Overall, we can concur with O’Donoghue and Dekkers (2018) who noted that alignment techniques (not a focus of this study) are so common in DMs that most reports do not even mention them, despite their significant impact on simulation results. This oversight is similar to the treatment of the MC method (see also Byrne, 2013; Lorscheid et al., 2012; Kennedy, 2019).

When analyzing the overall use of the MC approach, studies reveal differing perspectives on the objectives of modeling: some prioritize analyzing current systems without accounting for variations or forecasting goals, thus considering repeated simulations unnecessary (see e.g., Ben Jelloul et al., 2023; Flannery & O'Donoghue, 2011). In contrast, the majority (30 out of 44) employ the probabilistic method to understand system functionality under uncertainty. Studies focusing on individual behavior and future trends through predefined scenarios and single-point estimates may fail to capture the full spectrum of potential outcomes or convey the inherent uncertainty of modeled phenomena. Such approaches might overlook rare yet impactful events, whereas the MC accounts for these events and their potential consequences (see Fuchs et al., 2018; Gilbert et al., 2018; Marois & Aktas, 2021; O'Donoghue, 2014; Rutter et al., 2011).

2.2.2 Data Granularity and Timeliness

Considering data aspects, the shortcomings of the models running yearly intervals have been recognized. Salonen et al. (2021) highlight challenges in capturing gradual changes such as increase in pension age or short social security spells with a model allowing transitions in one year time intervals. For instance, based on data, the average duration of sickness and unemployment spells is one week, although these periods accumulate over an individual's life course (see also Perhoniemi et al., 2023; Zaidi and Rake, 2001). Although Kingston et al. (2018) provides forecasts in yearly intervals, they enhanced the accuracy of their simulation results by updating health behaviors and disease conditions on a monthly basis. Chen et al. (2019) acknowledges the limitation of not modeling shorter disease dynamics similarly than Andreassen et al. (2020), who suggest that with improved data access and today's computing power, monthly time units could be preferable in the MOSART model (renowned for evaluating the Norwegian pension system) to avoid aggregating data annually and potentially overlooking nuances.

To continue, in an ideal world, employing close-to-real time data for model calibration would reduce the risk of obsolete information affecting transition probabilities – an issue that is especially important when addressing rapidly evolving matters, such as changes in labour market status during economic crisis (see e.g., O'Donoghue & Sologon, 2023). To the best of understanding, no research efforts in this direction were found in this review except epidemiology models. The findings align also with O'Donoghue and Loughrey (2014), who observed that microsimulation models tend to be built on historical data (see also Klevmarken, 2008), limiting researchers' ability to analyze and monitor recent changes and developments.

However, it's important to note that not many phenomena require the daily/monthly forecast accuracy and frequent data calibration typical in pandemic research. Instead, "traditional" social policy models could aim to reduce the delay between data collection and utilization, moving from a lag of several years to using more recent statistics. This shift would better reflect contemporary issues, such as the interconnections between labour force participation and health status (see O'Donoghue & Sologon, 2023).

Admittedly, increased granularity and the use of more timely data to update transition probabilities (together with MC method) add to model complexity in regards of model calibration and computational demands. Nevertheless, many renowned models in the field already require substantial computing power and resources for maintenance due to their high modularity. Today's technological capabilities, such as cloud computing and big data analytics can help overcoming this issue (see Andreassen et al., 2020; O'Donoghue & Dekkers, 2018; Richiardi et al., 2023).

Looking forward, there may be a trend towards simpler models that allow for agile calibration with detailed, current data, albeit sacrificing some modularity (Harding, 2007; Li & O'Donoghue, 2013; Zaidi & Rake, 2001). For instance, localized projections (with ABM approach) are vital for addressing regional disparities and tailoring policies to specific areas. They enhance the relevance of simulations and allow for more detailed evaluations of policy impacts (see Ballas et al., 2005; Birkin et al., 2017; Ernst et al., 2023; Wu & Birkin, 2011). Agile calibrated models providing timely forecasts could potentially be recognized also at the tactical decision-making level.

2.2.3 Other Findings

Machine learning (ML) techniques, in addition to being utilized in model calibration tasks, can aid in addressing complexity arising from models' non-linearities, a topic of ongoing discussion (see e.g., Jiang & Li, 2024; Klevmarken, 2008; Kopasker et al., 2024; O'Donoghue & Dekkers, 2018; Wolfson & Rowe, 2014). The integration of ML could enable the development of more dynamic and predictive models, which could better address complex societal challenges and facilitate faster decision-making. These methods could uncover unobserved, detailed behavioural patterns among individuals thus improving simulation granularity and supporting e.g., ABM constructions (see discussion of Margetts & Dorobatu, 2023). That is, model structures where individual model components interact with each other instead of being passive and detached (see e.g., Axtell 2000). Although there are few applications within the reviewed works, Khalil et al. (2024) provide an innovative application of explainable artificial intelligence (xAI) with the aim to interpret ML models, elucidating input-output relationships in complex settings. This study can be regarded as a pioneering effort in integrating ML within DM schemes in the research domain. Other studies like Rodriguez et al. (2022) in healthcare and other ML-assisted models (see e.g., Shi et al., 2015) offer also insights into applying advanced methods, potentially inspiring social science research.

3 CONCLUSION

This paper presented a literature analysis on the use of probabilistic methods such as Monte Carlo simulation in dynamic microsimulation models and related reporting practices of probabilistic outcomes. This study, to the best of our knowledge, is the first review that addresses the use of such methods and related challenges in reporting the analysis findings in the given context.

It was shown that the current literature often lacks a statistical treatment of the model and if given, there are no standard practices on how a (MC) simulation is conducted and presented. As another important finding, we did not find evidence that attempts were made to develop DMs towards nowcasting with the help of extensive real-time datasets in other study contexts than epidemiology.

The results imply that population-based modeling studies, a predominant focus of the review conducted, could adopt probabilistic thinking to address the inherent uncertainty associated with complex socio-economic processes to make the modeling results more robust and reliable. Common guidelines for UA application and related communication/reporting practices could enhance the transparency of modeling insights as the vulnerability of results would become better communicated to policymakers and less weight could be put on single-point estimates. Also, transition probabilities calculated sub-annual periods can lead to more accurate simulations by incorporating finer temporal variations, e.g., monthly updates can capture short-term trends or immediate impacts of policy changes that yearly intervals might miss. In this regard, it was concluded that the research field could benefit from the development and application of smaller, more targeted models that could offer greater agility in terms of maintenance, particularly in incorporating updated data.

This paper has limitations, notably not being a fully comprehensive systematic review. Nonetheless, it provides some preliminary directions for future research efforts to improve probabilistic treatment of DMs in the context of demographic models. Additionally, future research could assess the role of emerging technologies, such as cloud computing, machine learning techniques, and big data analytics.

The final limitation of this research to be pointed out is the method used to assess data granularity, categorizing models as agent-based or spatial. Future evaluations could provide an extended analysis of variables like demographic precision. A more comprehensive review could delve deeper into whether the nature of the phenomena being modeled warrants more frequent updates to transition probabilities. The review focused only on the MC method with frequentist viewpoint, omitting Bayesian methods or distinguishing bootstrapping from MC approach. It also did not cover the alignment techniques used together with the MC, or other reporting practices such as goodness-of-fit or standard error statistics.

REFERENCES

Alho, J., & Lassila, J. (2023). Assessing components of uncertainty in demographic forecasts with an application to fiscal sustainability. Journal of Forecasting, 42(7), 1560-1568.

Andreassen, L., Fredriksen, D., Gjefsen, H. M., Halvorsen, E., & Stølen, N. M. (2020). The dynamic cross-sectional microsimulation model MOSART. International Journal of Microsimulation, 13(1), 92-113.

Aransiola, T. J., Cavalcanti, D., Ordoñez, J. A., Hessel, P., Moncayo, A. L., Chivardi, C., ... & Rasella, D. (2024). Current and Projected Mortality and Hospitalization Rates Associated With Conditional Cash Transfer, Social Pension, and Primary Health Care Programs in Brazil, 2000-2030. JAMA Network Open, 7(4), e247519-e247519.

Archer, L., Lomax, N., & Tysinger, B. (2021). A dynamic microsimulation model for ageing and health in england: The english future elderly model. International Journal of Microsimulation, 14(3), 2-26.

Astolfi, R., Lorenzoni, L., & Oderkirk, J. (2012). Informing policy makers about future health spending: a comparative analysis of forecasting methods in OECD countries. Health Policy, 107(1), 1-10.

Atella, V., Belotti, F., Kim, D., Goldman, D., Gracner, T., Piano Mortari, A., & Tysinger, B. (2021). The future of the elderly population health status: Filling a knowledge gap. Health economics, 30, 11-29.

Axtell, R. (2000). Why agents?: on the varied motivations for agent computing in the social sciences (Vol. 17). Washington, DC: Center on Social and Economic Dynamics.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B., & Rossiter, D. (2005). SimBritain: a spatial microsimulation approach to population dynamics. Population, Space and Place, 11(1), 13-34.

Ballas, D., Clarke, G. P., & Wiemers, E. (2005). Building a dynamic spatial microsimulation model for Ireland. Population, Space and Place, 11(3), 157-172.

Becker, J. E., Shebl, F. M., Losina, E., Wilson, A., Levison, J. H., Donelan, K., ... & Freedberg, K. A. (2024). Using simulation modeling to inform intervention and implementation selection in a rapid stakeholder-engaged hybrid effectiveness-implementation randomized trial. Implementation Science Communications, 5(1), 70.

Ben Jelloul, M., Bozio, A., Perdrix, E., Rain, A., & Toulemon, L. (2023). Dynamique du processus de perte d'autonomie dans les populations vieillissantes/Dynamic of the Disablement Process in Ageing Populations. Economie et Statistique, 538(1), 13-31.

Birkin, M., Wu, B., & Rees, P. (2017). Moses: dynamic spatial microsimulation with demographic interactions. New frontiers in microsimulation modelling, 53-77.

Bonin, H., Reuss, K., & Stichnoth, H. (2015). Life-cycle incidence of family policy measures in Germany: Evidence from a dynamic microsimulation model.

Brown, L., & Harding, A. (2002). Social modelling and public policy: Application of microsimulation modelling in Australia. Journal of Artificial Societies and Social Simulation, 5(4).

Burka, D., Mohácsi, L., Csicsman, J., & Soós, B. (2017). Supporting Pension Pre-Calculation With Dynamic Microsimulation Technologies. In ECMS (pp. 562-568).

Burgard, J. P., & Schmaus, S. (2019). Sensitivity analysis for dynamic microsimulation models (No. 15/19). Research Papers in Economics.

Burgard, J. P., Krause, J., Merkle, H., Münnich, R., & Schmaus, S. (2019). Conducting a dynamic microsimulation for care research: Data generation, transition probabilities and sensitivity analysis. In Stochastic Models, Statistics and Their Applications: Dresden, Germany, March 2019 14 (pp. 269-290). Springer International Publishing.

Burgard, J. P., Krause, J., & Schmaus, S. (2021). Estimation of regional transition probabilities for spatial dynamic microsimulations from survey data lacking in regional detail. Computational Statistics & Data Analysis, 154, 107048.

Byrne, M. D. (2013). How many times should a stochastic model be run? An approach based on confidence intervals. In Proceedings of the 12th International conference on cognitive modeling, Ottawa.

Böheim, R., Horvath, T., Leoni, T., & Spielauer, M. (2023). The Impact of Health and Education on Labor Force Participation in Aging Societies: Projections for the United States and Germany from Dynamic Microsimulations. Population Research and Policy Review, 42(3), 39. Simulation, 21(1).

Caro, J. J., Briggs, A. H., Siebert, U., & Kuntz, K. M. (2012). Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Medical Decision Making, 32(5), 667-677.

Chen, C., Lim, J. T., Chia, N. C., Wang, L., Tysinger, B., Zissimopoulos, J., ... & Yoong, J. (2019). The long-term impact of functional disability on hospitalization spending in Singapore. The Journal of the Economics of Ageing, 14, 100193.

Craig, P., Barr, B., Baxter, A. J., Brown, H., Cheetham, M., Gibson, M., ... & Bambra, C. (2022). Evaluation of the mental health impacts of Universal Credit: protocol for a mixed methods study. BMJ open, 12(4), e061340.

Crato, N. (2023). From lack of data to data unlocking: Computational and statistical issues in an era of unforeseeable big data evolution. In Handbook of Computational Social Science for Policy (pp. 125-139). Cham: Springer International Publishing.

Dekkers, G. (2015). The simulation properties of microsimulation models with static and dynamic ageing–a brief guide into choosing one type of model over the other. International Journal of Microsimulation, 8(1), 97-109.

Ernst, J., Dräger, S., Schmaus, S., Weymeirsch, J., Alsaloum, A., & Münnich, R. (2023). The Influence of Migration Patterns on Regional Demographic Development in Germany. Social Sciences, 12(5), 255.

Flannery, D., & O'Donoghue, C. (2011). The Life-cycle Impact of Alternative Higher Education Finance Systems in Ireland. Economic & Social Review, 42(3).

Fontana, M., & Guerzoni, M. (2022). Modelling Complexity with Unconventional Data: Foundational Issues in Computational Social Science. In Handbook of Computational Social Science for Policy (pp. 107-124). Cham: Springer International Publishing.

Fuchs, J., Söhnlein, D., Weber, B., & Weber, E. (2018). Stochastic forecasting of labor supply and population: An integrated model. Population Research and Policy Review, 37(1), 33-58.

Fukawa, T. (2011). Household projection and its application to health/long-term care expenditures in Japan using INAHSIM-II. Social Science Computer Review, 29(1), 52-66.

Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K. P., & Wilkinson, H. (2018). Computational modelling of public policy: Reflections on practice. Journal of Artificial Societies and Social

Harding, A. (2007, August). Challenges and opportunities of dynamic microsimulation modelling. In Plenary paper presented to the 1st General Conference of the International Microsimulation Association

Head, A., Birkett, M., Fleming, K., Kypridemos, C., & O'Flaherty, M. (2024). Socioeconomic inequalities in accumulation of multimorbidity in England from 2019 to 2049: a microsimulation projection study. The Lancet Public Health, 9(4), e231-e239.

Horvath, T., Leoni, T., Reschenhofer, P., & Spielauer, M. (2023). Socio-economic inequality and healthcare costs over the life course–A dynamic microsimulation approach. Public Health, 219, 124-130.

Jiang, Y., & Li, L. (2024). Projections of functional dependence among the late middle-aged and older population from 2018-2048 in China: a dynamic microsimulation. Global Health Research and Policy, 9(1), 15.

Kashyap, R., & Zagheni, E. (2023). Leveraging Digital and Computational Demography for Policy Insights. In Handbook of Computational Social Science for Policy (pp. 327-344). Cham: Springer International Publishing.

Keegan, M. (2011). Mandatory superannuation and self-sufficiency in retirement: an application of the APPSIM dynamic microsimulation model. Social science computer review, 29(1), 67-84.

Kennedy, M. C. (2019). Experimental design principles to choose the number of Monte Carlo replicates for stochastic ecological models. Ecological Modelling, 394, 11-17.

Khalil, M. A., Fatmi, M. R., & Orvin, M. (2024). Developing and microsimulating demographic dynamics for an integrated urban model: a comparison between logistic regression and machine learning techniques. Transportation, 1-35.

Klevmarken, A. (2008). Chapter 2 Dynamic Microsimulation for Policy Analysis: Problems and Solutions. In Simulating an ageing population: A microsimulation approach applied to Sweden (pp. 31-53). Emerald Group Publishing Limited.

Kingston, A., Robinson, L., Booth, H., Knapp, M., Jagger, C., & Modem Project. (2018). Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. Age and ageing, 47(3), 374-380.

Knoef, M., Alessie, R., & Kalwij, A. (2013). Changes in the Income Distribution of the D utch Elderly between 1989 and 2020: a Dynamic Microsimulation. Review of Income and Wealth, 59(3), 460-485.

Kopasker, D., Bronka, P., Thomson, R. M., Khodygo, V., Kromydas, T., Meier, P., ... & Katikireddi, S. V. (2024). Evaluating the influence of taxation and social security policies on psychological distress: A microsimulation study of the UK during the COVID-19 economic crisis. Social Science & Medicine, 351, 116953.

Lappo, S. (2015). Uncertainty in microsimulation. Master's Thesis, University of Helsinki.

Lawson, T. (2016). How the ageing population contributes to UK economic activity: A microsimulation analysis. Scottish Journal of Political Economy, 63(5), 497-518.

Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. Journal of the American Statistical Association, 89(428), 1175-1189.

Lee, J. T., Crettenden, I., Tran, M., Miller, D., Cormack, M., Cahill, M., ... & Xiang, F. (2024). Methods for health workforce projection model: systematic review and recommended good practice reporting guideline. Human Resources for Health, 22(1), 25.

Lee, R., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. Journal of the American Statistical Association, 89(428), 1175–1189.

van Leeuwen, E. (2016). Simulating the expenditures of Scottish households: a two-step microsimulation approach to the Cairngorms National Park. In New Pathways in Microsimulation (pp. 233-248). Routledge.

Li, Z., Xiong, G., Lv, Y., Ye, P., Liu, X., Tarkoma, S., & Wang, F. Y. (2024). An Urban Trajectory Data-Driven Approach for COVID-19 Simulation. IEEE

Li, J., & O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. International Journal of microsimulation, 6(2), 3-55.

Li, J., & O'Donoghue, C. (2016). An overview of binary alignment methods in microsimulation. New Pathways in Microsimulation, 233-248.

Lorscheid, I., Heine, B. O., & Meyer, M. (2012). Opening the 'black box'of simulations: increased transparency and effective communication through the systematic design of experiments. Computational and Mathematical Organization Theory, 18, 22-62.

Mahadevan, S., & Sarkar, S. (2009). Uncertainty analysis methods. US Department of Energy, Washington, DC, USA.

Margetts, H., & Dorobantu, C. (2023). Computational social science for public policy. In Handbook of Computational Social Science for Policy (pp. 3-18). Cham: Springer International Publishing.

Marois, G., & Aktas, A. (2021). Projecting health-ageing trajectories in Europe using a dynamic microsimulation model. Scientific reports, 11(1), 1785.

May, P., Normand, C., Matthews, S., Kenny, R. A., Romero-Ortuno, R., & Tysinger, B. (2022). Projecting future health and service use among older people in Ireland: an overview of a dynamic microsimulation model in The Irish Longitudinal Study on Ageing (TILDA). HRB Open Research, 5.

Milne, B., Lay Yee, R., McLay, J. M., Pearson, J., Von Randow, M., & Davis, P. (2015). Modelling the Early life-course (MELC): A microsimulation model of child development in New Zealand.

Münnich, R., Schnell, R., Brenzel, H., Dieckmann, H., Dräger, S., Emmenegger, J., ... & Stein, P. (2021). A population based regional Dynamic microsimulation of Germany: the MikroSim model. methods, data, analyses, 15(2), 24.

Nadeau, C., Wong, S. L., Flanagan, W. M., Oderkirk, J., Manuel, D., Wall, R., & Tremblay, M. S. (2013). Development of a population-based microsimulation model of physical activity in Canada. Health Rep, 24(10), 11-9.

Navicke, J., Rastrigina, O., & Sutherland, H. (2014). Nowcasting indicators of poverty risk in the European Union: A microsimulation approach. Social Indicators Research, 119, 101-119.

O'Donoghue, C. (2001). Dynamic microsimulation: A methodological survey. Brazilian Electronic Journal of Economics, 4(2), 77.

O'Donoghue, C. (Ed.). (2014). Handbook of microsimulation modelling. Emerald Group Publishing.

O'Donoghue, C., & Dekkers, G. (2018). Increasing the impact of dynamic microsimulation modelling. The International Journal of Microsimulation, 11(1), 61-96.

O'Donoghue, C., & Loughrey, J. (2014). Nowcasting in microsimulation models: a methodological survey. Journal of Artificial Societies and Social Simulation, 17(4), 12. Transactions on Computational.

O'Donoghue, C., & Sologon, D. M. (2023). The Transformation of Public Policy Analysis in Times of Crisis-A Microsimulation-Nowcasting Method Using Big Data.

Patxot, C., Solé Juvés, M., Souto Nieves, G., & Spielauer, M. (2018). The impact of the retirement decision and demographics on pension sustainability: A dynamic microsimulation analysis. International Journal of Microsimulation, 2018, vol. 11, num. 2, p. 84-108.

Perhoniemi, R., Blomgren, J., & Laaksonen, M. (2023). Identifying labour market pathways after a 30-day-long sickness absence–a three-year sequence analysis study in Finland. BMC Public Health, 23(1), 1102.

Rasella, D., Richiardi, L., Brachowicz, N., Jara, H. X., Hanson, M., Boccia, D., ... & Pizzi, C. (2021). Developing an integrated microsimulation model for the impact of fiscal policies on child health in Europe: the example of childhood obesity in Italy. BMC medicine, 19, 1-12.

Rephann, T. J., & Holm, E. (2004). Economic-demographic effects of immigration: results from a dynamic spatial microsimulation model. International Regional Science Review, 27(4), 379-410.

Richiardi, M., Bronka, P., van de Ven, J., Kopasker, D., & Katikireddi, S. V. (2023). SimPaths: an open-source microsimulation model for life course analysis. Available at SSRN 4808042.

Rodriguez, P. J., Veenstra, D. L., Heagerty, P. J., Goss, C. H., Ramos, K. J., & Bansal, A. (2022). A Framework for Using Real-World Data and Health Outcomes Modeling to Evaluate Machine Learning–Based Risk Prediction Models. Value in Health, 25(3), 350-358.

Rutter, C. M., Zaslavsky, A. M., & Feuer, E. J. (2011). Dynamic microsimulation models for health outcomes: a review. Medical Decision Making, 31(1), 10-18.

Sabelhaus, J., & Topoleski, J. (2007). Uncertain policy for an uncertain world: The case of social security. Journal of Policy Analysis and Management, 26(3), 507-525.

Salonen, J., Tikanmäki, H., & Lappo, S. (2021). Partition of the Life Course: An Extended Dynamic Microsimulation Analysis.

Sauerbier, T. (2002). UMDBS-a new tool for dynamic microsimulation. Journal of Artificial Societies and Social Simulation, 5(2).

Sharif, B., Wong, H., Anis, A. H., & Kopec, J. A. (2017). A practical ANOVA approach for uncertainty analysis in population-based disease microsimulation models. Value in Health, 20(4), 710-717.

Sharif, B., Kopec, J. A., Wong, H., Fines, P., Sayre, E. C., Liu, R. R., & Wolfson, M. C. (2012). Uncertainty analysis in population-based disease microsimulation models. Epidemiology Research International, 2012(1), 610405.

Shi, Y., Zeng, W., Wang, N., Wang, S., & Huang, Z. (2015). Early warning for human mental sub-health based on fMRI data analysis: an example from a seafarers' resting-data study. Frontiers in psychology, 6, 1030.

Smithson, M. (2003). Confidence intervals (No. 140). Sage.

Spielauer, M. (2007). Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: survey and review. International Journal of Microsimulation, 1(1), 35-53.

Spielauer, M., & Dupriez, O. (2019). A portable dynamic microsimulation model for population, education and health applications in developing countries. International Journal of Microsimulation, 12(3), 6-27.

Spooner, F., Abrams, J. F., Morrissey, K., Shaddick, G., Batty, M., Milton, R., ... & Birkin, M. (2021). A dynamic microsimulation model for epidemics. Social Science & Medicine, 291, 114461.

Tamborini, C. R., Reznik, G. L., Iams, H. M., & Couch, K. A. (2022). The Growing Socioeconomic Gap in Lifetime Social Security Retirement Benefits: Current and Future Retirees. The Journals of Gerontology: Series B, 77(4), 803-814.

Tikanmäki, H., Sihvonen, H., & Salonen, J. (2015). Distributional Effects of the Forthcoming Finnish Pension Reform: a Dynamic Microsimulation Approach.

Van Sonsbeek, J. M., & Gradus, R. H. (2006). A microsimulation analysis of the 2006 regime change in the Dutch disability scheme. Economic modelling, 23(3), 427-456.

Walker, A. (2004). Impact of health on the ability of older Australians to stay in the workforce-with possible contributions to economic sustainability.

Wolfson, M., & Rowe, G. (2014). HealthPaths: Using functional health trajectories to quantify the relative importance of selected health determinants. Demographic Research, 31, 941-974.

Wu, B. M., Birkin, M. H., & Rees, P. H. (2011). A dynamic MSM with agent elements for spatial demographic forecasting. Social Science Computer Review, 29(1), 145-160.

Xue, X., Chen, F., Zhou, D., Wang, X., Lu, M., & Wang, F. Y. (2021). Computational experiments for complex social systems—Part I: The customization of computational model. IEEE Transactions on Computational Social Systems, 9(5), 1330-1344.

Zaidi, A., & Rake, K. (2001). Dynamic microsimulation models: a review and some lessons for SAGE. Simulating Social Policy in an Ageing Society (SAGE) discussion paper, 2.

Zhang, Y., & Miller, E. J. (2024). Analyzing housing supply location choice: a comparative study of the modelling frameworks. Scientific Reports, 14(1), 1435.

Zhang, C., Zafari, Z., Slejko, J. F., Camelo Castillo, W., Reeves, G. M., & Dosreis, S. (2023). Impact of different interventions on preventing suicide and suicide attempt among children and adolescents in the United States: a microsimulation model study. Frontiers in psychiatry, 14, 1127852.

Zhao, Y., Jiang, C., Vega, M. A., Todd, M. D., & Hu, Z. (2023). Surrogate modeling of nonlinear dynamic systems: a comparative study. Journal of Computing and Information Science in Engineering, 23(1), 011001.

Knowledge Pyramid Perspective of the Political Data Ecosystem: A Case Study of Bhutan

Phub Namgay[1,2][a] and Pema Wangdi[2][b]

[1] *Department of Informatics and Media, Uppsala University, Sweden*

[2] *School of Data Science and Data Analytics, Sherubtse College, Royal University of Bhutan, Bhutan*

phub.namgay@im.uu.se, pemawangdi.sherubtse@rub.edu.bt

Keywords: Political Data Ecosystem, Knowledge Management, Knowledge Pyramid, Democratic Electoral Process, Bhutan.

Abstract: This study examines the dynamics of data management and knowledge flow in the political data ecosystem through the lens of the Knowledge Pyramid. We used open-government electoral documents and polling data for granular insights into how data, information, knowledge, and wisdom (DIKW) are managed in Bhutan's political data ecosystem. Bhutan's electoral stakeholders and political parties manage and use DIKW of varying types, sizes, and complexities. In particular, political parties use information systems, websites, and social media to manage data and construct and use knowledge for political activities. Democracy is still young and gaining a foothold in Bhutan. The political parties do not employ complex data technologies and rich human resources to manage DIKW emanating from the political data ecosystem. Thus, scope exists for electoral stakeholders and political actors to explore and adopt effective and efficient knowledge management infrastructures to deal with DIKW elements in the political arena, namely the complex dynamics of turning raw data into higher elements of the Knowledge Pyramid. In addition to contributing to the knowledge management literature through an in-depth account of the DIKW aspects in the political space, this paper demonstrated the analytical and explanatory power of the Knowledge Pyramid for discourse on the political data ecosystem.

1 INTRODUCTION

The conventional political environment is gradually transitioning to a data-driven political system. Political actors use data technologies to harness insights and knowledge from electorate data to boost campaign activities and electoral success, such as using big data in the US election (Ruppert et al., 2017). Cambridge Analytica (Dommett et al., 2023; Micheli et al., 2020; Ruppert et al., 2017) had a significant impact on the political stakeholders on exploiting the value of data (Miller and Mork, 2013; Lee, 2017) generated in the political space. The political actors also now recognise data as an intangible asset and use the so-called political technologies (Ruppert et al., 2017) to tap its value to drive campaign activities and enrich decision-making. For instance, during the national campaigns and electoral processes of countries such as the United States (US), data management and analytics were undertaken by a chief analytics officer and his hub of statisticians, engineers, and data scientists of political parties. Likewise, much of the data analysis tasks of the electoral activities are increasingly delegated to and driven by complex algorithmic processes (Mittelstadt et al., 2016). In the present work, the term *political* was prefixed to the definition of data ecosystem suggested by Oliveira and Lóscio (2018) to define a working definition of *political data ecosystem*—Organisational entities, infrastructural technologies, electoral activities, digital data, and contextual norms in the democratic political space, where stakeholders concerned, political actors, general citizens, and democratic values are in an interplay to create and capture value of sociopolitical data.

The literature on the use of data in democratic politics, such as data-driven campaign activities and targeting potential voters (Dommett et al., 2023; Bennett and Lyon, 2019), is growing, given the significant effect of data analytics, knowledge management, and social media on the electoral process. Dommett et al. (2023) highlight that the prior studies on data use in democratic politics focused on the US and es-

[a] https://orcid.org/0000-0001-6034-7274

[b] https://orcid.org/0009-0000-8857-7354

Namgay, P. and Wangdi, P.
Knowledge Pyramid Perspective of the Political Data Ecosystem: A Case Study of Bhutan.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 85-96
ISBN: 978-989-758-716-0; ISSN: 2184-3228

tablished democracies. In a sense, such US-centric information affects our understanding of how electoral stakeholders and political parties in different regions of the world, namely emerging democracies, use data and related technologies to construct, manage, and use knowledge in their political space. The use of data and analytics by political actors of countries relatively new to democratic values and ethos—that is, where democracy is recently established, to drive electoral process and political campaigns has not received much scholarly attention and is thus an understudied area. For instance, in Bhutan, a small country in South Asia (Bates, 2009; Metz, 2014), democracy was established only in 2008, and, on a political timeline, democracy is only 16 years old. Consequently, discourses on the use of data and analytics in such a political space are limited.

It is timely for literature to be augmented with insights into countries other than established democracies (Dommett et al., 2023) on how they use data and exploit knowledge to facilitate democratic electoral processes. In the present study, we examine Bhutan's political data ecosystem from the perspective of the Knowledge Pyramid (Jennex, 2017; Frické, 2009; Tuomi, 1999) to determine how electoral stakeholders and political parties deal with aspects of data, information, knowledge, and wisdom (DIKW) for informing electoral processes and political campaigns. Data increasingly impact the democratic election, hence data-driven election (Bennett and Lyon, 2019) and campaigning (Baldwin-Philippi, 2017). The ability to turn data into knowledge by contextualising, structuring, and giving meaning to data empowers political actors to use the value of data for various electoral or political activities. Thus, the Knowledge Pyramid (Jennex, 2017; Frické, 2009) is an apt framework to account for data and analytics practices in the political data ecosystem of Bhutan. We address the following research question: *How do electoral stakeholders and political parties use data, analytics, and knowledge to inform decisions and actions in the political data ecosystem?*

The paper is structured as follows: Section 2 discusses the prior literature on data, politics, and Knowledge Pyramid. Section 3 overviews the methodology adopted in the study, and Section 4 presents the study findings. In Section 5, we discuss the interpretation of the findings and connect it with the extant literature, along with contributions to research and practice, limitations of the study, and future work. Finally, Section 6 concludes the paper.

2 RELATED LITERATURE

2.1 Data and Politics

The resources and capabilities for harnessing the political value of data are fundamental in political endeavours. Likewise, intelligent use and related expertise to analyse and turn data emanating from the political space into meaningful information and valuable knowledge is critical to the success of campaign activities, such as Barack Obama's data-driven campaign work in the 2008 and 2012 elections (Baldwin-Philippi, 2017; Bennett and Lyon, 2019; Jin et al., 2015). The phenomenon of data-driven politics is in the mainstream of democratic elections, such as the hiring of Cambridge Analytica by Donald Trump's team during the 2016 US election (Schippers, 2020; Ruppert et al., 2017; Baldwin-Philippi, 2017). It is a real-world case of transforming data into value or extracting value from data (Micheli et al., 2020) and subsequent practical use of the value in the political arena. Similarly, instances of using big data during the US elections (Schippers, 2020; Ruppert et al., 2017; Dommett et al., 2023) and Brexit referendum (Ruppert et al., 2017) are examples of putting data value to use. Constructs such as data-driven campaigning, micro-targeting, voter profiling, email analytics, and data politics (Dommett et al., 2023; Baldwin-Philippi, 2017; Papakyriakopoulos et al., 2018; Ruppert et al., 2017) are also added to the literature. The extant literature discusses how political actors access and analyse data for insights into campaigns and streamline political activities (Dommett et al., 2023; Ruppert et al., 2017). Thus, in-depth knowledge of the processes requires an understanding of the complex nature of the political data ecosystem.

Furthermore, knowledge of the complexities of the political data ecosystems and the use of political technologies (Ruppert et al., 2017) to manage it is a fundamental intellectual capital (Quintas et al., 1997). The ubiquity of digitalisation and the intelligent use of social media (Baldwin-Philippi, 2017), namely analytics of social data (Olteanu et al., 2019), assist political actors in competing with each other during democratic elections by producing, analysing, and using DIKW within context and time frame. For example, Baldwin-Philippi (2017) highlights the novel use of data and analytics for political campaigns and how corresponding analytical insights inform content production and facilitate political advertising (Schippers, 2020; Ruppert et al., 2017; Bennett and Lyon, 2019). In Bhutan's context, the social media regulation of the Election Commission of Bhutan (ECB) also underlines social media as a channel to disseminate information, communicate content (Election Commission of Bhutan, 2018), and manage electoral compassing activities (Dommett et al., 2023). Data in the political space was also used to develop prediction models for voters' personality traits and the likelihood of turning out to vote (Bennett and Lyon, 2019). Dommett et al. (2023) suggest using data to formulate campaign strategies and evaluate the effectiveness of a campaign retrospectively.

However, despite growing political discourses on data-driven democratic election (Bennett and Lyon, 2019), searches on citation databases, such as Scopus, IEEE Xplore, and ACM Digital Library, with the search query — *((“Knowledge Pyramid” OR “Knowledge hierarchy” OR “Information Pyramid” OR “DIKW hierarchy” OR “Information hierarchy”) AND (“Political data” OR “politics data” OR “Political space” OR “Political data ecosystem*”))*— did not yield useful results[1]. We inferred it as a gap in knowledge on using the perspective of the Knowledge Pyramid (Jennex, 2017) to examine the management and use of DIKW aspects among electoral stakeholders and political parties for informing administrative decision-making and shaping campaign activities.

2.2 Knowledge Pyramid

Knowledge Pyramid is a widely used model to illustrate the logical relationship between DIKW based on meaning, context, and value (Tuomi, 1999; Frické, 2009; Jennex, 2017). The processes to transform data into higher elements of the Knowledge Pyramid reflect creating and building value, thus data value chain (Miller and Mork, 2013). Each step higher in the Knowledge Pyramid answers more questions about a phenomenon captured in the data. According to the Knowledge Pyramid, *data* is a raw or unorganised discrete collection of facts (Zins, 2007; Frické, 2009), such as polling day data of democratic election. The next level is *information*, which is contextual, cleansed, processed, and analysis-ready data (Bellinger et al., 2004; Frické, 2009). In the prior literature (Khatri and Brown, 2010), data and information are often used synonymously and do not differentiate between the two, and the same applies to information and knowledge (Wang and Noe, 2010). Zins (2007) argues that such perception is problematic for articulating data, information, and knowledge. Similarly, scholars also question whether information is data or knowledge (Quintas et al., 1997). *Knowledge* is the consolidation and application of disparate pieces of information, with underlying meanings and schemas, to perform tasks and achieve goals. When one uses data-driven information and knowledge to solve problems, it is an instance of the possession and use of wisdom (Frické, 2009), which is the apex of the Knowledge Pyramid (Rowley, 2007). In the knowledge hierarchy discourse, *wisdom* is succinctly defined as knowledge applied in action, with the caveat that one appreciates the fallible nature of knowledge (Frické, 2009).

The underlying epistemologies and philosophies of the elements of the Knowledge Pyramid have been criticised and critiqued by scholars (Frické, 2009; Tuomi, 1999; Jennex, 2017). However, Jennex (2017) suggests embracing the debate but does not encourage it. Scholars critique the inherent weakness of the traditional Knowledge Pyramid and propose a revised version of it (Jennex, 2017). The revised Knowledge Pyramid (Jennex, 2017) incorporates aspects of the Internet of Things, data analytics, and big data (Jin et al., 2015). Some critics of the Knowledge Pyramid argue that data is not the building block of the higher elements of the Knowledge Pyramid (Tuomi, 1999; Jennex, 2017). They assert that an individual's prior wisdom and knowledge play an essential role in understanding the world, which drives an individual to gather information to collect data about a phenomenon of interest. Their perspective implies that elements of the Knowledge Pyramid are intricately linked, and one could visualise hierarchy flowing downward rather than upward (inverted) (Tuomi, 1999) or in both directions within the context of the natural or real-world (Jennex, 2017). Some arguments for needing a revised Knowledge Pyramid are supported by mathematical facts and social reality (Tuomi, 1999; Frické, 2009). Given the datafied reality we are experiencing, individuals, organisations, and societies continually gather, process, and analyse DIKW by leveraging their insights and sense-making capabilities, further complemented by infrastructural technologies. It is a cyclical process in that one's prior wisdom and knowledge shape gathering information on what aspects of data to collect to accomplish a task. Although not expressly mentioned, we consider the revised Knowledge Pyramid in the present work.

3 METHODOLOGY

3.1 Study Setting

Bhutan, a country known for the concept of Gross National Happiness (Bates, 2009; Metz, 2014), is relatively new to democracy. The matrix of democratic values and ethos in Bhutan started only in 2008 (Elec-

[1] We adapted the search query across the citation databases, but the search keyword remained the same.

tion Commission of Bhutan, 2008). Bhutan is a small country $(38,394\,(14,824) km^2 (mi^2))$ located in South Asia with roughly about 750,000 population. Regarding governance and politics, Bhutan has a democratic, constitutional monarchy system with a bicameral parliament. The king (*Druk Gyalpo*) is the head of the state, and the Prime Minister is the head of the government. In the bicameral parliament, the upper house, the National Council, consists of 25 elected members (20 non-partisan elected plus five members appointed by the Druk Gyalpo). The lower house, the National Assembly, consists of 47 elected members from the ruling and opposition parties. The use of democratic principles to elect candidates and form a government is a recent phenomenon among Bhutanese citizens. Democracy empowers Bhutanese society to elect 47 candidates representing their constituency at the National Assembly. It also allows the voices and aspirations of the citizens to be heard by the highest legislative and executive bodies. For instance, the fourth democratic election concluded in January 2024, and Bhutan has a new democratically elected government.

The electoral stakeholders such as the ECB (Election Commission of Bhutan, 2024) have increasingly used technologies to streamline democratic electoral processes. In doing so, it facilitates the management of coherent and reliable electoral data for free democratic elections. Data from numerous sources, such as documents of political parties, demographic data of candidates, primary-round polling data (vote for political parties of one's choice), and general election polling data (vote for the candidates fielded by the two political parties that have made it through the primary round) is generated at a relatively significant scale. Meanwhile, Bhutanese political parties also use data to inform campaign activities and enrich decision-making, which resonates with Bennett and Lyon (2019) that modern democratic campaigns use data. To illustrate, the political activities, manifesto documents and party websites suggest that they perform document analysis and descriptive data analytics, with complex analytics delegated to external tools such as Facebook ad analytics (Baldwin-Philippi, 2017) during the parliamentary election, which is discussed further in the later section of this paper. It is worth exploring how Bhutan's electoral stakeholders and political parties manage and use DIKW aspects through the lens of the Knowledge Pyramid (Jennex, 2017; Frické, 2009; Tuomi, 1999).

3.2 Method

We used qualitative research design to analyse archival documents and empirical electoral data, albeit secondary data, to provide an account of data, analytics, and knowledge practices in Bhutan's political data ecosystem through the lens of Knowledge Pyramid (Jennex, 2017). It is worth mentioning that we did not conduct interviews or surveys in the present work. The snowballing technique gathered open-government electoral documents and support materials from the ECB (Election Commission of Bhutan, 2024) and related governmental agencies. These documents were made available on their website. The Independent Verification Committee (Independent Evaluation Committee (IEC), as per section 5.4 of Rules of Election Conduct, 2022) verifies these documents, looking at factual accounts of political ideologies and filtering out unfeasible promises. We also consulted the political parties' 4th Parliamentary National Assembly election manifestos in Table 2 for insights into the use of DIKW and related technologies and their implication on political activities. Likewise, to complement the inductive qualitative inquiry (Thomas, 2006) with quantitative insights and understandings, we manually scrapped the aggregated election result data from the ECB website[2] [3]. Indeed, different data sources were central to triangulating data and enriching the description of the democratic electoral process in Bhutan with statistical facts and figures. We also gathered public social media data for information on the profile of Bhutanese political parties across various social media platforms, such as the count of followers and type of posts during the upcoming democratic election.

3.3 Data Analysis

We used a general inductive approach (Thomas, 2006) to analyse the open-access textual documents and numeric polling data to provide an account of aspects of DIKW in the political data ecosystem and to reflect on the data-driven democratic electoral process of electing the 4th National Assembly member in Bhutan. The unit of analysis (Yin, 2014) is the political data ecosystem of Bhutan. Similarly, the unit

[2] Results of the 4th National Assembly Elections, 2023-2024 (primary round)— https://www.ecb.bt/declaration-of-results-of-the-4th-national-assembly-elections-2023-2024/

[3] Results of the 4th National Assembly Elections, 2023-2024 (general election)— https://www.ecb.bt/declaration-of-results-of-the-4th-national-assembly-elections-2023-2024-general-election/

Table 1: Documents consulted for insights into DIKW aspects in the political data ecosystem of Bhutan.

Documents	Source	Description
Election Act of the Kingdom of Bhutan, 2008	ECB	Serves as the overall guiding principle for all election-related activities, including constituency information, nomination processes and disqualifications, memberships and campaigning, advertising rules, and dispute settlements
Information, Communications and Media Act of Bhutan 2018	BICMA	Mandates the Bhutan Information Communication and Media Authority (BICMA) to maintain the required standards and prudent use of ICT and media facilities and mechanisms for data sharing and protection
ECB Social Media Rules and Regulations of the Kingdom of Bhutan, 2018	ECB	Documents rules and regulations for political advertising, the use of social media by the ECB and Election Officials, mentions the period of no campaign (48 hours), and provisions of reporting violations
Election Advertising Regulations of the Kingdom of Bhutan, 2018	ECB	Lists out election advertising particulars, online election advertising, the use of posters and banners, and highlights the use of the Internet or social media only for the election advertising
Electronic Voting Machine (EVM) Rules and Regulations of the Kingdom of Bhutan, 2018	ECB	Serves as a manual for the proper use of electronic voting machines to train polling personnel and voter education, and also outlines the procedures for safeguarding machines and handing-taking over election materials
Strategy for the Implementation of the Provisions Related to Election Advertising, 2021	ECB	Provides more clarity on the implementation of the election advertising provisions in the spirit of fairness and equality and also specifies the extent of time and ceiling amount allocated to the political parties for elections
Media Coverage of Elections Rules and Regulations of the Kingdom of Bhutan, 2021	ECB	Illustrates provisions for fair and equal access to paid election advertising and equal allocation of broadcasting time and space by both ECB and a political party, including restrictions imposed on each
Rules and Regulations on Content	BICMA	Outlines the content specifics to empower content providers and ensure accountability and encourage creativity and innovation
Rules and Regulations on ICT Facilities and Services in Bhutan	BICMA	Informs the general public of ICT facilities and services available and protocols/eligibility to obtain licenses for providing such facilities
e-Governance Policy for the Royal Government of Bhutan	GovTech Agency	Aims to leverage existing and emerging information technology for increasing competitiveness, enhancing productivity, and improving service delivery through online services and sustainable governance

of observation is the different information technologies and related techniques used by various electoral stakeholders and political actors to manage and use DIKW for the democratic electoral process. Since a democratic election is a high-stakes task, aspects of the electoral processes and associated data in Bhutan are well documented in the election-related acts, policies, and guidelines (Table 1). Thus, data sources for this study are analysis-ready for in-depth insights into the intricate dynamics of humans, technologies, and DIKW in the political data ecosystem. All the documents were closely read and analysed (Thomas, 2006) to explicate the phenomenon (Yin, 2014)—that is, the use of DIKW aspects among electoral stakeholders and political actors to facilitate the democratic electoral system in Bhutan. In doing so, the current work upholds the rigour and trustworthiness of a qualitative study (Yin, 2014; Nowell et al., 2017). Through the perspective of the Knowledge Pyramid (Jennex, 2017; Tuomi, 1999), we zoom in on how data emanating from the political space is managed, analysed, and turned into information, knowledge, and wisdom among electoral stakeholders and political actors. We performed qualitative coding (only open code) of the data set for information on action patterns and DIKW flows in Bhutan's political data ecosystem. In order to consolidate our analytic output, a flow chart was used to model the complex DIKW-driven interplay of human, technological, and organisational actors in the democratic electoral process in Bhutan.

4 FINDINGS

4.1 Political Data Ecosystem of Bhutan

The open-access documents and aggregated polling data (primary and general elections) from the ECB are infused with valuable information to examine the data-driven democratic electoral process in Bhutan. In fact, the activities of electoral stakeholders and political parties are well documented by capturing the subtleties of the electoral process, from the announcement of the upcoming National Assembly election to declaring primary and general election results and confirming the ruling and opposition parties. This end-to-end democratic electoral process is informed by various data such as party information, candidate profiles, polling data, and election results. Table 2 provides an overview of the registered political parties, the 2023-2024 National Assembly election results, and the use of various information technologies in the political data ecosystem. The collation, analysis, and interpretation of electoral data in the third and fourth columns and last row of the table are performed by stakeholders at ECB. The gathering of the polling data and corresponding rigorous analytics to transform data into higher elements of the Knowledge Pyramid, namely information (such as aspects of the respective constituency) and knowledge (such as electoral trends and patterns), of the 4th National Assembly general election can be viewed elsewhere (https://www.ecb.bt/geresults2024/). Regarding the wisdom aspects of the Knowledge Pyramid, upon rigorous analysis of electoral results and methodical assessment of the situation, the electoral stakeholders in Bhutan work on tasks of ensuring that the Bhutanese society has a new government and an opposition party, as illustrated in Figure 1.

Social media platforms have a significant impact on the political data ecosystem (Bennett and Lyon, 2019; Baldwin-Philippi, 2017). The political parties in Bhutan recognise the opportunities afforded by social media platforms, albeit wary of the downside of such tools on democracy. Bhutanese political parties use social media, such as Facebook, WhatsApp, and Messenger, extensively for crowdsourcing information on political activities, collecting constructive feedback on pledges, and gathering knowledge of community needs. For instance, during the 4th National Assembly election, the BTP conducted election-related ad campaigns on Facebook (Baldwin-Philippi, 2017). Meta Ad Library automatically calculates values for metrics such as platform (Facebook, Instagram, and Audience Network), categories (social issues, elections or politics), ad audience size, amount spent, impressions, gender, and location. The merit of such tools is that ad delivery information is presented well in a digestible format (in summaries, charts, and graphs), such as the 'Impressions' of the official BTP Facebook page was 700K–800K. The post mainly appeared in social media posts and feeds of people in Thimphu *dzongkhag* [4] (37%) and mostly covered people aged 18-44[5]—that is, the digital ads of BTP was displayed on Facebook users in the aforementioned region and demography (see *microtargeted Facebook ads* in Baldwin-Philippi, 2017, p. 629). Similarly, the political parties also use *Kuensel*'s[6] Facebook page for digital ads during the democratic election.

Additionally, from our lived experiences of democratic election reality in Bhutan, political parties use insights assembled from social media for microtargeting (Bennett and Lyon, 2019), albeit rudimentarily, through personalised text messages for potential electorates, audio chats in social media groups, and emails to potential individual electors (Papakyriakopoulos et al., 2018). All these activities to leverage social media for political endeavours are guided by the ECB's in-house social media policy (Election Commission of Bhutan, 2018). The policy also stipulates that if the security level of social media platforms and access privilege of the desired information is uncertain, persons with a stake in the electoral process should not share confidential information on such tools. Considering literacy aspects of potential electors and other factors such as geography, Bhutanese political parties generally prefer Facebook posts on political activities, expressly or implicitly gathering information and knowledge for political ends. They or their party workers are also active across social media groups for insights into and synthesis of the political debates and discussions among group members on needs and demands in the community. In a

[4] *Dzongkhag* means a district in Dzongkha language

[5] BTP conducted an ad campaign on Facebook during the 4th National Assembly Election in Bhutan—https://www.facebook.com/ads/library/?active_status=all&ad_type=all&country=ALL&view_all_page_id=101216219504499&search_type=page&media_type=all

[6] *Kuensel* is the national newspaper of Bhutan

Table 2: Overview of the political parties that contested in the 4th National Assembly election in Bhutan.

Political Party	Political Slogan	Primary Election*	General Election⁺	Remark (Use of data analytics, social platforms, and other technologies)
Bhutan Tendrel Party (BTP) (Founded: November 2022)	Your Voice, Your Hope	61,331 (19.58%)	Votes— 147,123 (45.02%) 17 (Opposition)	Conducted ad campaigns related to social issues, elections, or politics on Facebook; 31K followers on Facebook; 18 followers on X (formally Twitter); and 132 followers on Instagram
Druk Nyamrup Tshogpa (DNT) (Founded: 20 January 2013)	Putting Nation First	41,106 (13.12%)		41K followers on Facebook, 5.8K+ followers on X, and 52 followers on Instagram
Druk Phuensum Tshogpa (DPT) (Founded: 25 July 2007)	Economic Prosperity and Social Well-being; Development with Equity and Justice	46,694 (14.91%)		8.9K followers on Facebook and 1.6K+ followers on X
Druk Thuendrel Tshogpa (DTT) (Founded: 2 May 2022)	Sunomics: Buddhist Capitalism with the spirit of GNH	30,814 (9.83%)		34K followers on Facebook, no X account, and 459 followers on Instagram
People's Democratic Party (PDP) (Founded: 24 March 2007)	For a Better Drukyul. The Promise we will Deliver.	133,217 (42.53%)	Votes—179,652 (54.98%) 30 (Ruling party)	Maintains an active political party website —https://pdp.bt, 61K followers on Facebook, 7.3K+ followers on X, and 117 followers on Instagram

*: Eligible Voters—497,058; Total Votes—313,162; Voter Turnout—63% (EVM: 195,719; PB: 117,443)
⁺: Eligible Voters— 498,135; Total Votes— 326,775; Voter Turnout— 65.6% (EVM—218,273; PB—108,502)

sense, Bhutanese political parties use social media as an information-eliciting and knowledge-building platform.

4.2 Data Practices Among Political Stakeholders

Bhutan's electoral stakeholders and political parties manage data and associated knowledge for various political activities. Figure 1 illustrates a flow chart of the recurrent action patterns in the electoral process of political parties filing to contest in the upcoming democratic election at the ECB through the primary and general rounds of casting votes to the declaration of results to the formation of ruling and opposition parties. The data, that is, votes from EVM and postal ballots, is consolidated by following schema and parameters developed by ECB (see Fig 5.1 in Election Commission of Bhutan, 2013, p. 16). ECB gathers massive amounts of data during the election period, such as party registration, electorate registration, and in-country or overseas postal ballot voters. They also maintain an information system (https://dramig.ecb.bt/membership/check) for registering political party members. The party members can use the system to check their membership status. If it is valid, the system displays information about their political affiliation. ECB also has a web-based system (https://berms.ecb.bt/enrollment/check) whereby electorates can check their enrolment status to vote in the upcoming democratic election. In Bhutan, voting is not compulsory, and the electoral system does not permit proxy voting. If an electorate wants to cast a vote via postal ballot (in-country or overseas), the system records the details of elector information and allows changing postal addresses based on the voters' choice.

The political actors of Bhutan now recognise the importance of tapping the value of data generated in the political data ecosystem. It is evident from their campaign activities and manifesto document that they perform analytics of data on socioeconomic developments in Bhutan, such as descriptive statistics and figures, to inform campaign activities, drive decision-making, and develop party pledges. Such activities are consistent with Bennett and Lyon (2019) that all modern democratic election campaigns use data. The website of some political parties has a feature that allows potential electors to join as party members. For example, the PDP's website has a 'Join Us' feature. It asks for CID, name, contact no, *gewog* (county), and *dzongkhag*. After entering this information, the system automatically fetches the name of the party candidate in the constituency of an elector. In effect, such activities are instances of using information technology, such as voter relationship management system (Bennett and Lyon, 2019), and DIKW aspects for informational canvassing and steering political campaigns through targeting and testing (Baldwin-Philippi, 2017). Similarly, the DTT's website uses cookies to track web traffic and click streams, which could be used to analyse the digital footprint of potential electors. The actions and dealings with data suggest that the political parties expressly or implicitly elevate data gathered from democratic politics to the higher elements of the Knowledge Pyramid (Jennex, 2017). However, since democracy in Bhutan is relatively embryonic, the analytical maturity of political parties is rather low in that no political parties use complex data analytics technologies to analyse political data for building predictive models or scoring to target voters during the democratic election (Bennett and Lyon, 2019).

4.3 DIKW in the Political Data Ecosystem

The data-driven democratic electoral activities of the political actors indicate how they deal with the DIKW elements of the Knowledge Pyramid (Jennex, 2017). The strategy to manage knowledge in the data analytics processes is a critical success factor (Jennex, 2017) as it has implications on exploiting DIKW for campaign activities and decision-making, such as insights into electorate voting (EVM and postal ballot) pattern and election trend analysis. Fig. 1 illustrates the interplay and dynamics of various sociotechnical actors during the democratic electoral process in Bhutan. It also embodies a complex exchange of DIKW among numerous stakeholders, such as the ECB, political parties, and the Office of the Returning Officer. Data such as manifesto documents for registration to contest in the election, primary and general round polling day data, and comprehensive data of the electorate are generated, managed, communicated, and used during the democratic election period. The political parties also bootstrap available resources and technologies to gather, manage, and turn data into higher elements of the Knowledge Pyramid (see Table 2). For instance, they use websites, social media, and other information systems to manage and operationalise DIKW in Bhutan's political data ecosystem. Their social media posts of the political activities during the upcoming election, such as photos from community *zomdu* (translates to inclusive decision-making) and updates on personal meetings with party workers in the 47 constituencies, also illustrate that the predominant method that political parties use for gathering aspects of DIKW is in-person campaigning or door-to-door canvassing (Dommett et al., 2023; Baldwin-Philippi, 2017; Bennett and Lyon, 2019).

Furthermore, polling day is one instance where data must be as accurate as possible; otherwise, it would endanger the democratic electoral process. The demographic information of electorates is carefully vetted through the voter identity card at the 809 polling stations. ECB follows first-past-the-post voting, where the voter votes for a single candidate, and the candidate with the maximum votes wins the election. On the polling day of the primary and general elections (Fig. 1), data from 47 constituencies and 809 polling stations across the country are transferred to the Department of Election of ECB for tallying the votes. The stakeholders in the department collate, analyse, and add meaning and context to the polling day data, which is an instance of dealing with elements of the Knowledge Pyramid (Jennex, 2017). For example, analysis of the polling data from the primary round is used to declare the winning and runners-up parties where they are eligible to participate in the general election to form the government, as illustrated in Fig. 1. Overall, the flowchart captures the logic of how democratic election decisions are made in the primary and general rounds to form the ruling and opposition parties. These recurrent action patterns in the

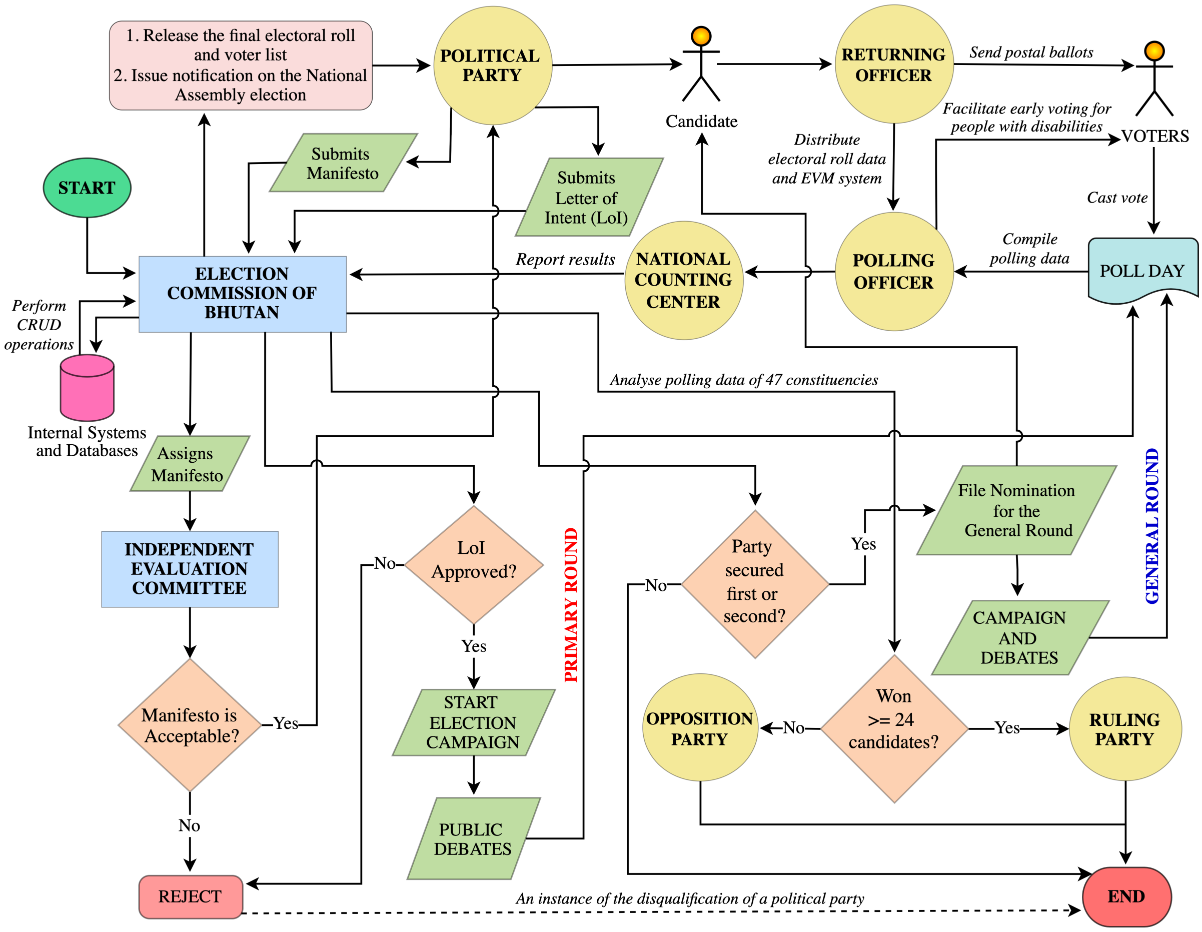


Figure 1: Dynamics and flow of DIKW in Bhutan's democratic electoral process.

flow chart every five years require electoral stakeholders and political parties to wrestle with DIKW tirelessly in the political data ecosystem of Bhutan.

5 DISCUSSION

5.1 DIKW-Driven Political Data Ecosystem

Considering the considerable influence of and discourses on data and analytics during the democratic election (Papakyriakopoulos et al., 2018; Baldwin-Philippi, 2017; Bennett and Lyon, 2019), electoral stakeholders and political actors must determine the actual and potential value of DIKW and its implications on political activities. Moreover, understanding DIKW (Jennex, 2017; Tuomi, 1999) would avoid duplicative efforts and ease the extraction of the value of data (Micheli et al., 2020) emanating from the political data ecosystem. It is also timely for the nascent political data ecosystem in emerging democracies, such as Bhutan, to determine barriers and opportunities to share and use political data for facilitating democratic election and avoid issues of information asymmetry (Mittelstadt et al., 2016)—that is, unequal distribution of or access to information. Indeed, doing so would enrich the open data attitude and culture in the political space, especially exploring possibilities for DIKW sharing and reuse among political actors. Furthermore, individuals entrusted with data analytics among political parties should conceive data-driven politics-related activities in terms of DIKW. Political parties with resource constraints need more technical know-how to perform complex data analytics and use analytical results to manage campaign activities and target electors during democratic election campaigns (Bennett and Lyon, 2019; Baldwin-Philippi, 2017). In Bhutan, given the smallness of the constituencies, geographic-based electoral targeting is a predominant method among political parties for voting intelligence of potential electors. They also gather DIKW of potential wins for their candidates in certain strongholds through their party workers as information conduits, albeit surprises are not uncommon during elections.

The political data ecosystem needs a systematic approach to managing the DIKW elements of the Knowledge Pyramid (Jennex, 2017; Frické, 2009). For instance, a contextual framework to manage DIKW (Jennex, 2017) is essential to facilitate the frictionless flow of data within political parties for decision-making and campaign activities (see Nissen, 2002 on the relational link between data, information, and knowledge). Likewise, electoral stakeholders and political parties must know what DIKW is needed and where it is available. Hence, providing support systems, such as developing capabilities for acquiring data from different sources and subsequent transformation into the higher levels of the Knowledge Pyramid (Jennex, 2017), is also crucial to tap the value of data emanating from various sources. Moreover, a coherent and reliable DIKW will positively affect electoral activities, such as DIKW-informed, analytics-driven, and cost-effective campaign activities (see Table 6 in Dommett et al., 2023). Additionally, mechanisms should be in place to protect DIKW from imitation by others (Gelhaar and Otto, 2020), which calls for proper inventorying of DIKW in the political environment (see Miller and Mork, 2013 on data inventory) to avoid the risk of accumulating ineffective DIKW and mistargeting electors. Regarding voter analytics and political micro-targeting (Bennett and Lyon, 2019), clear legislation is fundamental about the ethics and legitimacy of using electorate data and related databases for data analytics to shape political campaigns and inform decision-making (Dommett et al., 2023; Baldwin-Philippi, 2017). In Bhutan's context, some safeguards, such as acts, policies, and guidelines, as summarised in Table 1, are in place to deal with the misuse of so-called political technologies (Ruppert et al., 2017) during the democratic election.

5.2 Knowledge Management in the Political Data Ecosystem

The complex interplay of various individuals, organisations, and institutions is fundamental for fostering the growth of an ecosystem (Gelhaar and Otto, 2020) and creating and sharing knowledge. Thus, it is crucial to ensure that tacit (experiences and insights) and explicit knowledge (knowledge formalised and codified in electronic systems) (Choo, 1996; Nonaka, 2007; Nissen, 2002) is used effectively and efficiently in the political data ecosystem. For example, in Bhutan, the ECB gathers political documents (such as manifestos, charters, and party membership) and polling data and accordingly shares them on its website. Political parties could explore opportunities to encourage knowledge sharing and use through collaboration with other parties. It would provide an avenue to exchange services whereby political actors share and exchange skills, capabilities, and knowledge that benefit each other. Some political data have a short shelf life in that data value decays over time (Namgay et al., 2023; Lee, 2017). Hence, political actors must also deal with boundary paradox (Quintas et al., 1997), where boundaries should be open for DIKW flow among the relevant stakeholders, but at the same time, political parties also have to protect and nurture their DIKW and intellectual capital. Quintas et al. (1997) argue,'It is upon the dynamic preservation of the latter [protect and nurture knowledge base and intellectual capital] that survival depends.'

The issue in the political data ecosystem is how to manage knowledge using effective and efficient technologies, which tools and techniques to use for managing DIKW, and what DIKW related to politics is needed for decision-making and campaign activities. Moreover, knowledge generation and application is a cyclical process, which alludes to the critique of knowledge hierarchy (Frické, 2009; Tuomi, 1999; Jennex, 2017). It costs resources, effort, and time to analyse data and turn it into knowledge. Therefore, it is imperative to have a practical plan or strategy to deal with and work on DIKW Jennex (2017), especially knowledge in the political data ecosystem, such as identifying metrics for managing and gauging DIKW. If there is shareable data by de-identifying certain attributes, such as personally identifiable information, stakeholders such as ECB and other governmental organisations should share it with political parties. Doing so prevents the political parties from performing needless tasks such as setting up infrastructures to collect and analyse data for insights and knowledge. It is also imperative to incentivise political parties to use data analytics and foster a culture of sharing and using knowledge (Wang and Noe, 2010), albeit for sociopolitical good. Otherwise, in emerging democracies with limited resources, mobilising data and analytics at scale for political ends is resource intensive, and inaccessibility to rich DIKW risks data-driven democratic politics favouring established political parties (Bennett and Lyon, 2019) with means to use complex political technologies and employ human resources to manage aspects of DIKW in the political data ecosystem.

5.3 Contributions, Limitations, and Future Work

This study contributes to the literature on knowledge management via insights into DIKW aspects (Jennex, 2017; Tuomi, 1999) in the political data ecosystem of emerging democracies that use data and analytics for political activities. We answer the call by Dommett et al. (2023) for exposition on data practices in democratic elections in different countries and contexts, especially 'less US-centric studies'. The present work also contributes to the discourse on the evolution of the nascent body of knowledge on the political data ecosystem. It is also timely for researchers interested in data-driven politics (Bennett and Lyon, 2019; Baldwin-Philippi, 2017) to rethink how the Knowledge Pyramid elements (Jennex, 2017) are managed and used in the political space. The study likewise augments the theory and praxis of the Knowledge Pyramid in research. Regarding implications for data stakeholders as well as policymakers, this study provides a platform to reflect on their current data practices from the frame of the Knowledge Pyramid (Jennex, 2017; Tuomi, 1999; Frické, 2009) to unlock the actual and potential value of data emanating from the political data ecosystem. The in-depth description of facets of and relational links between DIKW also sheds light on how political actors can administer data-driven electoral processes and manage knowledge for decision-making and campaign activities.

We acknowledge that this study has some limitations. The current research is based on a country where democracy is not even two decades old. Likewise, the present work only used open-access governmental documents and aggregated polling data to provide an account of the political data ecosystem from the perspective of the Knowledge Pyramid (Jennex, 2017). The subjectivity of the DIKW elements of the Knowledge Pyramid (Frické, 2009) could also affect the interpretations of the findings and discussions in the present work. An area worth exploring is using empirical data such as interviews and surveys to examine data-driven democratic political activities via the lens of the Knowledge Pyramid. Another research direction is algorithmic processes that underlie the data analytics systems in political technologies, namely the handling of sensitive electorate data. Moreover, insights into the suggested study will illustrate how individuals, political actors, and institutions in other countries manage and use different facets of data for electoral and political activities without compromising privacy and security.

6 CONCLUSION

In this paper, we analysed Bhutan's political data ecosystem through the lens of the Knowledge Pyramid to examine the dynamics of DIKW generated in the political data ecosystem to facilitate democratic electoral activities and political campaigns. The democratic electoral process of the 4th National Assembly Election in Bhutan was unpacked for insights into the practices of transforming data into higher levels of the Knowledge Pyramid. In Bhutan, the electoral stakeholders and political parties deal with DIKW on a relatively small scale and with varying infrastructural complexities. The political parties use websites, information systems, and social media to manage aspects of DIKW, facilitate decision-making, and administer campaign activities. However, considering the nascency of democracy in Bhutan, no extensive data technologies or human resources are employed to harness the actual and potential value of data generated in the political space. In this regard, electoral stakeholders and political actors could explore opportunities to adopt robust data analytics technologies and knowledge management infrastructure to manage DIKW emanating from the political data ecosystem. This study advances the literature on knowledge management through a fine-grained account of the democratic election in a country experiencing the significant effect of political technologies to turn political data into higher elements of the Knowledge Pyramid for various ends in the political world.

REFERENCES

Baldwin-Philippi, J. (2017). The myths of data-driven campaigning. *Political Communication*, 34(4):627–633.

Bates, W. (2009). Gross national happiness. *Asian-Pacific Economic Literature*, 23(2):1–16.

Bellinger, G., Castro, D., and Mills, A. (2004). Data, information, knowledge, and wisdom.

Bennett, C. J. and Lyon, D. (2019). Data-driven elections: Implications and challenges for democratic societies. *Internet Policy Review*, 8(4).

Choo, C. W. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, 16(5):329–340.

Dommett, K., Barclay, A., and Gibson, R. (2023). Just what is data-driven campaigning? a systematic review. *Information, Communication & Society*, 27(1):1–22.

Election Commission of Bhutan (2008). *Election Act of the Kingdom of Bhutan, 2008*. Election Commission of Bhutan, Thimphu.

Election Commission of Bhutan (2013). *Handbook for Counting Supervisor*. Election Commission of Bhutan, Thimphu.

Election Commission of Bhutan (2018). *ECB Social Media Rules and Regulations of the Kingdom of Bhutan, 2018*. Election Commission of Bhutan, Thimphu.

Election Commission of Bhutan (2024). *Election Commission of Bhutan*. Election Commission of Bhutan, Thimphu.

Frické, M. (2009). The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science*, 35(2):131–142.

Gelhaar, J. and Otto, B. (2020). Challenges in the emergence of data ecosystems. In *23rd Pacific Asia Conference on Information Systems*, 1–14.

Jennex, M. E. (2017). Big data, the internet of things, and the revised knowledge pyramid. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 48(4):69–79.

Jin, X., Wah, B. W., Cheng, X., and Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, 2(2):59–64.

Khatri, V. and Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1):148–152.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3):293–303.

Metz, T. (2014). Gross national happiness: A philosophical appraisal. *Ethics and Social Welfare*, 8(3):218–232.

Micheli, M., Ponti, M., Craglia, M., and Suman, A. B. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2).

Miller, H. G. and Mork, P. (2013). From data to decisions: A value chain for big data. *IT Professional*, 15(1):57–59.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):1–21.

Namgay, P., Wangdi, P., Thinley, S., and Namgyel, T. (2023). Fairifying STEM data ecosystem to enhance data reuse. In *2023 IEEE Frontiers in Education Conference*, 1–9.

Nissen, M. E. (2002). An extended model of knowledge-flow dynamics. *Communications of the Association for Information Systems*, 8:251–266.

Nonaka, I. (2007). The knowledge-creating company. *Harvard Business Review*, 85(7/8):162–171.

Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1).

Oliveira, M. I. S. and Lóscio, B. F. (2018). What is a data ecosystem. In *19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1–10.

Olteanu, A., Castillo, C., Diaz, F., and Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2:1–33.

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., and Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for germany. *Big Data & Society*, 5(2):1–15.

Quintas, P., Lefrere, P., and Jones, G. (1997). Knowledge management: A strategic agenda. *Long Range Planning*, 30(3):385–391.

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2):163–180.

Ruppert, E., Isin, E., and Bigo, D. (2017). Data politics. *Big Data & Society*, 4(2):2053951717717749.

Schippers, B. (2020). Artificial intelligence and democratic politics. *Political Insight*, 11(1):32–35.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2):237–246.

Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, 1–12.

Wang, S. and Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review*, 20(2):115–131.

Yin, R. K. (2014). *Case Study Research: Design and Methods*. SAGE Publications, Inc., 5 edition.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4):479–493.

Support Learning Design and Analytics with EduP Knowledge Model

Thi Hong Phuc Nguyen[1,2], Ngoc Tram Nguyen-Huynh[1,2] and Thi My Hang Vu[1,2,*]
[1]*Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam*
[2]*Vietnam National University, Ho Chi Minh City, Vietnam*
{20120350, 20120388}@student.hcmus.edu.vn, vtmhang@fit.hcmus.edu.vn

Keywords: Knowledge Model, Ontology, Learning Design, Learning Analytics.

Abstract: Learning design (LD) have been a prominent topic in the academic community for many years. It aims at planning and organizing learning activities and resources to promote learning process and engage students in achieving learning outcomes. Learning analytics (LA) has matured in the education field and developed a strong connection with learning design. Learning analytics provides valuable insights to inform learning design decisions, while learning design serves as a means to turn learning analytics results into actionable strategies. Their alignment completes the big picture for enhancing teaching and learning. Despite numerous studies proposing means to support LD/LA and their alignment, both fields still face many challenges due to the lack of a consolidated framework for reflecting on the various types of knowledge essential for LD/LA. This paper aims at proposing a comprehensive framework, named EduP (Education-Domain-User-Pedagogy), that supports LD/LA by leveraging different types of knowledge. The main contributions of the framework include a knowledge model and an insight engine. The knowledge model helps clarify essential components for LD/LA and their relationships, while the insight engine addresses how this knowledge is accessible to teachers in the context of LD/LA. A brief discussion on the implications and future research is also presented.

1 INTRODUCTION

Our research focuses on the multidisciplinary aspects of learning design and learning analytics, and their alignment within a framework that supports them by leveraging knowledge-based solutions.

Learning Design (LD) has been a prominent topic in the academic community for many years. LD focuses on creating and refining learning scenarios, i.e., which consist of a sequence of learning activities and resources to engage learners in achieving specific learning outcomes (Koper & Bennett, 2008).

The creation of these scenarios is based on various pedagogical strategies (e.g., preparing activities and resources for problem-based learning differs from those used in inquiry-based learning). Many studies have been conducted to assist teachers in creating effective learning scenarios through various means, such as editing tools (Celik & Magoulas, 2016; Pozzi et al., 2020), modeling languages for specifying learning scenario elements (Botturi & Stubbs, 2008), or pedagogical design patterns (Eyal & Gil, 2020).

Learning Analytics (LA) is another area of interest in educational science, focusing on collecting, processing, and mining educational data to generate insights that support educational decision-making (Hernández-de-Menéndez et al., 2022; C. Romero & Ventura, 2020). With the advent of advanced data analytics tools and methods, learning analytics (LA) has emerged as a powerful tool for analyzing educational data to enhance learning and teaching (Mangaroska & Giannakos, 2019).

Many researchers in the analytics field focus on developing learning analytics dashboards to visualize learner performance and progression (Susnjak et al., 2022). Others concentrate on identifying associations within learning data to discover new insights, such as predicting dropout rates or identifying at-risk learners (Ouyang et al., 2023; Ramaswami et al., 2023). Another area of interest is personalization, which aims to provide learners with appropriate activities and resources based on their learning contexts (Chatti & Muslim, 2019; Romero et al., 2019).

Learning design and analytics have matured in their respective fields. However, these two topics

[*] Corresponding author

Nguyen, T., Nguyen-Huynh, N. and Vu, T.
Support Learning Design and Analytics with EduP Knowledge Model.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 97-107
ISBN: 978-989-758-716-0; ISSN: 2184-3228

have a strong convergence. Learning analytics provides valuable insights to inform learning design decisions. On the other hand, learning design serves as a means to turn learning analytics results into actionable strategies (Mangaroska & Giannakos, 2019). Together, these two fields offer a comprehensive approach to improving teaching and learning activities. As a result, a significant number of studies focus on the alignment of learning design and learning analytics (Ahmad et al., 2022; Bakharia et al., 2016).

Despite the numerous studies on proposing means to support LD/LA and their alignment, both fields still faces many challenges. The first challenge involves elaborating learning scenarios, which requires teachers to have a strong understanding of both pedagogical principles and learning domains, as well as how to integrate them effectively (Schmitz et al., 2017). A significant number of teachers lack this knowledge, which limits their ability to design effective learning scenarios (Lui & Bonner, 2016; Tatto et al., 2020). Therefore, developing learning scenarios remains a challenging task for teachers that requires additional support (Vu & Tchounikine, 2021). The second challenge involves the lack of a consolidated framework for learning analytics, which prevents data from being interpreted meaningfully. This makes it difficult to derive actionable insights from the data, thereby complicating their effective application to enhance teaching and learning (Ahmad et al., 2022).

To address these challenges, this study aims to propose a comprehensive framework, named **EduP** (**E**ducation-**D**omain-**U**ser-**P**edagogy), that supports learning design and analytics by leveraging different types of knowledge, such as pedagogical knowledge and learning domain knowledge. The framework also focuses on ensuring that this knowledge is accessible to teachers in the context of learning design and learning analytics.

To propose such supportive tools, knowledge should be clarified and structured efficiently. Additionally, to provide a consolidated framework, all stages—from collecting and organizing data to importing it into the knowledge base, to exploiting and disseminating the knowledge to teachers—should be well-defined. Ontologies, which provide formal representations of domain concepts and serve as powerful reasoning tools, are essential for structuring knowledge (Vu et al., 2023). The state-of-the-art reveals numerous types of ontologies for modeling learning activities, learning outcomes, learning domain knowledge, learner profiles, or generic ontologies that can be applied across various domains (Rahayu et al., 2022; Wang & Wang, 2021).

In the subsequent sections, the paper presents in more detail an ontology-based framework for LD/LA. These sections are organized as follows. Section 2 presents the methodology adopted to target the objectives of this study. Section 3 introduces related works by first providing a brief summary of essential topics on knowledge modeling and exploiting in education, followed by a review of the state-of-the-art research in these areas. Section 4 clarifies the first output of the paper, which is the definition of essential knowledge types that assist in LD/LA. Section 5 introduces a knowledge structure to organize these knowledge types as the second output of the paper. Subsequently, Section 6 provides a method/process for discovering this knowledge through the use of a reasoning engine as the last output of the paper. Section 7 focuses on the validation of the propsed framework through some real-world scenarios in higher education. Finally, Section 8 concludes by the implications and limitations of the study and suggests directions for future research.

2 METHODOLOGY

This section outlines a methodology based on the Design Science Research (DSR) methodology to conduct the research presented in this paper (Dresch et al., 2015). DSR emphasizes the creation of innovative artifacts to solve specific problems.

DSR's artifacts can be: *constructs* providing fundamental concepts for describing a specific problem and its solutions; *models* linking the constructs in a real-world situation; *methods* providing guidelines/processes for solving problems; and *instantiations* demonstrating how the theoretical constructs, models, and methods can be applied in practice (Peffers et al., 2007).

This research aims at proposing a knowledge-based framework for enhancing learning design and analytics (**EduP** Framework). The artifacts for the framework are created through the following phases.

Problem Identification. This phase focuses on identifying the research questions to be addressed for building EduP framework. Two key questions are identified: RQ#1: *What types of essential knowledge can support learning design and analytics*? And RQ#2: *How can the knowledge be elaborated and used effectively*?

Solution Definition. This phase defines the objectives of a solution to solve the identified problem, which requires the EduP framework: define essential knowledge types in LD/LA and determine effective methods for reasoning and disseminating this knowledge. To define these objectives, a brief literature review is conducted to summarize the current state-of-the art in knowledge modeling and reasoning for the education sector.

Design and Development. This phase involves creating EduP artifacts. These artifacts are classified in constructs, models, methods, and instantiation, according to DSR methodology (Peffers et al., 2007).

- EduP **Constructs** and EduP **knowledge model** are proposed to address RQ#1. The constructs define key knowledge components in LD/LA, while the knowledge model outlines how these components are related to one another.
- EduP **Insight Engine** is proposed as a method/process within the framework to response to RQ#2. The method defines multiple modules for representing, elaborating, and reasoning about knowledge, aiming to generate insights in LD/LA.
- Two **Instantiations** are also created to validate the framework in a subsequent phase. The first one is an ontology based on EduP knowledge model. The second one is a web-based reasoner built upon EduP insight engine. The reasoner serves as a prototype for reasoning with the created ontology through a simple interface.

Demonstration and Evaluation. This phase involves validating the proposed framework in real-world situations. The ontology and web-based reasoner developed in the previous phase are used to address various case studies in the higher education context.

3 RELATED WORKS

This section provides an overview of current research on ontology-based solutions for knowledge modeling and reasoning in education, addressing both the structural and behavioral aspects of these solutions.

3.1 Structural Aspect

The structural aspect focuses on the types of knowledge represented in ontologies. An ontology is a specification that defines concepts within a domain and their relationships in a structured, formal, and explicit manner (Gruber, 1993). In education, an ontology is defined as "a system of primitive vocabularies/concepts for constructing a tutoring system" (Mizoguchi et al., 1996). In technology-enhanced learning, ontologies are considered effective tools for modeling the learning and teaching domain due to their formal expressiveness, support for sharing, and reasoning capabilities.

In General, ontologies can be used to model a wide variety of information types. The classification of these ontologies can follow different criteria, such as their levels of abstraction (domain-independent and domain-specific ontologies) (Guarino et al., 2009), their intentions (domain ontologies modeling a target domain, task ontologies modeling generic problems and their solutions, and application ontologies dedicated to activities within a specific application), (Al-Yahya et al., 2015; Mizoguchi et al., 1996). From the perspective of smart systems, ontologies can be classified in five major types: ontologies for modeling generic concepts across domains, domain ontologies, user ontologies, context ontologies, and merged ontologies combining multiple ontology types to provide a comprehensive reasoning (Chimalakonda & Nori, 2020; Vu et al., 2023).

In Education, most ontologies are *domain ontologies*, which are used to describe the concepts of specific learning domains such as mathematics, physics, and programming (e.g., Iatrellis et al., 2019; Lalingkar et al., 2014; Ramesh et al., 2016). Other research also focuses on modeling *pedagogical elements* such as curriculum/syllabus, learning scenarios, learning activities, and outcomes (Hyunsook & Jeongmin, 2016; Katis et al., 2018; Reynolds et al., 2023). *Learner profiles* specifies user data such as user profiles, interest, needs; which is typically employed for learning recommendations and personalization (e.g., Pelap et al., 2023; Romero et al., 2019). *Context ontologies* are another type that emerged with the evolution of smart systems, supporting retrieving the most relevant knowledge, according to a specific learning context (Aguilar et al., 2018; Cabrera et al., 2017; Ouissem et al., 2021; Perera et al., 2014).

3.2 Behavioral Aspect

The behavioral aspect examines the processes and methods used to develop and reason with knowledge. This involves two key components: elaboration and reasoning.

Elaboration, also known as ontology building, involves the extraction of data and its integration into a predefined knowledge model or ontology structure, as outlined in the structural aspect. which can be done manually, semi-automatically, and automatically.

Manual approaches are performed by domain experts who analyze the specific domain, annotate data, and manually integrate it into an ontology structure (e.g., Verdú et al., 2017). This method is costly and prone to errors, especially with large datasets. However, human interaction and expert domain analysis can ensure the rationality of the extracted data, which is essential for complex domains such as education.

Automatic approaches employ natural language processing (NLP), data mining, or machine learning algorithms to extract information from unstructured data, such as text, and integrate it into an ontology structure without human intervention (e.g., Aguilar et al., 2018; Lacasta et al., 2018; Wei & Shao, 2022). These methods facilitate efficient and cost-effective knowledge extraction from large datasets. However, the absence of expert monitoring can lead to the generation of knowledge that may be unreasonable or inaccurate within the domain.

Semi-automatic approaches involve both domain experts and algorithms (e.g., (Cano-Benito et al., 2021; Chang et al., 2020; Ghazal et al., 2020; T. M. H. Vu & Tchounikine, 2021)). Several parts of the process are performed or supervised by experts, while others are carried out by algorithms. These approaches benefit from AI techniques to automate part of the extraction process, while also leveraging the collaboration with domain experts to minimize the risk of constructing an incomplete or potentially incorrect knowledge base

The second key component of the behavioral aspect is **reasoning**. The purpose of reasoning is to derive insights, make inferences, and effectively utilize the knowledge within ontologies to address issues and answer questions. This can be done through various methods such as query languages, built-in reasoners, and user-defined inference rules and algorithms.

Ontology query languages and built-in reasoners are typical solutions for *ontology-based reasoning*. The most widely used ontology query language is SPARQL, which provides a formal syntax for extracting and manipulating data within ontologies (e.g., Lacasta et al., 2018; LeClair et al., 2022; Wen et al., 2022). Another solution involves using built-in reasoners such as HermiT and Pellet, which are integrated into ontology editors like Protégé (e.g., Andrade et al., 2019). The reasoners enable automatic deduction and consistency checking. However, these approaches have a limitation: they require additional interfaces to be user-friendly, as query languages and the Protégé interface are too complex for non-technical users.

Another common solution involves using *human-defined rules* to discover knowledge from ontologies (e.g., Bensassi et al., 2019; Ghazal et al., 2020; L. Romero et al., 2019) or proposing *custom solutions and algorithms* tailored to specific applications and purposes (e.g., Demaidi et al., 2018). These approaches can be resource-intensive, requiring significant input from experts to define rules and program algorithms, and may not fully leverage the inherent benefits of ontology support.

3.3 Research Gap Identification

From a brief summary and analysis of related works, we have identified several research gaps that highlight the motivation behind our research and provide a clear direction for how our work can address the current limitations in the existing literature. These limitations are outlined as follows.

There is Insufficient Focus on Aligning Multiple Ontologies to Enhance Teaching and Learning. Improving LD/LA requires teachers to have a strong knowledge not only in their specific teaching domain but also in pedagogical strategies to create effective learning scenarios. *Aligning these ontologies should be clearly defined and integrated into a unified framework* that can holistically facilitate their implementation and application in LD/LA.

There is a lack of user-friendly tools for utilizing knowledge from the end-user perspective. Numerous types of ontologies have been proposed, but we still lack an efficient means to deliver the knowledge contained within these ontologies to users. Current methods do not pay enough attention to bridging the gap between the complex, structured data within ontologies and the practical, accessible insights needed by end users. This highlights a *critical need for developing user-friendly interfaces and tools* that can facilitate the efficient extraction and application of knowledge from ontologies.

There is a Lack of a Unified Framework That Incorporates a Knowledge-Based Approach to Promoting LD/LA. Such a framework would facilitate a more cohesive and systematic application of ontological principles and could serve as a blueprint for the future conception and development of these services. The framework needs to *encompass all phases, from collecting data and integrating it into knowledge bases to delivering the knowledge effectively to end users through efficient means*.

4 KNOWLEDGE COMPONENTS

This section introduces EduP constructs in detail to clarify the key knowledge components involved in knowledge-based learning design and analytics.

These constructs are organized into a multi-level structure to facilitate reusability and future extension. The abstract level adopts the 5W1H model (**w**ho, **w**hat, **w**hy, **w**hen, **w**here, and **h**ow) as proposed by (Jang & Woo, 2012). This model enables general reasoning to anwser the question such as "who achieves what in which context (when, where)?".

Inherited from the abstract level, concepts at lower levels are tailored specifically to the education sector. This level encompasses three core knowledge types in a KB-based LD/LA framework: **p**edagogical knowledge (learning goal, learning outcome, learning level, activity); learning **d**omain knowledge (topic); **u**sers (individual learner, group). Additionally, contextual knowledge is defined to facilitate connections among these three knowledge types (see Figure 1 for the proposed constructs and their relationships).

5 KNOWLEDGE MODEL

This section defines a generic knowledge structure that we use to organized the proposed constructs.

Pedagogy-Pedagogy Linking involves connecting various aspects of teaching and learning to ensure that educational strategies are aligned with learning goals (the relations *"includes", "achieves", "targets", "involves"*). It encompasses two main components: first, defining the learning goals, which are the specific objectives students are expected to achieve; and second, developing learning activities that help students meet these goals through measurable learning outcomes. The connection between learning outcomes and activities is mediated by the context in which learning occurs,

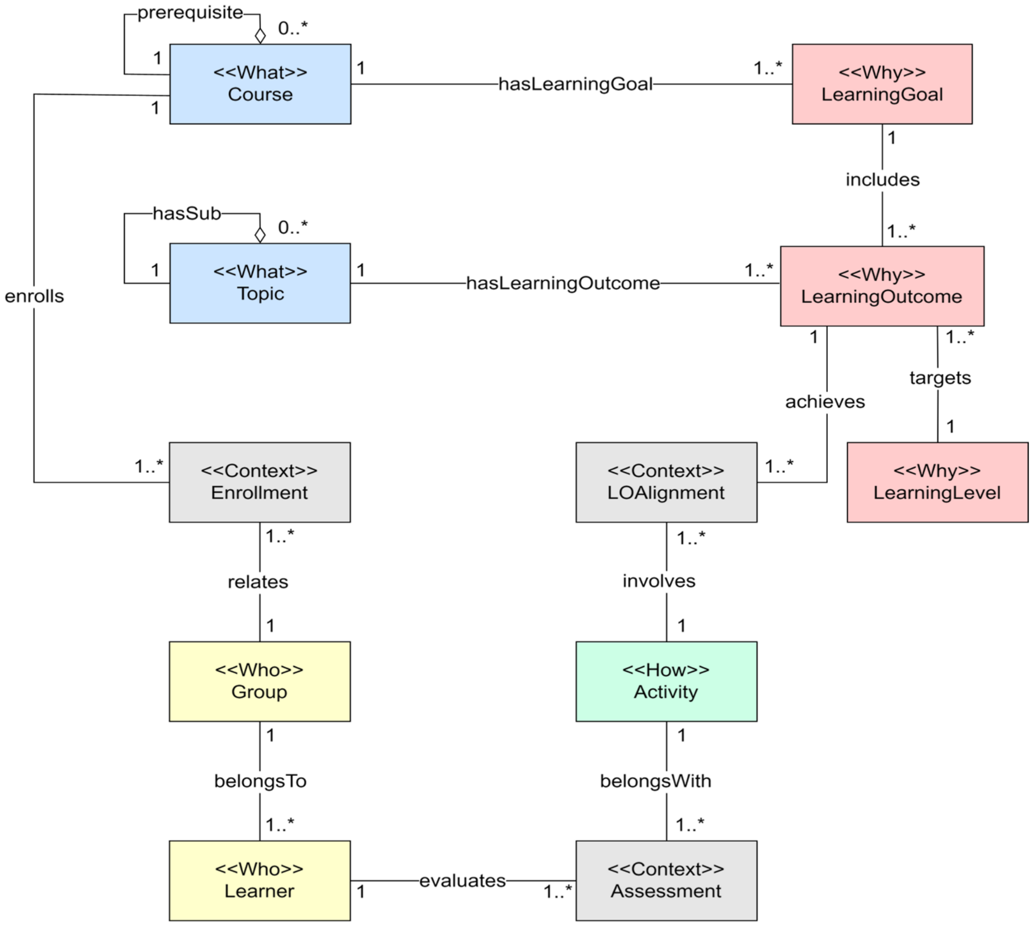


Figure 1: Knowledge Model.

emphasizing that the design of activities should be adapted to the specific "when" and "where" of the learning environment. For instance, the activities chosen to achieve a particular learning outcome may vary between in-class sessions and homework assignments, reflecting the need for context-sensitive pedagogical approaches.

Domain-Pedagogy Linking involves integrating pedagogical knowledge with learning domain content to enhance learning analytics (the relations *"hasLearningGoal", "hasLearningOutcome"*). This process connects various elements of pedagogy—such as learning outcomes, learning goals, and instructional activities—with the content-specific knowledge relevant to a course. By linking these pedagogical components with the learning domain knowledge, educators can better track and analyze what topics learners have mastered. This alignment supports the effective measurement of student progress and achievement, facilitating more accurate and actionable insights into learning outcomes. It helps in identifying which specific topics students have successfully learned and which areas may require additional focus, thereby enabling targeted interventions and improved educational strategies.

Domain-Pedagogy-Learner Linking involves integrating pedagogical and learning domain knowledge with individual learner profiles to enhance educational experiences. This approach aligns learning outcomes, goals, and activities with learner assessments and enrollment contexts. This integration also enables sophisticated analytics to track which topics learners have mastered and at what level of proficiency. Furthermore, it allows for the creation of personalized tools that enable learners to monitor their own performance and progres.

6 INSIGHT ENGINE

This section details the EduP insight engine, proposed as a method/process within the EduP framework to address RQ#2: how can knowledge be elaborated and used effectively? The method defines multiple modules for representing, elaborating, and reasoning about knowledge, aiming to generate insights in Learning Design (LD) and Learning Analytics (LA).

To propose a unified approach that covers the entire process from handling raw data to delivering insights and to highlight the transitions of data to knowledge and knowledge to insights, the process proposed here relies on the DIKW (**D**ata, **I**nformation, **K**nowledge, **W**isdom) model, which is hierarchical framework that allows structuring processes in a systematic way (Rowley, 2007).

Accordingly, EduP Insight Engine composes of the three main components, as presented in Figure 2.

- **Data Module:** Organize and preprocess data to be imported into the ontology.
- **Information Module:** Create an ontologie from the data processed in the Data Module.
- **Knowledge Module:** Reason with the ontology to generate knowledge for predefined specific use cases (**Wisdom**).

This is a semi-automatic process that involves some actions performed automatically by programs (denoted by rectangles in light red) and others requiring user intervention (denoted by rectangles in light blue). The details are presented below.

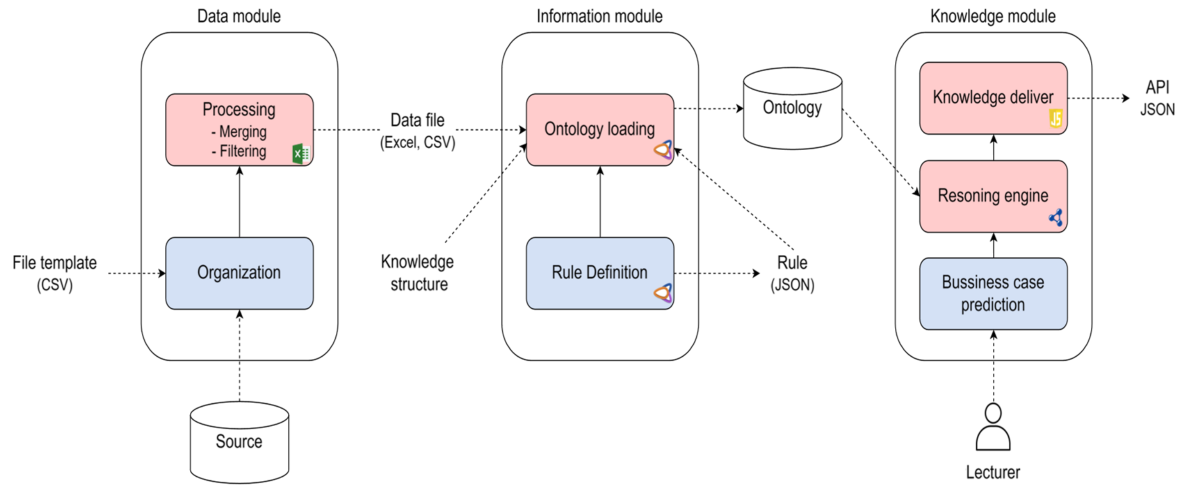


Figure 2: Insight Engine.

6.1 Data Module

The objective of this module is to collect, process, and transfer data from different sources into data files in CSV format, based on predefined structures (file templates). These data files will then be imported into the proposed ontology structure in the information module.

Data Source. Three primary data sources are considered here: *syllabus documents* used to identify pedagogical knowledge (learning outcomes, learning goals, learning activities) and a portion of the learning domain knowledge (topics to be acquired by learners); *data from LMS* used to acquire user history and interaction; *databases* from institutions providing learner assessment information.

Organization. This submodule aims to transfer data from various sources into CSV files. Some parts of this submodule require intervention from educators or teachers, particularly in identifying essential pedagogical knowledge from textual syllabus.

Processing. This submodule involves preprocessing the data to prepare it for import through Python programs. Tasks include handling missing values, removing duplicates, and merging or splitting files if necessary. The output is a properly formatted spreadsheet that can be automatically imported into the ontologies.

6.2 Information Module

This module is responsible for creating ontologies (also known as knowledge bases) from the data processed in the data module. This ensures that the data is structured in a way that facilitates reasoning in the knowledge module. Protégé is used as the editor for constructing these ontologies.

Ontology Loading. This submodule begins by creating the ontology's abstract structure based on the EduP knowledge model in Protégé. This structure includes the main concepts and their relationships and remains nearly unchanged throughout the ontology's lifecycle. In subsequent steps, ontology data from CSV or Excel files will be imported into this abstract structure.

Rule Definition. These are JSON-based rules that define the mapping between the content of CSV/Excel files and the ontology structure. They specify how each part of the data files can be identified and mapped into a component of the ontology structure. Since the data is complex and large, automatic loading using these predefined rules is essential.

6.3 Knowledge Module

This module focuses on exploring ontologies to generate insights and subsequently delivering these insights to other systems through APIs.

Reasoning Engine. This submodule consists of a set of SPARQL queries that can be used to reason the ontology created from the information module. The query system is defined based on predefined use cases analyzed and proposed from the perspectives of educators and teachers.

Business Case Prediction. These predefined use cases are declared as elements in the wisdom module. For each specified use case, the required knowledge is identified, and appropriate queries are invoked. This module is responsible for mapping the predefined use cases to suitable knowledge. This task is currently conducted through collaboration between educators/teachers and programmers.

Knowledge Deliver. This submodule manages the interaction between the EduP Insight Engine and other applications, such as web-based interfaces, through APIs. It handles receiving requests and responding to them, enabling the delivery of insights to educators and teachers. This allows for visualizing results in an accessible format, facilitating informed decision-making and enhancing the educational process.

6.4 Wisdom Module

This module is not a software component. Instead, it results from requirement analysis from the perspectives of educators and teachers. The predefined use cases capture essential knowledge for common requirements in learning design and analytics (see Figure 3 for more details).

Use case	Sub use case	Description
Data Management	Import	Lecture imports data from local
	Modify	Lecture modifies data
	Export	Lecture exports data
Knowledge Reasoning	Course Reasoning	Analyze and reason about Course
	Topic Reasoning	Analyze and reason about Topic
	Group Reasoning	Analyze and reason about Group
	Learner Reasoning	Analyze and reason about Learner
	Activity Reasoning	Analyze and reason about Activity
	LO Reasoning	Analyze and reason about Learning Outcome
Learning Analytics	Course Analytics	Analyze data related to Course
	Topic Analytics	Analyze data related to Topic
	Group Analytics	Analyze data related to Group
	Learner Analytics	Analyze data related to Learner
	Activity Analytics	Analyze data related to Activity
	LO Analytics	Analyze data related to Learning Outcome

Figure 3: Wisdow Management Module.

Data Management. This group enables end-users import, export, and modify ontology data through user-friendly interfaces.

Knowledge Reasoning. This group supports teachers and educators in reflecting on various types of knowledge. Each case study outlines potential outputs based on the input knowledge type. For instance, given a specific learning outcome for a course, the related data might include the associated learning goal, relevant topics, the required learning level according to Bloom's Taxonomy, and the learning activities designed to achieve this outcome. Understanding these types of knowledge is crucial for assisting teachers and educators in learning design. However, the conception and development of learning design tools is beyond the scope of this research.

Learning Analytics. This group focuses on data analytics based on knowledge types defined in ontologies. For example, analyzing the connection between learning outcomes and learner assessment results can generate statistics on which learners achieve or do not achieve the outcomes.

7 INSTANTIATIONS

This section presents two instantiations developed to validate the EdUP framework. First, the EduP knowledge structure and process are applied to create an ontology for a database course, demonstrating the framework's applicability. Second, the EduP web-based reasoner is constructed to illustrate how the ontology can support learning design and learning analytics through real-world scenarios. The reasoner processes the constructed ontologies to generate insights for teachers through interactive user interfaces.

As illustrated in Figure 4, the reasoning process provides results related to specific learning outcomes. By default, all relationships associated with a specified learning outcome are presented in a simple table format, which facilitates easy consultation for teachers. This user-friendly presentation allows educators to quickly access and interpret relevant information, aiding in the evaluation and refinement of teaching strategies. The interactive nature of the interface enhances the usability of the insights, making it easier for teachers to leverage data in their decision-making processes.

Figure 5 displays the interface for descriptive analytics, which highlights the number of students who have completed specific activities within a course. This straightforward statistical overview demonstrates how the knowledge extracted from ontologies can be effectively applied in learning analytics. By providing clear metrics on student participation and achievement, the interface showcases the practical value of integrating ontological knowledge into educational analysis. This integration not only aids in tracking student progress but also illustrates how such knowledge can be used to derive actionable insights for improving course design and instructional strategies.

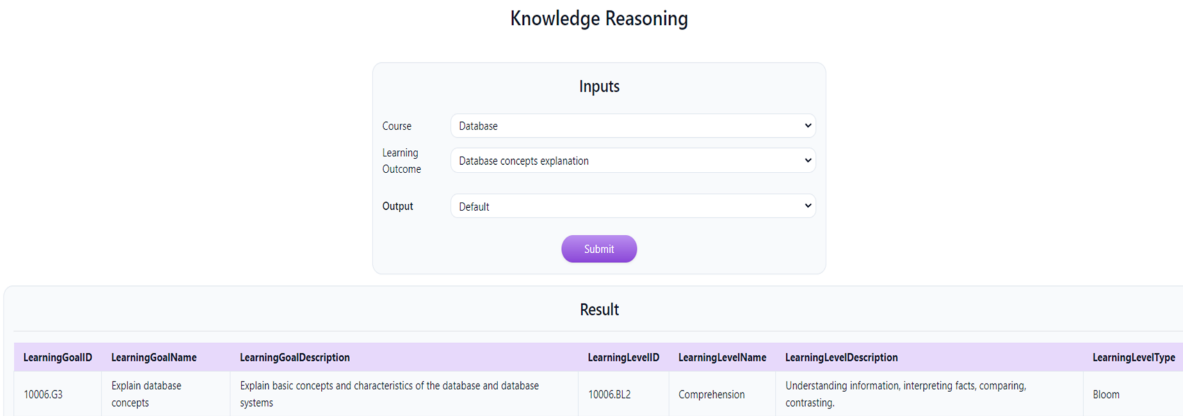


Figure 4: Knowledge Reasoning Interface.

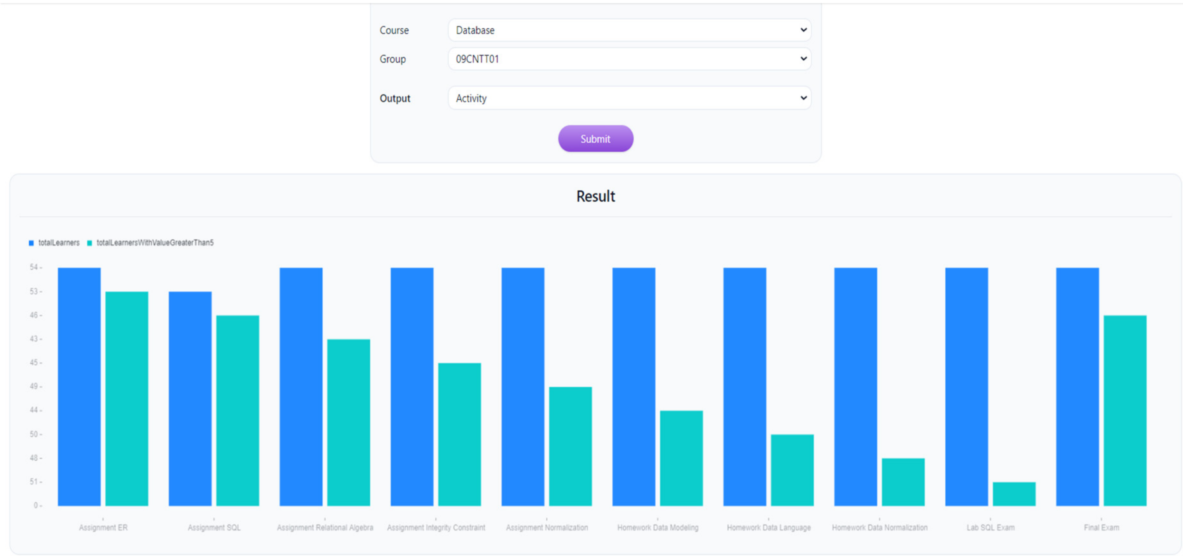


Figure 5: Learning Analytics Interface.

8 CONCLUSIONS

This section summarizes the contributions of the paper and offers suggestions for future research directions.

In terms of **contribution**, the paper first presents a comprehensive knowledge model that integrates various types of knowledge within the context of LD/LA. This model provides teachers with a holistic overview of how domain-specific knowledge can be acquired through various pedagogical strategies. The second contribution is a method that defines the main phases and associated components to facilitate reasoning on knowledge bases. By linking multiple knowledge types and providing a structured method for reasoning on knowledge bases, this paper offers valuable tools for educators and researchers. The case studies used for validation highlight the potential for implementing the proposed framework in the future.

In terms of **future research**, since the framework is a proof of concept, the knowledge model is currently simple and needs further development to meet the requirements of LD/LA. Additionally, some components of the proposed method can be automated to reduce costs, leveraging advancements in AI, for example, automatically identifying learning topics from syllabus. Finally, more research is needed to explore how knowledge can be translated into measurable indicators within learning analytics.

ACKNOWLEDGEMENTS

This research is funded by University of Science, VNU-HCM under grant number CNTT 2023-09.

REFERENCES

Aguilar, J., Jerez, M., & Rodríguez, T. (2018). CAMeOnto: Context awareness meta ontology modeling. *Applied Computing and Informatics*, *14*(2), Article 2. https://doi.org/10.1016/j.aci.2017.08.001

Ahmad, A., Schneider, J., Griffiths, D., Biedermann, D., Schiffner, D., Greller, W., & Drachsler, H. (2022). Connecting the dots – A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*, jcal.12716. https://doi.org/10.1111/jcal.12716

Al-Yahya, M., George, R., & Alfaries, A. (2015). Ontologies in E-Learning: Review of the Literature. *International Journal of Software Engineering and Its Applications*, *9*(2).

Andrade, L. O. M., Valter, R., Ramos, R., Vidal, V., Andrade, D., & Oliveira, M. (2019). LARIISA: An intelligent platform to help decision makers in the brazilian health public system. *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, 501–504. https://doi.org/10.1145/3323503.3362122

Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 329–338. https://doi.org/10.1145/2883851.2883944

Bensassi, I., Elyusufi, Y., & El Mokhtar, E.-N. (2019). Smart Connection of User Profiles in a Big Data Context. *Proceedings of the 4th International Conference on Big Data and Internet of Things*, 1–8. https://doi.org/10.1145/3372938.3372977

Botturi, L., & Stubbs, T. (Eds.). (2008). *Handbook of Visual Languages for Instructional Design: Theories and Practices*. IGI Global. https://doi.org/10.4018/978-1-59904-729-4

Cabrera, O., Franch, X., & Marco, J. (2017). Ontology-based context modeling in service-oriented computing: A systematic mapping. *Data & Knowledge Engineering*, *110*, 24–53. https://doi.org/10.1016/j.datak.2017.03.008

Cano-Benito, J., Cimmino, A., & García-Castro, R. (2021). Toward the Ontological Modeling of Smart Contracts: A Solidity Use Case. *IEEE Access*, *9*, 140156–140172. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3115577

Celik, D., & Magoulas, G. D. (2016). A Review, Timeline, and Categorization of Learning Design Tools. In D. K. W. Chiu, I. Marenzi, U. Nanni, M. Spaniol, & M. Temperini (Eds.), *Advances in Web-Based Learning – ICWL 2016* (Vol. 10013, pp. 3–13). Springer International Publishing. https://doi.org/10.1007/978-3-319-47440-3_1

Chang, M., D'Aniello, G., Gaeta, M., Orciuoli, F., Sampson, D., & Simonelli, C. (2020). Building Ontology-Driven Tutoring Models for Intelligent Tutoring Systems Using Data Mining. *IEEE Access*, *8*, 48151–48162. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2979281

Chatti, M. A., & Muslim, A. (2019). The PERLA Framework: Blending Personalization and Learning Analytics. *The International Review of Research in Open and Distributed Learning*, *20*(1). https://doi.org/10.19173/irrodl.v20i1.3936

Chimalakonda, S., & Nori, K. V. (2020). An ontology based modeling framework for design of educational technologies. *Smart Learning Environments*, *7*(1), 28. https://doi.org/10.1186/s40561-020-00135-6

Demaidi, M. N., Gaber, M. M., & Filer, N. (2018). OntoPeFeGe: Ontology-Based Personalized Feedback Generator. *IEEE Access*, *6*, 31644–31664. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2846398

Dresch, A., Lacerda, D. P., & Antunes Jr, J. A. V. (2015). *Design Science Research: A Method for Science and Technology Advancement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-07374-3

Eyal, L., & Gil, E. (2020). Design patterns for teaching in academic settings in future learning spaces. *British Journal of Educational Technology*, *51*(4), 1061–1077. https://doi.org/10.1111/bjet.12923

Ghazal, R., Malik, A. K., Qadeer, N., Raza, B., Shahid, A. R., & Alquhayz, H. (2020). Intelligent Role-Based Access Control Model and Framework Using Semantic Business Roles in Multi-Domain Environments. *IEEE Access*, *8*, 12253–12267. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2965333

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Guarino, N., Oberle, D., & Staab, S. (2009). What Is an Ontology? In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 1–17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3_0

Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C. A., & Ramírez Mendoza, R. A. (2022). Learning analytics: State of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *16*(3), 1209–1230. https://doi.org/10.1007/s12008-022-00930-0

Hyunsook, C., & Jeongmin, K. (2016). An Ontological Approach for Semantic Modeling of Curriculum and Syllabus in Higher Education. *International Journal of Information and Education Technology*, *6*(5), 365–369. https://doi.org/10.7763/IJIET.2016.V6.715

Iatrellis, O., Kameas, A., & Fitsilis, P. (2019). EDUC8 ontology: Semantic modeling of multi-facet learning pathways. *Education and Information Technologies*, *24*(4), 2371–2390. https://doi.org/10.1007/s10639-019-09877-4

Jang, S., & Woo, W. (2012). *5W1H: Unified User-Centric Context*. 2.

Katis, E., Kondylakis, H., Agathangelos, G., & Vassilakis, K. (2018). Developing an Ontology for Curriculum and Syllabus. In A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, & M. Alam (Eds.), *The Semantic Web: ESWC 2018 Satellite Events* (Vol. 11155, pp. 55–59). Springer International Publishing. https://doi.org/10.1007/978-3-319-98192-5_11

Koper, R., & Bennett, S. (2008). Learning Design: Concepts. In H. H. Adelsberger, Kinshuk, J. M. Pawlowski, & D. G. Sampson (Eds.), *Handbook on Information Technologies for Education and Training* (pp. 135–154). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74155-8_8

Lacasta, J., Lopez-Pellicer, F. J., Espejo-García, B., Nogueras-Iso, J., & Zarazaga-Soria, F. J. (2018). Agricultural recommendation system for crop protection. *Computers and Electronics in Agriculture*, *152*, 82–89. https://doi.org/10.1016/j.compag.2018.06.049

Lalingkar, A., Ramnathan, C., & Ramani, S. (2014). Ontology-based smart learning environment for teaching word problems in mathematics. *Journal of Computers in Education*, *1*(4), 313–334. https://doi.org/10.1007/s40692-014-0020-z

LeClair, A., Jaskolka, J., MacCaull, W., & Khedri, R. (2022). Architecture for ontology-supported multi-context reasoning systems. *Data & Knowledge Engineering*, *140*, 102044. https://doi.org/10.1016/j.datak.2022.102044

Lui, A. M., & Bonner, S. M. (2016). Preservice and inservice teachers' knowledge, beliefs, and instructional planning in primary school mathematics. *Teaching and Teacher Education*, *56*, 1–13. https://doi.org/10.1016/j.tate.2016.01.015

Mangaroska, K., & Giannakos, M. (2019). Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Mizoguchi, R., Sinitsa, K., & Ikeda, M. (1996). Task Ontology Design for Intelligent Educational/Training Systems. *Workshop on Architectures and Methods for Designing Cost-Effective and Reusable*.

Ouissem, B., Lamia, M., & Hafidi, M. (2021). A Proposed Ontology-Based Generic Context Model for Ubiquitous Learning. *International Journal of Web-Based Learning and Teaching Technologies*, *16*(3), 47–64. https://doi.org/10.4018/IJWLTT.20210501.oa4

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 4. https://doi.org/10.1186/s41239-022-00372-4

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), Article 3. https://doi.org/10.2753/MIS0742-1222240302

Pelap, G. F., Fitime, L. F., & Mbaiossoum, B. L. (2023). *Ontology-Based Modeling of the Learner in a Web Educational System: Towards Learning Analytics and Adaptive Learning*. *8*(7).

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, *16*(1), 414–454. https://doi.org/10.1109/SURV.2013.042313.00197

Pozzi, F., Asensio-Perez, J. I., Ceregini, A., Dagnino, F. M., Dimitriadis, Y., & Earp, J. (2020). Supporting and representing Learning Design with digital tools: In between guidance and flexibility. *Technology, Pedagogy and Education*, *29*(1), 109–128. https://doi.org/10.1080/1475939X.2020.1714708

Rahayu, N. W., Ferdiana, R., & Kusumawardani, S. S. (2022). A systematic review of ontology use in E-Learning recommender system. *Computers and Education: Artificial Intelligence*, *3*, 100047. https://doi.org/10.1016/j.caeai.2022.100047

Ramaswami, G., Susnjak, T., Mathrani, A., & Umer, R. (2023). Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. *Technology, Knowledge and Learning*, *28*(3), 959–980. https://doi.org/10.1007/s10758-022-09613-x

Ramesh, R., M., S., & Iyer, S. (2016). Integrating the Learning Objectives and Syllabus into a Domain Ontology for Data structures Course. *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*, 266–271. https://doi.org/10.1145/2899415.2899453

Reynolds, S., Pate, W. C., & Ochoa, O. (2023). An Ontology and Management System for Learning Outcomes and Student Mastery. *2023 IEEE Frontiers in Education Conference (FIE)*, 1–5. https://doi.org/10.1109/FIE58773.2023.10343171

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1355.

Romero, L., Link to external site, this link will open in a new window, Saucedo, C., Ma, L. C., & Gutiérrez, M. (2019). Supporting self-regulated learning and personalization using ePortfolios: A semantic approach based on learning paths. *International Journal of Educational Technology in Higher Education*, *16*(1), 1–16. https://doi.org/10.1186/s41239-019-0146-1

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180.

Schmitz, M., Van Limbeek, E., Greller, W., Sloep, P., & Drachsler, H. (2017). Opportunities and Challenges in Using Learning Analytics in Learning Design. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data Driven Approaches in Digital Education* (Vol. 10474, pp. 209–223). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_16

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12. https://doi.org/10.1186/s41239-021-00313-7

Tatto, M. T., Rodriguez, M. C., & Reckase, M. (2020). Early career mathematics teachers: Concepts, methods, and strategies for comparative international research. *Teaching and Teacher Education*, *96*, 103118. https://doi.org/10.1016/j.tate.2020.103118

Verdú, E., Regueras, L. M., Gal, E., de Castro, J. P., Verdú, M. J., & Kohen-Vacs, D. (2017). Integration of an intelligent tutoring system in a course of computer network design. *Educational Technology Research and Development*, *65*(3), 653–677. https://doi.org/10.1007/s11423-016-9503-0

Vu, T. M. H., & Tchounikine, P. (2021). Supporting teacher scripting with an ontological model of task-technique content knowledge. *Computers & Education*, *163*, 104098.

Vu, T.-M.-H., Le Dinh, T., Dam, N. A. K., & Pham-Nguyen, C. (2023). *Context-aware Knowledge-based Systems: A Literature Review*. https://hdl.handle.net/10125/103190

Wang, Y., & Wang, Y. (2021). A Survey of Ontologies and Their Applications in e-Learning Environments. *Journal of Web Engineering*, *20*(6), 1675–1720. Journal of Web Engineering. https://doi.org/10.13052/jwe1540-9589.2061

Wei, N., & Shao, X. (2022). A Semantic Model of Internet of Things for Intelligent Translation and Learning. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/1651288

Wen, Y., Zhu, X., & Zhang, L. (2022). CQACD: A Concept Question-Answering System for Intelligent Tutoring Using a Domain Ontology With Rich Semantics. *IEEE Access*, *10*, 67247–67261. IEEE Access.

Crafting the Future: Developing and Evaluating a Digital Mindset Competence Model for the Industrial Craft Sector

Seyma Kocak[1,2][a] and Jan Pawlowski[1,2][b]

[1]*University of Jyväskylä, Faculty of Information Technology, Jyväskylä, Finland*

[2]*University of Applied Sciences Ruhr West Institute of Computer Science, Bottrop, Germany*

seyma.kocak@hs-ruhrwest.de, jan.pawlowski@hs-ruhrwest.de

Keywords: Digital Mindset, Design Science Research, Industrial Craft Sector, Digital Competence.

Abstract: The recent development of digitization has significantly influenced various sectors of the economy, and the Industrial Craft Sector is no exception. The transition to digital technologies and processes is inevitable and holds the potential for increasing efficiency and creating competitive advantages. This research used the Design Science methodology to develop a Digital Mindset Competence Model. This model comprises eight dimensions specifically tailored to the requirements and challenges of the Industrial Craft Sector. These dimensions aim to promote and strengthen the digital mindset among professionals in the Industrial Craft Sector. To ensure the validity and relevance of this model, experts from the Industrial Craft Sector were involved in a qualitative methodology. The combination of scientific methodology and practical experience ensures a comprehensive perspective and guarantees the applicability of the developed model. The results of this research underscore the importance of digital transformation in the Industrial Craft Sector and the necessity of a digital mindset. The developed Digital Mindset Competence Model provides a targeted approach to promoting digital competencies in the Industrial Craft Sector and guides future developments in this area. It becomes evident that an appropriate digital mindset is essential to optimally leverage the potentials of digitization in the Industrial Craft Sector and successfully navigate continuous change. This scholarly contribution contributes to raising awareness of the significance of a digital mindset in the Industrial Craft Sector. It forms a basis for further investigations and practical applications within digital transformation.

1 INTRODUCTION

Digitalization is a critical challenge that craft businesses must actively promote to remain competitive and not lose touch with the rapidly evolving business world (Timchuk & Evloeva, 2020). Craft enterprises provide specialized craft activities and offer products or services based on technical knowledge and tradition. They cover a broad spectrum, from repair and maintenance work to manufacturing customized products (Abel, 2007). Although craft businesses know digitalization (Rohleder & Schulte, 2020; Veltkamp & Schulte, 2020), its importance is often not sufficiently recognized (Welzbacher et al., 2015). Digitalizing the Industrial Craft Sector is necessary to enable efficiency gains, optimize work processes, and improve customer communication (Aghimien et al., 2022). This requires employees and managers to develop the essential skills and knowledge to use new technological tools and processes, which can improve productivity, innovation, and competitiveness (Nikmehr et al., 2021). A fundamental problem in the Industrial Crafts Sector is the development of digital skills and attitudes (Ezeokoli et al., 2016; Kocak & Pawlowski, 2022). To successfully implement digital transformation in the Industrial Craft Sector, a model that captures the level of an organization's digital mindset is needed. Appropriate measures can be defined and adapted only by understanding the digital mindset (Kratochvil et al., 2021). Although digital mindset concepts and models exist in the literature (Hildebrandt & Beimborn, 2021; Aliabina, 2020; Young et al., 2020), these mainly focus on general sectors and are less specific to the Industrial Craft Sector. Only a few studies specifically address

[a] https://orcid.org/0000-0002-1282-8740

[b] https://orcid.org/0000-0002-7711-1169

Kocak, S. and Pawlowski, J.
Crafting the Future: Developing and Evaluating a Digital Mindset Competence Model for the Industrial Craft Sector.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 108-119
ISBN: 978-989-758-716-0; ISSN: 2184-3228

digitalization, digital skills, and digital mindset in the Industrial Craft Sector (Ezeokoli et al., 2016; Aghimien et al., 2022; Parusheva, 2019; Kocak & Pawlowski, 2023). This study addresses the following research question:

How can a Competency Model for a Digital Mindset in the Industrial Craft Sector be Developed and Assessed?

The study aims to develop and evaluate a model of digital competencies and attitudes specifically tailored to the needs of the Industrial Craft Sector. It aims to close a research gap and expand the theoretical understanding of the application of digital technologies in traditional sectors. This will provide new insights into specific industries' digitalization processes and address the Industrial Craft Sector's unique challenges. The developed model will help craft enterprises improve their digital skills and attitudes, optimize work processes, and increase productivity. This is essential for competitiveness and full exploitation of digitalization's benefits. It also promotes innovation, enables the development of new products and services, and supports adaptation to changing market demands. To answer the research question, the design science approach (Hevner & Chatterjee, 2010; Peffer et al., 2007) was chosen to develop the model and evaluate it qualitatively with experts from the craft sector (Myers, 2019). First, the theoretical part and the development of the model are presented, followed by a detailed description of the methodology.

2 THEORY

The Industrial Craft Sector comprises independent activities in material processing aimed at meeting the individual needs of companies (Abel, 2007). According to the Crafts Code, craft enterprises are service, production, and commercial businesses (Buschfeld et al., 2011). Since there is no EU-wide definition, this description is used. The work process in the construction industry is also strongly characterized by craftsmanship and is considered industrial craft (Diego et al., 2020). The global digital revolution is propelling the Industrial Craft Sector towards a promising future with digital transformation. Enhancing digital technologies can potentially revolutionize the sector, offering more efficient services and operations. Understanding the different phases of digital transformation and developing specific strategies for each phase is crucial, paving the way for a brighter future (Vial, 2019). The Industrial Craft Sector embraces digitalization globally, with researchers playing a crucial role in advocating for technological progress. This shift is seen as an opportunity, with technological advancements being harnessed in the sector. Researchers' efforts are steering the construction industry toward digital transformation, inspiring a wave of change (Yang et al., 2022). Although the Industrial Craft Sector is open to digital transformation, it rarely shows initiative (Rohleder & Schulte, 2020; Veltkamp & Schulte, 2020; Welzbacher et al., 2015). Digital technologies can be a decisive success factor, but the Industrial Craft Sector is reluctant to start the digital transformation (Vasiliki et al., 2020). This reluctance is due to numerous organizational, technical, and individual barriers and a lack of skills and attitudes (Yang et al., 2022; Kocak & Pawlowski, 2022). Consequently, the Industrial Craft Sector has not yet fully exploited the potential of digital transformation (Sriyolia et al., 2021). Industrial Craft Companies should develop a clear digital strategy to overcome organizational barriers and adapt internal processes accordingly (Vogelsang et al., 2019). Leadership should receive digital transformation and change management training to promote acceptance and implementation (Kocak & Pawlowski, 2022). Implementing modern IT infrastructures and regular updates is crucial. Partnerships with technology providers can help overcome technical hurdles and ensure access to the latest technologies (Rohleder & Schulte, 2020). Training employees in new technologies and fostering a digital mindset are essential (Veltkamp & Schulte, 2020). Mentoring programs and peer learning can also help build confidence in digital technologies (Oesterreich & Teuteberg, 2016). A tailored training framework that integrates digital skills, attitudes, and personality traits is urgently needed (Ezeokoli et al., 2016). Continuing education programs should promote technical, social, and emotional competencies to develop holistic digital competence (European Construction Sector Observatory, 2020). Another aspect is adopting a digital mindset (Pammer et al., 2021; Kocak & Pawlowski, 2022). A digital mindset describes "patterns of thinking embodied in people's cognitive processes, filters, and core beliefs, composed of cognitive mechanisms and knowledge structures that influence and promote the use and application of digital technologies and the management of their consequences in the context of individuals, organizations, or society." From this, it can be inferred that, in addition to attitudes, competencies also play an essential role in realizing the digital mindset. This paper uses the definition by Hildebrand

and Beimborn (2022). Theories suggest that by applying and building a digital mindset, bridging the digital gap within the organization, and developing lasting innovative capabilities, organizations can foster the growth of innovative developments locally, among competitors, and across industries (Fisher, 2022). Therefore, it can be concluded that digital capabilities combine a digital mindset, knowledge, competencies, and attitude (Gekara et al., 2017). A tailored framework that integrates digital skills, attitudes, and personality traits is urgently required to establish a digital mindset in craft organizations, as specific personality traits are crucial for shaping a digital mindset (Ezeokoli et al., 2016; European Observatory, 2020). Existing models, such as those by Bredendiek Knorr (2020), focusing on openness, agility, proactivity, creativity, customer orientation, and fault tolerance, and by Hildebrandt and Beimborn (2021), emphasizing thinking patterns like risk-taking and resilience, lack specific dimensions and are designed for general organizations. Kocak and Pawlowski (2021) identified relevant digital skills and attitudes for managers and employees but did not explicitly investigate craft enterprises; thus, analyzing these existing models helps address the digital transformation challenges in the Industrial Craft Sector. These approaches were used to develop the Digital Mindset Competence Model for the Industrial Craft Sector. The digital transformation in the Industrial Craft Sector has significant potential. However, it is currently hindered by organizational, technical, and individual barriers, which can be overcome through a clear strategic approach, regular training, promotion of a digital mindset, and the development of tailored training programs to enhance digital competence and technology acceptance, ultimately leading to successful digital transformation by applying and adapting proven models to the sector's specific needs.

3 DEVELOPMENT OF THE DIGITAL MINDSET COMPETENCE MODEL

The Digital Competence Mindset Competence Model was developed using the design science approach (Hevner & Chatterjee, 2010; Peffer et al., 2007) and began with a systematic literature review and concept matrix analysis (Webster & Watson, 2002) to identify and analyze existing models in the industrial craft sector. The second phase involved using a concept matrix and units of analysis to examine further academic approaches to digital mindset models, dimensions, competencies, and technologies, facilitating the adaptation of innovative approaches for the industrial crafts sector. Aliabina (2020) used a quantitative approach to study digital culture, competence, and knowledge in banking, retail, and telecommunication sectors, noting strengths like a large sample size and sector diversity but lacking detailed descriptions and validation procedures. Young et al. (2020) and Knorr (2020) examined digital learning aptitude, literacy, and entrepreneurial mindset, providing detailed descriptions and identifying interrelations among variables. However, neither study lacked formal scientific methodologies or evaluated models. Kollmann's (2020) research on digital execution, skills, and mindset highlighted lifelong learning and openness to new technologies but also suffered from a lack of formalized models and scientific methodologies, similar to Hildebrandt & Beimborn's (2021) literature review on digital innovation and thinking, which identified thinking patterns but lacked a scientific methodology, evaluation framework, and detailed descriptions. In the Industrial Craft Sector, Kocak and Pawlowski utilized a mixed-method approach in their 2023 study, exploring Technological and Professional competencies. Strengths included the application of a taxonomy and the classification of competencies and attitudes. However, the study lacked a formal model and an unspecified scientific methodology. The qualitative research conducted by Kocak & Pawlowski in 2021 within the general sector focused on Communication and Information Processing Competencies. Strengths encompassed the classification and sorting of factors. Inadequacies in the research were identified, including lacking a formal model, scientific methodology, and detailed descriptions of elements (Neeley & Leonardi, 2022). While strengths were found in recognizing essential factors like customer centricity and digital competence, the research often lacked a structured approach and comprehensive factor descriptions (Salvetti et al., 2022). Similarly, Joseph et al.'s 2022 study lacked specific dimensions and characteristics, borrowing factors from other models without providing a formal framework or detailed methodology. In Lessiak's 2020 literature review, growth orientation and collaboration emphasis were explored, supplemented by qualitative interviews. Strengths included the development of a model, the description of main dimensions, and the classification of sizes. Inadequacies encompassed the lack of a common taxonomy, undefined sub-factors, an unevaluated model, and no information about the analysis of items in the questionnaire construct.

The summarized studies present various approaches to understanding the digital mindset, highlighting unique methods, domains, and dimensions. Many studies, such as those by Aliabina (2020), Kollmann (2020), Kocak & Pawlowski (2023), and Joseph et al. (2022), lack a developed model and standardized scientific methodology like ADR or DSR, affecting reliability and comparability. Studies like Young et al. (2020) and Salvetti et al. (2022) focus only partially on digital mindset dimensions, potentially overlooking crucial aspects, while Kocak & Pawlowski (2021) and Salvetti et al. (2022) lack a classification of characteristics or a model, hindering systematic understanding. These gaps underline the need for a comprehensive digital mindset competency model, providing a structured framework, detailed dimensions, standardized methods, and a specific taxonomy for the industrial crafts sector to understand better, cultivate, and assess essential digital competencies. The approaches of Kocak and Pawlowski (2023) and Bredendiek and Knorr (2020) were used to develop the digital mindset and competence model. For the development of the dimensions, Information processing Competence, Communication Competence, Technological Competence, Personal Competence, and Personal Traits were taken from Kocak & Pawlowski (2021), as they have already classified many competencies, attitudes, and personality traits for the Industrial Craft Sector. Bredendiek and Knorr (2020) adopted the dimension of entrepreneurial orientation, as many competencies identified in the literature are necessary to drive digital transformation in the industrial craft sector.

Teamwork, cooperation, collaboration, customer relationship, and technical communication competencies are essential in the Industrial Craft Sector due to the team-based nature of the work and the high priority on customer communication (Kocak & Pawlowski, 2023). Essential personality traits include openness (Bredendiek & Knorr, 2020), communication, flexibility, curiosity, and compromise (Kocak & Pawlowski, 2023), which are crucial as artisans often have limited time due to numerous orders. Additionally, agility, result-oriented mindset, openness to learning, self-restraint, and self-confidence (Bredendieck & Knorr, 2020; Kocak & Pawlowski, 2023) were integrated into the digital attitude dimension, and critical faculties, taking responsibility, and lifelong learning (Kocak & Pawlowski, 2023) were added to the personal competencies dimension, reflecting the continuous need for learning and adapting to new technologies in the digital transformation. In total, we have eight main dimensions for our Digital Competence Mindset Competence Model:

Communication Competence (CC). Ability to communicate effectively, both verbally and nonverbally, in personal, professional, and digital contexts (Salleh, 2008).

Information Processing Competence (IPC). Ability to gather, understand, evaluate, organize, and use information from various sources, critical for making informed decisions in the digital era (Tahvanainen & Luoma, 2018).

Entrepreneurial Orientation (EO). Ability and willingness to identify opportunities, take risks, drive innovation, and strive for growth (Xu & Xu, 2012; Tahvanainen & Luoma, 2018).

Technological Competence (TC). Capability to effectively use, understand, and manage technological tools and systems across professional, educational, and personal contexts (Tahvanainen & Luoma, 2018).

Development Competence (DC). Ability to continuously learn, adapt, and foster personal growth, crucial in the era of digitalization (Kocak & Pawlowski, 2021; Tahvanainen & Luoma, 2018).

Personal Competence (PC). Ability to interact harmoniously with others, essential in social, professional, and individual contexts (Tahvanainen & Luoma, 2018).

Digital Attitude (DA). Mindset towards embracing digital opportunities and adapting to modern demands (Hildebrandt & Beimborn, 2022; Kocak & Pawlowski, 2021).

Personality Traits (PT). Traits that influence attitudes and reactions, shaping individual interactions and actions (Mekhaznia et al., 2021).

Figure 1: Digital Mindset Competence Model.

Description of the Digital Mindset Competence Modell Factors

The Digital Mindset Competence Model for the Industrial Craft Sector has eight dimensions. Bloom's subdimension descriptions were partly taken from various sources or formulated according to an established taxonomy (Krathwohl, 2002). A total of 57 sub-dimensions were identified and defined. The table below shows the first two sub-dimensions for each central dimension.

Table 1: Description of the Sub-Factors.

Sub-factors	Description
Enthusiasmus (PT)	The trait of possessing passionate zeal, a heightened interest in a specific topic or task (Von Ohain, 2019)
Accuracy (PT)	The trait of appreciating the thoughts, feelings, and behavior of myself or others
Communication (CC)	Ability to communicate constructively, effectively, and consciously (Kocak & Pawlowski, 2022)
Digital Communication (CC)	Ability to use online tools such as email, social media messaging, and texting to reach others (Bordi, Okkonen, Mäkiniemi, & Heikkilätammi, 2018).
Flexible Mindset (DA)	Property to adapt to change and, above all, the anticipation of future innovation (Kocak & Pawlowski, 2022)
Visioning (DA)	Attitude to imagine the future, plan, develop a vision, turn ideas into reality, and create future scenarios to guide efforts and actions (Mihardio & Sasmoko, 2019).
Decision Making (EO)	Ability to make decisions whose outcome is not specific (Kocak & Pawlowski, 2022)
Quality Management (EO)	Ability to develop and implement quality planning, assurance, quality control, and improvement. (Kim, 2020).
Data collection (IPC)	Ability to collect and measure data from multiple sources to get a complete and accurate picture of an area.
Evaluating data, information, and digital content (IPC)	Ability to analyze, compare, and critically evaluate the credibility and reliability of data sources, information, and digital content. To analyze, interpret, and critically evaluate the data, information, and digital content (Riina et al., 2016)
Technology Application (TC)	Ability to utilize digital technologies to improve the strategic positioning of a company.
ICT-Safety (TC)	Ability to monitor and control confidential information (Kemendi, 2021; Riina, Yves, Staphanie, & Van Den Brande, 2016).
Programming Skills (DC)	Ability to know and apply the syntax and code of a programming language (Riina, Yves, Staphanie, & Van Den Brande, 2016).
Work with AI (DC)	Ability to know and use different programming languages, signal processing techniques, and neural network architectures (Schuur, Rezazade Mehrizi, & Ranschaert, 2021).
Innovation Skills (PC)	Ability to identify opportunities to improve performance by changing methods, processes, products, and services (Kocak & Pawlowski, 2022)
Creativity (PC)	Ability to develop different ideas and opportunities to create value, combine knowledge and resources to achieve valuable effects, research, and experiment with innovative approaches (Kocak & Pawlowski, 2022)

Five Stage Model

In addition to the Digital Mindset Competence-Model developed in this phase of the design science research methodology, we also developed a stage model to help Industrial Craft Companies assess the current stage or maturity level of digital mindset in their company. The stage model represents five stages. Within the Digital Mindset Competence Model framework, various development stages are described that reflect the change in the digital mindset in organizations. These stages are explained below:

Table 2: Description of the Five-Stage Model.

Stage	Description
Stage 1: Digital Newcomers	Digital newcomers are organizations that lack the characteristics of the digital mindset competency model and need to develop an essential awareness of the importance and impact of digital transformation.
Stage 2: Digital Starters	The Digital Starter phase begins the learning process where the importance of digital transformation is gradually understood, with initial signs of a digital mindset and competencies emerging. However, many additional skills and attitudes are still needed to reach an advanced level.
Stage 3: Digital Intermediate	The "Digital Intermediary" phase shows a deep understanding of digital transformation and a comprehensive digital mindset but still requires continuous growth to fully develop all necessary skills and attitudes.
Stage 4: Digital Professionals	"Digital professionals" have a highly developed digital mindset, mastering and applying all relevant characteristics for a successful digital transformation.
Stage 5: Digital Pioneers	The "Digital Pioneer" is at the top of the digital competence model, mastering its attributes, embodying the highest level of digital competence, and serving as a role model by putting digital principles into practice.

The competence levels are determined by summarizing and averaging items from established instruments across eight dimensions, then dividing the total mean values by the number of items to obtain the final competence level.

4 METHOD

The present study is based on the DSR approach (Design Science Research) (Peffer et al., 2007; Hevner & Chatterjee, 2010) and follows a qualitative research methodology (Myers, 2019) that addresses current problems not comprehensively discussed in the existing literature. The applied research methodology begins with a systematic literature review (Webster & Watson, 2002), the starting point for developing the research question and methodology. This process enables a thorough analysis of existing knowledge. It identifies research gaps, leading to the definition and description of key concepts and variables relevant to studying the digital mindset in the industrial craft sector.

Our study is based on organizational culture theory, which posits that collectively shared beliefs, values, norms, and practices significantly influence the behavior and performance of an organization's members (Schein, 1990). This theory is applied to a Digital Mindset Competence Model for Industrial Craft Organizations, where establishing a culture that emphasizes digital competence and a willingness to learn promotes employees' digital skills and engagement, thereby overcoming barriers and supporting successful digital transformation (Mohammadi, 2020; European Construction Sector Observatory, 2020). Furthermore, the core theory of the work is based on information theory, which requires craft organizations to ensure their members possess the necessary digital skills and technological knowledge to manage digital tools effectively, contributing to the improvement of organizational culture (Ezeokoli et al., 2016; Fisher, 2022).

Problem Identification

The digitization process has now become an omnipresent phenomenon and extends to almost all sectors of the economy, including the Industrial Craft Sector. The companies involved in the Industrial Craft Sector show a fundamental openness and interest in digitization. Still, it is striking that they rarely take proactive steps and attach comparatively little importance to digitization in their companies (Überbacher et al., 2020). This reluctance is in marked contrast to the dynamic nature of digital transformation, which nowadays represents a decisive competitive advantage. A critical issue limiting Industrial Craft Enterprises in their efforts to embrace digitalization is the existence of many digitalization barriers (Aghimien et al., 2022; Kocak & Pawlowski, 2022). In particular, the digital competencies of employees and the implementation of a digital mindset within the organization stand out as critical challenges (Kocak & Pawlowski, 2022). Regarding digital competencies, many Industrial Craft Companies lack sufficient expertise in using digital tools and technologies. Efficient use of digital resources and implementation of digital innovations can be hindered by employees' lack of digital skills and knowledge, requiring targeted training and qualification measures. Additionally, embedding a digital mindset in traditionally minded Industrial Craft Businesses is challenging due to a lack of understanding of digitization opportunities and resistance to adapting business models, necessitating cultural change and fostering innovation and flexibility to remain competitive and future-ready.

The Objective of a Solution, Design and Development, and Demonstration

Understanding the required digital skills and attitudes is essential to driving digital transformation in the Industrial Craft Sector, enabling the development of targeted measures. A model for digital competence and mindset quantifies the degree of digital mindset, forming the basis for future training programs, and is developed through a thorough literature review to identify existing concepts. These concepts are analyzed using a concept matrix, a proven methodology according to Webster and Watson (2002). This analysis identifies relevant approaches and forms the basis for structured model development, presented systematically using a virtual Miro Board to illustrate the model's intricate relationships. A methodical approach with critical questions and evaluation questionnaires ensures a comprehensive understanding and aims to validate the model's effectiveness and alignment with the Industrial Craft Sector needs. Implementing this digital skills model strategically accelerates digital transformation by systematically recording and quantifying digital skills and attitudes, enabling targeted measures to prepare the Industrial Craft Sector for future challenges.

Evaluation and Communication

A qualitative research methodology was chosen for the evaluation because it offers profound insights into the complexity of human behavior and social phenomena that cannot be adequately captured using quantitative approaches alone (Myers, 2019). The philosophical basis of this work is the interactionist approach, which aims to develop a deep understanding and subjective experience (Myers, 2019). The qualitative methodology is based on grounded theory, which builds an approach by collecting data without prior assumptions to develop patterns and concepts from the data. We conducted semi-structured interviews with experts that dominated this study (Myers, 2019). These experts set the framework for evaluation and answered questions about comprehensibility, completeness, impact, and potential additions to the main dimensions, sub-dimensions, and the 5-step model. Expert interviews, conducted in person or virtually, are integral to the results and communication in this scientific work.

Sample

A homogeneous sampling method was used to select individuals with similar characteristics, focusing on experts from the craft sector and academic circles with relevant specialist knowledge. The experts were chosen based on their technical, process-related, and explanatory expertise in their field and the context of digitalization to ensure diverse perspectives. Considering a homogeneous population (Guest et al., 2006), a sample size of about 12 participants is recommended for qualitative interviews, while phenomenological studies usually envisage 3-10 interviews. The sample size of the current research was N=10, with experts selected based on their experience and expertise in crafts and digitalization. The study was conducted from January to October 2022-2023, with each interview lasting approximately 1.5 hours. Six male and four female experts took part in the interviews. Of the ten experts, seven had two years of professional experience, two had more than five years of experience, and one had three years of experience. The average age of the experts was 29 years *(M*= 29.9). Most experts were from the industrial craft sector (5 out of 10), with others from a university specializing in digital transformation in craft organizations, industrial construction and production, the energy sector, and geotechnics and environment. Their professional positions varied widely, including research assistants, civil engineers, environmental engineers, auditors, project managers, and specialized electrical and geological engineers experienced in digital transformation.

Procedure and Data Analysis

In the study's first phase, the Industrial Craft Sector experts evaluated the Digital Mindset Competence Model to ensure its comprehensibility and completeness and suggested additional dimensions. They also reviewed the five-stage model to assess its maturity and the clarity of each stage's definitions and descriptions. All interviews, which lasted approximately one and a half hours, were recorded and transcribed to evaluate the competencies, attitudes, and personality traits, focusing on clarity, importance, and potential improvements.

Construct and semantic validity were tested for reliability, with construct validity reflecting experts' interpretations and semantic validity categorizing similar text meanings (Bryman et al., 2008). Using Mayring's (2015) method for qualitative analysis, the study followed Myers's (2019) approach, incorporating participant feedback, confirmation tests, deductive coding, and selective coding to categorize data and address the research question. Finally, the transcripts were analyzed using a structured content analysis, testing construct, and semantic validity to ensure significance and reliability, with construct validity reflecting experts' interpretations and semantic validity categorizing similar text meanings, supported by participant feedback and confirmation checks, following Myers's (2019) approach (Bryman et al., 2008).

5 EVALUATION OF THE DIGITAL MINDSET COMPETENCE MODEL

The results of an expert survey confirm the general acceptance and comprehensibility of the model presented, which is tailored to its applicability in the Industrial Craft Sector. Several experts (E1, E5, E8) emphasize that the dimensions and sub-factors of the model (E2, E3, E8) are particularly well suited to skilled trades professions. A vital advantage of the model is promoting a digital mindset, which supports companies by simplifying work and communication between employees and customers (E5). In addition, digitalization makes cost savings possible (E5). One expert emphasizes the need to implement the concept in craft businesses to drive digital change (E2). Another expert points out that the relevant dimensions can vary depending on the sector, with different aspects possibly being at the forefront in the construction industry rather than in other industrial craft sectors (E3). About the dimensions, it is argued that apart from personality traits, all other skills and attitudes are crucial for a digital mindset (E3). It is recommended that additional information on the model be provided and the terminology be illustrated with verbs (E6, E9). The integration of information processing skills is considered particularly important (E9). The experts see a significant benefit of the model in improving interaction and communication within the craft organization (E1). The five-level model for assessing digital skills is considered relevant, but it should be noted that some terms should be clarified, such as "entrepreneurial skills" and "entrepreneurial orientation" (E9). The experts' opinions show that the model covers many important aspects but still has room for improvement (E4, E9). The importance of the model for the industrial craft sector is particularly emphasized, as it is seen as a contribution to promoting digitalization in companies (E9). Finally, a common taxonomy for the five-step model is proposed (E2, E9, E10). At the same time, one expert considers entrepreneurial skills irrelevant to the model, as she sees them as innovation in the company and not essential for a digital mindset (E5). Communication and flexibility are highlighted as crucial skills, mainly because of teamwork and the importance of communication in craft organizations (E2). Another expert notes that personal skills are appropriate but somewhat complex (E10). Finally, one subject matter expert reiterates the concept's validity and argues that adopting a digital mindset will help craftspeople manage tasks more efficiently and facilitate communication within the workforce and with customers (E5).

5.1 Results of the Claritiy, Importance, and the Impact of the Factors

In the academic context, two codes were developed for the category "Clarity": "Description" and "Assignment of characteristics to categories." Experts (E5, E9) recommended revising definitions such as personality traits and digital mindset. The coding "Assignment of characteristics to categories" identified various factors that should be integrated into the corresponding dimensions. In particular, it was suggested that emotional intelligence be assigned to personal skills, with Expert 9 considering adaptability as a personality trait that should be integrated into this dimension. Other characteristics, such as intercultural competence, were assigned to personal competence, while teamwork was given to communication competence, knowledge sharing to digital mindset, and critical thinking to entrepreneurial orientation (E9). In the "Priority or importance" category, the experts (E6) emphasized that communication skills, cultural aspects, and openness are crucial to driving digital change in industrial craft businesses. The experts also analyzed the interactions between the factors. It was highlighted that transparency has a positive effect on different ways of thinking (E9), adaptability has positive effects (E1), and communication skills have positive impacts on cooperation, collaboration, and willingness to take risks (E5, E9). Expert 9 emphasized that emotional intelligence has a positive influence on trust. Problem-solving skills, in turn, promote factors such as critical thinking, creativity, innovative thinking, and motivation (E9). Regarding adverse effects, the view was expressed that openness could impair conflict resolution (E9). One expert (E2) said that flexibility was essential in the construction industry to work on weekends. Communication was a critical characteristic of digitalization (E2, E3), as was openness, as open people tend to be willing to learn new information (E2). Another expert emphasized the importance of cybersecurity, as digitalization brings challenges such as hacker attacks and data protection (E10). An additional essential aspect is the culture of error in organizations, as mistakes should be seen as learning opportunities for the future (E7).

6 DISCUSSION

This study develops and evaluates a digital mindset competence model tailored to the Industrial Craft Sector using the design science approach and involving experts from the Industrial Craft Sector. After a literature review analyzed using a concept matrix, the model is based on Kocak and Pawlowski's (2023) and Knorr's (2020) methods. The study meticulously traversed the steps outlined in the design science methodology, culminating in the refinement and enhancement of the model based on the acquired results. The imperative of developing a Digital Mindset Competence Model designed explicitly for the Industrial Crafts Sector is underscored by the escalating digitalization and technological metamorphosis in today's professional milieu. The Industrial Crafts Sector, traditionally characterized by manual dexterity and experiential knowledge, grapples with the challenge of harnessing the potential of digitalization. A specialized competency model facilitates the identification of requisite skills and attitudes indispensable for artisans to navigate successfully within a digitalized milieu. The pertinence of the Digital Mindset Competence Model for the Industrial Crafts Sector lies in its capacity to augment individual competitiveness and empower proficient craft organizations to navigate digital transformation proactively. This advances efficiency and quality in the craft processes and fortifies the industry's competitive stance holistically. In our study, we conducted an exhaustive literature review, expanding upon the findings of Kocak and Pawlowski (2021). A focal point of our research was the investigation of existing digital mindset concepts, with diverse concepts and models subjected to comprehensive analysis for understanding and comparison. The outcomes of this analysis underscore that, despite the existence of digital mindset concepts, no specific models were identified for the Industrial Crafts Sector (Aliabina, 2020; Hildebrandt & Beimborn, 2021; Knorr, 2020; Kollmann, 2020; Kocak & Pawlowski, 2021). While these analyzed concepts offer valuable insights into digital mindsets, there exists a conspicuous need for a Digital Mindset Competence Model tailored to the requisites of the Industrial Crafts Sector. These findings accentuate the relevance of our research, which endeavors to establish a practice-oriented and industry-specific foundation for fostering a digital mindset in the Industrial Crafts Sector.

The results confirm the model's comprehensibility and completeness for the Industrial Crafts Sector, although some factors needed re-categorization into the correct dimensions, which was successfully done. The stage model was refined through improved definitions and terminology, enhancing iteration. Experts highlighted the significance of communication skills, cultural competencies, and openness, which are crucial for successful digitalization. The model adapts individual competencies and organizational requirements, bridging traditional craftsmanship and digital innovation.

The Digital Mindset Competence Model encompasses eight dimensions, further subdivided into factors. While literature encompasses concepts and models (Hildebrandt & Beimborn, 2022; Knorr, 2020; Kocak & Pawlowski, 2021) developed for the general sector, these models lack the dimensions crucial for the digital mindset. In contrast, our model encompasses digital attitudes, digital competencies, personality traits, entrepreneurial orientation, and personal competencies, all pivotal for the digital mindset. Another distinctive contribution of our work lies in the detailed descriptions of the factors, employing a standard taxonomy to elucidate elements within respective dimensions, some of which are echoed in the literature (Hildebrandt & Beimborn, 2021; Kocak & Pawlowski, 2023). An additional contribution is our five-stage model, which facilitates determining organizational levels by applying the model in craft organizations. This model is instrumental in guiding further research to develop measures tailored to the craft sector—a level model not present in existing literature on digital mindset models in the craft sector. This research significantly fills the industrial crafts sector gap by developing and evaluating a digital competencies and attitudes model tailored to its unique needs and challenges. It provides novel insights into digital technology implementation in traditional sectors. Using a design science approach and a mixed methodology, the study offers precise guidelines for industrial craft enterprises to enhance their digital skills and attitudes, which are essential for maintaining competitiveness and realizing the potential of digitalization.

Implementing the developed model will optimize work processes, increase productivity, and improve customer communication, leading to excellent customer and employee satisfaction. Developing digital skills and a digital mindset enhances daily efficiency and innovation in craft businesses, allowing them to create new products and adapt to market changes. While the study offers valuable insights, it has limitations, such as the need for further evaluation of factor descriptions and customization for various facets of craftsmanship. Future work should include developing survey items, conducting structural equation modeling to understand factor correlations, and exploring the model's impact on job performance and digital organizational culture in the Industrial Crafts Sector.

7 CONCLUSION

This study aims to develop and evaluate a digital mindset competence model for the Industrial Crafts Sector using the design science approach to address practice-relevant problems and contribute to knowledge development. Experts from various Industrial Crafts sectors with digitalization experience evaluated the model for completeness, understandability, and correct allocation of factors. The results indicated that some terms and definitions need modification, particularly in the model and the five-stage framework. The model serves as the first approach for identifying the digital mindset and skills in Industrial Craft Sector and was improved through expert evaluation. Future research should include further evaluation, ideally through focus group discussions within specific areas of the Industrial Crafts Sector.

REFERENCES

Abel, R. (2007). *Erwerbsregulierung in klein Handwerksunternehmen* [Acquisition regulation in small craft enterprises]. München und Mering: Rainer Hampp Verlag

Aliabina, E. (2020). The Concept of Digital Mindset in the Context of Entrepreneurship. *Sciences of Europe, 61(2),* pp. 15–19.

Bordi, L., Okkonen, J., Mäkiniemi, J. P., & Heikkilätammi, K. (2018). Communication in the digital work environment: Implications for wellbeing at work. *Nordic Journal of Working Life Studies, 8(3*), pp. 29-48. https://doi.org/10.18291/njwls.v8iS3.105275

Bryman, A., Becker, S., & Sempik, J. (2008). Quality criteria for quantitative, qualitative and mixed method research: A view from social policy. *International journal of social research methodology, 11(4),* pp. 261-276. https://doi.org/10.1080/13645570701401644

Buschfeld, D., Dilger, B., Hess, L., Schmid, K., & Voss, E. (2011). Identification of future skills needs in micro and craft (-type) enteprises up to 2020. Cologne, Hamburg, Vienna.

Diego, C., Mêda, P., Chichorro Goncalves, M., & Sousa, H. (2020). Worker 4.0: The future of sensored construction sites. *Buildings 10 (10),* pp. 1-22. https://doi.org/10.3390/buildings10100169

European Construction Sector Observator. (2020). Improving the human capital basis. Analytical Report. Retrieved from https://ec.europa.eu/docsroom/documents/4126

Ezeokoli, F. O., Okolie, K. C., Okoye, P. U., & Belonwu, C. C. (2016). Digital transformation in the Nigeria construction industry: The professionals' view. *World Journal of Computer Application and Technology, 4(3)*, pp. 23-30. https://doi.org/10.13189/wjcat.2016.040301

Fisher, D. S. (2022). Unfreezing and Refreezing the Digital Mindset of Business. *International Journal of Innovative Science and Research Technology, 7(3),* pp. 901–905.

Fitriani, H. and Ajayi, S. (2023). Investigation of requisite measures for enhancing sustainable construction practices in Indonesia, *Engineering, Construction and Architectural Management, 30 (6),* pp. 2602–2620. https://doi.org/10.1108/ECAM-11-2021-1051

Gekara, V., Molla, A., Snell, D., Karanasios, S., & Thomas, A. (2017). Developing appropriate workforce skills for Australia's emerging digital economy.

Grosso, S. L., Meinero, D., & Volonta, M. (2019). Digitalising the construction sector- Unlocking of with the chain value data potential findings. CECE. Retrieved from https://www.cece.eu/publications/digital-reports

Hevner, A., & Chatterjee, S. (2010). Design Research in Information Systems. *Design research in information systems*, pp. 9-22.

Hildebrandt, Y., & Beimborn, D. (2021). The Intangible Key for Digitalization: Conceptualizing and Measuring the "Digital Mindset". *Proceedings of the 2021 on Computers and People Research Conference*, pp. 89-91.

Hildebrandt, Y., & Beimborn, D. (2022). A Cognitive Conveyor for Digital Innovation-Definition and Conceptualization of the Digital Mindset. *Wirtschaftsinformatik 2022 Proceedings 12*.

Johannes, K., Voordijk, H., Wakkee, I. and Aranda-Mena, G. (2024). Implementing organisational change in a digitalising facilities management organisation through stewardship interventions. *Journal of Facilities Management, 22(1)*, pp. 144-159. https://doi.org/10.1108/JFM-11-2021-0139

Joseph, C., Norizan, S., Enggong, T. S., Rahmat, M., & Nyet, C. A. (2022). Realizing sustainable development goals via entrepreneurial digital mindset: Resource-based view perspective. *International Journal of Accounting, 7(42).* pp. 1-12. https://doi.org/10.55573/IJAFB.074201

Kemendi, A. (2021). E-Commerce safety and security in the Industry 4.0 era. *Military National Security Service*, p. 195.

Kim, G. S. (2020). The effect of quality management and Big Data management on customer satisfaction in Korea's public sector. *Sustainability, 12(13),* p. 5474. https://doi.org/10.3390/su12135474

Knorr, J. (2020, 03 08). Digital Mindset zur Steigerung der Wettbewerbsfähigkeit von Unternehmen [Digital Mindset to increase the competitiveness of companies]. In J. Knorr, Digitale Transformation in der Unternehmenspraxis: Mindset- Leadership-Akteure-Technologien (pp. 45-60). Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from https://www.dci.digital/

Ko, D. G., Kirsch, L. J., & King, W. R. (2005). Antecedents of knowledge transfer from consultants to clients in enterprise system implementations. *MIS Quarterly, 29 (1),* pp. 59-85. https://doi.org/10.2307/25148668

Kocak, S., & Pawlowski, J. (2021). A qualitative study on the categorisation and priorisation of digital competencies and attitudes for mangers and employees. *International joint conference on knowledge discovery, knowledge engineering, and knowledge management. SCITEPRESS-Science and Technology Publications*. https://doi.org/10.5220/0010674700003064

Kocak, S., & Pawlowski, J. (2022). Characteristics in Digital Organizational Culture: A Literature Review. Proceedings of the 14th International *Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management- KMIS,* pp. 31-42. https://doi.org/10.5220/0011530500003335

Kocak, S., & Pawlowski, J. (2022). Digitalization Breakers: A Mixed Method For Finding The Concrete Barriers In The Craft Sector. *PACIS 2022 Proceedings, 53,* pp. 1-16.

Kocak, S., & Pawlowski, J. (2023). Digital Transformation in the Craft Sector: Identifying Digital Competencies and Attitudes for Managers and Employees. SIMIS CPR2023

Kollmann, T. (2020). *Das Digital Mindset* [The Digital Mindset]. In T. Kollmann, Digital Leadership (pp. 43-78). Wiesbaden: Springer-Gabler.

Krathwohl, D. R. (2022). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41 (4)*, pp. 212-218. https://doi.org/10.1207/s15430421tip4104_2

Kratochvil, R., Khanamiryan, M., Soylu, A., & Kvålshaugen, R. (2021). Organizational context factors that stimulate digital mindsets to unfold in digital transformation. *Academy of Management Proceedings*. https://doi.org/10.5465/AMBPP.2021.10024abstract

Lessiak, C. (2020). The Digital Mindset as a prerequisite for successful Digital Transformation: Definition, Attributes and Approach for Measurement. *Workshop specifického výzkumu 2020*, p. 120.

Mårtensson, M. (2000). A critical review of knowledge management as a management tool. *Journal of knowledge management*, pp. 204–216. https://doi.org/10.1108/13673270010350002

Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. *Business & Information systems engineering, 57 (5)*, pp. 339–343. https://doi.org/10.1007/s12599-015-0401-5

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. Approaches to qualitative research in mathematics education: Examples of methodology and methods, pp. 365–380. https://doi.org/10.1007/978-94-017-9181-6_13

McGuinnes, C. (2005). Behind the acquisition metaphor: conceptions of learning and learning outcomes in TLRP school-based projects. *Curriculum Journal, 16(1)*, 31–47. https://doi.org/10.1080/0958517042000336791

Mekhaznia, T., Djeddi, C., & Sarkar, S. (2021). Personality Traits Identification Through Handwriting Analysis. Pattern Recognition and Artificial Intelligence. *Communications in Computer and Information Science, 1322*, pp. 155-169. https://doi.org/10.1007/978-3-030-71804-6_12

Mihardio, L. W., & Sasmoko, S. (2019). Digital Transformation: Digital Leadership Role in Developing Business Model Innovation Mediated by Co-Creation Strategy for Telecommunication Incumbent Firms. Strategy and Behaviors in the Digital Economy. https://doi.org/10.5772/intechopen.82517

Mohammadi, S. (2020). Organizational culture and its impact on organizational productivity. *International Journal of Human Capital in Urban Management*, 5 (3)

Myers, M. D. (2019). Qualitative research in business and management. *Qualitative research in business and management*, pp. 1–364.

Neeley, T., & Leonardi, P. (2022). *Developing a Digital Mindset*. Harvard Business Review, 100(5-6).

Nikmehr, B., Hosseini, M. R., Martek, I., Zavadskas, E. K., & Antucheviciene, J. (2021). Digitalization as a Strategic Means of Achieving Sustainable Efficiencies in Construction Management: A Critical Review. . *Sustainability, 13*, p. 5040. https://doi.org/10.3390/su13095040

Oesterreich, T. D., & Teuteberg, F. (2016). Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computer in Industry, 83*, pp. 121-139. https://doi.org/10.1016/j.compind.2016.09.006

Osmundsen, K. (2020). Competences for digital transformation: Insights from the Norwegian energy sector. *Proceedings of the 53rd Hawaii international conference on system sciences*.

Pammer, M., Aufschnaiter, C., & Waldegger, J. (2021). Vernetzte Arbeitswelt der digitalen Arbeitnehmer [Networked working world of digital employees]. In Handwerk digital: Kulturwandel als Erfolgsfaktor (pp. 61-77). Wiesbaden: Springer.

Parusheva, S. (2019). Digitalization and Digital Transformation in Construction-Benefits and Challenges. *Information and Communication Technologies in Business and Education*, pp. 126–134.

Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information system research. *Journal of Management Information Systems, 24(3)*, pp. 45-77. https://doi.org/10.2753/MIS0742-1222240302

Punie, Y., Zinnbauer, D., & Cabrera, M. (2006). A review of the impact of ICT on learning. *European Commission Brussels, 6 (5)*, pp. 635-650.

Puolitaival, T., Davies, K., & Kähkönen, K. (2019). Digital technologies and related competences in construction management in the era of fast-paced digitalisation. *Proceedings of CIB World Building Congress*.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interviews. *Qualitative Research in Accounting & Management, 8(3)*, pp. 238-264. https://doi.org/10.1108/11766091111162070

Riina, V., Yves, P., Staphanie, C. G., & Van Den Brande, G. (2016). DigComp 2.0: The Digital Competence Framework for Citizens. *Update Phase 1: the Conceptual Reference Model*. https://doi.org/10.2791/520113

Rohleder, B., & Schulte, K. (2022, 03 21). Digitalisierung im Handwerk [Digitalisation in the Industrial Craft Sector]. Retrieved from https://www.bitkom.org/sites/default/files/pdf/Presse/Anhaenge-an-PIs/2017/03-Maerz/Bitkom-ZDH-Charts-zur-Digitalisierung-des-Handwerks-02-03-2017-final.pdf

Salleh, L. M. (2008). Communication competence: A Malaysian perspective. *Pacific and Asian Communication Association, 11(3)*, pp. 303-312.

Salvetti, F., Galli, C., Bertagni, B., Gardner, R., & Minehart, R. (2022). A digital mindset for the society 5.0: Experience an online escape room. Innovations in Learning and Technology for the Workplace and Higher Education. *The Learning Ideas- Conference 2021*, pp. 290-303. https://doi.org/10.1007/978-3-030-90677-1_28

Sánchez, O., Revuelta, M. P., Gómez-Cabrera, A., & Salazar, L. A. (2023). Papre Planes for Teaching Construction Production Systems Based on Lean Tools: Continuous Improvement Cells and 5S. *Buildings, 13(2)*, p. 558. https://doi.org/10.3390/buildings13020558

Schein, E. H. (1990). Organizational culture, *American Psychological Association*, 45(2), 109-119. https://doi.org/10.1037/0003-066X.45.2.109

Schuur, F., Rezazade Mehrizi, M. H., & Ranschaert, E. (2021). Training opportunities of artificial intelligence (AI) in radiology: a systematic review. *European Radiology, 31*, pp. 6021-6029. https://doi.org/10.1007/s00330-020-07621-y

Sriyolia, Z., Harwin, N., & Yahya, K. (2021). Barriers to implement building information modeling (BIM) in construction industry: A critical review. IOP Conference Series: *Earth and Environmental Science , 738 (1)*, pp. 1-9. https://doi.org/10.1088/1755-1315/738/1/012021

Tahvanainen, S., & Luoma, E. (2018). Examining the competencies of the chief digital officer. *Americas Conference on Information Systems*, pp. 1-10.

Timchuk, O. G., & Evloeva, M. V. (2020). Difficulties in transforming the construction industry under the digital economy. *IOP Conference Series: Materials Science and Engineering*, p. 880. https://doi.org/10.1088/1757-899X/880/1/012082

Timoshevskaya, O., Londikov, V., Andreev, D., Samsonenkov, V., & Klets, T. (2021). Digital data Processing Based on wavelet transforms. International Scientific and Practical Conference, 2, pp. 174-180. https://doi.org/10.17770/etr2021vol2.6634

Überbacher, R., Brozzi, R., & Matt, D. T. (2020). Innovation in craftsmanship family Smes in times of digitalization. *Piccola Impresa/ Small Business, (1*). https://doi.org/10.14596/pisb.345

Vasiliki, C., Evertz, L., & Süß, S. (2020). Digitale Transformation im Handwerk: Eine qualitative Analyse der individuellen Veränderungsbereitschaft [Digital transformation in the Industrial Craft Sector: A qualitative analysis of individual willingness to change]. ZfKE-Zeitschrift für KMU und Entrepreneurship 68 (3-4), pp. 149-168.

Veltkamp, N., & Schulte, K. (2022, 03 21). Digitalisierung des Handwerks. Retrieved from https://www.bitkom.org/sites/default/files/2020-03/200304_prasentation_digitaleshandwerk_final.pdf

Vogelsang , K., Liere-Netheler, K., Packmohr, S., & Hoppe, U. (2019). Barriers to Digital Transformation in Manufacturing: Development of a Research Agenda. *Proceeding of th e52nd Hawaii International*.

Von Ohain, B. P. (2019). Leader Attributes for Succesful Digital Transformation . *Fortieth International Conference on Information Systems*, pp. 1-17.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Welzbacher, C., Pirk, W., Ostheimer, A., Bartelt, K., Bille, J., & Klemmt, M. (2015). *Digitalisierung der Wertschöfungs- und Marktprozesse-Herausforderungen und Chancen für das Handwerk-Eine Vorstudie im Rahmen der Konzeption eines Demostrations - und Kompetenzzentrums im Handwerk* [Digitalisation of value creation and market processes - challenges and opportunities for the skilled crafts sector - a preliminary study as part of the conception of a demonstration and competence centre in the Industrial Craft Sector]. Hannover.

Xu, T., & Xu, Y. (2012). A literature review of relationship between entrepreneurial orientation and firm performance. *2012 International Symposium on Management of Technology (ISMOT*), pp. 132-135. https://doi.org/10.1109/ISMOT.2012.6679442

Yang, K., Sunindijo, R. Y., & Wang, C. C. (2022). Identifying Leadership Competencies for Construction 4.0. *Construction 4.0*, pp. 1-18. https://doi.org/10.3390/buildings12091434

Young, R., Wahlberg, L., Davis, E., & Abhari, K. (2020). Towards a theory of digital entrepreneurship mindset: The role of digital learning aptitude and digital literacy. *Americas Conference on Information Systems*, pp. 1-10.

Using Formal Concept Analysis for Corpus Visualisation and Relevance Analysis

Fabrice Boissier[1] [a], Irina Rychkova[2] and Bénédicte Le Grand[2]
[1]*LRE, EPITA, 14-16 Rue Voltaire, Le Kremlin-Bicêtre, France*
[2]*CRI, Université Paris 1 Panthéon, Sorbonne, 90 Rue de Tolbiac, Paris, France*
fabrice.boissier@epita.fr, {irina.rychkova, benedicte.le-grand}@univ-paris1.fr

Keywords: Corpus Visualization, Relevance Visualization, Topic Modeling, Formal Concept Analysis, Text Mining, Natural Language Processing.

Abstract: Corpora analysis is a common task in digital humanities that profits from the advances in topic modeling and visualization from the computer science and information system fields. Topic modeling is often done using methods from the Latent Dirichlet Allocation (LDA) family, and visualizations usually propose views based on the input documents and topics found. In this paper, we first explore the use of Formal Concept Analysis (FCA) as a replacement for LDA in order to visualize the most important keywords and then the relevance of multiple documents concerning close topics. FCA offers another method for analyzing texts that is not based on probabilities but on the analysis of a lattice and its formal concepts. The main processing pipeline is as follows: first, documents are cleaned using TreeTagger and BabelFy; next, a lattice is built. Following this, the mutual impact is calculated as part of the FCA process. Finally, a force-based graph is generated. The output map is composed of a graph displaying keywords as rings of importance, and documents positioned based on their relevance. Three experiments are presented to evaluate the keywords displayed and how well relevance is evolving on the output map.

1 INTRODUCTION

In the age of data, knowledge is an essential factor that increases the capacity to make the best decisions (North and Kumta, 2018). Data visualization remains a difficult task for knowledge extraction as numerous visualizations are available and each business and/or application face their own challenges (Andrienko et al., 2020)(Padilla et al., 2018)(Engebretsen and Kennedy, 2020). It requires a perfect understanding of the goal to achieve and the manipulated data.

Visualizing text corpora is a cross-domain application of interest for information systems and digital humanities (Jänicke et al., 2015). Improvements in visualization lead to better analysis of multiple types of texts like historical newspapers (Menhour et al., 2023) or even poetry (De Sisto et al., 2024). However, few work has been done to visualize the relevance between documents based on their topics. *Topic modeling* also contributes to texts' analysis as a research orientation of the information retrieval field by "*uncovering latent text topics by modeling word associations*" (Hambarde and Proenca, 2023). Topics are simply a list of words that share a similar theme: either each word is strongly/directly expressing the theme, or the collection of words illustrates an abstract theme by their semantic links (but alone, they would not make sense). Multiple topic modeling methods exist (Alghamdi and Alfalqi, 2015)(Kherwa and Bansal, 2019). They usually consider documents as a *bag-of-words* where the order of words is not important; only their occurrence in each document is important. The main trend in topic modeling consists of probabilistic methods where topics are built based on the probability of each term being part of a specific topic. Recent advances in natural language processing (NLP) also introduced neural networks combined with traditional methods, allowing the capture of the context of words within documents and reusing it to analyze newer documents.

In this paper, we explore the use of *Formal Concept Analysis* (FCA) (Ganter and Wille, 2012) instead of more traditional topic modeling methods, and we propose a visualization of the main keywords of a corpus and documents' relevance on a force-based graph. The strength of FCA resides in the fact that it an-

[a] https://orcid.org/0000-0002-0067-6524

Boissier, F., Rychkova, I. and Le Grand, B.
Using Formal Concept Analysis for Corpus Visualisation and Relevance Analysis.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 120-129
ISBN: 978-989-758-716-0; ISSN: 2184-3228

alyzes the relationships within data and produces a lattice that can be used for calculating useful measures like similarities. Instead of trying to find back memberships of an object between multiple distributions or the best matching weight of a parameter in a neural network, FCA highlights the existing relationships between objects based on their attributes. FCA is known as a viable text mining method (Carpineto and Romano, 2004) and is a good candidate for multiple applications in the knowledge field (Poelmans et al., 2013). FCA has been used in conjunction *with* a topic modeling method (Akhtar et al., 2019) but not *instead* of it.

The paper is organized as follows: First, we explain the main methods currently used in topic modeling as well as topic and relevance visualizations. Second, we present our processing pipeline from the input documents to the output map. Then, we present three experiments, the text materials used, their topics, and their relevance. Finally, we discuss the results and conclude.

2 RELATED WORK

2.1 Topic Modeling

Analyzing documents implies creating statistics on the used terms. *Term Frequency - Inverse Document Frequency* (TF-IDF) (Salton, 1983) calculates a ratio from the frequency of each term to the total number of documents. Documents are therefore seen as a ratio of words independent of their ordering, like a *bag-of-words*. TF-IDF is not exactly a topic modeling method, but it shows the importance and uniqueness of terms within the corpus.

Latent Semantic Analysis (LSA) (Deerwester et al., 1990) transforms documents into a *latent semantic space* from which multiple outputs can be analyzed. The main method behind LSA is the *singular value decomposition*, which produces three matrices based on a parameter K given by the user: a matrix of *terms per K features*, a matrix of *K features per K features*, and a matrix of *K features per documents*. Multiple analyses can be done on those matrices, but in our case, the *terms per K features* one is the most important as it allows us to know how well each term is linked to each feature (that in fact represents latent topics). One problem induced by LSA is that it can't manage polysemy: each term is used as the same entity in any document. Homonyms, like synonyms and even various forms of the same word, can produce an inconsistency because of the missing context of each document. Standardization of the input, like stemming or even lemmatization, can partially leverage this problem.

Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999) is an upgrade of LSA that introduces a probabilistic point of view by building a generative model for each text corpus. Because topics are scattered within documents, probabilities help find the terms that compose them. PLSA relies on an *aspect model* (the probabilities between terms, documents, and the latent topics) and a *mixture decomposition* that obtains similar results as the singular value decomposition, allowing PLSA also to build the three matrices of LSA.

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is also a generative probabilistic model that aims to model text corpora. It aims to improve the PLSA mixture decomposition by using a hierarchical Bayesian model. LDA allows not only the finding of topics of words within documents, in the case of text corpora, but also its usage with more confidence than PLSA as a probabilistic generative model in multiple domains.

LDA can be combined with *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al., 2018) in order to increase the quality of its results (Peinelt et al., 2020)(George and Sumathy, 2023). BERTs are a collection of pre-trained neural networks designed to help researchers in NLP by considering words and their neighbors before and after them. By their nature, BERTs do not produce a list of topics but can be used to generate a summary. Similarly to BERTs, it can be noted that *Generative Pre-trained Transformers* (GPT) (Brown et al., 2020) achieve numerous NLP tasks. Still, they also share the drawback of not explaining the weights within the neural network to build their answers.

Correlated Topic Model (CTM) (Blei and Lafferty, 2006) is derived from LDA and uses another distribution in order to better capture topics and their relations within documents. Indeed, when a document concerns a theme, it usually talks about some neighbor topics: a text about travel might talk about tourism, beaches, and airplanes, but probably not about fighter jets. Topics are uncorrelated in LDA because of the Dirichlet distribution, whereas in CTM, thanks to the logistic normal distribution, topics are correlated and present links to one another. The presence of a topic triggers the possibility of finding one or multiple other topics. CTM also proposes a visualization of topics: each topic is represented in a bubble of words, and each bubble is linked with the other correlated topics.

2.2 Visualization of Topics and Relevance

The presentation of topics is important as reading a list might confuse a human: a list forces an order of reading, which is not always the best one depending on the context. Visualization of topics is usually made of tag clouds (Singh et al., 2017)(Lee et al., 2010), which is convenient but shows terms as a whole block. Some visualizations like (Singh et al., 2017) or (Gretarsson et al., 2012) show the distinction between topics and the terms composing them. However, if a document used in the corpus is irrelevant or contains too many irrelevant parts, the results might be altered without the user being informed. In addition, each document must be deeply reviewed to find the discrepancy. Few work has been done to explicit the relevance of documents one to another based on their topics.

(Assa et al., 1997) presents a relevance map between topics and keywords in the case of web search queries. Their proposal uses dimension reduction to create a 2D map and gravitational forces for placing nodes. Similarly, VIBE (Olsen et al., 1992)(Olsen et al., 1993) and its variants (Ahn and Brusilovsky, 2009) create *Points of Interest* (POI) that act precisely like topics around which documents are placed based on relevance. The main drawback is that it does not highlight the most important common words and topics but only places resources based on their relevance to all the topics found.

(Fortuna et al., 2005) proposed a map of terms and documents based on *MultiDimensional Scaling* (MDS) methods. Terms and documents are placed on a two-dimensional map, and the background color varies based on the density. The main drawback of this contribution is the difficulty of getting a clear overview of the documents and terms.

In (Newman et al., 2010), the authors proposed a topic map after studying topic modeling and 2D projection. First, they compare three topic modeling methods and end by using LDA. Next, they compare three methods of projection, namely *Principal Component Analysis* (PCA), *Local Linear Embedding* (LLE), and *Force Directed Layout* (FDL) which is the best. The topic map presents documents as nodes colored by their most important topic. Their position depends on their relevance to one another. However, the authors concluded that the evaluation of visualization is complex and could be made only by human judgment; in addition, they also stated that maps with a dozen documents are probably the most accurate and valuable in understandability and navigation for a human.

TopicViz (Eisenstein et al., 2012) proposes a visualization of documents and topics by nodes with a force-directed layout and, more importantly, interaction with the user. The topic modeling method used is LDA. The user can pin topic nodes in the workspace, making the documents float based on their relevance to each topic. Such a map allows one to distinguish which document is more relevant to some or more topics based on its position. The user can also pin document nodes, making the topics float between them. This visualization is particularly interesting because document and topic nodes can be pinned, allowing it to show relevance. However, the user must pin the nodes himself in order to see the relevance. Based on the number of detected topics, deciding where to pin each topic to see the documents' relevance better might be difficult.

PaperViz (di Sciascio et al., 2017) is a dedicated tool for researchers during the paper-gathering step. It offers multiple views for multiple contexts: tree hierarchy for search queries, a tag cloud of the 20 most frequent terms, the strength of the relationship between documents and a search query or a collection, and references management. The main strength of PaperViz is its completeness in the user interface.

3 VISUALIZATION PIPELINE WITH FORMAL CONCEPT ANALYSIS

The visualization pipeline comprises two main phases (see Figure 1): *semantic pre-processing*, where documents are analyzed in order to produce an occurrence matrix of terms per documents, and *structural analysis*, where the matrix is analyzed in order to create a graphical representation of the relevance. Globally, the pipeline relies on natural language processing (NLP) methods in the first phase and formal concept analysis (FCA) methods in the second phase.

3.1 Semantic Pre-Processing

The *semantic pre-processing* phase aims to extract the most important terms and concepts from each document and gather them within a matrix representing the whole text corpus. This phase is composed of 5 steps as follows:

(PI.0) **Selection of Documents by the User:** the user selects the documents for analysis and comparison. Two requirements must be fulfilled for the best results: each document must have enough content, and the content must be mainly textual.

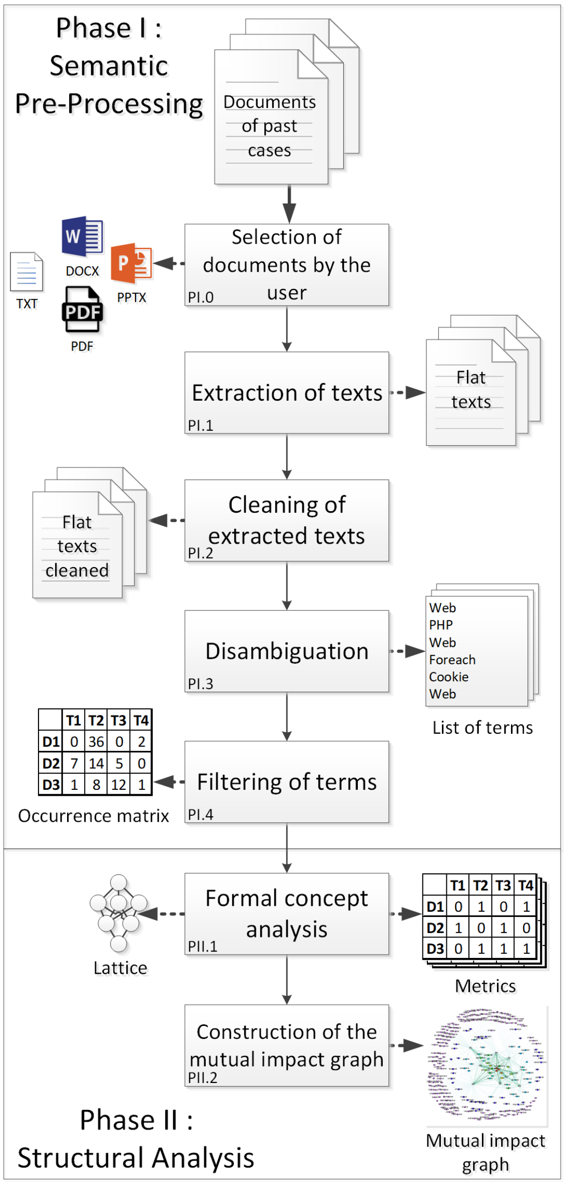


Figure 1: The main steps of the pipeline.

(PI.1) **Extraction of Texts:** each document is transformed into a regular flat text file. This step relies on optical character recognition (OCR) methods. In our experiments, we used PDFtoText[1] as the OCR.

(PI.2) **Cleaning of Extracted Texts:** each text is cleaned in order to increase its quality and reduce its size, typically by removing the useless spaces and artifacts that the OCR would have created and, eventually, some of the stop words. In our experiments, we used TreeTagger (Schmid, 1994)(Schmid, 1995) with a custom list of words to keep.

[1]https://www.xpdfreader.com/pdftotext-man.html

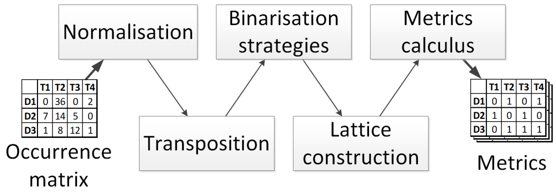


Figure 2: The formal concept analysis (PII.1) sub-steps.

(PI.3) **Disambiguation:** each cleaned text is transformed into a list of named entities by resolving polysemy and synonymy problems. Advanced NLP methods are required for this task. In our experiment, we used BabelFy (Moro et al., 2014) as it understands multiple languages and calculates three scores for each recognized named entity. The named entities are also transformed into unique references from BabelNet (Navigli and Ponzetto, 2012), allowing us to manipulate the exact same named entities in all documents, whatever the input languages are.

(PI.4) **Filtering of Terms:** the most irrelevant named entities are removed based on the *coherence score* attributed by BabelFy in the previous step. In our experiments, we require a coherence score of at least 0.05 to keep a named entity. This score was empirically chosen because it removes way more irrelevant named entities than relevant ones.

3.2 Structural Analysis

The *structural analysis* phase comprises two steps that calculate metrics in order to produce the mutual impact graph showing the relevance of documents.

(PII.1) The *formal concept analysis* is the first step, divided into five sub-steps (see Figure 2). Its objective is to produce the mutual impact matrix between terms and documents from the occurrence matrix in order to evaluate the relevance of documents.

- *Normalisation:* occurrences of terms per documents are transformed into proportions in order to reduce the length disproportions between documents. Absolute values are converted into percentages per line. Thus, all documents are treated equally.
- *Transposition:* the matrix is transposed in order to change the point of view from documents characterized by occurrences of terms into terms characterized by their appearances within documents.
- *Binarisation Strategy:* building the lattice requires a *formal context*, or, in other words, a binary matrix. Multiple strategies of binarization exist (also called *refinement strategies*) (Jaffal et al., 2015). In our case, we use the most simple

one: the *direct strategy* that transforms all values > 0 as 1 and keeps 0 as 0.

- *Lattice Construction:* the formal context representing terms within documents is used for building a lattice and its *formal concepts* (Belohlavek, 2008). A formal concept is a node containing objects and attributes (at least one of them). In our case, the objects are terms, and the attributes are the documents.
- *Metric Calculus:* the lattice is analyzed, and the *mutual impact* (Jaffal et al., 2015) metric is calculated by comparing the appearances of couples of terms and documents within each formal concept. The mutual impact shows how strong the relationship is between each term and each document. This bond is calculated for each term and each document with the formula:

$$MI(O_i, A_j) = \frac{\textit{formal concepts containing } O_i \text{ and } A_j}{\textit{formal concepts containing } O_i \text{ or } A_j}$$

where O_i represents a term and A_j represents a document. The output is a mutual impact matrix with a value representing the bond between each term and each document.

(PII.2) The *construction of the mutual impact graph* is the final step. Its objective is to visualize on a map the terms and their importance within the corpus, as well as the documents and their relevance. The visualization uses the mutual impact matrix as an adjacency matrix and produces a graph of terms and documents. We used Gephi (Bastian et al., 2009) with the ForceAtlas (first version) spatialisation algorithm. Nodes are moved until they find their optimal position thanks to attractive and repulsive forces. Nodes are repulsing each other, and edges attract nodes based on the values of the edges. Because of the input format, the visualization is a bipartite graph: a set of nodes represents documents, and another set represents terms. As presented in Figure 3, the nodes of documents are colored in grey and are linked to numerous nodes of terms. Unlike, these term nodes are colored based on the number of neighbors (the warmer the color, the more the node of term is linked to different nodes of documents). When focusing on the nodes of terms, we can find terms from every document in the center of the map and terms from fewer documents scattered around (terms, present in only one document, are forming specific groups for each document far from the central set). The central set of terms is, in fact, a global view of the text corpus with its keywords. When focusing on the nodes of documents, we can visually see the relevance of documents within the corpus by checking how close documents are to the central set of terms.

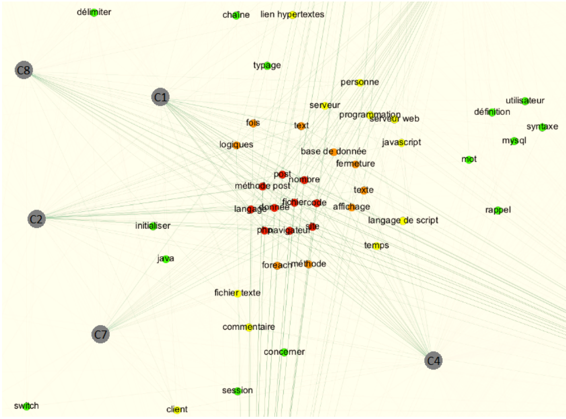


Figure 3: The output map of terms and documents.

4 EXPERIMENT

4.1 Scenarios of Evaluation

A proof of concept (Elliott, 2021) has been realized with a specific scenario to check the pipeline's validity and properties. A prototype has been developed[2] and used in three proof of concept demonstrations. A first case expects to visualize the content of 9 PHP courses. The mutual impact graph is visualized to check the corpus' main keywords and the documents' relevance. A second case introduces a Java course in the corpus. The validity of the mutual impact graph is checked by comparing the first and second cases: the Java course should be the most irrelevant because it is not specialized in PHP. A third case is presented in order to check how correcting a document impacts the results.

In the regular case, 9 PHP courses in French are processed through the pipeline. These courses present web development with HTML, PHP, and MySQL for beginners. They are denoted as C1-C9 in the following figures. 6 of these courses are made of slides (C1, C2, C4, C7, C8, C9), and 3 are made of regular texts (C3, C5, C6). A Java course (CJA) in text format is later introduced to check how the mutual impact graph behaves when some errors appear. This course is also in French and in a text format. The third experiment only works on courses in text format. Therefore, we introduced 4 more PHP courses in text format (C11, C15, C17, C19) to avoid disproportion by keeping a close number of input documents as in the other cases.

[2]https://github.com/metalbobinou/CREA-phd

4.2 General View of the Corpus' Content

The documents are processed through phase I, and the corpus is transformed into a matrix of occurrences. The mutual impact graph is generated using the *direct strategy* in step PII.1.

Figure 4 shows terms as nodes colored following a cold-warm schema. Red nodes are terms occurring in all the documents, orange nodes are terms occurring in all the documents minus one, etc. In the regular case, the terms in red nodes are: *post*, *nombre* (number), *méthode post* (post method), *fichier* (file), *code*, *donnée* (data), *langage* (language), *site*, *php*, *navigateur* (browser). These terms are typical of a course on web development with PHP. They are extended with the terms in orange nodes like *base de donnée* (database) or even *foreach*, which are also typical for a website in PHP that uses a database. Terms that are present in fewer documents but still in more than half of the corpus (the yellow and green nodes) are also typical of web development for nearly all of them (*session*, *mysql*, *utilisateur* (user), ...).

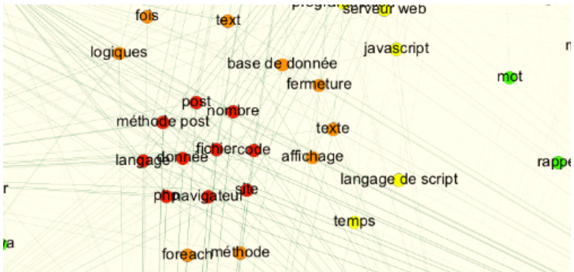


Figure 4: Central set terms of the mutual impact graph in regular case.

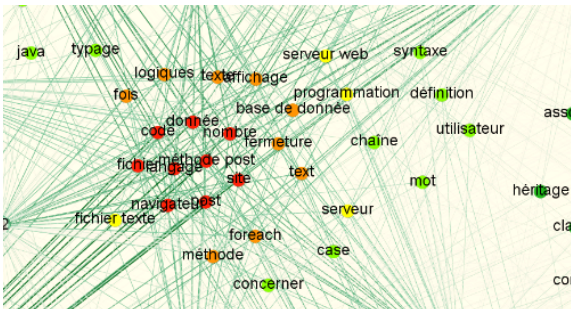


Figure 5: Central set terms of the mutual impact graph in Java case.

In Figure 5, the Java course is added to the corpus. The terms in red nodes are the same as in Figure 4, except *php* which becomes an orange node. This behavior is expected as the Java course does not discuss PHP; therefore, one document does not include it. The terms in other colors are still relevant as they mainly concern client-server programming, OOP vocabulary, and similar topics. It can be noted that another color is introduced in order to show that an additional document is present. Nearly all of the terms in green in the first case are now light green. In the entire graph, some of the terms in green are repulsed into dark green nodes (meaning they are missing from one more document). However, a majority of terms from the first case are still present in the second case with the same number of document edges.

4.3 Relevance of Documents

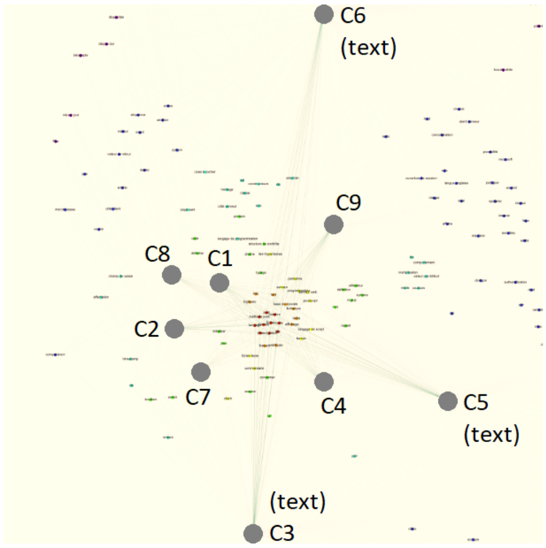


Figure 6: Relevance of documents in regular case.

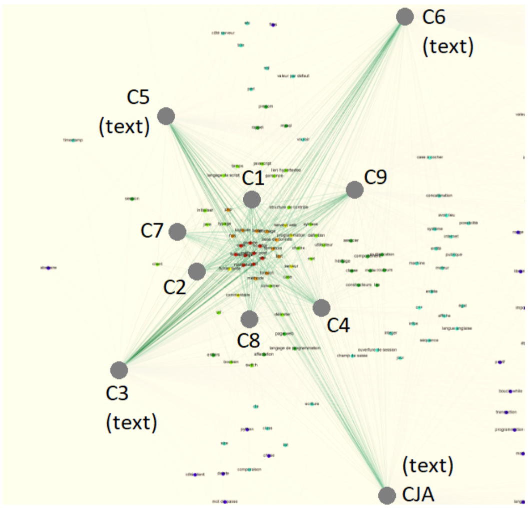


Figure 7: Relevance of documents in Java case.

Figures 6 and 7 represent the relevance of documents in two cases: the regular case with only PHP courses, and the Java case with the additional Java course. The relevance's visualization is also produced from the mutual impact graph, except the focus is mainly on the grey nodes representing documents.

In the regular case of PHP courses (Figure 6), the documents in the slide format are closer to the central set than the text ones. In the Java case (figure 7), the Java document (CJA) is the most distant of the central set. It can be noted that C6 is as distant from the central set as CJA. This discrepancy is explained by the fact that the C6 document contains not only a PHP course but also reports of students' projects in more than half of the document. These reports discuss various business problems that require a website (online shoe store, online music store, etc). Therefore, the document is not exactly a pure PHP course like the others.

In order to test how the mutual impact graph reacts when a document is corrected, we compared multiple cases of courses while correcting one of them. Document C6 is written in three parts with nearly the same amount of pages: the regular PHP course, the reports of the students' projects, and an advanced PHP course. For the experiment, we used PHP courses in text format only (in order to avoid the effect of mixing slide and text formats), and we corrected document C6 by removing the students' projects first and, later, the advanced chapters. The experiment was also done twice, with and without the Java course, in order to have a better view of the effect of correction on a corpus with and without an irrelevant course.

In the regular case without any modification (Figure 8), C6 is the most outlier in both regular and Java cases. In the Java case, CJA is the sole document nearly as far as C6. After removing the students' projects from C6 (Figure 9), it becomes one of the closest documents from the central set in both cases, indicating that it became way more relevant to the corpus than previously. In the Java case, CJA becomes the most outlier, keeping its irrelevance. Finally, when the advanced chapters are also removed from C6 (Figure 10), it becomes the closest to the central set in the Java case, and the second closest in the regular case. C11 and C15 become the most outliers in the regular case. In the Java case, CJA and C11 are the most outliers.

In each figure from 8 to 10, document C6 becomes closer to the central set thanks to the corrections. It can be noted that CJA stays the most outlier in every case, and the other documents move away a bit. Therefore, the corrections show improvements in the positioning of C6 while keeping the irrelevance of CJA.

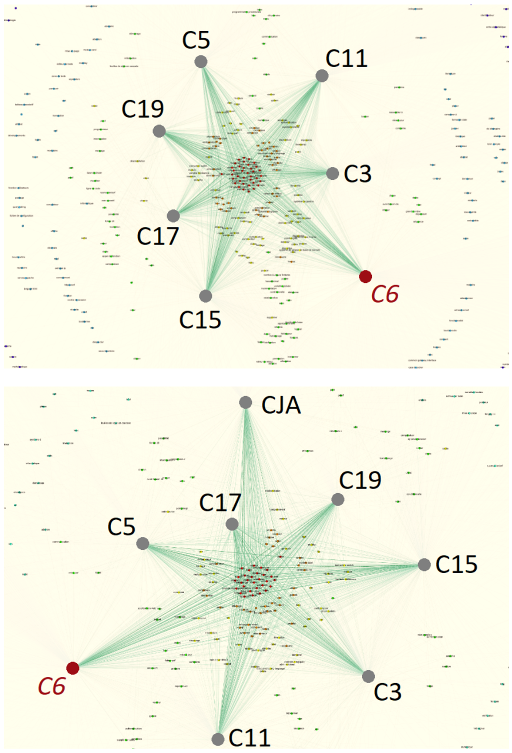


Figure 8: Relevance of text documents (C6 intact).

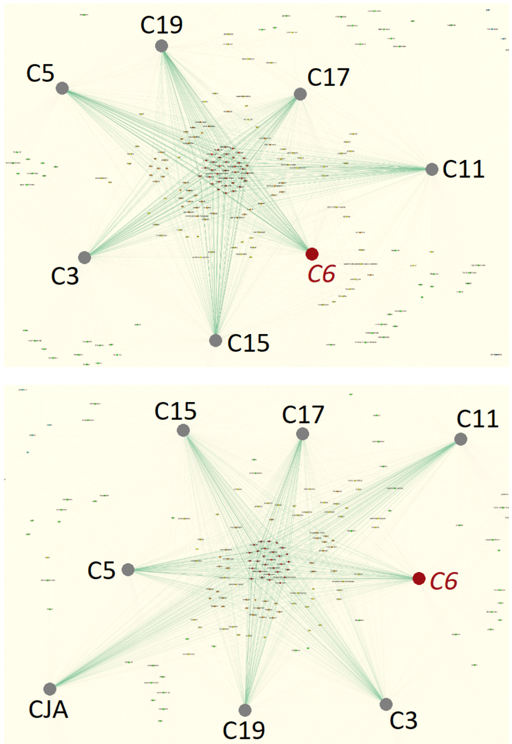


Figure 9: Relevance of text documents (no student projects).

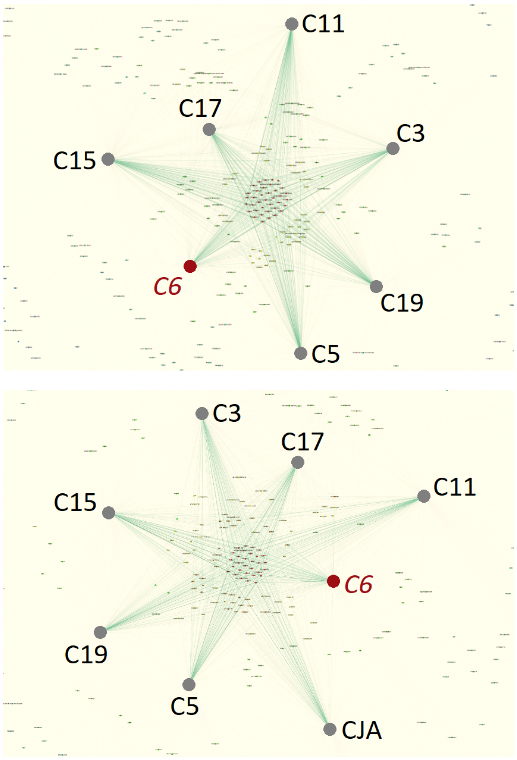


Figure 10: Relevance of text documents in the regular case (no student projects, nor advanced chapters).

5 DISCUSSION

In the mutual impact graph of the regular case (Figure 4), the set of terms present in all of the documents at the center of the graph, shown as red nodes, are relevant to the content of the text corpus, and they also produce a clear summary of the main keywords. Each ring of colors around the central set adds more terms relevant to the corpus. When the Java course is introduced (Figure 5), the terms in all the documents are nearly the same: only *php* is a bit repulsed. As the Java course also mainly talks about programming and partially about web development with dedicated frameworks, the results are nearly unchanged, which is the expected behavior. The mutual impact graph with the terms lets a user quickly get an idea of the subject and keywords composing the corpus. It can be used to quickly discover a new academic field and find the keywords that best describe it. Another usage for this graph would be to help build a book's index: the keywords are highlighted, and the author selects which words to keep or remove.

Concerning the relevance of documents in Figure 7, the Java course is the most distant one with C6 (which contains a lot of unwanted content) in the first experiment. This behavior is perfect for the current case: the teacher who would like to use existing documents is warned that C6 and CJA should be checked more precisely in order to detect if their content is relevant. Even if the text documents are distant from the slide documents, the most irrelevant ones are far away, allowing the user to measure the relevance visually. In the third experiment, the relevance of document C6 is greatly improved because of the corrections applied while keeping the irrelevant document CJA far away. It must be noted that the shape of the graph changes because it reflects the relevance of each document relating to the general relevancy of the whole corpus. The mutual impact graph is a picture of the corpus as a whole: it is not a graph about the relevance of one or some documents against one or some other documents. Multiple use cases can be derived from the mutual impact graph. The graph would help users select the best documents about one or more topics or remove the most irrelevant ones. Another usage would be for a teacher to compare its own course with the existing ones or even with research articles to check how close it is to the state of the art.

The global results show that FCA, with the mutual impact measure, can highlight a corpus' main terms and even show its documents' relevance. It does not create a list of topics nor calculate the probability of each term being included in a topic like LDA, but it does reflect the importance of each term for the whole corpus. This behavior is expected by the nature of FCA: it "*differs from statistical data analysis in that the emphasis is on recognizing and generalizing structural similarities, such as set inclusion relation from the data description, and not on mathematical manipulations of probability distributions*" (Carpineto and Romano, 2004). However, we acknowledge that BabelFy does participate actively in the process of topic modeling by recognizing the named entities and evaluating their relevance to each portion of the text. It must be noted that the construction of the formal context also filters some terms. FCA (formal context's construction + metrics calculus) and BabelFy must be used together, or at least, FCA with an entity linking tool. Another limitation concerns the fact that the documents used in the experiments were written in French and with the goal of teaching. In our opinion, language shouldn't be a problem as BabelFy is well trained for recognizing a lot of languages, but mixing different documents written with distinct goals or distinct formats might affect the quality of the resulting map in a similar way that can be seen in Figure 6 where the documents in the slide format are closer to the central set than the text ones.

6 CONCLUSION

In this paper, we proposed a visualization pipeline for textual corpora analysis based on FCA instead of the usual LDA for the topic modeling step. Mutual impact was used within FCA in order to produce a matrix for the force-based graph algorithm. The pipeline produces a map that can be used in two ways:

- The main keywords are placed by order of importance, allowing the reader to quickly get an idea of the topics contained in the corpus
- Documents are placed based on their relevance to the keywords found, allowing the reader to see an eventual discrepancy in the chosen texts.

The map presents a visualization of the corpus as a whole. Removing a document impacts the visualization because of the absence of a node and because the topic modeling step does not work on the same texts. To evaluate these claims, we presented a case study on multiple PHP courses and an intruder Java course (also about programming). First, a map displayed the most important keywords and the variations with and without the Java course. Then, one of the PHP courses containing more than half of its text about out-of-scope topics was corrected, showing a significant upgrade in the output.

We consider FCA an exciting method for topic modeling and expect to try other metrics on the lattice in order to find more possible usages. Multiple usages and combinations have been already proposed in (Poelmans et al., 2013), but we expect to use the conceptual similarity metric (Jaffal et al., 2015) for an even more precise combination of terms. Also, a deeper comparison with LDA and other newer methods like neural networks might be interesting as the construction of the results does not rely on probabilities and is perfectly transparent thanks to the set theory behind FCA. Numerous applications in digital humanities can be considered with FCA because transparency, and therefore explainability, are already part of its design by highlighting relationships between objects.

REFERENCES

Ahn, J.-w. and Brusilovsky, P. (2009). Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization*, 8(3):167–179.

Akhtar, N., Javed, H., and Ahmad, T. (2019). Hierarchical summarization of text documents using topic modeling and formal concept analysis. In *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2018, Volume 2*, pages 21–33. Springer.

Alghamdi, R. and Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1).

Andrienko, G., Andrienko, N., Drucker, S. M., Fekete, J.-D., Fisher, D., Idreos, S., Kraska, T., Li, G., Ma, K.-L., Mackinlay, J. D., et al. (2020). Big data visualization and analytics: Future research challenges and emerging applications. *BigVis 2020: Big data visual exploration and analytics*.

Assa, J., Cohen-Or, D., and Milo, T. (1997). Displaying data in multidimensional relevance space with 2d visualization maps. In *Proceedings. Visualization'97 (Cat. No. 97CB36155)*, pages 127–134. IEEE.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.

Belohlavek, R. (2008). Introduction to formal concept analysis. *Palacky University, Department of Computer Science, Olomouc*, 47.

Blei, D. and Lafferty, J. (2006). Correlated topic models. *Advances in neural information processing systems*, 18:147.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Carpineto, C. and Romano, G. (2004). *Concept data analysis: Theory and applications*. John Wiley & Sons.

De Sisto, M., Hernández-Lorenzo, L., De la Rosa, J., Ros, S., and González-Blanco, E. (2024). Understanding poetry using natural language processing tools: a survey. *Digital Scholarship in the Humanities*, page fqae001.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

di Sciascio, C., Mayr, L., and Veas, E. (2017). Exploring and summarizing document colletions with multiple coordinated views. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, pages 41–48.

Eisenstein, J., Chau, D. H., Kittur, A., and Xing, E. (2012). Topicviz: Interactive topic exploration in document collections. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 2177–2182.

Elliott, S. (2021). Proof of concept research. *Philosophy of Science*, 88(2):258–280.

Engebretsen, M. and Kennedy, H. (2020). *Data visualization in society*. Amsterdam university press.

Fortuna, B., Grobelnik, M., and Mladenic, D. (2005). Visualization of text document corpus. *Informatica*, 29(4).

Ganter, B. and Wille, R. (2012). *Formal concept analysis: mathematical foundations*. Springer Science & Business Media.

George, L. and Sumathy, P. (2023). An integrated clustering and bert framework for improved topic modeling. *International Journal of Information Technology*, 15(4):2187–2195.

Gretarsson, B., O'donovan, J., Bostandjiev, S., Höllerer, T., Asuncion, A., Newman, D., and Smyth, P. (2012). Topicnets: Visual analysis of large text corpora with topic modeling. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(2):1–26.

Hambarde, K. A. and Proenca, H. (2023). Information retrieval: recent advances and beyond. *IEEE Access*.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57.

Jaffal, A., Le Grand, B., and Kirsch-Pinheiro, M. (2015). Refinement strategies for correlating context and user behavior in pervasive information systems. *Procedia Computer Science*, 52:1040–1046.

Jänicke, S., Franzini, G., Cheema, M. F., and Scheuermann, G. (2015). On close and distant reading in digital humanities: A survey and future challenges. *EuroVis (STARs)*, 2015:83–103.

Kherwa, P. and Bansal, P. (2019). Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, 7(24).

Lee, B., Riche, N. H., Karlson, A. K., and Carpendale, S. (2010). Sparkclouds: Visualizing trends in tag clouds. *IEEE transactions on visualization and computer graphics*, 16(6):1182–1189.

Menhour, H., Şahin, H. B., Sarıkaya, R. N., Aktaş, M., Sağlam, R., Ekinci, E., and Eken, S. (2023). Searchable turkish ocred historical newspaper collection 1928–1942. *Journal of Information Science*, 49(2):335–347.

Moro, A., Raganato, A., and Navigli, R. (2014). Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

Navigli, R. and Ponzetto, S. P. (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Newman, D., Baldwin, T., Cavedon, L., Huang, E., Karimi, S., Martinez, D., Scholer, F., and Zobel, J. (2010). Visualizing search results and document collections using topic maps. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(2-3):169–175.

North, K. and Kumta, G. (2018). *Knowledge management: Value creation through organizational learning*. Springer.

Olsen, K. A., Korfhage, R. R., Sochats, K. M., Spring, M. B., and Williams, J. G. (1993). Visualization of a document collection: The vibe system. *Information Processing & Management*, 29(1):69–81.

Olsen, K. A., Williams, J. G., Sochats, K. M., and Hirtle, S. C. (1992). Ideation through visualization: the vibe system. *Multimedia Review*, 3:48–48.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., and Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. *Cognitive research: principles and implications*, 3:1–25.

Peinelt, N., Nguyen, D., and Liakata, M. (2020). tbert: Topic models and bert joining forces for semantic similarity detection. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 7047–7055.

Poelmans, J., Kuznetsov, S. O., Ignatov, D. I., and Dedene, G. (2013). Formal concept analysis in knowledge processing: A survey on models and techniques. *Expert systems with applications*, 40(16):6601–6623.

Salton, G. (1983). Introduction to modern information retrieval. *McGraw-Hill*.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *New methods in language processing*, page 154.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to german. In *In Proceedings of the ACL SIGDAT-Workshop*, pages 47–50.

Singh, J., Zerr, S., and Siersdorfer, S. (2017). Structure-aware visualization of text corpora. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 107–116.

Building Atlas of Knowledge Maps: Towards Smarter Collaboration

Anna Kuznetsova[1][a], Tatiana Gavrilova[2][b] and Olga Alkanova[2][c]
[1]*Lyceum No. 408, St. Petersburg, Russian Federation*
[2]*Graduate School of Management, St. Petersburg State University, St. Petersburg, Russian Federation*
anna.romantseva@gmail.com, gavrilova@gsom.spbu.ru

Keywords: Ontology, Knowledge Maps, Knowledge Atlas, Visualization System, Knowledge Transfer.

Abstract: The paper discusses the possibilities and prospects for creating corporate atlas of knowledge maps – a visual guide of diagrams describing the intellectual assets of the enterprise. The discussed case is based on the university business school. Mapping or visualization provides information transparency of communications in universities making collaboration smart and effective. The walls of universities are opaque, and visualization provide a higher level of teaching, research, consulting and administration. The paper presents the preliminary results of the project "Methodology and technology for developing digital knowledge maps for education and research teams'' and proposes and describes specific features of a systematic repository of diagrams, that is called an atlas of knowledge maps. We developed a set of diagrams to describe knowledge, created an ontology of the properties of such maps and suggested considering the most popular ones as a kind of atlas from which decision makers can select relevant maps for their work. The survey is preceded by the use of ontologies - conceptual models of areas of knowledge and professional activities of the teacher. In general, the approach can be adapted to business companies and government organizations if they are interested in disclosing their intellectual capital.

1 INTRODUCTION

Business and academic work require cooperation. Learning includes access to influencers and experts. It can be difficult to find colleagues and potential partners in an overloaded world of redundant and contradictory information.

But companies and universities are in no hurry to share their intellectual assets, and often companies themselves do not know about their "treasures". Acquisition and systematization of such information resources are useful primarily for the companies themselves, in addition, they are invaluable in the market. The paper discusses the possibilities and prospects for creating the atlas of corporate knowledge maps - visual guide to the intellectual assets of the enterprise based on the case of a university business school.

Visual knowledge maps are a powerful tool for enhancing understanding and fostering collaboration in a company setting. These maps can be used to visually represent information, ideas, and relationships in a clear and concise manner, making it easier for faculty and students to grasp complex concepts and share knowledge with their colleagues.

Visualization allows to present complex data and identify patterns, trends, and structures, which facilitates deeper exploration of the data. Diagrams allow all the employees and newcomers to expand less cognitive energy deciphering the meaning of the text they are reading, which means they will have more cognitive energy available for the critically important tasks of understanding, assessment and reflection (Miller 2023, Moody 2007). The main benefits of knowledge visualization are related to: stakeholder engagement, flexibility, knowledge transfer, signalling role, agility and interactivity (Troise, 2022). Using knowledge representation and mapping help to organize the smarter collaboration. The term was coined by H. Gardner (Gardner, 2017) when she described the need for highly-specialised experts to come together in order to tackle more complicated issues than any of them could do on their own.

[a] https://orcid.org/0000-0002-3612-6014
[b] https://orcid.org/0000-0003-1466-8100
[b] https://orcid.org/0000-0002-2530-6765

Kuznetsova, A., Gavrilova, T. and Alkanova, O.
Building Atlas of Knowledge Maps: Towards Smarter Collaboration.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 130-136
ISBN: 978-989-758-716-0; ISSN: 2184-3228

The paper discusses some preliminary results of the METAKARTA project (MEthodology and Technology for developing digital Knowledge mAps for education and Research TeAms) where we developed the methodology visualizing teaching and research activity of the faculty members.

The paper structure is as follows: the current section 1 provides the motivation for creating a new approach, section 2 presents a brief literature review and highlights the existing research gap, the atlas' attributive ontology design is described in section 3, while section 4 provides a demonstration of this approach in a decision-making process.

2 KNOWLEDGE MAPPING

Knowledge maps are powerful information visualization techniques that allow describing knowledge assets, connecting experts, accessing knowledge over time, existing knowledge resources and knowledge gaps (Faisal et al., 2019). The main tools that are widely used in knowledge mapping, require the participation of both experts and analysts who develop visual diagrams reflecting

- Sources of knowledge;
- Location of knowledge elements;
- Owners of knowledge elements;
- Links and relations between them, etc.

Knowledge maps are closely related to competency maps and employee competency management, which are denoted as skills and competencies in corporate decisions (Anthony, 2021). Such maps turn enterprise data into valuable and insightful information. Knowledge maps are one of the tools used in knowledge engineering for organizing and presenting knowledge, forming a graphical framework and landscape in visualizing complex concepts, decision support, knowledge sharing, etc. (Balaid et al., 2016).

However little attention is paid to the development of a well-structured set of visual representations of key concepts, relationships, knowledge owners of a knowledge domain of the organization encouraging the employees to see the big picture, promote collaboration, and improve organization and focus.

The development of knowledge maps starts with the definition of goals and stakeholders. For each level, a basic atlas (visual set) of types of knowledge maps was created.

2.1 Definition of Goals

In the field of management, the following goals may be solved using the developed knowledge maps:

- optimization and activation of resources, including the formation of project teams or working groups taking into account the principle of complementarity, ensuring the transfer of knowledge from experts to employees who have gaps (Liebowitz, 2005) (in this case, an employee development plan is formed based on such tools as coaching and mentoring) and strategic planning for the development of assets (Zack, McKeen, Singh, 2009) (based on the analysis of the map for various areas of knowledge, a decision is made to close gaps or change the focus of activity);
- identification of the hidden potential of employees. The principle of completeness, implemented in the construction of ontologies of subject areas, provides a comprehensive analysis and allows for the formalization of those areas of knowledge that were previously not in the field of view when assessing employees. By discovering previously unknown competencies and publications of subordinates, a manager can make a more informed (and therefore less risky) decision about developing new areas of activity (Butt et al. 2021).

2.2 A Stakeholder Analysis

Before the knowledge mapping study, a stakeholder analysis was conducted. Stakeholders who influence academic and research teams and benefit in one way or another from access to the knowledge map data may include both external and internal users and can be divided into three categories: managers (administrators), experts, and ordinary employees, including newcomers (Pereira et al., 2023). The METAKARTA project expanded the traditional classification and identified another category: external experts. In modern universities, the roles described above are represented by internal stakeholders: administration (managers), research and teaching staff, including young scientists and postgraduates (experts and ordinary employees). Based on the fundamental differences between these three groups of knowledge map recipients, a classification was proposed at three levels: general, focused, and detailed, as described in previous publications. These three levels in the described case correspond to:

- institution level,
- department level,
- individual level.

The next two figures illustrate school and department levels for shaping the research activity my mapping the bibliometric data extracted from Google Scholar.

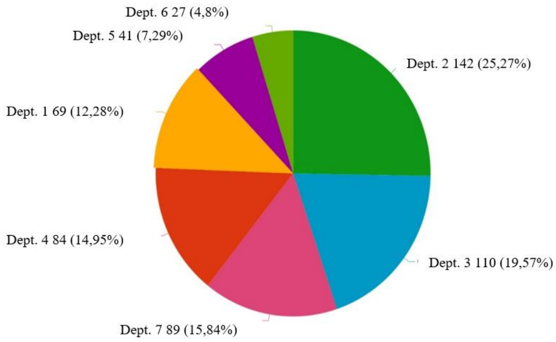


Figure 1: Distribution of all the publications among the school departments.

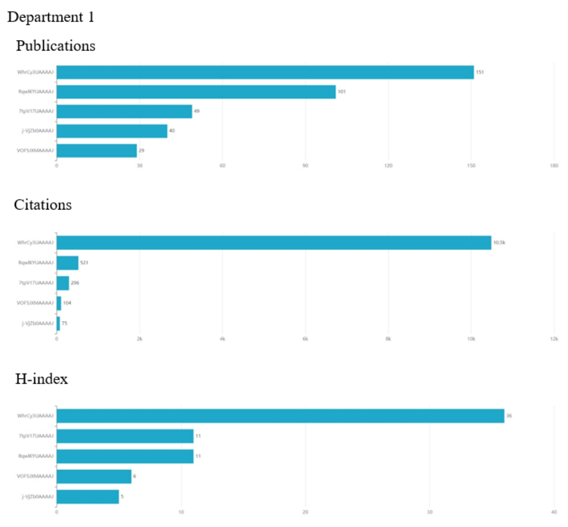


Figure 2: Portrait of department X.

Figure 1 shows the general level distribution of all the publications listed in the database among the university school departments. Here the information on the percentage of the total amount of publications of the university school departments is stated. That helps to evaluate the more and less active departments in terms of publications.

Figure 2 shows focused and detailed levels of generalization describing a portrait of the faculty from department X and gives the information on the number of publications for each of the faculty members, their H-index, the number of citations of each of the teacher.

3 DATA COLLECTION METHODOLOGY

The METAKARTA Project results include visual representations of two information landscapes – teaching and research. The essential part of the project was devoted to data collection.

Bibliometric data was acquired from Google Scholar, while teaching information needs two data processing stages. Two surveys with self-assessment questions were conducted where the faculty assess their teaching competences and expertise. Initially a set of secondary data was used from 2019-2020 larger project initiated as part of the internal self-assessment of the targeted school full-time faculty.

Organization of the data was as follows: each respondent answered a series of binary questions whether they consider themselves as an expert in a particular area of knowledge from the predefined set of ontologies. In case of a positive answer for a particular area a set of questions regarding teaching, research and applied consulting experience followed. Consequently, the dataset was organized in a "matrix" logic – the assessment of experience in each type of activity was carried out for each area of knowledge noted by the employee.

The analysis of results of the first survey helps to prepare the second updated one.

Data for the new questionnaire were collected in the middle of the 2023/2024 academic year from the current full-time faculty of the same school as for the first data collection (all full-time faculty members who teach at least 1 course per year were surveyed). Total sample size was 56 qualified faculty members. The retention rate of the full-time staff between the two datasets was 68.3%, which, provided that the data from the first and second surveys are brought to a single coding, makes it possible to build not only maps reflecting the development of employees, but also maps of changes by departments – a new set of maps that show the dynamics of the internal knowledge.

As a result of the two datasets comparison, we found out that:

Time spent for questionnaire fulfillment decreased by around 30-50% (depending on the number of areas of expertise – the effect was higher for employees with more areas of expertise.

The average number of knowledge areas reported by employees as areas of expertise increased from 2.5 to 3.5, thus providing a more detailed picture of knowledge in specific scientific areas (assessed across employees who participated in two data collections). A random check for deviations showed that, overall, additional areas of expertise were supported by objective experience existing already by the time of the first data collection, which proves increased data accuracy when a sequential approach is used.

4 AN ATLAS' ATTRIBUTIVE ONTOLOGY DESIGN

Based on the extensive experience in working with questionnaires in social sciences (Aithal & Aithal, 2020) a combination of primary and secondary data was used to the maximum extent possible while the development of knowledge maps for the atlas. It was found that when building knowledge maps based on primary information, there is a risk of obtaining unreliable data. In terms of completeness of information, the optimal solution is to combine primary and secondary data to build knowledge maps. In this case, different implementation scenarios are possible:

- from secondary data to primary (building an employee profile based on secondary data, then verifying this profile by the employee as part of collecting primary data with the ability to additionally collect self-assessment data);
- from primary data to secondary (primary data are verified based on available secondary data - the expertise declared by the employee is confirmed based on available objective information);
- independent assessment of the employee profile - building individual maps based on subjective assessment and secondary data with subsequent generalization on the expert profile dashboard.

Atlas systematizes the significant properties of knowledge maps bringing the connections among them. We borrowed the term from classical definition: Atlas is "a bound collection of maps often including illustrations, informative tables, or textual matter" (Merriam-Webster).

When creating an atlas of knowledge maps that describes modern diagram templates and recommendations for their use, work was carried out to generalize and structurally describe the existing diagrams. Information design in knowledge maps aims to avoid confusion by presenting data in a way that's easy to understand. Based on the study of researchers Lenger and Eppler who compiled a table similar to the periodic table, consisting of more than 100 different visualization techniques, divided by type of use (Lengler, Eppler, 2007) we include more than 20 visual diagrams into the atlas. Also atlas systematizes the recommendations for their use, it describes modern diagram templates and structurally summarize describe the existing diagrams in a form of a table.

The conceptual representation of atlas may be defined as an attributive ontology (Fig.3).

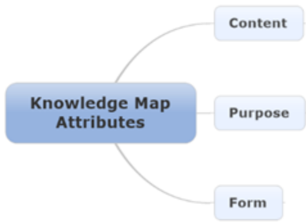


Figure 3: Structure of attriBUTive OnTology of kNowledge maps: upper level.

The BUTTON Ontology (attriBUTive OnTology of kNowledge maps) is a generalization and systematic description of various characteristics (attributes) of knowledge maps used to visualize the information landscape of companies and universities. This ontology summarizes many characteristics of knowledge maps into three categories:

- content;
- format; and
- purpose of map.

The format of this conference paper does not allow to show all the BUTTON ontology framework.

Using of the developed atlas create an additional advantage for all its users and the project stakeholders.

5 KNOWLEDGE ATLAS

The atlas of the knowledge maps presents systemic vision of possible diagrams that scaffolds the understanding of university intellectual assets from a range of perspectives. The paper tries to provide comprehensive insight into the ways in which university and faculty members visualize their bibliometric and teaching intellectual assets.

5.1 Classification Based on Content of Knowledge Maps

The content of knowledge maps plays a key role in determining their effectiveness and applicability in different contexts. It includes two main aspects: the carrier and the elements of knowledge.

In the context of an educational institution, a carrier (teacher) is an entity responsible for the accumulation, transfer and acquisition of knowledge. The main feature of the attribute "carrier" is its "potential". The potential of the carrier (teacher) reflects his or her cumulative knowledge, skills and experience in certain areas and includes the depth and breadth of the teacher's expertise. The dynamics of the carrier reflect the changes in his potential over time: the teacher's self-improvement, his participation in professional training and education, as well as the continuous updating of knowledge. Relationships between carriers are a network of interactions, exchange of knowledge and experience. This includes various forms of cooperation, such as the exchange of educational materials, joint research projects, etc.

In the context of an educational institution, a carrier (teacher) is an entity responsible for the accumulation, transfer and acquisition of knowledge. The main feature of the attribute "carrier" is its "potential". The potential of the carrier (teacher) reflects his or her cumulative knowledge, skills and experience in certain areas and includes the depth and breadth of the teacher's expertise.

The knowledge elements on the map include specific learning materials and information elements belonging to the carriers. Elements can be organized into different structures, have priorities, locations, and formats. By the structure of knowledge elements, we understand ways of classifying, organizing and linking individual elements of knowledge to ensure their accessibility and understanding. The structure helps to navigate in the set of knowledge, understand their interrelations and find the necessary information. Knowledge elements are prioritized subjectively by managers and reflect their understanding of the importance and relevance of knowledge components in the context of a particular area or task. Prioritization allows you to identify aspects that should be paid attention to when planning training programs, courses and human resources. The location of knowledge elements includes the geographical location of the teacher (for example, in a branch of the university), the academic unit (department, faculty) and the program.

5.2 Classification Based on Format of Knowledge Maps

The shape of knowledge maps is an important aspect of their visual representation, determining the way information is displayed. For knowledge maps, we have considered graphs, tables, charts, as well as metaphor drawings as shown in Figure 4.

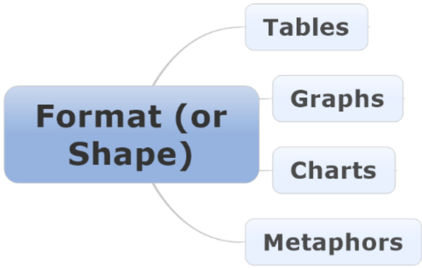


Figure 4: Possible Formats of knowledge maps in the atlas

The atlas is designed in the form of the table with a description of the difficulty level of the diagram, the preview of the pictogram, its design and the main characteristics and purpose. The maps include four major types as shown in Figure 4:

- Tables;
- Graphs;
- Charts;
- Metaphors.

Tables include one-level, multilevel and nested tables as shown in Figure 5.

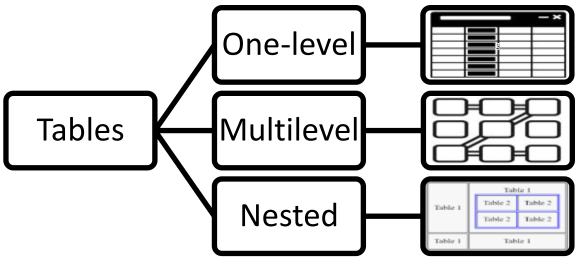


Figure 5: Types of tables in the atlas.

Heat map tables serve graphical representation of data using color and size to encode text tables for easier comparison of data values.

Field of knowledge	Employee 13	Employee 19	Employee 3	Employee 30	Employee 35	Employee 48	Employee 63	Employee 8	Total
B. International Management and International Business							3	4	7
C. Information Systems and Information Management	4	2	3	4	7	3	7	5	35
D. Entrepreneurship and Innovation			1					5	6
E. Marketing and Sales					1		2		3
F. Organizational Behavior and Human Resource Management								1	1
G. Economics, Ecosystems, Markets, and Institutions								4	4
H. Data Analysis and Decision Making Methods				1	3		2		6
I. Urban and Regional Management and Planning				2				1	3
J. Strategic Management and Organizational Development			5	3				4	12
K. Operations Management and Project Management					4	1			5
L. Interdisciplinary and Other Fields of Knowledge	5	1	6	3				4	19
Total	9	3	15	13	15	4	14	28	101

Figure 6: A heat map table: overview of one department's knowledge.

The heat map table in Figure 6 presents an overview of the level of expertise of professors in one of the departments of the university based business school. The collected data in the table presents the coded names of professors (columns) and the corresponding fields of knowledge (rows).

The capacity of the level of expertise of each professor is determined by the sum of spheres of competence within multiple fields of knowledge, which are presented in the corresponding cells.

The resulting table provides an efficient and easy-to-read presentation of the level of expertise of each professor according to their self-assessment.

The table can be used to identify areas of strengths and weaknesses among professors, and to allocate resources more effectively based on each professor's expertise. Further research could explore the relationship between the level of expertise of professors and the quality of their teaching and research outcomes.

There are two types of graphs as shown in Figure 7:

- Hierarchical
- Network undirected.
-

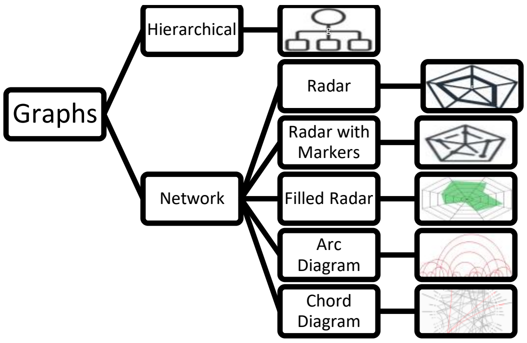


Figure 7: Types of graphs in the atlas.

Hierarchical tree is a graphical representation of hierarchically organized data in the form of a tree.
Network maps include five types:

- Radar - a graphical representation of data in the form of petals, typically used to compare different categories or aspects.
- Radar with Markers - a radar chart where in addition to the petals, markers are also displayed to indicate values.
- Filled Radar – a radar chart in which the areas between the petals are filled with color for better visualization.
- Arc diagram is useful to reveal the overlap of data.
- Chord diagram reveals the relationship between the objects inside the organization.

Charts include one-, two- and three-dimensional ones. Metaphors include images that are used as universal metaphors to visually organize the information (physical landscape, pyramid, fishbone, etc.)

5.3 Classification Based on Purpose

The characteristics of a purpose include

- purpose itself,
- focus,
- stakeholders,
- level of generalization.

The purpose of knowledge maps plays a key role in their creation and can be considered in different contexts, e.g. - decision-making / market positioning / raising general awareness within the company.

"Focus" refers to the main focus of using a knowledge map. Within the framework of the study of the experience of teachers in three areas, the following types of focus can be distinguished: academic work / research / consulting and projects.

Stakeholders in our study include external and internal users. External ones include applicants, business partners and customers who can use knowledge maps to obtain information about the educational institution, projects and employees. Internal users are administration, teachers, researchers and students.

The level of generalization of knowledge maps, which is determined by the purpose and task of mapping, is also important. It can be either universal or specialized. Universal knowledge maps are applicable in various fields of knowledge and disciplines, showing the general picture of what is happening (for example, a faculty science citation map). Specialized knowledge maps can be focused on specific areas or levels, such as faculty, graduate school, department, or individual faculty.

6 CONCLUSION

The information space of organizations is overloaded, there is a need to find convenient assistants that facilitate the processing of information for users. The most difficult and labour-intensive part of working with information is associated with its search, structuring, and compression. The visual approach is one of the possible ways to scaffold the information flow.

The paper discusses the developing of the prototype of an atlas of knowledge maps describing the intellectual assets of the university business school. This prototype of the atlas includes invariant representations of knowledge maps of educational, scientific and consulting activities, depending on the stakeholder, the task itself, the selected level of generalization for mapping (institute/department/individual).

One of the key benefits of using this atlas is that the maps from it help students, faculty and administration see the big picture of the academic life. By mapping out key concepts and their interconnections, employees can gain a better understanding of how different pieces of information fit together, who are the experts and how they contribute to the overall goals of the university unit in teaching and research. This can help employees make more informed decisions and work more effectively towards shared objectives and smarter collaboration.

Systematic analysis of corporate, administrative and scientific knowledge creates the potential to significantly improve the quality of information support, creating knowledge management systems for more effective interaction between various groups of organization employees and external users and stakeholders.

Ultimately, using visual knowledge maps can lead to smarter decision-making, more innovative solutions, and a more efficient and effective company. It is a step to visual organization (Bell, Warren, & Schroeder, 2014).

ACKNOWLEDGEMENTS

Authors thank Dr. Dmitry Kudryavtsev, who was the initiator of the project and started the research design, and Miroslav Kubelsky for the digital support of bibliometric maps. The work was carried out by Gavrilova T.A., Alkanova O.N. as part of METAKARTA PROJECT grant No. 23-21-00168, https://rscf.ru/project/23-21-00168/.

REFERENCES

Aithal, A., Aithal, P. (2020). Development and Validation of Survey Questionnaire & Experimental Data—A Systematical Review-Based Statistical Approach. International Journal of Management, Technology, and Social Sciences (IJMTS), 5, 233-251.

Anthony, Jr, B. (2021). Information flow analysis of a knowledge mapping-based system for university alumni collaboration: a practical approach. Journal of the Knowledge Economy, 12(2), 756-787.

Balaid, A., Abd Rozan, M.Z., Hikmi, S.N., and Memon, J. (2016). Knowledge Maps: A Systematic Literature Review and Directions for Future Research, International Journal of Information Management, Vol 36, No. 3, 451–475.

Bell, E., Warren, S., and Schroeder, J. (2014). Introduction: The visual organization. In The Routledge companion to visual organization (pp. 1-16). Routledge.

Butt, S. A. (2021). et al. A knowledge map for ICT integration in the silver economy. Procedia Computer Science, (181), 93-701.

Faisal, H., Rahman, A., and Zaman, G. (2019). Knowledge Mapping for Research Papers. International Journal of Computer Science and Network Security, 19 (10), 158–164.

Gardner, H. (2017). Smart Collaboration: How Professionals and Their Firms Succeed by Breaking Down Silos. Harvard Business Review Press.

Lengler, R., Eppler, M.J. (2007). A Periodic Table of Visualization Methods. URL: https://www.visual-literacy.org/periodic_table/periodic_table.html

Lengler, R., Eppler, M.J. (2007). Towards a periodic table of visualization methods of management.

Liebowitz, J. (2005). Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. J.of knowledge management,9,76–86.

Merriam-webster dictionary (https://www.merriam-webster.com/dictionary/atlas#:)

Miller, F. The power of Diagrams. https://www.francismiller.com/articles/

Moody, D. (2007). What makes a good diagram? improving the cognitive effectiveness of diagrams in is development. In Advances in Inform. Systems Development: New Methods for the Networked Society. Boston, MA: Springer US. 481-492.

Pereira, V., Bamel, U., Temouri, Y., Budhwar, P., Del Giudice, M. (2023). Mapping the evolution, current state of affairs and future research direction of managing cross-border knowledge for innovation. Intern. Business Review. – Vol. 32, No. 2, 101834.

Troise, C. (2022). Exploring knowledge visualization in the digital age: an analysis of benefits and risks. Management Decision, Vol. 60 No. 4, pp. 1116-1131. https://doi.org/10.1108/MD-01-2021-0086

Zack, M., McKeen, J., and Singh, S. (2009). Knowledge management and organizational performance: An exploratory analysis. Journal of Knowledge Management, 13(6), 392–409. https://doi.org/10.1108/13673270910997088

Benchmarking of Retrieval Augmented Generation: A Comprehensive Systematic Literature Review on Evaluation Dimensions, Evaluation Metrics and Datasets

Simon Knollmeyer[1,*][a], Oğuz Caymazer[2,*][b], Leonid Koval[1][c], Muhammad Uzair Akmal[1][d], Saara Asif[1][e], Selvine G. Mathias[1][f] and Daniel Großmann[1][g]

[1]*Technische Hochschule Ingolstadt, AImotion Bavaria, Esplanade 10, Ingolstadt, Germany*

[2]*University of Münster, Department of Information Systems, Münster, Germany*

{Simon.Knollmeyer, Leonid.Koval, MuhammadUzair.Akmal, Saara.Asif, SelvineGeorge.Mathias, Daniel.Grossmann}@thi.de, o_caym01@uni-muenster.de

Keywords: Large Language Model, Retrieval Augmented Generation, Evaluation Dimensions, Evaluation Metrics, Datasets, Systematic Literature Review.

Abstract: Despite the rapid advancements in the field of Large Language Models (LLM), traditional benchmarks have proven to be inadequate for assessing the performance of Retrieval Augmented Generation (RAG) systems. Therefore, this paper presents a comprehensive systematic literature review of evaluation dimensions, metrics, and datasets for RAG systems. This review identifies key evaluation dimensions such as context relevance, faithfulness, answer relevance, correctness, and citation quality. For each evaluation dimension, several metrics and evaluators are proposed on how to assess them. This paper synthesizes the findings from 12 relevant papers and presents a concept matrix that categorizes each evaluation approach. The results provide a foundation for the development of robust evaluation frameworks and suitable datasets that are essential for the effective implementation and deployment of RAG systems in real-world applications.

1 INTRODUCTION

The rapid evolution of Artificial Intelligence (AI) especially in the field of Large Language Models (LLMs) attracts widespread attention due to their groundbreaking achievements in solving complex problems even surpassing the performance of humans in certain fields (Benbya et al., 2024; Bubeck et al., 2023; OpenAI et al., 2023). The speed of AI development in the area of LLMs outpaced methods to assess their performance and accuracy, leading to major flaws in existing traditional benchmarks to evaluate LLMs output through reliable metrics impeding their adoption (Hammond, 2024).

Despite their potential, LLMs face substantial challenges, particularly in fully grasping contextual factors such as unique technical requirements within a specific industries, yet understanding these factors is essential for effective decision-making (Benbya et al., 2024). Even the most powerful models such as GPT-4 struggle with hallucinations, lack of the ability to update itself, and limited context (Bubeck et al., 2023; OpenAI et al., 2023). Several researchers point out that these LLMs seem to rather memorize frequently occurring information encountered during their pre-training and struggle with infrequent information, i.e., which would typically occur in a specific industry (Kandpal et al., 2022; Mallen et al., 2022).

[a] https://orcid.org/0009-0002-1429-6992
[b] https://orcid.org/0009-0003-4096-3784
[c] https://orcid.org/0000-0003-4845-6579
[d] https://orcid.org/0009-0007-3961-1174
[e] https://orcid.org/0009-0006-1284-5635
[f] https://orcid.org/0000-0002-6549-0763
[g] https://orcid.org/0000-0002-7388-5757
* co-authors of the paper, contributed equally

Knollmeyer, S., Caymazer, O., Koval, L., Akmal, M., Asif, S., Mathias, S. and Großmann, D.
Benchmarking of Retrieval Augmented Generation: A Comprehensive Systematic Literature Review on Evaluation Dimensions, Evaluation Metrics and Datasets.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 137-148
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Most promising and common to solve this problem is to augment LLM with non-parametric less common knowledge by providing them retrieved text chunks from an external database (Asai et al., 2024; Y. Gao et al., 2023; Mallen et al., 2022; Wang et al., 2023; Zhang et al., 2023). This approach is known as Retrieval Augmented Generation (RAG), and there are several different RAG paradigms, ranging from so-called naïve RAG to more advanced ones (Asai et al., 2023; Y. Gao et al., 2023; Ma et al., 2023). Research results suggest that LLM RAGs outperform LLMs, particularly in long-tail knowledge questions (Asai et al., 2023; Izacard et al., 2022; Ma et al., 2023; Mallen et al., 2022; Wang et al., 2023).

However, there is still uncertainty about the accuracy of such approaches due to the lack of comprehensive evaluation frameworks to provide evaluation dimensions and metrics to assess RAG LLMs (Y. Gao et al., 2023; Wang et al., 2023). Thus, we address the following research questions (RQ):

- RQ1: How to evaluate a RAG-enhanced LLM comprehensively across different dimensions and metrics?
- RQ2: What type of datasets are available for applying the dimensions and metrics?

Therefore, the research contribution of this paper lies in addressing the current research gap by providing a systematic overview of how to evaluate RAG pipelines comprehensively, offering insights into the development of robust evaluation dimensions, metrics and possible datasets (Y. Gao et al., 2023; Hammond, 2024; Wang et al., 2023).

The paper is structured as follows: The next section introduces RAG. Section three explains the chosen research method. Section four presents the evaluation framework. Finally, the last section provides the conclusions.

2 RETRIEVAL AUGMENTED GENERATION

Traditional pre-trained LLMs such as GPT and BERT encode knowledge within their parameters, but struggle with tasks requiring specific factual knowledge which is not present in their parameters (Y. Gao et al., 2023; Lewis et al., 2020). This problem is evident in the fact that even the most powerful models such as GPT-4 struggle with made-up facts known as hallucinations, a lack of ability to self-update and limited context (Bubeck et al., 2023; OpenAI et al., 2023). Even further increasing the size of their parameters, i.e., the training dataset, in which the knowledge appears to be stored to include more information will be likely insufficient to address the issue of long-tail knowledge (Kandpal et al., 2022; Mallen et al., 2022).

Therefore, Lewis et al. (2020) proposed RAG by combining non-parametric memory with parametric memory that uses a dense vector index of external documents such as Wikipedia articles that can be dynamically accessed using a retriever. Research results comparing the performance of RAG LLM with standalone LLM, suggest for the former superior performance (Asai et al., 2023; Izacard et al., 2022; Lewis et al., 2020; Ma et al., 2023; Mallen et al., 2022; Wang et al., 2023).

A naïve RAG pipeline involves three main steps. Firstly, the documents containing specific information are indexed (Lewis et al., 2020). The most common method is to split the documents into smaller sections so-called chunks and store their embedding in a vector database (Y. Gao et al., 2023). In the second step, a given input query is likewise embedded and then compared with the passages in the vector database by calculating the similarity, returning a set of top-ranked chunks that are most relevant for the query (Y. Gao et al., 2023; Karpukhin et al., 2020). In the final step, the retrieved content and the query are combined and prompted into an LLM so that it can provide a coherent answer (Y. Gao et al., 2023; Lewis et al., 2020).

This naïve setup can be modified by applying different advanced methods relating to pre- or post-retrieval (Asai et al., 2023; Y. Gao et al., 2023; Ma et al., 2023). For instance, Ma et al. (2023) propose query rewriting as an advanced pre-retrieval method and report performance improvements.

Despite the use of advanced methods, the RAG approach can still be divided into the outlined steps of a naïve RAG for evaluation. However, there is a lack of evaluation dimensions and metrics on how to analyse and assess such systems, e.g., which evaluation dimensions to consider and what kind of metrics to calculate for which step (Y. Gao et al., 2023).

3 RESEARCH METHOD

The Systematic Literature Review (SLR) is a well-known and established research method within Information Systems (IS) research for reviewing scientific articles based on a search process (Bell et al., 2019; Paré et al., 2016). The term "systematic" means that the research steps should be understandable, reproducible, and grounded in a structured process that minimizes potential researcher bias by providing a clear audit trail for decisions and conclusions (Bell et al., 2019). SLR is especially valuable when summarizing and comparing fragmented knowledge on a certain topic (Bell et al., 2019). Existing literature reveals a significant gap in comprehensive evaluation frameworks for assessing RAG systems (Y. Gao et al., 2023; Wang et al., 2023). Therefore, given the emerging and unexplored research on evaluation dimensions for RAG, the SLR is particularly appropriate.

This paper ensures rigorous documentation by using the literature search process proposed by Brocke et al. (2009), extended with the recommendations of Paré et al. (2016). It also follows the recommendation of Webster and Watson (2002) to use a concept matrix for structuring and comparing the results. The complete adopted literature search process is illustrated in Figure 1.

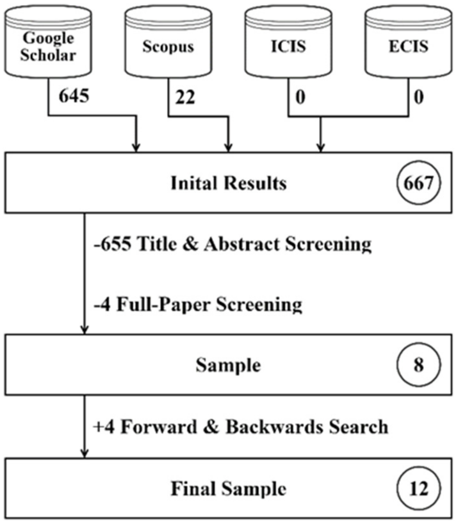


Figure 1: Systematic Literature Review process.

The search process included conducting a pre-study on Google Scholar by a detailed screening of titles, abstracts and full texts to identify papers specifically addressing evaluation metrics for RAG systems (Y. Gao et al., 2023; Wang et al., 2023). These insights resulted in the following search string:

- TITLE-ABS-KEY (“Retrieval Augmented Generation” AND “Evaluation Metric”)

The following inclusion and exclusion criteria were applied to filter relevant papers from the search results in Google Scholar, Scopus and the IS conferences ICIS and ECIS:

Inclusion Criteria:

- Papers from academic journals, conferences, or gray literature.
- Published in English.

Exclusion Criteria:

- Duplicates across databases.
- Minimal relevance to evaluation metrics, or lack of focus on RAG systems.
- No guidelines on metric implementation or application at the different RAG steps.

The search process commenced with an abstract screening of the initial results, followed by a full-text review of the selected papers. Forward and backward citation searches were subsequently performed on relevant studies to identify additional literature. This comprehensive approach yielded a final sample of 12 papers, as illustrated in Figure 1.

In the final step, overarching categories from the final sample were synthesized into a concept matrix (Webster & Watson, 2002). Relevant concepts on evaluation dimensions were identified and mapped to the RAG steps (cf. Section 2) to provide an accurate overview on how to evaluate each RAG phase.

The concept matrix is shown in Table 1. It categorizes the sampled papers based on predictive evaluation criteria and dataset characteristics. The former includes the columns "Retrieval" and "Generation" relating to the RAG steps. The "Evaluator" column indicates whether lexical matching, semantic similarity or LLM as a judge is used for evaluation. The dataset characteristics include single-hop and multi-hop reasoning tasks, synthetic datasets (triples), and open-domain question answering (QA). Each paper is marked ("X") to show the criteria it addresses, providing a comprehensive overview of their focus areas. In addition, the frequency of occurrence in the literature can be used to determine how widespread or accepted a metric is. Each proposed metric for the evaluation dimensions depending on the evaluator and the requirements for the data to calculate it are summarized in Table 2.

4 RESULTS

This section starts with examining the first column of the concept matrix: **predictive evaluation**, which involves assessing the performance of the RAG system in retrieving accurate context and effectively utilizing it to generate responses (Guinet et al., 2024).

Table 1: Concept Matrix with Evaluation Dimensions and Datasets.

Reference	Predictive Evaluation								Dataset			
	Retrieval	Generation				Evaluator			Retrieval		Generation	
	Context Relevance	Faithfulness	Answer Relevance	Correctness	Citation Quality	Lexical Matching	Semantic Similarity	LLM as a Judge	Single-Hop	Multi-Hop	Synthetic Dataset (Triple)	Open-Domain QA
(Adlakha et al., 2023)		X		X		X	X	X	X	X		X
(Es et al., 2023)	X	X	X					X	X	(X)	X	
(T. Gao et al., 2023)				X	X	X			X	X		X
(Guinet et al., 2024)				X		X			X		X	
(Hu et al., 2024)		X				X	X	X	X	X		X
(Min et al., 2023)		X						X	X			(X)
(Rackauckas et al., 2024)		X				X	X	(X)	X	X		X
(Rau et al., 2024)						X	X	X	X	X		X
(Ravi et al., 2024)		X						X	X			X
(Saad-Falcon et al., 2023)	X	X	X					X	X		X	
(Yu et al., 2024)	X	X	X	X		X			X			
(Zhang et al., 2024)	X			X		X			X			

Table 2: Summary of proposed Evaluation Dimensions, the corresponding Metrics and Dataset Requirements.

Evaluation Dimension	**Definition**	**Evaluator**	**Evaluation Metric**	**Dataset Requirement**	**Reference**
Context Relevance	Relates to the retrieval step and measures the extent to which the retrieved context contains only the information required to answer the query and as little irrelevant information as possible.	Lexical Matching	$\text{Recall@k} = \frac{\lvert\text{Relevant Passages} \cap \text{k-Passages}\rvert}{\lvert\text{Relevant Passages}\rvert}$	Context, Reference Retrieval (Golden-Retrieval)	(Yu et al., 2024)
			$\text{MRR@k} = \frac{1}{\lvert Q\rvert}\sum_{i=1}^{\lvert Q\rvert}\frac{1}{\text{rank}_i}$		(Rackauckas et al., 2024)
		LLM as a judge	$\text{CRS} = \frac{\text{Number of extracted sentences } S_{ext}}{\text{Total Number of Sentences in } c(q)}$	Query-Context	(Es et al., 2023)
Faithfulness	Refers to the generation step and measures the degree to which the LLM response is grounded in the retrieved context.	Lexical Matching	$\text{K-Precision} = \frac{\text{Matched Tokens}}{\text{Response Tokens}}$	Context-Response	(Adlakha et al., 2023)
		LLM as a judge	$\text{FC} = \frac{\sum_{i=1}^{N} \text{Faithfulness Score}_i}{\text{Total Number of Responses}}$		(Adlakha et al., 2023; Hu et al., 2024; Min et al., 2023; Ravi et al., 2024)
			$\text{FS} = \frac{\lvert V\rvert}{\lvert S\rvert}$	Triple (Query-Context-Response)	(Es et al., 2023)
Answer Relevance	Assesses whether the LLM response is directly addressing the actual query.	LLM as a judge	$\text{ARS} = \frac{1}{n}\sum_{i=1}^{n}\text{sim}(q, q_i)$	Query-Response	(Es et al., 2023; Saad-Falcon et al., 2023)
Correctness	Refers to the generation step and evaluates whether the LLM response answers the query accurately by matching with the expected answer, often referred to as "golden-passage" provided by human annotators.	Lexical Matching	$\text{Recall} = \frac{\text{Relevant Tokens in Response}}{\text{Total Relevant Tokens in Gold-Passage}}$	Reference Answer (Golden-Passage), Response	(Adlakha et al., 2023)
		LLM as a judge	$\text{CC} = \frac{\sum_{i=1}^{N} \text{Correctness Score}_i}{\text{Total Number of Responses}}$	Query, Reference Answer, Response	(Adlakha et al., 2023; Rackauckas et al., 2024)
Citation Quality	focuses on assessing whether an LLM correctly cites its sources when generating text.	LLM as a judge	Citation recall and citation precision calculated by the judging LLM	Context-Response	(T. Gao et al., 2023)

The proposed evaluation pipeline outlines (cf. Figure 2) the process of assessing the RAG approach across various evaluation dimensions and focuses on the retrieval and generation stages of a typical RAG system. The evaluation process starts with the **retrieval step**, emphasizing **context relevance** as a critical dimension to assess how effectively relevant information is retrieved.

Subsequently, the focus then shifts to the **generation step**, examining the evaluation dimensions of **answer relevance, correctness, faithfulness,** and **citation quality** to determine the accuracy and reliability of the generated responses. Each evaluation dimension is carefully defined, and quantifying metrics are proposed according to the sampled papers.

How and what type of metric is ultimately used to calculate the respective evaluation dimension depends on the chosen **evaluator**. **Lexical matching** metrics focus on exact word matching and simple statistical calculations, such as keyword frequency or position-based measures like Mean Reciprocal Rank (MRR), i.e., these metrics assess how closely the words in the documents match the query without considering deeper meanings.

Semantic similarity metrics, on the other hand, go beyond surface-level text comparison to understand the underlying meanings and concepts by comparing their semantical similarity based on context and conceptual relationships between words and sentences. This approach captures the intent of the query and the documents, evaluating relevance through semantic similarity rather than just keyword occurrence.

Finally, **LLM as a judge** uses a LLM to evaluate content by making it context-aware, prompting the model to consider the coherence, factuality, and relevance of the information based on its comprehensive understanding of language and knowledge. Therefore, the choice of the evaluator determines the type of metric applied to assess each evaluation dimension, depending on whether the focus is on exact word matching, conceptual similarity, or a nuanced, context-aware judgment by an advanced LLM.

The upcoming sub-sections follows the rationale of first explaining the evaluation dimension and then the respective evaluator by detailing how to calculate the metric proposed in the sampled papers.

4.1 Context Relevance

The evaluation dimension of **context relevance** pertains to the retrieval step and assesses the degree to which the retrieved context contains only the necessary information to answer the query, reducing computational costs and improving efficiency by minimizing irrelevant content (Es et al., 2023; Saad-Falcon et al., 2023; Yu et al., 2024). Additionally, when retrieved passages are too LLMs often struggle to effectively utilize the information, particularly if the relevant details are embedded in the middle of the passage (Es et al., 2023). Hence, concise query-relevant passages significantly improve the LLM generation quality (Es et al., 2023; Yu et al., 2024).

Recall@k and **MRR@k** are key **lexical metrics** for evaluating the retrieval performance in RAG systems (Rackauckas et al., 2024; Yu et al., 2024). Each metric provides a different perspective on the effectiveness of the retrieval process. **Recall@k**

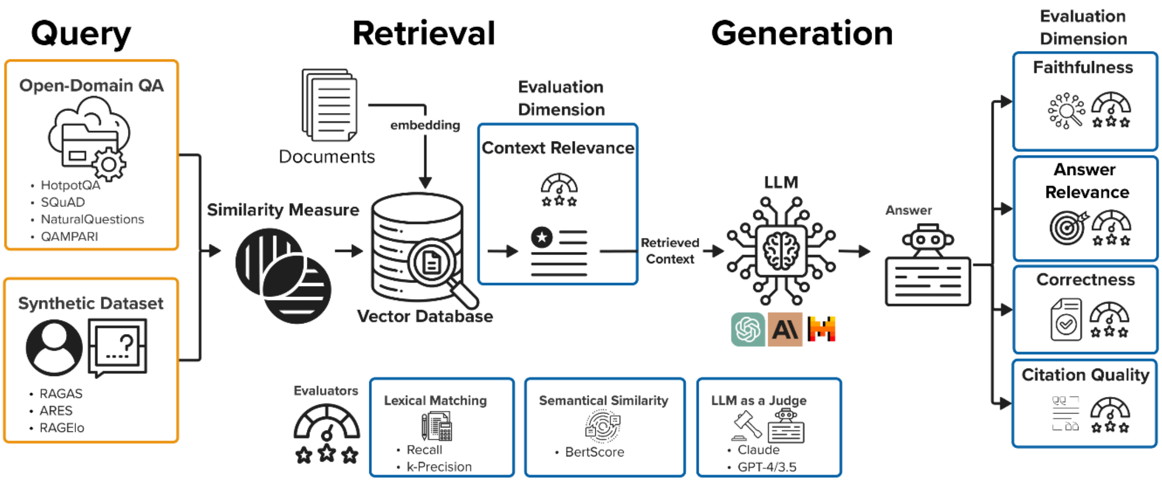


Figure 2: Evaluation Pipeline with Evaluation Dimensions in a naïve RAG setup.

measures how many relevant passages are captured within the top k retrieved chunks, even if some irrelevant ones are included. The formula is as follows:

$$\text{Recall@k} = \frac{|\text{Relevant Passages} \cap \text{k-Passages}|}{|\text{Relevant Passages}|} \quad (1)$$

MRR@k calculates context relevance by emphasizing the rank of the first relevant passage across multiple queries (Rackauckas et al., 2024). If a relevant passage appears in the top k results, its contribution to the MRR@k score is the inverse of its rank, e.g., a passage ranked two contributes as ½ with MRR@5 (Rackauckas et al., 2024). If no relevant passage is found in the top k the contribution is zero. The formula for MRR@k is:

$$\text{MRR@k} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\text{rank}_i} \quad (2)$$

By using the **LLM as a judge**, it is possible to calculate an estimated context relevance score. Given a query q and its retrieved context c(q), the LLM is prompted to extract a subset of sentences (S_{ext}) from c(q) that are relevant to answering q by using the following *prompt*:

"*Please extract relevant sentences from the provided context that can potentially help answer the following question. If no relevant sentences are found, or if you believe the question cannot be answered from the given context, return the phrase 'Insufficient Information'. While extracting candidate sentences, you're not allowed to make any changes to sentences from the given context.*" (Es et al., 2023)

The prompt instructs the LLM to select only the sentences that it considers relevant to q without changing the content. The **Context Relevance Score (CRS)** is calculated by dividing the relevant sentences extracted (S_{ext}) from the context c(q) by the total number of sentences. This can be expressed with the following formula (Es et al., 2023):

$$\text{CRS} = \frac{\text{Number of extracted sentences } S_{ext}}{\text{Total number of sentences in c(q)}} \quad (3)$$

The CRS indicates the proportion of the context that is relevant. A higher score indicates that a greater proportion of the retrieved context is focused and relevant for answering the query, while a lower score indicates that much of the retrieved context contains irrelevant information (Es et al., 2023; Saad-Falcon et al., 2023; Yu et al., 2024).

4.2 Faithfulness

The evaluation dimensions **faithfulness** refers to the generation step and evaluates how well an LLM's response is grounded in the retrieved context, i.e., all information in the response can be directly inferred from it (Adlakha et al., 2023; Es et al., 2023; Hu et al., 2024). This evaluation dimension is crucial to identify possible hallucinations in the answer of LLMs ensuring factual correctness (Adlakha et al., 2023; Es et al., 2023; Ravi et al., 2024). For instance, Adlakha et al. (2023) found that GPT-4 had the highest agreement with human annotations, followed by GPT-3.5 and k-precision.

As the **primary lexical metric** Adlakha et al. (2023) propose **k-precision** to evaluate the degree of faithfulness since it has the highest agreement with human judgements. It can be calculated as the proportion of tokens in the LLMs response that are present in the retrieved context, i.e., it is the overlap of matching tokens with the retrieved context divided by the total number of tokens in the response (Adlakha et al., 2023). Hence, the formula is as follows:

$$\text{K-Precision} = \frac{\text{Matched Tokens}}{\text{Response Tokens}} \quad (4)$$

Another way of calculating the faithfulness is by using **LLMs** such as GPT-4/3.5 as **evaluators**. The LLMs are prompted to judge whether the response and the retrieved context match from an ordinal scale of "fully", "partially", or "not at all" (Adlakha et al., 2023). In the same way, Ravi et al. (2024) propose to assess the responses through an evaluator LLM whether the responses are supported, contradicted or not supported by the retrieved context. In order to quantify this ordinal scale, we can assign numerical scores depending on the degree of support and take the average of all individual response scores, i.e., assign 1 for "fully", 0.5 for "partially", and 0 for "not at all". The formula for calculating the **Faithfulness Coefficient (FC)** is as follows:

$$\text{FC} = \frac{\sum_{i=1}^{N} \text{Faithfulness Score}_i}{\text{Total number of responses}} \quad (5)$$

A similar method is used by Hu et al. (2024), who propose a framework that extracts "claim triplets" (subject, predicate, object) to represent fine-grained knowledge assertions within the LLM response. The purpose of this extraction is to break down the answer into specific atomic claims that can be checked individually (Hu et al., 2024; Min et al., 2023). A judging LLM evaluates each triplet as "entailment" (supported), "contradiction" (contradicted), or "neutral" (unsupported) (Hu et al., 2024). Other authors also follow this approach of breaking down the statements from the LLM response into atomic facts to obtain a fine-grained measure of the faithfulness degree (Min et al., 2023).

Es et al. (2023) propose a process in which the answer $a_s(q)$ is considered faithful to the context $c(q)$ if the statements in the LLM response can be directly inferred from the retrieved context. The process begins by using a **LLM as a judge** to decompose the LLM response into a set of statements, $S(a_s(q))$, which involves breaking down longer sentences into shorter ones (Es et al., 2023). For each statement s_i in S, the judging LLM verifies if it can be inferred from the given context $c(q)$ using a verification function $v(s_i, c(q))$ (Es et al., 2023). The judging LLM assesses whether each statement is supported by the information in the retrieved context and provides a "yes" or "no" verdict for each statement (Es et al., 2023). **The Faithfulness Score** (**FS**) is then calculated as the ratio of the number of supported statements V to the total number of statements S, which can be expressed as follows:

$$\mathrm{FS} = \frac{|\mathrm{V}|}{|\mathrm{S}|} \tag{6}$$

4.3 Answer Relevance

The evaluation dimensions **answer relevance** refers to the generation step and assesses whether the LLM response is directly addressing the query (Es et al., 2023; Rackauckas et al., 2024; Saad-Falcon et al., 2023; Yu et al., 2024). This evaluation dimension penalizes incomplete or redundant answers, **regardless of factuality** (Es et al., 2023; Rackauckas et al., 2024).

By using **LLM as a judge** it is possible to calculate an estimation of the answer relevance (Es et al., 2023; Rackauckas et al., 2024; Saad-Falcon et al., 2023). Given a generated answer $a_s(q)$, the judging LLM is prompted to generate n potential questions q_i that could be answered by using $a_s(q)$ (Es et al., 2023). This is done using the following prompt:

> *Generate a question for the given answer: answer:[answer]* (Es et al., 2023)

Subsequently, text embeddings for all generated questions q_i and the original query q are created to calculate in the next step the cosine similarity between their embeddings (Es et al., 2023). The **Answer Relevance Score (ARS)** is obtained by averaging the similarity between the generated questions q_i and the original query q using the following formula:

$$\mathrm{ARS} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{sim}(\mathrm{q}, \mathrm{q_i}) \tag{7}$$

where $sim(q, q_i)$ represents the cosine similarity between the embeddings of the generated questions $\mathrm{q_i}$ and the original query q (Es et al., 2023). This metric effectively measures how well the generated answer matches the intent and content of the original question (Es et al., 2023).

4.4 Correctness

The evaluation dimension **correctness** refers to the generation step and evaluates whether the LLM's response accurately matches the "golden passage" provided by human annotators (Adlakha et al., 2023; T. Gao et al., 2023). This metric **focuses on the factual accuracy** of the information by comparing the LLM response with a reference answer (Adlakha et al., 2023; T. Gao et al., 2023; Guinet et al., 2024; Rackauckas et al., 2024).

As a **primary lexical metric** Adlakha et al. (2023) propose using **recall** as it correlates well with human annotations. Traditional metrics like Exact Match (EM), F1, and ROUGE are often too strict due to their focus on exact word matching (Adlakha et al., 2023). Recall measures how much of the reference answer's essential content is captured in the model's response without penalizing additional information (Adlakha et al., 2023).

Some authors find that **semantic similarity** metrics like BERTScore is less effective than recall for correctness due to lower alignment with human annotations (Adlakha et al., 2023; T. Gao et al., 2023). These metrics do not account for factual accuracy or logical consistency, as responses can be textually similar but factually incorrect (Adlakha et al., 2023; T. Gao et al., 2023).

By using **LLMs** such as GPT-3.5 and GPT-4 **as evaluators to judge** the correctness of responses by prompting them with the question, the reference answer, and the LLMs response to determine whether the model response is correct, partially correct, or incorrect (Adlakha et al., 2023; Rackauckas et al., 2024). This approach seems to yield the highest agreement with human annotations (Adlakha et al., 2023; Rackauckas et al., 2024). Correctness can be quantified by assigning scores: 1 for "fully correct", 0.5 for "partially", and 0 for "incorrect". The

Correctness Coefficient (CC) is the average of all response scores. Therefore, the formula is as follows:

$$CC = \frac{\sum_{i=1}^{N} \text{Correctness Score}_i}{\text{Total number of responses}} \quad (8)$$

4.5 Citation Quality

The evaluation dimension **citation quality** focuses on assessing whether an LLM correctly cites its sources when generating text (T. Gao et al., 2023). Citation quality is calculated using two metrics: **citation recall** and **citation precision**. Citation recall ensures that every piece of information in the generated response is fully supported by the cited passages, while citation precision checks whether all cited passages are relevant and necessary for the statements made (T. Gao et al., 2023).

To perform this evaluation automatically, the **LLM acts as a judge**, which is prompted with a chain-of-thought method instead of simple lexical matching (T. Gao et al., 2023). The LLM checks if the concatenated text from the cited passages semantically supports the generated statements (T. Gao et al., 2023). For citation recall, it evaluates whether all generated statements are substantiated by the citations. A statement receives a recall score of 1 if it is fully supported by at least one citation, otherwise, it receives a score of 0 (T. Gao et al., 2023). For citation precision, the LLM identifies any "irrelevant" citations, i.e., those that do not independently support a statement or are not necessary when other citations already provide full support (T. Gao et al., 2023). A citation receives a precision score of 1 if it is relevant and contributes to the statement's support and 0 if it is irrelevant (T. Gao et al., 2023).

The robustness of these metrics is validated by their **strong correlation** with **human judgements** (T. Gao et al., 2023). By averaging the citation recall and precision scores across all statements and citations in the generated response, an overall citation quality score can be calculated, providing a comprehensive measure of how accurately and appropriately an LLM uses citations in its output.

5 DATASETS & APPLICATION

Building on the concept matrix's evaluation dimensions presented above, it's essential to consider how different datasets relate to the evaluators of these dimensions and its corresponding evaluation metrics. The choice between open-domain QA datasets and synthetic datasets like RAGAS or ARES, along with the type of reasoning required (single-hop vs. multi-hop), plays a crucial role in ensuring the robustness and reliability of these evaluations.

Open-domain QA datasets such as SQuAD2.0, HotpotQA, and Natural Questions are based on Wikipedia articles. These are mainly suitable for **lexical matching** enable comparisons across RAG systems (Kwiatkowski et al., 2019; Rajpurkar et al., 2018; Yang et al., 2018). These datasets often include “golden passages” which make them ideal for evaluating **correctness** and **faithfulness** by providing a factual reference (Kwiatkowski et al., 2019; Rajpurkar et al., 2018; Yang et al., 2018). For instance, SQuAD2.0 includes unanswerable queries requiring RAG systems to recognize when there is insufficient information to provide a valid answer (Rajpurkar et al., 2018; Rau et al., 2024). Also calculating **context relevance** is straightforward because the datasets provide clear ground truth in terms of which passages are relevant, making them ideal for recall-oriented metrics like **Recall@k** or **MRR@k** (Adlakha et al., 2023; Hu et al., 2024). However, evaluating **answer relevance** and **citation quality** is more challenging with open-domain QA datasets since these typically focus on finding a single correct answer rather than assessing nuanced citation practices or multi-source relevance (Kwiatkowski et al., 2019; Rajpurkar et al., 2018; Yang et al., 2018).

Synthetic datasets such as RAGAS and ARES are specifically designed to evaluate the effectiveness of RAG systems by minimizing reliance on human annotations (Es et al., 2023; Saad-Falcon et al., 2023). These frameworks often use synthetic datasets that only require query-context-response triples, making them suitable to evaluate every evaluation dimension (Es et al., 2023; Hu et al., 2024; Min et al., 2023; Saad-Falcon et al., 2023). Synthetic datasets combined with LLM judges align well with human annotations, outperforming lexical and semantic similarity metrics (Adlakha et al., 2023; Saad-Falcon et al., 2023). Additionally, this approach is model-agnostic, allowing flexible use across different LLMs and setups (Es et al., 2023; Saad-Falcon et al., 2023).

This adaptability ensures that RAG systems can be effectively assessed and fine-tuned for diverse and complex queries, enhancing their performance in practical, real-world setting. Table 3 summarizes the key differences between the datasets by comparing them.

In terms of reasoning, **single-hop** and **multi-hop** queries require different approaches and datasets **Single-hop** reasoning involves deriving an answer from a single piece of evidence, i.e., one retrieved

Table 3: Comparing Open Domain QA Datasets with Synthetic Datasets for evaluating RAG.

Aspect	Open Domain QA Dataset	Synthetic Dataset
Examples	SQuAD2.0, HotpotQA, Natural Questions, QAMPARI	RAGAS, ARES
Reasoning Type	Single-hop (e.g., SQuAD2.0, Natural Questions) Multi-hop (e.g., HotpotQA, QAMPARI)	Single-hop and adaptable to Multi-hop
Evaluation Dimensions	Correctness, Faithfulness, Context Relevance (basic)	Correctness, Faithfulness, Context Relevance (multi-retrieval), Answer Relevance, Citation Quality
Evaluators	Lexical Matching	Semantic Similarity, LLM as a Judge
Strengths	High reliability for correctness and faithfulness due to golden-passages	Model and vendor agnostic, adaptable for various queries and rapid evaluation of RAG without the need for human annotations or gold-passages
Limitations	Less effective in evaluating multi-source citations, complex context relevance, and answer relevance in multi-hop	Associated costs related to token usage and potential latency due to LLM judging, different performances depending on the employed LLM model

passage, and is well-suited for the evaluation dimensions **correctness**, **faithfulness**, and **basic context relevance**. Popular single-hop datasets are Natural Questions (Kwiatkowski et al., 2019) and SQuAD2.0 (Rajpurkar et al., 2018), where the relevant information is contained within a single passage, allowing the calculation of metrics predominantly with lexical matching or simple semantic similarity, e.g., focusing on metrics such as precision and recall (Adlakha et al., 2023; Ravi et al., 2024). In contrast, **multi-hop** reasoning requires to connect multiple pieces of retrieval, usually from different documents or distant parts of the same document, to be combined in order to obtain a correct answer. This approach is better suited for evaluating **context relevance** for multiple retrieval, **answer relevance**, i.e., how the combined context informs the answer, and **citation quality** that correctly attributes information to multiple sources (Adlakha et al., 2023; Es et al., 2023; T. Gao et al., 2023; Hu et al., 2024). Open-domain QA datasets like HotpotQA (Yang et al., 2018) or QAMPARI (Amouyal et al., 2022) are specifically designed for multi-hop reasoning. These require synthesizing information from multiple retrieved contexts, which involves understanding complex connections and contextual relevance that go beyond surface-level comparisons. Employing LLMs as a judge for this evaluation is the most suitable option since LLMs can comprehend the combination of multiple contexts better by making more nuanced judgement than simple lexical matching or semantical similarity of concepts (Es et al., 2023; T. Gao et al., 2023; Min et al., 2023; Saad-Falcon et al., 2023).

6 PRACTICAL APPLICATION

In order to apply evaluation dimensions and select the appropriate datasets, it is necessary to understand dataset requirements. For instance, the evaluation dimension faithfulness requires data regarding the retrieved passages of the RAG to be tested and a reference retrieval (golden retrieval), but these are often not available in real-world operations. Therefore, synthetic datasets that are applicable reference-free would be more suitable for practical operation, and they also have a high alignment with human annotations (Es et al., 2023; Saad-Falcon et al., 2023).

It is also necessary to select suitable metrics according to the objective of optimizing the RAG system. For this purpose, a distinction should first be made as to whether the performance of the retrieval or generation step should be considered. A suitable metric is then selected according to a specific problem. For example, if the factuality of the answer is to be increased, correctness is a more suitable metric than faithfulness.

Finally, it is important to build an automated, robust and reliable evaluation pipeline that can be used to evaluate the RAG system (Es et al., 2023).

7 CONCLUSION

This paper proposes a comprehensive evaluation framework specifically for RAG by conducting an SLR and providing an extensive overview of currently existing evaluation approaches. Since the introduction of RAG in 2020 (Lewis et al., 2020), it has taken considerable time for methods to be developed and established in the literature for evaluating these approaches. Despite the rigorous methodology, it remains possible that some papers were overlooked due to the rapid pace of developments in the field.

With reference to RQ1, our evaluation framework introduces robust evaluation dimensions and metrics to assess the different steps within RAG. Moreover, this paper advances the understanding of RAG evaluation by providing reliable dimensions and metrics (Y. Gao et al., 2023; Wang et al., 2023). Furthermore, Section 5 gives a comprehensive summary about RQ2 by providing an overview of the available datasets to apply the proposed evaluation dimensions and metrics and what kind of requirements to consider.

As a future avenue of research, a practical application of these metrics should be conducted to validate their use and alignment with human preferences. In addition, the provision of all the metrics presented could be examined within a framework to simplify practical application.

ACKNOWLEDGEMENTS

The presented paper was produced as part of the research project MoFaPro. This project is funded by the AUDI AG. The present approach was developed within the institute "AImotion Bavaria" at the Technische Hochschule Ingolstadt. This work is part of Oğuz Caymazer's master's thesis and was conducted during his internship at the AUDI AG.

REFERENCES

Adlakha, V., Behnamghader, P., Lu, x. H., Meade, n., & Reddy, S. (2023). *Evaluating correctness and faithfulness of instruction-following models for question answering.* Https://doi.org/10.48550/arxiv.2307.16877

Amouyal, S. J., Wolfson, T., Rubin, O., Yoran, O., Herzig, J., & Berant, J. (2022). *QAMPARI: An Open-domain Question Answering Benchmark for Questions with Many Answers from Multiple Paragraphs.* https://doi.org/10.48550/arXiv.2205.12665

Asai, A., Wu, Z., Wang, Y [Yizhong], Sil, A., & Hajishirzi, H. (2023). *Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection.* https://doi.org/10.48550/arXiv.2310.11511

Asai, A., Zhong, Z., Chen, D [Danqi], Koh, P. W., Zettlemoyer, L., Hajishirzi, H., & Yih, W. (2024). *Reliable, Adaptable, and Attributable Language Models with Retrieval.* https://doi.org/10.48550/arXiv.2403.03187

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (Fifth edition). Oxford University Press.

Benbya, H., Strich, F., & Tamm, T. (2024). Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. *Journal of the Association for Information Systems*, *25*(1), 23–36. https://doi.org/10.17705/1jais.00861

Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In European Conference on Information Systems (Chair), *17th European Conf. on Information Systems*. https://www.wi.uni-muenster.de/research/publications/3069

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y [Yi]. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4.* https://doi.org/10.48550/arXiv.2303.12712

Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2023). *RAGAS: Automated Evaluation of Retrieval Augmented Generation.* https://doi.org/10.48550/arXiv.2309.15217

Gao, T., Yen, H., Yu, J., & Chen, D [Danqi]. (2023). *Enabling Large Language Models to Generate Text with Citations.* https://doi.org/10.48550/arXiv.2305.14627

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y [Yi], Sun, J., Wang, M., & Wang, H. (2023). *Retrieval-Augmented Generation for Large Language Models: A Survey.* https://doi.org/10.48550/arXiv.2312.10997

Guinet, G., Omidvar-Tehrani, B., Deoras, A., & Callot, L. (2024). *Automated Evaluation of Retrieval-Augmented Language Models with Task-Specific Exam Generation*. https://github.com/amazon-science/auto-rag-eval https://doi.org/10.48550/arXiv.2405.13622

Hammond, G. (2024, April 10). Speed of AI development stretches risk assessments to breaking point. *Financial Times*. https://www.ft.com/content/499c8935-f46e-4ec8-a8e2-19e07e3b0438

Hu, X [Xiangkun], Ru, D., Qiu, L., Guo, Q., Zhang, T., Xu, Y., Luo, Y., Liu, P., Zhang, Y [Yue], & Zhang, Z [Zheng]. (2024). *RefChecker: Reference-based Fine-grained Hallucination Checker and Benchmark for Large Language Models.* https://doi.org/10.48550/arXiv.2405.14486

Izacard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., Dwivedi-Yu, J., Joulin, A., Riedel, S., & Grave, E. (2022). *Atlas: Few-shot Learning with Retrieval Augmented Language Models.* https://doi.org/10.48550/arXiv.2208.03299

Kandpal, N., Deng, H., Roberts, A., Wallace, E., & Raffel, C. (2022). *Large Language Models Struggle to Learn Long-Tail Knowledge.* https://doi.org/10.48550/arXiv.2211.08411

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D [Danqi], & Yih, W. (2020). *Dense Passage Retrieval for Open-Domain Question Answering.* https://doi.org/10.48550/arXiv.2004.04906

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., Epstein, D., Polosukhin, I., Devlin, J., Lee, K., Toutanova, K., Jones, L., Kelcey, M., Chang, M.-W., Dai, A. M., Uszkoreit, J., Le, Q., & Petrov, S. (2019). Natural Questions: A Benchmark for Question Answering Research. *Transactions of the Association for Computational Linguistics*, *7*, 453–466. https://doi.org/10.1162/tacl_a_00276

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.* https://doi.org/10.48550/arXiv.2005.11401

Ma, X., Gong, Y., He, P., Zhao, H., & Duan, N. (2023). *Query Rewriting for Retrieval-Augmented Large Language Models.* https://doi.org/10.48550/arXiv.2305.14283

Mallen, A., Asai, A., Zhong, V., Das, R., Khashabi, D., & Hajishirzi, H. (2022). *When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories.* https://doi.org/10.48550/arXiv.2212.10511

Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W., Koh, P., Iyyer, M., Zettlemoyer, L., & Hajishirzi, H. (2023). FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 12076–12100). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.741

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., . . . Zoph, B. (2023). *GPT-4 Technical Report.* https://doi.org/10.48550/arXiv.2303.08774

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems*, *25*(6), 493–508. https://doi.org/10.1057/s41303-016-0020-3

Rackauckas, Z., Câmara, A., & Zavrel, J. (2024). *Evaluating RAG-Fusion with RAGElo: an Automated Elo-based Framework.* https://doi.org/10.48550/arXiv.2406.14783

Rajpurkar, P., Jia, R., & Liang, P. (2018). *Know What You Don't Know: Unanswerable Questions for SQuAD.* https://doi.org/10.48550/arXiv.1806.03822

Rau, D., Déjean, H., Chirkova, N., Formal, T., Wang, S., Nikoulina, V., & Clinchant, S. (2024). *BERGEN: A Benchmarking Library for Retrieval-Augmented Generation.* https://doi.org/10.48550/arXiv.2407.01102

Ravi, S. S., Mielczarek, B., Kannappan, A., Kiela, D., & Qian, R. (2024). *Lynx: An Open Source Hallucination Evaluation Model*. https://arxiv.org/abs/2407.08488

Saad-Falcon, J., Khattab, O., Potts, C., & Zaharia, M. (2023). *ARES: An Automated Evaluation Framework for Retrieval-Augmented Generation Systems.* https://doi.org/10.48550/arXiv.2311.09476

Wang, C., Liu, X., Yue, Y., Tang, X., Zhang, T., Jiayang, C., Yao, Y., Gao, W., Hu, X [Xuming], Qi, Z., Wang, Y [Yidong], Yang, L., Wang, J [Jindong], Xie, X., Zhang, Z [Zheng], & Zhang, Y [Yue]. (2023). *Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity.* https://doi.org/10.48550/arXiv.2310.07521

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. https://www.jstor.org/stable/4132319?seq=1#metadata_info_tab_contents

Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R., & Manning, C. D. (2018). *HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering.* https://doi.org/10.48550/arXiv.1809.09600

Yu, H., Gan, A., Zhang, K., Tong, S., Liu, Q., & Liu, Z. (2024). *Evaluation of Retrieval-Augmented Generation: A Survey.* https://doi.org/10.48550/arXiv.2405.07437

Zhang, Z [Zihan], Fang, M., & Chen, L. (2024). *RetrievalQA: Assessing Adaptive Retrieval-Augmented Generation for Short-form Open-Domain Question Answering.* https://doi.org/10.48550/arXiv.2402.16457

Zhang, Z [Zihan], Fang, M., Chen, L., Namazi-Rad, M.-R., & Wang, J [Jun]. (2023). *How Do Large Language Models Capture the Ever-changing World Knowledge? A Review of Recent Advances.* https://doi.org/10.48550/arXiv.2310.07343

SHORT PAPERS

A Model for Designing Personalized and Context-Aware Nudges

Randi Karlsen and Anders Andersen
Department of Computer Science, UiT The Arctic University of Norway, Tromso, Norway
{*randi.karlsen, anders.andersen*}*@uit.no*

Keywords: Smart Nudging, Digital Nudging, Nudge Design Model, Personalization and Context-Awareness, Adaptive Nudge Design, Behavioral Change.

Abstract: Nudging is a popular approach used to influence people to change their behavior towards a desirable goal. To be effective, nudges should be tailored to the user's specific needs based on user profile, current behavior, and context information. In this paper, we target the questions of *what to nudge for* and *when to nudge*, and how nudges can be automatically designed at a time when the user needs nudging. We present a model for personalized and context-aware nudge design that is adaptive in that it continuously tailors the set of relevant activities and the time for creating a nudge to the user's needs for behavioral change. The model follows a just-in-time approach, where nudges are created at the time when nudging is needed, based on the user's current situation.

1 INTRODUCTION

Nudging is a technique used to influence people to change behavior towards a desired outcome, e.g., to adopt healthier or more environmentally friendly habits. Behavioral change is the process of adopting new habits or modifying existing ones to achieve a specific goal, and nudging is an effective tool for promoting behavioral change involving psychological factors that influence human behavior without restricting their freedom of choice (Thaler and Sunstein, 2008).

Our focus is on personalized and context-aware digital nudges, called *smart nudges*, where digital technology is used to influence or guide people's behavior (Karlsen and Andersen, 2019). *Personalized nudging* refers to the use of tailored nudges based on the user's specific characteristics, activity history, and preferences, while *context awareness* uses people's situation and environment to identify opportunities, limitations, and obstacles, to further tailor nudging to the needs of the user.

To facilitate behavioral change, nudging will challenge the user by motivating for activities that brings the user closer to the nudging goal. To do so, a nudging system will continuously keep track of the user's activities, preferences, achievements, and environment, to design relevant and timely nudges that are presented to the user, e.g., on a mobile device.

The idea behind tailored nudging is that people are more likely to respond positively to nudges that are relevant and meaningful to them, rather than using generic or "one-size-fits-all" nudges (Schneider et al., 2018; Peer et al., 2020; Mills, 2022). However, a challenge when designing tailored nudges is to determine what to nudge for and when to nudge.

This paper presents a practical model for how to design smart nudges based on the user's current behavior and situation. We focus on *what to nudge for* and *when to nudge*, and suggest a general approach to select activity and time frame for a nudge. We exemplify how the proposed solution can be used in two different use cases; for physical activity nudging and green transportation nudging.

In the following, we first present smart nudges, nudge design, and the two use cases. We describe several factors that influence nudge design, including level of behavior, behavioral progress, activity patterns, reactions to previous nudges, capability, and opportunity. This is followed by a description on how activity and time frame for nudges are selected. Finally, we discuss our approach and conclude.

2 BACKGROUND

This section describes smart nudges, principles for smart nudge design, and two use cases that will exemplify how our generic model for nudge design can be applied to different nudging systems.

Karlsen, R. and Andersen, A.
A Model for Designing Personalized and Context-Aware Nudges.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 151-162
ISBN: 978-989-758-716-0; ISSN: 2184-3228

2.1 Smart Nudges and Nudge Design

A *smart nudge* is a personalized and context-aware digital nudge, described through a set of components, including an *activity* selected for the nudge, a *time frame* for when the activity is suggested done, some *influence* that motivates the user to do the activity, and some practical information useful when following the nudge.

Every component of a smart nudge can be tailored to the user's need for nudging. A user profile (including, e.g., user interests, capabilities, activity history, and reactions to previous nudges) and context information (describing the user's environment and surroundings) are used to determine which activity and practical information to include, which time frame to target, and how to influence the user.

Nudges are typically offered to the user through an application on a mobile phone, and a combination of back-end and edge computing determines when and what to nudge for through monitoring, selecting, and analysing data from various sources, e.g., IoT sensors, third-party data sources, mobile phone sensors, and stored user data (Andersen et al., 2018).

Designing a nudge involves several tasks, as illustrated in Figure 1. In this paper, we focus on the details concerning selecting activity and time frame for nudges. In paper (Karlsen and Andersen, 2024) we describe the task of triggering smart nudges, paper (Dalecke and Karlsen, 2020) focus on the influence part of nudge design, while we are currently working on details concerning content selection and nudge presentation.

The design tasks are influenced by a set of nudge design principles, presented in Table 1 and previously described in (Karlsen and Andersen, 2022), that guide the selection of activity and time frame.

Table 1: Principles for smart nudge design.

Challenge	Nudging must challenge the user to choose activities that improve the user's behavior.
Consolidate	After improving behavior, the user must be stabilized on the new level of behavior. For a period the user is not challenged.
Progress	Challenging the user will continue after a consolidate period. Given there is still room for improvement.
Variation	The user should over time be given a variety of nudges, to make nudging interesting and introduce new activities.
Timeliness	A nudge should be given at a time when the user can react to it and when the nudge can be effective.
Safety and feasibility	A nudge must avoid activities that are dangerous or impossible for the user to do.

We distinguish between generic and specific nudges. A *generic nudge* suggests an activity (such as walking or cycling), and leaves it to the user to decide how and where the activity is performed. The nudge can optionally suggest a time frame, distance, or duration of the activity. An example nudge is: *"It's time for a walk. The weather is nice and you have time this afternoon."*.

A *specific nudge* suggests a trail, that has a location, route, and destination, and is characterized by properties such as activity type, distance, estimated duration, and difficulty. For example, a nudge can suggest a hike to Mount X, which has a 5.8 km long trail, elevation gain of 553 m, an estimated duration of 2.5 hours, classified as hard, and has a spectacular view from the top. Description and properties of such trails can be found on online sources such as AllTrails.com[1], and can be used for creating specific nudges.

2.2 Use Cases

2.2.1 Physical Activity Nudging

Nudging people to be sufficiently physically active is a desirable goal since regular physical activity is a known protective factor for the prevention and management of physical and mental diseases (WHO, 2016).

A nudge will present the user with a suggestion for an activity and a time frame for doing it. For the user to improve, nudging must suggest behavior that goes beyond the user's current behavior, e.g., by nudging the person to (i) be active more often, (ii) be active for longer periods, or (iii) engage in more challenging activities.

2.2.2 Green Transportation Nudging

The goal of green transportation nudging is to motivate the user to choose environmentally friendly transportation means. This is desirable because of the urban challenges of increased traffic, congestion, air pollution, and global-scale issues of climate change and global warming (Comission of the European Communities, 2007).

When the user needs to move between two locations, origin (O) and destination (D), nudging can motivate the user to choose alternatives to, e.g., private car usage. Relevant alternatives depend on, e.g., the availability of public transportation, whether the

[1] https://www.alltrails.com

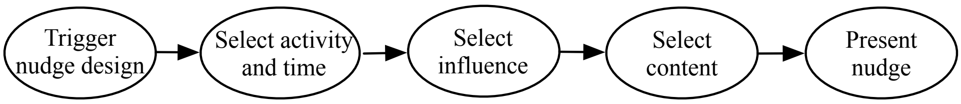


Figure 1: Tasks in a nudge design process.

distance between *O* and *D* allows walking or cycling, and the user's ability to use the different transportation means.

Alternatives to using the private car include walking, cycling, public transportation, and the use of motorbikes and electric scooters. An alternative can also be to optimize and limit private car usage through carpooling, incentive parking (park and ride), and finding the most efficient driving route to the destination.

Available transportation means are ranked according to how environmentally friendly they are, and a nudge will suggest that the user selects a transportation alternative that is better than what the user normally prefers between *O* and *D*.

2.2.3 Related Work

The goal of nudging is to motivate people to change their behavior in some direction. As motivation is an important feature of nudging, much focus has been on the influence component, as seen in, e.g., (Caraban et al., 2019; Forberger et al., 2019; Caraban et al., 2020; Villalobos-Zúñiga and Cherubini, 2020; Bergram et al., 2022). However, it is equally important to focus on other components of a nudge. In this paper, we target the questions of *what to nudge for* and *when to nudge*, and how timely nudges can be automatically designed. A *just-in-time nudge design* approach is necessary since we tailor nudges to the user's behavior, situation, and environment, which is known only at the time of nudging.

Different authors have described digital nudge design through theoretical models or frameworks intended for designers or practitioners (Meske and Potthoff, 2017; Mirsch et al., 2018; Schneider et al., 2018; Purohit and Holzer, 2019). In contrast, we describe a model that targets automatic and adaptive nudge design in that it continuously tailors nudges to the user's current needs for behavioral change.

The main focus for nudging has up until recently been on one-size-fit-all nudges (Anagnostopoulou et al., 2020; Peer et al., 2020). However, the importance of personalization in designing effective nudges has been recognized by several authors, e.g., (Schöning et al., 2019; Peer et al., 2020; Mills, 2022), and experiments (Peer et al., 2020; Mills, 2022) show that personalized nudges can lead to more effective nudging compared to non-personalized nudges. Personalized nudging is a topic in need of more research (Caraban et al., 2019; Jesse and Jannach, 2021; Bergram et al., 2022), and we target this important topic through our model for nudge design.

Our approach to personalized and context-aware nudging is different from other approaches we are aware of, in that nudges are *automatically designed* and *adaptively tailored* to the current need of the user, based on continuous monitoring of user behavior and environment. The user is gently challenged to improve, based on an *adaptive goal*, and the nudging system will continuously distinguish between impossible and possible activities for the user.

3 FACTORS INFLUENCING NUDGE DESIGN

Smart nudging requires knowledge about the user's current behavior and behavioral change. This is obtained by monitoring user activities relevant to the nudging goal. This section describes how the user's Level of Behavior is measured, how Behavioral Progress is determined, how Activity Patterns are useful for nudge design, and how reactions to previous nudges influence nudging. Nudge design is also influenced by user capability (i.e., the ability to do an activity) and opportunity (i.e., how context can make an activity possible or not). Table 2 summarises these factors and describes how they influence smart nudge design.

3.1 Level of Behavior and Behavioral Progress

This section presents a general description of Level of Behavior and Behavioral Progress, followed by examples of how the concepts are used in the two use cases.

3.1.1 General Description

Level of Behavior (LoB) refers to how well a user behaves with respect to the nudging goal, e.g., how physically active or environmentally friendly the user is. LoB is measured using an ordinal scale, e.g., *Very Low, Low, Medium, Good, Excellent*.

The user's current LoB is determined by monitoring and assessing all activities relevant to the nudg-

Table 2: Factors influencing the selection of activity and time frame during nudge design.

Factors	Description	Effects on nudge design
Level of Behavior (LoB)	A measure of how well the user behaves regarding the nudging goal.	LoB is used as a basis for setting an *ActivityGoal*, determining *behavioral progress*, and keeping track of how user behavior changes over time.
Behavioral progress	As the user's behavior change, the progress is described as *improved*, *stable*, or *decreased*.	Determines if the nudging system should challenge the user to improve or help the user to stay on the current activity level.
Activity pattern	Describes recurring activities that happen in a systematic way.	Can detect activities the user is predominantly doing, when and under which circumstances activities are normally done. Identifies preferred activities and time frames, and can guide how to improve behavior.
Reactions to nudges	Monitor reactions to nudges and register which nudges the user accepts and rejects.	Knowing which activities, time frames, and influence types the user responded positively to, contributes when selecting components for new nudges.
User capability	The individual's physical and psychological capacity to engage in an activity.	Distinguishes between activities the user is unable to do (i.e., impossible activities) and activities the user is capable of doing.
Opportunity	The factors that lie outside the individual that make an activity possible or not	An activity can be impossible because of permanent/long-lasting circumstances (e.g., lack of equipment) or temporary circumstances (e.g., challenging weather conditions).

ing goal. As behavior changes over time, LoB is determined for discrete time periods (e.g., per week or month), and LoB for the latest period determines if the user needs to improve or if the best LoB is reached and the user should be nudged only to continue on the same behavioral level.

User activity is in our use cases measured using a numerical value. Each activity is given a value that reflects how well it supports the nudging goal and $UserActivity(P)$ represents the total value of all the user's activities in period P. Level of Behavior for P, i.e., $LoB(P)$, is determined by mapping the $UserActivity(P)$ value to the LoB scale.

To follow the principle of *challeng the user*, the nudging system automatically sets and adjusts a goal (denoted *ActivityGoal*) for the next time period P. If the goal is met for period P_i, *ActivityGoal* is slightly increased for the next period P_{i+1}.

Behavioral Progress guides the "aggressiveness" of the system, by determining if the system should challenge the user to improve or help the user to stay on the current activity level. Three states describe behavioral progress, i.e., *improved, stable*, and *decreased*, and enables the system to design nudges so that the principles of *consolidate* and *progress* are followed.

When the user reaches a higher LoB (e.g., when going from *Medium* to *Good* on the LoB scale), the progress state, *PrState*, is set to *improved* and the system suspends, for a period, the automatic adjustment of the goal, as nudging will focus on consolidating the user on the newly reached level. After a while, the user enters a *stable* stage and the system can again challenge the user by gently increasing the *ActivityGoal* value. If user behavior declines, the state is set to *decreased*, the value of *ActivityGoal* remains unchanged, and the user will be challenged in order to improve.

Formula 1 describes how the *ActivityGoal* value is automatically adjusted with an improvement factor *Impr*. The *ActivityGoal* is increased only if the user has reached the activity goal the last time period, and the user is not in an *improved* state.

$$\begin{aligned} If\, UserActivity(P) \geqslant ActivityGoal \\ AND\, PrState \neq \text{“}improved\text{”} \\ then\, ActivityGoal = ActivityGoal + Impr \end{aligned} \quad (1)$$

To keep track of how user behavior changes over time, historic values of Level of Behavior are stored in a *History of Behavior*. The history can, e.g., show whether user behavior improves over time or if the time of year has any impact on behavior.

3.1.2 Physical Activity Nudging

Physical activity can be measured using step count and , $UserActivity(P)$ represents the total number of steps for all activities performed during P. For activities such as walking or running, a pedometer is used for simply counting steps. Other types of activities (e.g., swimming or climbing) can be manually registered by the user and converted to steps using a step conversion factor. In Formula 2 the number of steps are calculated for an activity a.

$$Steps(a) = StepConversionFactor(a) * Duration(a) \quad (2)$$

The number of steps obtained from a pedometer does not measure user effort or exercise intensity. Using number of steps per minute or complementing the pedometer with a heart rate monitor are two approaches to get indications of effort or intensity. If effort or intensity is available, number of steps can be normalized, by multiplying with an effort factor before being added to the $UserActivity(P)$ value, as described in Formula 3.

$$NormalizedSteps(a) = steps(a) * effort(a) \quad (3)$$

The *ActivityGoal* is also given in number of steps, and the user can be challenged by expecting more steps the next period. To reach the new goal, the system may nudge for longer walks, more challenging exercises where user effort increases, or nudge more often to increase the frequency of physical activity.

Based on health organizations' recommendations for physical activity, mapping between $UserActivity(P)$ value and the LoB scale can be adapted according to, e.g., age and capabilities of the user.

3.1.3 Green Transportation Nudging

Determining *Level of Behavior* regarding green transportation includes monitoring or registering the type of transportation chosen in every travel situation.

A travel is described through a set of properties, including (*O, D, Time, TMeans*), where *O* and *D* represent origin and destination, *Time* represents departure time, and *TMeans* the transportation means used for the travel. Each type of transportation means is characterized by its environmental friendliness (EF) using an EF-value in the range [0,1], where the higher the value, the more environmentally friendly the transportation means is.

$UserActivity(P)$ represents the average EF-value over every transportation means used in period *P*. This is expressed in Formula 4, using the set $Travel(P)$ that includes all EF-values of travels made by the user in *P*.

$$UserActivity(P) = \frac{\sum_{x \in Travel(P)} x}{|Travel(P)|} \quad (4)$$

$Travel(P) = \{x | x \text{ is the EF-value of a travel in } P\}$

Both $UserActivity(P)$ and *ActivityGoal* represent EF-values in the range [0,1]. To help the user reach the goal, nudging can suggest more environmentally friendly transportation means (with a better EF-value) compared to what the user has previously chosen.

The user may have different habits when traveling between different locations, e.g., the user may prefer to walk to work, while using the car when traveling with kids or going shopping. Knowing the user's LoB for recurring (O,D) pairs can be helpful for identifying where improvements can be suggested.

Consequently, user activity is determined per (O,D) pair (i.e., $UserActivity_{(O,D)}(P)$) using Formula 5, where $Travel_{(O,D)}(P) \subseteq Travel(P)$. This makes it possible to register LoB per (O,D) pair and have an (O,D) specific goal, i.e., $ActivityGoal_{(O,D)}$.

$$UserActivity_{(O,D)}(P) = \frac{\sum_{x \in Travel_{(O,D)}(P)} x}{|Travel_{(O,D)}(P)|} \quad (5)$$

$Travel_{(O,D)}(P) = \{x | x \text{ is EF-value of a travel between O and D in } P\}$

Generally, taking the bus or a car gives a lower EF-value compared to walking and cycling. However, if the distance is too long and walking/cycling is not practical or possible, the bus (or other public transportation means) can be considered the greenest choice. Also, the time of day may determine what to expect from the user and what the nudging system can recommend. Walking late at night may not be safe, and therefore not an option. Taking the bus, or perhaps even a taxi, will in this situation be considered the greenest alternative. In each situation, the greenest alternative is given the best EF-value.

3.2 Activity Patterns

This section presents a general description of activity patterns used in nudging, followed by examples of how activity patterns are detected in the two use cases.

3.2.1 General Description

Activity Patterns represent information on user activities detected through monitoring and/or manual registration of activities. Generally, an activity pattern describes recurring activities that happen in a systematic and predictable way and, with respect to nudging, targets activities that are relevant to the nudging goal.

As nudging should support the user to improve behavior, we expect activity patterns to change over time. Therefore, activity patterns are detected for different time periods (e.g., weekly or monthly patterns), where differences between patterns can uncover changes in behavior.

An activity pattern identifies activities the user is predominantly doing, characteristics of these activities, when and under which circumstances (e.g., weather conditions) activities are normally done. Knowing the user's current activity pattern and level of behavior is the key to knowing how to nudge the user. The activity pattern is used as *baseline* when selecting activities for a nudge while avoiding activities or circumstances that represent a radical change for the user.

User *preferences* can to some extent be favored in nudge design, given that it does not conflict with the goal of improving behavior. The user may, for example, be inclined to accept a nudge suggesting a familiar activity or an activity with properties that do not deviate too much from the user's current behavior.

3.2.2 Physical Activity Nudging

Activity patterns provide information on which activities the user prefers, habits concerning when to exercise, and the duration and/or distance of previous activities. When designing new nudges, this is used as a basis to determine which activities to nudge for and when to nudge.

To identify activity patterns, we focus on activities (or exercises) that last more than a certain amount of time (e.g., 15 minutes). This is different from determining Level of Behavior, where we include all steps during a time period.

Exercises can be identified by analyzing i) *step count data*, where an *exercise* is identified as continuous activity (i.e., steps) exceeding a certain amount of time, and ii) *manually registered exercises*.

To manually register exercises, many tracking devices (such as FitBit[2], Oura ring[3], and Apple Watch[4]) allow users to select an activity type (such as walking, cycling, swimming, or climbing) and register start time and duration. An activity tracker can also support the user by automatically detecting certain types of exercises.

The two complementary approaches; step count analysis and analysis of manually registered exercises, may both find exercises not detected by the other. For example, step count analysis will not detect swimming and climbing, but may detect (walking) exercises overseen or forgotten by the user, and therefore not manually registered. Also, some exercise types (such as walking and running) may be detected by both approaches. As part of a data cleaning process, duplicate exercises are identified and merged by comparing the time property.

An exercise is described through a set of properties, including *activity type*, number of *steps*, *distance*, *duration*, *elevation*, and *time* of day/week when the exercise took place. Activity patterns can provide frequency of activity types (to identify favorites or activities the use never do), at which time the user tends to be active (indicating when it may be useful to nudge), and relations between activity and time (e.g., which activities are preferred at certain times of week/year).

For each type of activity, distance, duration, and elevation can be described using statistical values such as average, distribution, minimum, and maximum, where minimum and maximum identify the currently easiest and most challenging experience (e.g., shortest and longest distance), while average and distribution are used for identifying the user's normal activity.

Activity patterns can be combined with context data (e.g., weather conditions or level of pollution) to detect which activities are predominantly done at a given condition.

3.2.3 Green Transportation Nudging

Activity patterns provide information on recurring travels between (O,D) pairs, including preferred transportation means and departure times. Recurring travels represent traveling habits, such as going to work at approximately the same time every weekday, picking up kids at school, or going to the gym at certain times during the week. As users may have different habits when traveling between different locations, an activity pattern for each recurring (O,D) pair (denoted $ActivityPattern_{(O,D)}$), is identified.

The user's movements, from one location to another, departure, and traveling time can be monitored using some activity tracking devices. Also, transportation means can to some extent be inferred based on monitored data. However, to obtain accurate travel information, the user must manually register the selected transportation means and departure time.

The activity patterns are used as a basis for selecting transportation means and time frame when designing new nudges, where the user can be challenged to use more environmentally friendly transportation means compared to the preferred choice, while patterns concerning departure time indicate when a nudge can be useful.

3.3 Reactions to Nudges

By registering reactions to nudges, the system can learn which nudges were accepted or rejected by the user. Accepted nudges reveal which nudges the user found useful, while rejected nudges reveal what the user did not find tempting or useful, where, e.g., the timing was wrong, or the activity too challenging or not interesting. Since there are many reasons for rejecting a nudge, it is useful with a response from the user on what caused the rejection.

[2]https://fitbit.com

[3]https://ouraring.com

[4]https://www.apple.com/watch/

For both physical activity and green transportation nudging the set of accepted nudges may over time reveal patterns of which activities tempted the user, and which influence types the user reacted positively to. The set of rejected nudges may, on the other hand, identify activities the user is not interested in, and possibly also influence types that do not motivate the user. An activity that is repeatedly rejected should be suspended from nudging for a period.

3.4 Capability and Opportunity

The work of (Michie et al., 2011) describes a *behavior system*, where capability, opportunity, and motivation interact to generate behavior. Capability and opportunity are factors that influence motivation, while behavior is influenced by all three components. Motivation is in our work represented by the influence component in a nudge, while capability and opportunity impact the selection of activity and time frame for a nudge.

3.4.1 Capability

Capability is defined as the individual's physical and psychological capacity to engage in a suggested activity, and includes having the necessary knowledge and skills to do the activity (Michie et al., 2011).

In a nudging system, the user can register personal capabilities and/or capability constraints, such as physical disabilities and lack of skills. The user's status regarding capability makes certain activities permanently or temporarily impossible. Activities can become possible if the capability status change, e.g., if the user recovers from a temporary disability or obtains the necessary skills. User capability contributes to distinguishing between activities that are impossible for the user, and activities that can be included in a nudge.

3.4.2 Opportunity

Opportunity is defined as all the factors that lie outside the individual that make a behavior possible or not (Michie et al., 2011). This includes physical circumstances (such as access to resources or facilities), environmental and social factors that support or hinder activities. Opportunity covers a large number of different circumstances that affect the user's ability to perform an activity, and therefore affect the choice of activity when designing a nudge.

We identify an *obstacle* as a situation which makes it impossible to go through with an activity. The reason can, e.g., be the lack of equipment (such as bicycle or skis), lack of facilities or services (such as safe outdoor spaces or public transportation), or harmful or hazardous situations (such as severe weather conditions) that make activities potentially dangerous.

An obstacle can be long-lasting or temporary. A long-lasting obstacle is removed only if some fundamental change is made, e.g., if new equipment or facilities are made available, or if the user change location where circumstances are different. A temporary obstacle ends as soon as the harmful situation is over. Long-lasting obstacles can be manually registered by the user, while temporary obstacles are detected based on monitoring the user's surroundings.

4 SELECTING ACTIVITY AND TIME FRAME FOR THE NEXT NUDGE

This section describes factors relevant when selecting a suitable activity and time frame during nudge design. In the following, we describe i) how the need for nudging is detected, ii) how the set of possible activities for a nudge is identified, and iii) how the selection of activity and time frame can be done.

4.1 Identifying an Activity Gap

4.1.1 General Description

A nudge can be triggered by certain situations that indicate the need for a gentle push towards the nudging goal. An important reason for nudging is that the user's current Level of Behavior is insufficient to reach the $ActivityGoal$.

Formula 6 calculates $ActivityGap(P)$, which represents a measure of the activity needed to reach the goal for the current (unfinished) time period, P.

$$ActivityGap(P) = ActivityGoal - PredActivity(P) \quad (6)$$

$PredActivity(P)$ represents the amount of activity predicted to be done in P, including already *completed activities* and *expected activities* in P. An expected activity can be a recurring activity, detected through the user's activity pattern, or an activity the user has committed doing in P.

The $ActivityGap(P)$ can be calculated from the first day of P, and as both completed and expected activities will change during P, $PredActivity(P)$ and $ActivityGap(P)$ must be regularly recalculated to reflect the user's achievements during P.

Knowing the $ActivityGap(P)$, makes it possible to plan and design a set of nudges for P that collectively helps the user to reach the activity goal. As the

$ActivityGap(P)$ value is updated during P, the set of planned nudges must also be updated.

The generic Formula 6 applies to both use cases. However, as described in the following Sections 4.1.2 and 4.1.3, $ActivityGap(P)$ and $PredActivity(P)$ are calculated differently in the two cases.

4.1.2 Physical Activity Nudging

For physical activity nudging, the $ActivityGap$ determines the additional number of steps needed for the user to reach the $ActivityGoal$.

Predicted activities, $PredActivity(P)$, is calculated as seen in Formula 7, where $CompActivity(P)$ and $ExpActivity(P)$ represent the number of steps for completed and expected activities in P, respectively.

$$PredActivity(P) = CompActivity(P) + ExpActivity(P) \quad (7)$$

The $ActivityGap(P)$ is calculated using Formula 6, and nudging is needed if $ActivityGap(P) > 0$.

4.1.3 Green Transportation Nudging

For green transportation nudging, $ActivityGoal$, $ActivityGap$, and all travels are measured using the EF-value. Predicted activities, $PredActivity(P)$, is calculated as a combined average EF-value over all completed and expected travels, see Formula 8. The calculation is based on two sets of EF-values, representing the completed travels, $CompTr(P)$, and the expected travels, $ExpTr(P)$.

$$PredActivity(P) = \frac{\sum_{x \in CompTr(P)} x + \sum_{y \in ExpTr(P)} y}{|CompTr(P)| + |ExpTr(P)|} \quad (8)$$

$CompTr(P) = \{x | x$ is the EF-value of a completed travel in period $P\}$
$ExpTr(P) = \{y | y$ is a predicted EF-value of an expected travel in period $P\}$

The expected transportation means, and thus the expected EF-value, for a travel between O and D is based on the user's previous behavior, such as the preferred transportation means for (O,D) the last few time periods.

Nudging is needed if $ActivityGap(P) > 0$, and, at the end of P, the $ActivityGoal$ is reached if $ActivityGap(P) \leqslant 0$.

When using (O,D) specific activity goals, i.e., $ActivityGoal_{(O,D)}$, activity gaps can be determined per (O,D) pair (i.e., $ActivityGap_{(O,D)}(P)$) using Formula 9, where $PredActivity_{(O,D)}(P)$ is calculated as in Formula 8 but in this case using the two sets $CompTr_{(O,D)}(P)$ and $ExpTr_{(O,D)}(P)$ which only includes completed and expected travels between (O,D) in period P.

$$ActivityGap_{(O,D)}(P) = ActivityGoal_{(O,D)} - PredActivity_{(O,D)}(P) \quad (9)$$

4.2 Identifying Possible Activities

4.2.1 General Description

In (Karlsen and Andersen, 2022) we classify activities as *impossible*, *possible*, *unlikely* (to be accepted by the user), and *probable* (that have the potential to be accepted by the user). Table 3 describes each class of activity, while Formula 10 shows the relation between the classes, where U represents a user and A all available activities.

Table 3: Classification of activities.

Impossible	Activities the user is not capable of doing. Identified based on user capability and opportunity.
Possible	Activities the user is capable of doing. $Possible(A,U) = Probable(A,U) \bigcup Unlikely(A,U)$
Probable	Activities the user has done in the past and is likely to do in the future. Likely to be accepted when suggested in a nudge.
Unlikely	Activities the user has never done, challenging and/or repeatedly rejected activities. Identified based on the user's activity pattern and reactions to previous nudges.

$$\begin{aligned} Possible(A,U) &= A - Impossible(A,U) \\ Probable(A,U) &= Possible(A,U) - Unlikely(A,U) \end{aligned} \quad (10)$$

The classification is user-specific, based on the user's activity pattern and reactions to previous nudges, where a challenging or repeatedly rejected activity can cause the activity to be classified as unlikely. Also, capability and opportunity can make an activity temporarily or permanently impossible.

4.2.2 Physical Activity Nudging

For the physical activity use case, activity patterns provide frequency of activity types (such as walking, cycling, and swimming), and identifies for each activity type, minimum, average, and maximum values for properties such as distance, duration, and elevation.

Physical activities can be conditionally possible, depending on properties of the activity. For example, a user may have a disability that makes short walks possible, while longer walks are impossible.

To distinguish between *probable* and *unlikely* activities, we identify activities that are *challenging* for the user. For each activity property, p, minimum (min) and maximum (max) values are used to distinguish between normal and challenging activities for user U, see Formula 11.

$$\begin{aligned} Normal(p,U) &= (min_p, max_p - d] \\ Challenge(p,U) &= (max_p - d, limit) \end{aligned} \quad (11)$$

$Normal(p,U)$ is an interval reflecting the user's previous achievements with respect to property p. For example, if p represents walking distance, max_p and min_p represent the longest and shortest distance, respectively, walked by user U. $Challenge(p,U)$ is an interval for p which represents efforts that are possible for the user, but currently beyond the user's maximum. An absolute *limit* for p can be registered by the user, and if p of an activity exceeds the limit, the activity is classified as *impossible*.

The d in $(max_p - d)$ represents a deviation from the max_p value and is used to adjust the upper boundary for normal activities. For example, if max_p is far above the average value for p, max_p may be too high to count as normal activity. Deviation d may be adapted to the user's current activity pattern by, e.g., being calculated as a function of average.

An *unlikely* activity is an activity the user has never done, has been repeatedly rejected when suggested in a nudge, or has one or more properties that are within the $Challenge(p,U)$ interval. An activity is *impossible* if capability or opportunity issues hinder it, or if at least one activity property is above the *limit*. Otherwise, it is a *probable* activity.

4.2.3 Green Transportation Nudging

As the user may have different habits when traveling between different (O,D) pairs, we identify *probable, unlikely*, and *impossible* transportation means per (O,D) pair. Walking may, e.g., be classified as unlikely for user U when going shopping, while it is possible when going to work.

Impossible transportation means are identified based on user capability and opportunity, and $impossible(A_{(O,D)},U)$ can, e.g., include i) *private car*, if the user does not have a driver's license, ii) *train*, if there are no train services for (O,D), or iii) *cycling*, if the user does not have a bicycle.

Transportation means where the user is physically active (e.g., walking or cycling), can be conditionally possible. Similarly to the approach in Section 4.2.2, an absolute *limit* for an activity property p can be registered by the user. If walking distance between O and D is above the limit, walking is impossible for (O,D).

To determine what is physically challenging for a user, a collective activity pattern covering all (O,D) pairs is used to determine how frequent physical activity is used as transportation means, and average and maximum values for distance and/or duration properties. These values determine similar intervals to the ones presented in Formula 11, and is used as a basis for classifying activities, as described in Section 4.2.2.

$Unlikely(A_{(O,D)},U)$ includes possible transportation means that U has never used, repeatedly rejected for (O,D), or, if it is a physical activity transportation means, has a property p within $Challenge(p,U)$.

4.3 Selecting Activity and Time Frame

4.3.1 General Description

To help the user reach the *ActivityGoal*, a nudge can either suggest a new activity that contributes to filling the *ActivityGap*(P) or suggest an improvement of an expected activity (e.g., suggest a longer walk or improve a predicted transportation means).

An *unlikely* activity may be selected to challenge the user to do something new or more demanding, while a *probable* activity, representing the familiar or preferred, can be selected to make it more likely that the user accepts the nudge.

Time frame and activity are closely connected, as the time frame for a nudge determines which activities are possible to suggest. An obvious requirement is that the selected activity must be possible to do within the targeted time frame. For some nudges, the time frame is first selected, and an activity that fits the time frame is subsequently chosen. For other nudges, the order is reversed, and activity is chosen before the time frame.

To fill the *ActivityGap*(P), the activity must take place in P, meaning that the time frame for the nudge must be within P. If a nudge is given with a time frame outside P and the nudge is accepted, the suggested activity will later be included as an expected activity in a future time period.

4.3.2 Physical Activity Nudging

The *ActivityGap*(P) represents the number of steps that remains in period P to reach the *ActivityGoal*. To fill the gap, the user needs to be more active, and nudging will suggest additional activities to the user.

Different activities (e.g., frequent short walks or a long hiking trip) can be equally useful as long as the suggested activity is possible for the user to do. That is, if a nudge suggests an activity with certain properties p, such as distance or duration, each property must be within the user's $Normal(p,U)$ or $Challenge(p,U)$ interval.

What to nudge for also depends on the user's situation (such as available time) and preferences, and which nudges the user reacts positively to. Therefore, when selecting an activity from either of the two sets $Probable(A,U)$ and $Unlikely(A,U)$, these factors must be considered. Additionally, the principle of *variation* (described in Section 2.1) can be supported, by, over time, nudging for a variety of activities.

When selecting a time frame for a nudge, the time since the last activity and time frame for committed activities are important factors. Time frames should not overlap, and there must be a sufficiently large interval between the end of one activity and the time-frame for the next.

4.3.3 Green Transportation Nudging

While physical activity nudging can fill an $ActivityGap(P)$ by adding more activities, green transportation nudging can only reach the $ActivityGoal$ by improving the EF-value of activities, i.e., selecting a transportation means that is more environmentally friendly than what the user normally prefers. The number of travels cannot be adjusted to fill an $ActivityGap(P)$. Travels are triggered only by the user's need to change location.

A location change is detected by i) monitoring the user's activity pattern to detect regular location changes, ii) using calendar information to detect appointments that require location change, and iii) letting the user register a location change.

The goal is to nudge the user to choose transportation means so that, when period P is over, $ActivityGap(P) \leqslant 0$ (i.e., the $ActivityGoal$ is reached). A nudge should suggest a transportation means that brings a positive $ActivityGap(P)$ value closer to 0, or keeps $ActivityGap(P) \leqslant 0$.

The system must detect which (O,D) pair the user can improve, and determine which improvement to suggest. If user U has previously traveled (O,D), there exists an expectation of which transportation means the user will choose and a corresponding predicted EF-value. To improve, a transportation means with a better EF-value must be selected from either of the sets $Probable(A_{(O,D)}, U)$ or $Unlikely(A_{(O,D)}, U)$.

Available transportation means are partially ranked based on their EF-values. This means that some transportation means (e.g. walking and cycling) are equally environmentally friendly and can have the same EF-value.

5 DISCUSSION

This section discusses some observations made during the work on smart nudge design. It includes how relevant activities can be determined ahead of nudge design, to make the design process more efficient, and how a system can create plans for nudging, either as a set of individual nudges or a succession of linked nudges. This section also shows how the smart nudge design fulfills the design principles presented in Section 2.1.

5.1 Predetermining Relevant Activities

To simplify the task of choosing an activity, a set of only relevant activities (denoted $RelActivities$) can be identified. To make nudging more efficient, the content of $RelActivities$ can be determined in advance, before nudge design time. Activities that are permanently impossible for the user, are the most obvious activities to be excluded from $RelActivities$.

A nudge must include an activity that represents an improvement, or at least a status quo, with respect to user behavior. In general, this means that all activities that represent a decrease in behavior should never be nudged for and are consequently excluded from $RelActivities$.

For green transportation nudging, the partial ranking of activities makes it possible to identify activities that represent a decline in user behavior and disregard them as not relevant for nudging. For example, if the user is predominantly walking to work, a nudge can suggest walking or cycling, but will never suggest public transportation or carpooling (since this represents a decline in behavior).

For physical activity nudging, where every activity contributes to reaching the $ActivityGoal$, it may be more difficult to discard activities as not relevant. However, for a relatively active user, the system may choose never to nudge for short or very easy activities, as these may not be considered a push towards improved behavior.

5.2 Creating a Nudging Plan

The $ActivityGap(P)$, described in Section 4.1, represents a prediction of how much the user must improve to reach the $ActivityGoal$. When predicting a deficiency in user behavior, the system can set up a plan for nudging so that the user can reach the goal during period P.

A nudging plan consists of a set of nudges $\{n_1, \ldots, n_m\}$, where each nudge can be predesigned, including a tentative activity, time frame, and other nudge components. When it is time to nudge the user, the final nudge design is done, and tentative components may be replaced if the user situation at nudge design time makes it necessary or beneficial. The nudging plan may be adjusted during P as the user situation may change or the $ActivityGap(P)$ is updated and the prediction changed.

A physical activity nudging plan can be created by detecting periods, during P, when the user is available for being active, and planning a nudge for each selected time frame. The planned nudges are spread over the period, and design and presentation of a nudge are done close to the selected time frame.

A green transportation nudging plan can be created by identifying planned travels (i.e., (O,D) pairs) during *P*, where the user's transportation habits can be improved, and planning a nudge for each selected (O,D) pair.

5.3 A Sequence of Nudges

Ranking of activities can be used as a basis for designing a sequence of linked nudges, where the first nudge suggests a higher-ranked activity than the next nudge in the list. When one of the linked nudges is accepted, the following nudges in the list are discarded.

For example, when nudging for transportation means between *O* and *D*, the system can first nudge for walking, and later, if the user did not walk, issue a new nudge for taking the bus. This gives a list of linked nudges, where walking has a better EF-value compared to public transportation.

5.4 Fulfilling Nudge Design Principles

The nudge design approach presented in this paper, follows the design principles described in Section 2.1. The principles of challenge, consolidate, and progress describes how nudging should stepwise challenge the user to improve behavior. The need to *challenge* the user is supported by automatically increasing the *ActivityGoal* value as the user's behavior improves. The *consolidate* and *progress* principles are handled using the three behavioral progress states (*improved, stable, decreased*), where the *consolidate* principle is followed when the user is in the *improved* state, while the *progress* principle is used for the states *stable* and *decreased*.

The *variation* principle is supported by keeping a set of possible activities (i.e., $Possible(A,U)$) and recognizing that an activity can be selected from either of the two sets $Probable(A,U)$ and $Unlikely(A,U)$, as long as the activity represents an improvement or a status quo with respect to user behavior.

Timeliness is supported by detecting time frames when nudging is needed or can be effective for the user. For physical activity nudging, an opening in the calendar can be utilized for exercising, while for green transportation nudging an arrival time at the destination, together with the selected activity, sets a required time frame for a nudge.

Feasibility is supported through classification of activities, where impossible activities will never be nudged for, while *safety* is supported by recognizing harmful or hazardous situations as obstacles that identify an activity as potentially dangerous and classify it as impossible.

6 CONCLUSION

This paper presents a model for adaptive nudge design, that provides personalized and context-aware nudges tailored to the user's current need for a gentle push towards a desirable change in behavior. The model follows a just-in-time nudge design approach, where tailoring of nudges is based on the user's behavior, situation, and environment at the time of nudging.

The design process adapts according to the user's change in behavior, by continuously challenging the user to improve based on an adaptive activity goal, which is automatically adjusted as user behavior improves. Activity and time frame for a nudge is selected based on what is needed for the user to reach the activity goal, and what currently is possible for the user.

This paper targets the questions of *what to nudge for* and *when to nudge*. Other aspects of nudge design, such as details regarding which influence type to select and how to present the nudge are left for future work.

The model is described through a general approach to selecting activity and time frame for a nudge, followed by examples of how the proposed solution can be used in two different use cases; physical activity nudging and green transportation nudging. Future work includes applying the general approach to other use cases. Presenting practical experiments using the design model is also left for future work.

REFERENCES

Anagnostopoulou, E., Urbančič, J., Bothos, E., Magoutas, B., Bradesko, L., Schrammel, J., and Mentzas, G. (2020). From mobility patterns to behavioural change: leveraging travel behaviour and personality profiles to nudge for sustainable transportation. *Journal of Intelligent Information Systems*, 54(1):157–178.

Andersen, A., Karlsen, R., and Yu, W. (2018). Green transportation choices with IoT and smart nudging. *Handbook of Smart Cities: Software Services and Cyber Infrastructure*, (1):331–354.

Bergram, K., Djokovic, M., Bezençon, V., and Holzer, A. (2022). The Digital Landscape of Nudging: A Systematic Literature Review of Empirical Research on Digital Nudges. *Conference on Human Factors in Computing Systems - Proceedings*.

Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P. (2019). 23 Ways to Nudge: A review of technology-mediated nudging in human-computer interaction. *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–15.

Caraban, A., Konstantinou, L., and Karapanos, E. (2020). The nudge deck: A design support tool for technology-mediated nudging. *DIS 2020 - Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 395–406.

Comission of the European Communities (2007). COM(2007) 551 final: Green Paper - Towards a new culture for urban mobility. *Green paper*.

Dalecke, S. and Karlsen, R. (2020). Designing Dynamic and Personalized Nudges. *ACM International Conference Proceeding Series*, Part F1625:139–148.

Forberger, S., Reisch, L., Kampfmann, T., and Zeeb, H. (2019). Nudging to move: A scoping review of the use of choice architecture interventions to promote physical activity in the general population. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):1–14.

Jesse, M. and Jannach, D. (2021). Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3(January):100052.

Karlsen, R. and Andersen, A. (2019). Recommendations with a Nudge. *Technologies*, 7(2):45.

Karlsen, R. and Andersen, A. (2022). The Impossible, the Unlikely, and the Probable Nudges: A Classification for the Design of Your Next Nudge. *Technologies*, 10(6):1–17.

Karlsen, R. and Andersen, A. (2024). How to Trigger Personalized and Context-Aware Nudges. *Presented at the 28th Int. Database Engineered Applications Symposium (IDEAS 2024), Bayonne, France, 28-31 August, 2024.*

Meske, C. and Potthoff, T. (2017). The dinu-model – A process model for the design of nudges. *Proceedings of the 25th European Conference on Information Systems, ECIS 2017*, 2017:2587–2597.

Michie, S., van Stralen, M. M., and West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1):1–11.

Mills, S. (2022). Personalized nudging. *Behavioural Public Policy*, 6(1):150–159.

Mirsch, T., Lehrer, C., and Jung, R. (2018). Making digital nudging applicable: The digital nudge design method. *International Conference on Information Systems 2018, ICIS 2018*, (2009):1–16.

Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., and Frik, A. (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior*, 109(November 2019):106347.

Purohit, A. K. and Holzer, A. (2019). Functional digital nudges: Identifying optimal timing for effective behavior change. *Conference on Human Factors in Computing Systems - Proceedings*, (May).

Schneider, C., Weinmann, M., and vom Brocke, J. (2018). Digital Nudging: Guiding Online User Choices through Interface Design. *Communications of the ACM*, 61(7):67–73.

Schöning, C., Matt, C., and Hess, T. (2019). Personalised nudging for more data disclosure? On the adaption of data usage policies format to cognitive styles. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2019-Janua(January):4395–4404.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Villalobos-Zúñiga, G. and Cherubini, M. (2020). Apps That Motivate: a Taxonomy of App Features Based on Self-Determination Theory. *International Journal of Human Computer Studies*, 140(April):102449.

WHO (2016). *WHO Guidelines on physical activity and sedentary behaviour*. WHO.

Automatic Transcription Systems: A Game Changer for Court Hearings

Alan Lyra[1], Carlos Eduardo Barbosa[1,2], Herbert Salazar[1], Matheus Argôlo[1], Yuri Lima[1], Rebeca Motta[1] and Jano Moreira de Souza[1]

[1]*Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*

[2]*Centro de Análises de Sistemas Navais (CASNAV), Marinha do Brasil, Rio de Janeiro, Brazil*

{alanlyra, eduardo, herbertsds, matheusargolo, yuriodelima, rmotta, jano}@cos.ufrj.br

Keywords: Automated Transcription, Court Hearings, Computing, Legal Documentation, Decision-Making Support.

Abstract: Lawsuits typically require a long time for resolution, and many court hearings may occur during the trial process. Legally, both parties must transcribe them and be open to the public if desired by the court. In court proceedings, a transcript is a record of all judges' decisions, the spoken arguments by the lawyers, and the depositions of the defense and witnesses. The scenario in Brazil is that for a long time, this process was manual, with a person responsible for the typing transcription. Today, with the electronic process, the court does not provide typed transcriptions anymore, but instead, the audio or video recordings of the hearings. In our work, we developed an automatic transcription solution for court hearings to obtain the best possible transcription considering current technologies' limitations, recording quality, participants' diction, and commonly used jargon in the legal sphere. With this work, we expect to ease this burdensome task with technical support and have a direct contribution to the legal environment.

1 INTRODUCTION

In the legal sphere, several testimonies and hearings resolve lawsuits. Historically, specialized individuals manually transcribed, making the process expensive and time-consuming. Technology can have many applications in court proceedings and facilitate a comprehensive end-to-end process, whether generating audiovisual media records or even processing information – transcribing (Nadaraj & Odayappan, 2020).

Organizing and managing these digital records can potentially contribute to court proceedings since this type of media can deliver better results, with the possibility of capturing subtle insights that cannot be written on paper, such people reactions.

We believe that transcribing hearing content allows easier assimilation of all procedural content by those who come to work with the proceeding. This approach does not exclude audiovisual recording but can highlight significant sections in procedural documents, including citations that can be easily referenced and classified (Gomes & Furtado, 2017).

Courts have applied new technologies and services to manage information and records of vast legal content. Services such as e-Litigation ("Integrated Electronic Litigation System," 2020) and e-Discovery ("Electronic Discovery," 2020) are examples that contribute to this approach. With the COVID-19 pandemic, the digital court is being experimented with and evolving daily to a more digital style (Nadaraj & Odayappan, 2020).

This work aims to organize the data generated from court hearings through a knowledge management process and transcribe the media using an automatic transcription system. We also try to refine this automated process to reach the highest level of accuracy through collaboration mechanisms.

This transcript, associated with the trial process and the corresponding media, will help significantly unfold a court lawsuit. In the current scenario – in Brazil's legal sphere – reports are generated after court hearings for decision-making and motions. However, reports and minutes – the common strategies to keep the memory of a court hearing – are formal and often limited, mostly when the institution handles extensive multi-stakeholder legal processes.

According to Chiu et al. (2001), the minutes of a trial process are summaries of it, constituting a part of the procedural memory. Soon after a hearing, looking at the minutes to review and act based on that

Lyra, A., Barbosa, C., Salazar, H., Argôlo, M., Lima, Y., Motta, R. and Moreira de Souza, J.
Automatic Transcription Systems: A Game Changer for Court Hearings.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 163-174
ISBN: 978-989-758-716-0; ISSN: 2184-3228

information is often helpful. Even during a hearing, it may be useful to refer to something from an earlier point, for example, by asking a question about a previous event in the legal process. For this reason, we argue that proposing a system that presents information from the legal process with mechanisms that add value to the decision-making can contribute to advancements in the legal system.

One of the challenges faced in the solution is adapting the existing Portuguese language technologies in a world where English has more significant scientific potential and technology options. The second challenge is regarding transcription correctness. Regardless of the efficiency, errors will appear due to the recording quality, participants' diction, and even jargon commonly used in the legal field. Another crucial challenge for developing the transcription system is voice identification through participants' voice biometry, which enables the automatic creation of a dialog in text format from the hearing inputs.

2 LITERATURE

This section overviews the Brazilian legal process and how the resulting media are processed. We also discuss the essential elements of knowledge management.

2.1 Legal Process in Brazil

Understanding how the trial process and hearings work, with motions, judges, and different courts, is not simple for those not from the legal field. The step by step of the legal process can vary according to the matter involved (civil law, criminal law, tax, etc.). Generally, a legal process consists of a request from the author to resolve a conflict. From this, the judge will determine the presentation of reasons and the production of evidence and will decide to recognize the right of one of the parties.

In the Brazilian legal process, the court communicates all steps to the parties through publications in the *Diário Oficial* (the official journal of the Brazilian Government). Nowadays, lawyers and stakeholders can monitor these proceedings online, facilitating quicker and more efficient processing. Additionally, legal software like ProJuris (2024), available on the market, can optimize the court's progress.

The court hearing is a unique and significant event that defines the fate of a lawsuit. It is a complex, agile, and dynamic event and usually results in the lawyer's only chance to demonstrate the Constitution (Brasil, 1988), the impediment, modification, or extinction of a right.

Figures 1, 2, and 3 present a simplified business process model (BPM) (OMG, 2011) of the Brazilian Legal Process, showing the generic steps from beginning to end.

Court hearings are the primary audio/video media producer during the legal process steps. The proposed system aims to use that media to automatically transcribe its contents to support the law worker, enabling them to perform faster and more efficient work. Analyzing the BPM, we highlight that court hearings often occur more than once in a lawsuit. According to Rio de Janeiro's General Public Defender, Dr. André Luis Machado de Castro, and the General Public Subdefender, Dr. Denis de Oliveira Praça (personal communication, Ago 07, 2018), analyzing the transcription of court hearings offers a straightforward and fast way to make their decisions, presenting a significant improvement in the working dynamics in the legal environment.

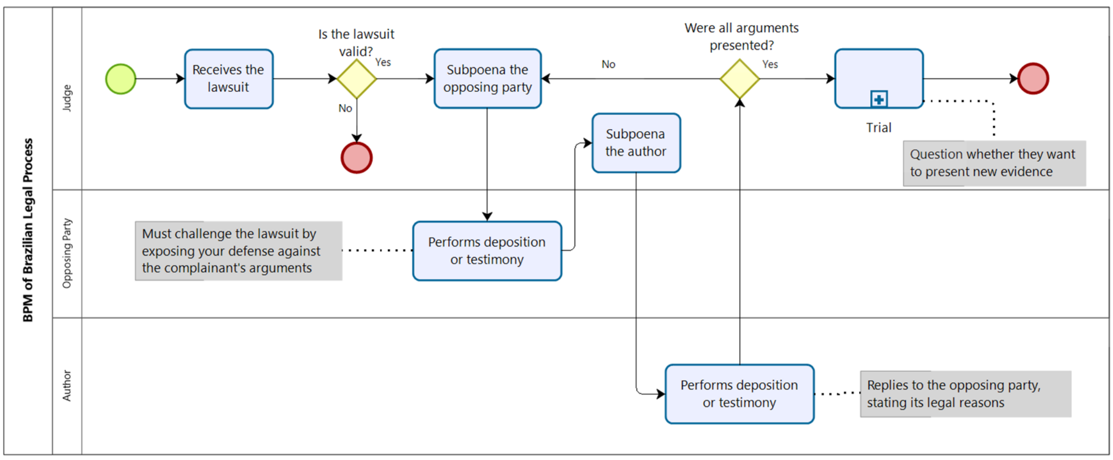


Figure 1: BPM of Brazilian Legal Process.

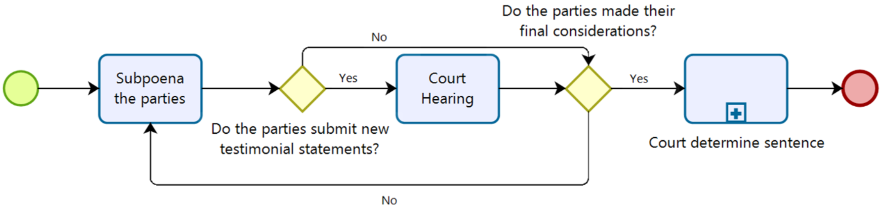


Figure 2: BPM of Brazilian Generic Subpoena Subprocess.

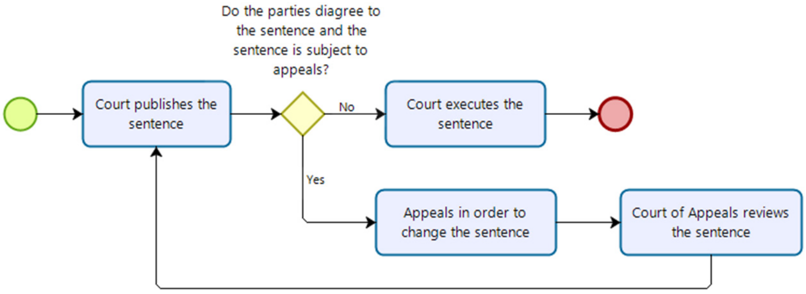


Figure 3: BPM of Brazilian Generic Verdict Subprocess.

Courts typically store media files from hearings in a digital lawsuit process. However, many legacy lawsuits still run on paper, with court hearing data stored in CD/DVD media. The most common type of lawsuit in Brazil that records court hearings is criminal. Civil cases rarely have their court hearings recorded.

2.2 Knowledge Management

Knowledge Management (KM) is a broad research topic, and as a discipline, it has several definitions, perceptions, and approaches. We understand KM as retrieving and organizing an organization's data and information to make it accessible to support decision-making and achieve strategic objectives through the discovered knowledge. The interest in and use of KM are increasingly expanding as more and more data are being produced. The KM is an area that permeates the companies' administration, information systems, management, and data science, with many research and practical and actual case initiatives for KM solutions (Patel, 2012).

Knowledge is a justified belief that enhances an entity's ability for effective action (Huber, 1991; Nonaka, 1994). Alavi and Leidner (2001) propose six different perspectives for observing knowledge: 1) Data and Information, 2) A state of mind, 3) An object, 4) A process, 5) A condition to have access to information, and 6) A capacity.

The first perspective of Data and Information shows only facts, raw numbers, processed information, and interpreted data. The second perspective of knowledge as a state of mind focuses on enabling individuals to expand their knowledge and apply it to the organization's needs. The third view defines knowledge as an object and postulates that knowledge can be seen as something to be stored and manipulated. Alternatively, knowledge can be seen as a simultaneous process of knowing and acting (Carlsson et al., 1996; McQueen, 1998; Zack, 1998), focusing on applying knowledge (Zack, 1998). The fifth view of knowledge is a condition of access to information (McQueen, 1998). According to this view, organizations must organize knowledge to facilitate access and retrieval of content, extending the knowledge base. Finally, in the latter perspective, knowledge can be viewed as a capacity that can influence future actions (Carlsson et al., 1996).

Along with understanding how we can observe knowledge, it is also paramount to have alternatives for managing it. Regardless of the definition, KM presents some core activities and central factors that structure any KM model (Stollenwerk, 2001). The generic KM process has the following activities (Alegbeleye, 2010; Dhamdhere, 2015; Mutula & Mooko, 2008):

- **Identification:** this activity focuses on strategic issues, such as identifying which competencies are relevant to the organizational context.
- **Capture:** in this activity, the objective is to acquire knowledge, skills, and experiences necessary to create and maintain the core competencies and areas of knowledge selected and mapped.

- **Selection and Validation:** this activity aims to filter knowledge, evaluate its quality, and synthesize it for future application.
- **Organization and Storage:** the goal is to ensure fast, easy, and correct knowledge recovery through effective storage systems.
- **Sharing:** this activity aims to ensure information and knowledge access to a more significant number of people that otherwise would remain restricted to a small group of individuals. Also, knowledge distribution refers to implementing a mechanism capable of automatically disseminating knowledge to various stakeholders to share new knowledge with those who need it quickly.
- **Application:** here, the objective is to put into practice the knowledge disseminated by the previous process, recording lessons learned from the use of knowledge, the benefits, and the challenges to be overcome.
- **Creation:** creating new knowledge involves learning, externalization, lessons learned, creative thinking, research, experimentation, discovery, and innovation. Many organizational activities can contribute to the creation of new knowledge. According to Nonaka (1994), knowledge creation is related to continuously transforming and adapting different types of expertise, such as practice and interactions.

Based on the objectives of this research area, the understanding and use of KM can support knowledge treatment in several places. In this work, we used these principles to guide this research related to the hearings of legal processes.

2.3 Related Work

Similar work was proposed by Huet (2006) when he created the Transcript Coding Scheme (TCS), which was a tool used to encode transcripts of corporate meetings – meetings – through the analysis of audio recordings, which remains the same.

Huet (2006) states that the data collected and explained in the transcription scheme offer the reader a vast amount of information, some of which may be considered critical and not captured using traditional reading techniques. However, processing and interpreting these highly detailed records require significant time and effort. Experiments have shown that accurately transcribing and encoding 30 minutes of recording takes approximately 10 hours.

The scheme proposed by Huet, to be better understood, can be divided into two parts:

- **Transcriber:** This step explicitly shows the transcription basics, stating the user identification, the speech transcription, and the timestamp when the speech occurred.
- **Coding Scheme:** This step provides information that aims to complement the transcript to offer more details to facilitate the decision-making process: type of speech; the purpose of speech; type of information provided; the argument; type of argument; participant’s knowledge area; description; and origin of speech.

Another work concerning transcriptions is by Williams et al. (2011), which proposes crowdsourcing techniques in complex transcriptions that allow explicit exchanges between precision, recall, and cost. These techniques include automatic transcription, incremental collaboration redundancy, and a regression model to predict the transcription’s reliability. The main idea is to use the crowd to assist in the final result of a transcription performed by an automatic transcription system (Parent & Eskenazi, 2010) based on requests for Amazon’s MTurk (Crowston, 2012). The techniques of redundancy and regression are defined by refining the process to validate a collaboration through the redundancy of correct answers through algorithms.

3 THE PROPOSED KM PROCESS

The developed KM model manages the information associated with the court hearing and produces a faster and better understanding for law workers. Analyzing the KM activities related to lawsuit trials, we understand that knowledge creation occurs during court hearings (opening statements, witness examination, closing arguments, and jury verdicts). We propose a Knowledge Management Model for Court Hearings, inspired by the study of Motta et al. (2022), which is focused on identifying, acquiring, and storing knowledge, as presented in Figure 4.

We detail each step in the proposed process:

- **Hearing Preparation:** The court hearing occurs when there is a need to collect information about the lawsuit from the parties or witnesses. The judge decides when and who will attend the court hearing.
- **Hearing Conduction:** This is a hearing instance. In a lawsuit, several hearings are made and recorded. This work relies on hearing the recording.

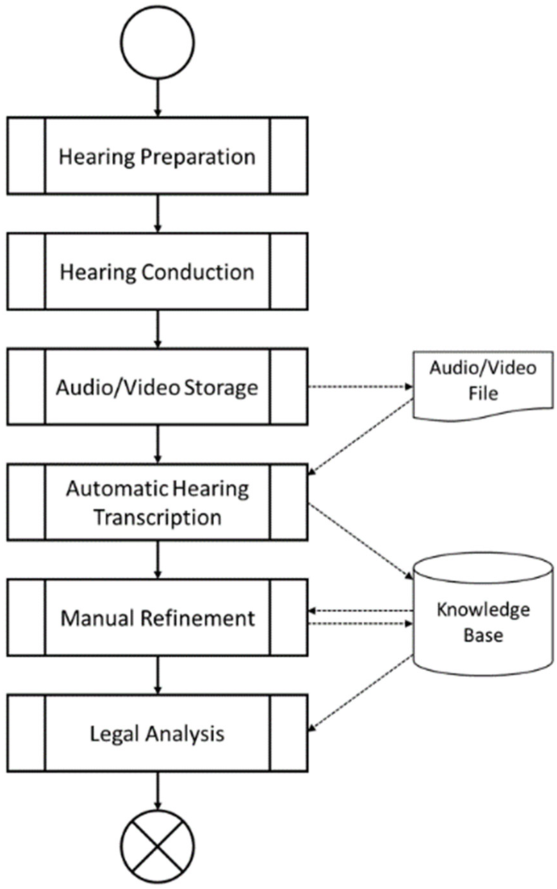


Figure 4: Knowledge Management Process for Court Hearings. Adapted from Motta et al. (2022).

- **Audio/Video Storage:** the default way to capture the hearing is to record audio or video media. Such an approach is simple to implement and frees the members of the accuracy of the stored hearing data – audio/video media is precise to solve any doubt about a hearing.
- **Automatic Hearing Transcription:** In this step, the system provides an automatized transcription of the hearing and determines the speaker of each sentence from the hearing audio/video. All knowledge will be stored in a shared database, allowing legal memory (processes, hearings, transcripts, and the source of the significant decisions).
- **Manual Refinement:** after the transcript generation, the result becomes available to be evaluated and fixed by any member of the judicial apparatus or the population in general through collaborative tools. This step fixes errors due to poor audio quality, poor diction, or legal jargon.
- **Legal Analysis:** based on the transcription, independent of the manual refinement, the law worker (judge, prosecutor, or lawyer, for example) may proceed in the hearing analysis using the text instead of the audio/video. Using specific keywords, they can find particular moments from the hearing without watching the whole media, in many cases, more than once. Such ability enables faster analysis of the court hearings. Besides, since the transcription is temporally associated with the media, it may be seen as a proxy to the media, allowing searching a video for the keyword spoken, for example. Although the transcription may have errors, the source (audio/video) is accurate.

4 THE PROPOSED TRANSCRIPTION SYSTEM

The developed system aims to contribute to the knowledge management of hearings by their automatic transcription. This solution enables a party to visualize the legal process and its transcription, locate other related audiences, and conduct a textual search of the available hearings. The system seeks to facilitate the work of the legal professional who will analyze the process, providing features like:

- Display of the media corresponding to the audience or hearing;
- Automatic transcription of this media along with the participant identification;
- Caption inserted in the media for direct accompaniment in the video;
- Search for words associated with the legal process; and
- Edit the transcript manually.

The system automatically generates the transcription scheme at the end of a hearing, providing the transcription and identifying the participants. The system also associates the data with the legal process, participant role (i.e., defense), and trial results to enrich the hearing. The hearing process is recorded in a media file (audio or video) during its execution and saved in a database. The system automatically generates and indexes a transcript for each individually submitted hearing.

In investigating which transcription method to use, we considered the following factors: accuracy, cost, and support for the Portuguese language. We initially evaluated two open-source transcription software – Kaldi (Povey et al., 2011) and

Pocketsphinx (Huggins-Daines et al., 2006) – but they only support English. Next, we assessed the Google Cloud Speech-to-Text API (Google, 2020a), which supports many languages – including Portuguese. The Google API presented a lower error rate in our tests, and we selected it for our solution. We use Java (Arnold et al., 2005) and FFmpeg (Tomar, 2006) technologies to complement the Google API to implement the full solution.

4.1 Diarization

We implemented the speaker identification algorithm. Firstly, we developed a speaker diarization system based on a convolutional neural network capable of identifying speech in a speaker's frontal video without using the associated audio wave for use cases in which the latter is either low quality, noisy, or outright missing. For this purpose, we extracted facial landmarks from each video frame using a facial identifier present in the Dlib (King, 2009) library based on the Histogram of Oriented Gradients (HOG). We fed them into the neural network using Tensorflow (Abadi et al., 2016), as presented in Figure 5.

With the onset of the COVID-19 pandemic and societal restrictions such as the mandatory use of face masks, the technique developed to detect facial landmarks became unfeasible, rendering diarization impractical. Considering this, we shifted our approach by creating an audio-based diarization model.

Speaker diarization involves identifying different speakers in multimedia content to temporally separate them, defining who spoke when, and producing a script. Traditionally, audio analysis has tackled this problem by extracting features as I-vectors (Dehak et al., 2011) and subsequently clustering them. Typically, these audio processing algorithms rely on prior knowledge of the number of speakers involved in the audio.

Google researchers proposed the utilization of a Long Short-Term Memory (LSTM) for creating feature vectors, termed D-vectors (Wang et al., 2017). They applied a clustering technique called Spectral Clustering, which involves constructing an affinity matrix among the examples and applying specific refinement operations to consider the temporal locality of the data, aiming to smooth, normalize, and remove noise from the data. It enabled a dynamic identification of the number of speakers in the audio.

We trained the model using video datasets from public hearings. We extracted and converted the audio signals into 25-millisecond frames with a 10-millisecond step. During the testing phase, we used the Voice Activity Detection (VAD) algorithm to identify spoken segments in the audio. Extracting audio signals is the first step in preparing them for the I-vector learning algorithm due to the complexity of raw signals. Initially, filter banks of the signals are computed to isolate different frequency components. We chose to extract Log Mel Frequency Energy filter banks, which involve performing a Fourier Transform, mapping the spectrum to the Mel scale, and other post-processing steps, as shown in Figure 6.

We implemented two models for feature extraction: Gaussian Mixture Model (GMM) and LSTM. Traditional models for I-vector extraction typically use dimensionality reduction of the GMM super-vector.

The final step of our implemented model involves clustering the feature vectors (I-vectors or D-vectors) extracted in the previous step. The chosen clustering models were: Spectral Clustering (Wang et al., 2017), Density-based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), Agglomerative Clustering (Hastie et al., 2009), and Self Organizing Maps (SOM) (Kohonen, 1990) followed by K-Means (MacQueen, 1967). We used different libraries for each clustering model. Figure 7 illustrates the performed activities. We fine-tuned each clustering algorithm.

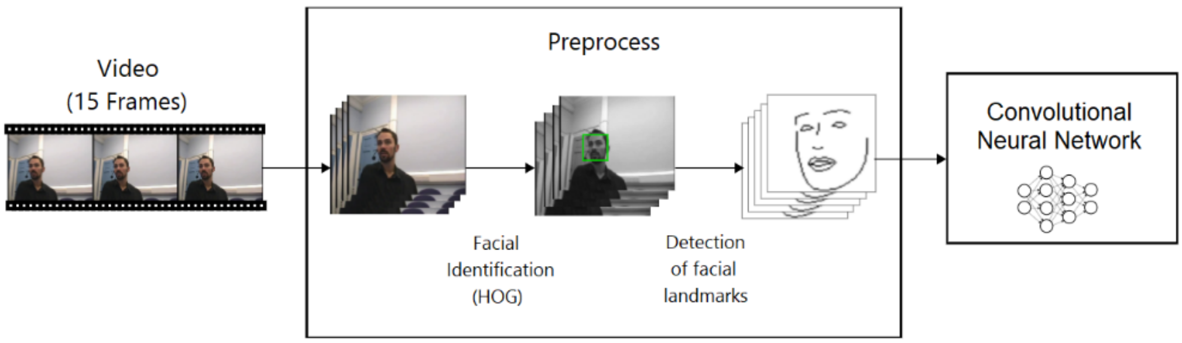


Figure 5: The sequence of data loading and model training.

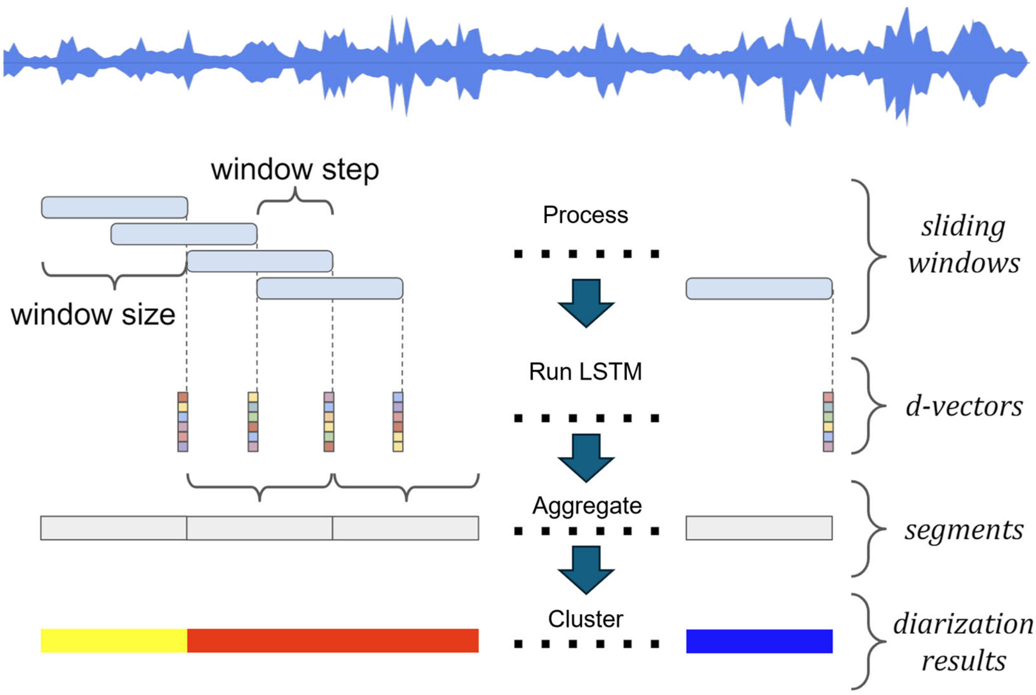


Figure 6: Flowchart of d-vector based diarization system, adapted from Wang et al. (2017).

The fine-tune process included:

- **Spectral clustering:** The percentile *p* for the Row-wise Thresholding and the *sigma* value for the Gaussian Blur refinement operation.
- **DBSCAN:** The *epsilon* represents the maximum distance between two samples for one to be considered in the other's neighborhood and the minimum number of samples in a sample to be considered a core point.
- **Agglomerative Clustering:** The criteria for cluster union and the metric used to compute this union.
- **SOM + K-Means:** Parameters such as algorithm initialization, training type, neighborhood type, and grid map format.

4.2 The Transcriptor Architecture

The architecture of the transcription system is shown in Figure 8, presenting the technologies used, interacting with the Database Management System (DBMS), and the system interface. The proposed transcription system continuously monitors the transcription, besides being responsible for collecting the metadata generated.

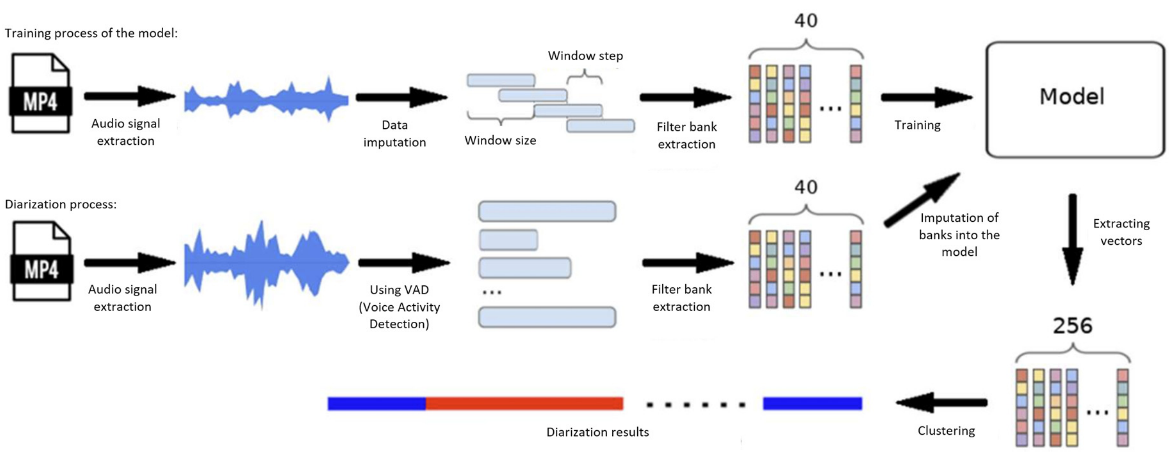


Figure 7: Flowchart of all steps of diarization system development.

The transcription begins when a user uploads new hearing audio/video to the system. The Transcriptor System detects the new file, extracts the audio, and uploads it to Google Cloud Storage. It then uses the Google Speech-to-Text API to transcribe the file and process the transcription results, storing them in the Transcriptor Data database.

Next, the Transcriptor System applies the Speaker Diarization Algorithm to identify the speaker for each transcribed word, storing the resulting metadata in the Transcriptor Data database. This database indexes the text and includes metadata such as the hearing identifier, the timestamp of each word, and the confidence level of the transcription.

The Transcriptor Interface provides user-friendly access to the transcriptions. Users can view the media (audio or video) with the transcription as subtitles, read the full transcription of the hearing, and perform searches within the text.

We highlight that Google's cloud for transcription through Google Speech API requires uploading audio files that are longer than one minute, which differs from the other two transcription methods, which are processed locally.

5 THE TRANSCRIPTOR IN USE

Developed as a legal system module in Brazil, the system stores digital lawsuit processes, stakeholders, and resolutions. Users can upload hearings in video format to obtain transcriptions. Users manually insert the description, date, participants, and media files on the upload screen. If users upload multiple media files for the same hearing, they can combine them into a single video. Users can submit videos from their computer or the Brazilian Judicial Court system.

After the file is uploaded, the transcription processes start automatically, running in the background, and take a few minutes. Depending on the video's size, the user can monitor the transcription status (in progress, error, or complete). The user can view the transcription after the processing step completion, as seen in Figure 9 (left).

The Visualize Hearing screen has the transcript of the selected hearing and the corresponding media side-by-side. The system presents the transcription as a caption attached to the video. The user can click on the word and be directed to the exact moment of the video when it was said. Also, by selecting editing mode, the alternative words defined in the transcript are displayed next to the option to correct the word when clicking on a word. The user can improve the transcription by selecting a present alternative or editing it as flowing text. The system also interpolates the metadata associated with each new word to compose a new transcription.

The user can improve speaker identification and add bookmarks to the transcript. Both features are accessible by selecting a word or sentence and the corresponding checkbox, as shown in Figure 9 (right).

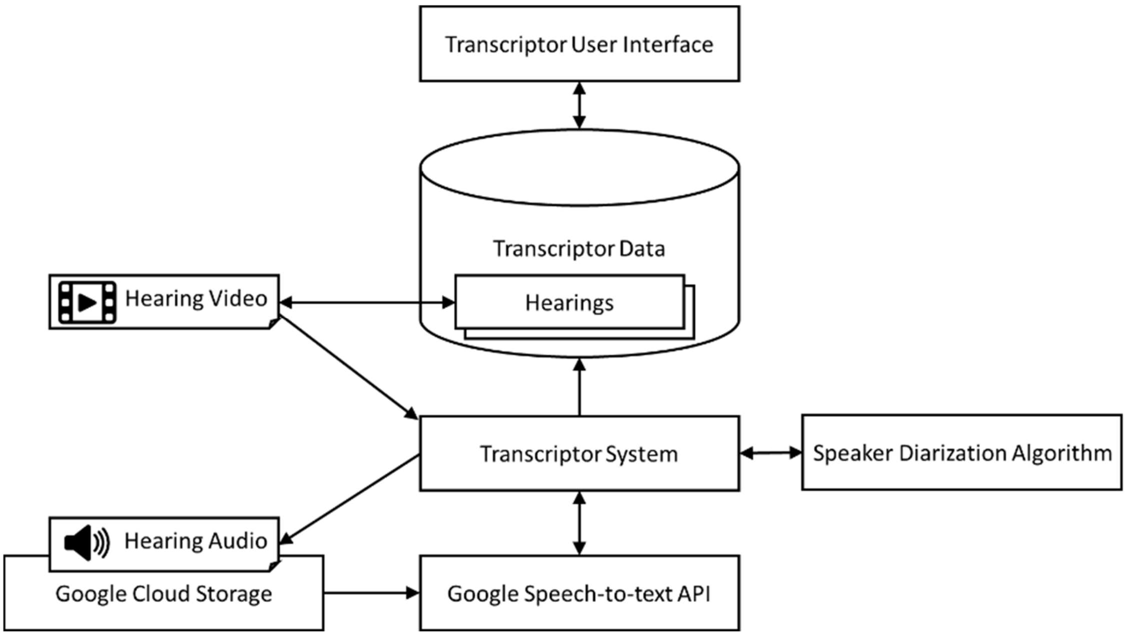


Figure 8: Transcription system of court hearings architecture.

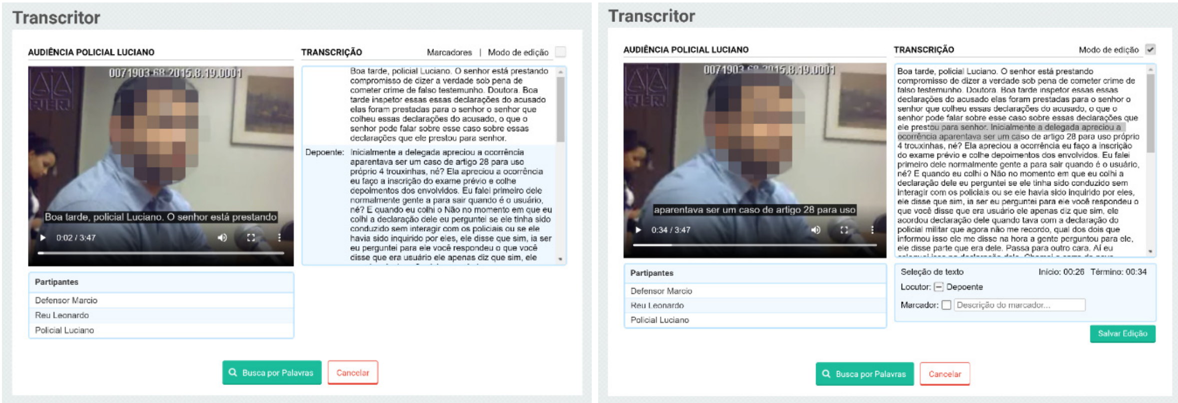


Figure 9: The visualization (left) and the editing mode (right) of a public Hearing transcribed.

Another feature available in the system is the word search. With it, the user can perform a textual search for any word present in the hearings transcribed by the system. The search can be general or filtered by a trial process.

The result lists the hearings in order, including the context in which they appear—predecessor and successor words—and the minute when the word was said in the media. The user can click on the "View full transcript" link to return to the main screen, and the system will present the complete transcript of the audience in question from the beginning. The user can also click the "View transcript at the specified time of the word" link to return to the main screen. The system will present the media and the complete transcript of the audience, starting a few seconds before the word searched is spoken. If the user searches for an alternative word, the result will display the description of an alternative word and the word transcribed by the transcription system. This functionality can facilitate the search and understanding of a hearing, for example, enabling a user to return quickly in the specific minute a sentence was given.

6 EXPERIMENTS AND RESULTS

Although the evaluation included three transcription methods, the implementation used the Google Cloud Speech-to-Text API (Google, 2020a) to provide a lower error rate and meet the Portuguese language without needing modifications. Moreover, the speech recognition technology provided by Google is continuously improving. The word recognition error rate in the company's last disclosure was 4.9% (Pichai, 2017).

Word recognition supported by Google machine learning achieved 95% accuracy in May 2017 for English. This rate also limits human accuracy (Meeker, 2017).

In tests performed using Google's API, the average confidence rate observed was ~ 0.92 for audios extracted from real audiences with good recording quality. The API itself provides this confidence rate for each transcription excerpt performed and assigned to each word corresponding to the excerpt.

The API provides more than one transcription alternative for certain words. Thus, corrections can be made based on the user's feedback through the system interface, using the alternatives provided or manual insertions. A list of other options must be stored in the DBMS with the corresponding section's indexation to return in a search for words or when requested in the system interface. Thus, the analysis of all knowledge obtained in the hearings can return the desired term, even if the transcription mechanism has not effectively transcribed that term. The beta version of the API also offers automatic scoring features (Google, 2020b) to improve the transcription result.

In the data transcription, the first issue was the language: Portuguese. Most algorithms are fine-tuned for English usage, and Portuguese models have demonstrated lower precision. We solved that issue by changing our transcription method to the Google Cloud Speech-to-Text API. Even so, Google's API presents issues: besides being a paid product, it demands uploading files in more than one minute to Google Cloud Storage – which brings the discussion about data security and privacy. We experimented with dividing the files into one-minute chunks to keep files locally. However, the splitting algorithm cut the audio abruptly, often cutting words, which impacted the transcription quality. Developing a smarter audio splitter was out of the question due to its complexity.

In speaker identification, before we develop our diarization model, we test the beta functionality of diarization of Google's Speak-to-Text API (Google, 2020a), which labels different voices identified in the audios by IDs. Voice training or prior registration is unnecessary since it does not recognize the person but changes the speaker or returns to the same. However, this functionality of Google's API needs to estimate how many people speak in a hearing to process the audio correctly. Using the incorrect value leads to errors in speaker identification. We do not have such data to provide to the algorithm.

We research other frameworks for speaker identification, such as MARF (Mokhov, 2008) and Microsoft Azure's Speaker Identification API (Microsoft Azure, 2020), which are based on neural networks to identify voice patterns. For this reason, these frameworks require a pre-trained set of possible speakers, and such algorithms associate the identified voice with the closest registered one. These algorithms cannot handle unknown voices; thus, every speaker must register in the system, and their voice must be trained previously.

In diarization, we focused our analysis on the vectors created by the extraction techniques and the performance of the implemented approaches using the diarization metric.

Initially, a dimensionality reduction was performed on the vectors generated by both chosen approaches to only two principal components, enabling their visualization in two dimensions to show the distance between each vector visually. Figure 10 (left) shows a close cluster in the D-vectors generated, making efficient clustering unfeasible. Conversely, the I-vectors – Figure 10 (right) – are more evenly distributed and farther apart. We highlight that the graphs shown in Figure 10 have different scales.

The Diarization Error Rate (DER) represents the fraction of total diarization time incorrectly attributed, whether due to False Alarms, Misses, Speaker Overlaps, or Confusion (Ryant et al., 2019).

$$DER=\frac{\text{False Alarm} + \text{Miss} + \text{Overlap} + \text{Confusion}}{\text{Time}} \quad (1)$$

Table 1 illustrates the performance of each implemented approach on the provided dataset. We also compare it with the previously obtained result using only speaker video images.

Table 1: Performance of each approach on the dataset.

Model	GMM	LSTM
Speaker video	32.5%	32.5%
Spectral Clustering	**21.79%**	80.90%
DBSCAN	57.24%	72.51%
Agglomerative	28.49%	84.47%
SOM + K-Means	65.39%	**67.85%**

As observed in Table 1, significant results were not obtained with the LSTM, suggesting that the model may not be learning the D-vectors correctly. As a neural network, LSTM is a complex model and more challenging to train, which may explain the difficulty in achieving significant results. On the other hand, the GMM generated I-vectors with substantial differences from each other, enabling efficient clustering and achieving a DER of 21.79% using Spectral Clustering. This result represents a 32.95% improvement over the previously designed model using only video images.

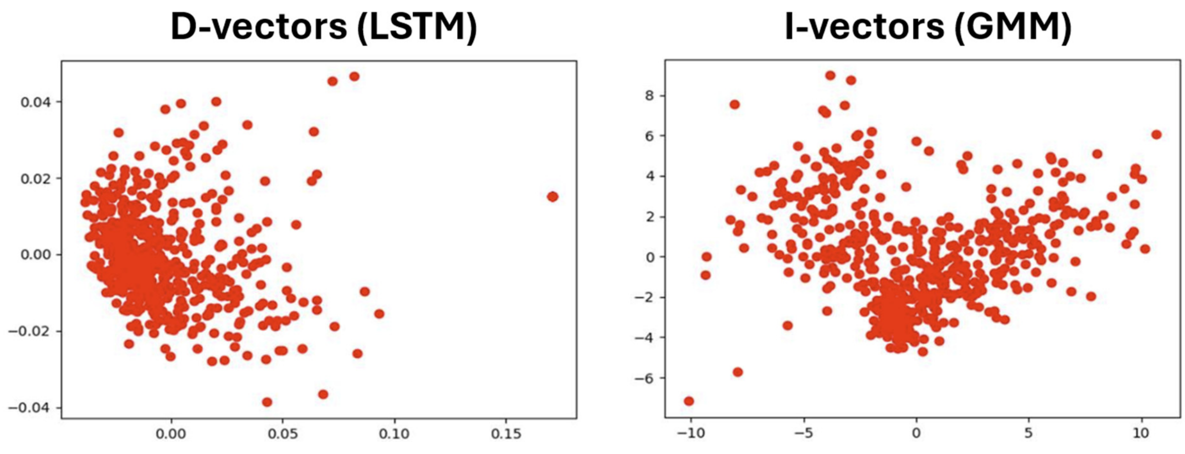


Figure 10: Dimensionality reduction of the D-vectors (LSTM) (left) and I-vectors (GMM) (right).

7 CONCLUSIONS

Legal proceedings involve conducting several testimonies and hearings. When performed, the transcriptions of these media are part of an expensive and slow process done manually by specialized people. Thus, few processes have their media records transcribed. In this work, an automatic transcription system was proposed, capable of organizing the media content generated in legal depositions/hearings based on a knowledge management model. The system allows collaboration to refine the automated process to achieve higher accuracy.

The system underwent an evaluation by legal members and proved to be quite helpful for its proposed use. It is currently in use in the Defense Office of Rio de Janeiro, with almost one hundred hours of court hearings transcribed. It can significantly contribute to the progress of legal processes and their decision-making. From a technical point of view, the transcripts had an average hit rate of 92% for audio extracted from real audiences with good sound quality. An automatic transcription system significantly contributes to the legal scope, streamlines the decision-making process, and maintains a legal memory that can be analyzed more quickly and efficiently later.

The limitations of this work are due to the existing technology for automatic transcriptions, which, for Portuguese, is still limited, in addition to features in Beta versions, such as automatic punctuation and announcer diarization. The latter contributes significantly to distinguishing interleaved speeches but still does not solve the problem of complete identification of the user, which implies voice training previously registered in the system to work for any user present in the hearings. This question is not viable for the type and quantity of users who participate in Brazilian legal processes. The quality of the generated media files also significantly impacts the result of the transcription.

In conclusion, the advancement of technology, such as better recording media quality and the Google API improvements for Portuguese, will enable a higher success rate in the future and the implementation of features still in beta. We highlight that this system must stay consistent and integrated into the legal system to effectively use and provide the desired legal memory easily accessible to law workers.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467 [Cs]*. http://arxiv.org/abs/1603.04467

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Alegbeleye, B. (2010). Old wine in new bottle: A critical analysis of the relationship between knowledge and library and information science. *Nigeria Library Association National Conference and AGM*, 18–23.

Arnold, K., Gosling, J., & Holmes, D. (2005). *The Java programming language*. Addison Wesley Professional.

Brasil, S. F. do. (1988). *Constituição da República Federativa do Brasil*. Senado Federal, Centro Gráfico.

Carlsson, S., Sawy, O. E., Eriksson, I., & Raven, A. (1996). *Gaining Competitive Advantage Through Shared Knowledge Creation: In Search of a New Design Theory for Strategic Information Systems.* 1067–1076.

Chiu, P., Boreczky, J., Girgensohn, A., & Kimber, D. (2001). *LiteMinutes: An Internet-based system for multimedia meeting minutes*. 140–149. https://doi.org/10.1145/371920.371971

Crowston, K. (2012). Amazon Mechanical Turk: A Research Tool for Organizations and Information Systems Scholars. In A. Bhattacherjee & B. Fitzgerald (Eds.), *Shaping the Future of ICT Research. Methods and Approaches* (Vol. 389, pp. 210–221). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35142-6_14

Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P. (2011). Front-End Factor Analysis for Speaker Verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(4), 788–798. https://doi.org/10.1109/TASL.2010.2064307

Dhamdhere, S. N. (2015). Knowledge Management Strategies and Process in Traditional Colleges: A Study. *International Journal of Information Library and Society*, *4*(1), Article 1.

Electronic discovery. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Electronic_discovery&oldid=969855680

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Knowledge Discovery and Data Mining*. https://api.semanticscholar.org/CorpusID:355163

Gomes, T. C., & Furtado, B. S. (2017, July 22). *A Justiça deve transcrever audiência de caso complexo*. Consultor Jurídico. http://www.conjur.com.br/2017-jul-22/opiniao-justica-transcrever-audiencia-complexo

Google. (2020a). *Cloud Speech-to-Text Documentation*. Cloud Speech-to-Text API. https://cloud.google.com/speech-to-text/docs/

Google. (2020b). *Getting automatic punctuation*. Cloud Speech-to-Text API. https://cloud.google.com/speech-to-text/docs/automatic-punctuation

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction. In *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed, pp. 520–528). Springer.

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, *2*(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Huet, G. (2006). *Design transaction monitoring: Understanding design reviews for extended knowledge capture* [PhD Thesis]. University of Bath.

Huggins-Daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M., & Rudnicky, A. I. (2006). Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices. *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, *1*, I-185-I–188. https://doi.org/10.1109/ICASSP.2006.1659988

Integrated Electronic Litigation System. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Integrated_Electronic_Litigation_System&oldid=935231346

King, D. E. (2009). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, *10*, 1755–1758.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. https://api.semanticscholar.org/CorpusID:6278891

McQueen, R. (1998). *Four views of knowledge and knowledge management.* 609-11.76.

Meeker, M. (2017, May 31). *Internet Trends 2017 Report*. 22nd edition of the Internet Trends Report at the Code Conference, Rancho Palos Verdes, California. https://www.slideshare.net/kleinerperkins/internet-trends-2017-report

Microsoft Azure. (2020). *Speaker Recognition | Microsoft Azure*. https://azure.microsoft.com/en-us/services/cognitive-services/speaker-recognition/

Mokhov, S. A. (2008). Introducing MARF: A Modular Audio Recognition Framework and its Applications for Scientific and Software Engineering Research. In T. Sobh (Ed.), *Advances in Computer and Information Sciences and Engineering* (pp. 473–478). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8741-7_84

Motta, R., Barbosa, C. E., Lyra, A., Oliveira, J., Zimbrão, G., & De Souza, J. M. (2022). Extracting Knowledge from and for Meetings. *2022 12th International Conference on Software Technology and Engineering (ICSTE)*, 82–90. https://doi.org/10.1109/ICSTE57415.2022.00019

Mutula, S., & Mooko, N. (2008). *Knowldge Management IN Information and Knowldge Management in the digital age:concepts, Technologies and African Perspectives*.

Nadaraj, P., & Odayappan, S. (2020). *Digital Courts: Are We Really Availing Infinite Possibilities Of Technology?* https://www.outlookindia.com/website/story/opinion-digital-courts-are-we-really-availing-infinite-possibilities-of-technology/351800

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37.

OMG. (2011). *Business Process Model and Notation (BPMN), Version 2.0* (p. 538) [Technical report]. Object Management Group.

Parent, G., & Eskenazi, M. (2010). Toward better crowdsourced transcription: Transcription of a year of the Let’s Go Bus Information System data. *2010 IEEE Spoken Language Technology Workshop*, 312–317. https://doi.org/10.1109/SLT.2010.5700870

Patel, D. S. (2012). *Management Discovery Of Knowledge Management In*.

Pichai, S. (2017). *Google I/O*. Google’s Annual Developer Conference, Montain View, CA.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motl´ıcek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. http://kaldi-asr.org/doc/

ProJuris. (2024). *ProJuris- Software para Escritórios de Advocacia de alto desempenho*. ProJuris. https://www.projuris.com.br

Ryant, N., Church, K., Cieri, C., Cristia, A., Du, J., Ganapathy, S., & Liberman, M. (2019). *The Second DIHARD Diarization Challenge: Dataset, task, and baselines*. https://doi.org/10.48550/ARXIV.1906.07839

Stollenwerk, M. de F. L. (2001). Gestão do conhecimento: Conceitos e modelos. In K. Tarapanoff, *(Org.) Inteligência organizacional e competitiva* (pp. 143–163).

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, *2006*(146), 10.

Wang, Q., Downey, C., Wan, L., Mansfield, P. A., & Moreno, I. L. (2017). *Speaker Diarization with LSTM*. https://doi.org/10.48550/ARXIV.1710.10468

Williams, J. D., Melamed, I. D., Alonso, T., Hollister, B., & Wilpon, J. (2011). *Crowd-sourcing for difficult transcription of speech*. 535–540. http://ieeexplore.ieee.org/document/6163988/

Zack, M. (1998, September). *An architecture for managing explicated knowledge*.

Data Governance to Be a Data-Driven Organization

Carlos Alberto Bassi[1][a], Solange Nice Alves-Souza[2][b] and Luiz Sergio de Souza[3][c]
[1]*Information System Post-Graduation Program (PPgSI), Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brazil*
[2]*Departamento de Engenharia de Computação e Sistemas Digitais, Escola Politécnica, Universidade de São Paulo, São Paulo, Brazil*
[3]*Faculdade de Tecnologia de São Paulo, São Paulo, Brazil*
cabassi@usp.br, ssouza@usp.br, luiz.souza71@fatec.sp.gov.br

Keywords: Data Governance, Data-Driven Organization, Data Governance Challenges, Data-Driven Challenges, Data Governance Culture, Data-Driven Culture.

Abstract: Many organizations have been trying to become data-driven which their business decisions, their relationships with customers / suppliers, the innovation of their products / services, the improvement in their performance and their growth are based on the collection and analysis of an increasing volume of data. To reach this level, organizations need to overcome a series of challenges related to the way they govern data and creating a data-driven culture. The main challenges to be overcome are directly related to data culture and the culture of the organization itself. These paper presents the results of a survey performed among 67 professionals with experience in Data Governance (DG) in which was possible to identify the main challenges to establish a DG program and data-driven culture in organizations, besides priories actions to face these challenges. The challenges and necessary actions to implement a DG program are shown and discussed. Addressing these challenges is fundamental to raising the organization culture and maturity in DG and, consequently, becoming a Data-driven organization.

1 INTRODUCTION

Data-driven Culture (DDC) is interpreted as a pattern of behaviors and practices by a group of people who share a belief that having, understanding, using data and information plays a crucial role in the success of their organizations (Chaudhuri et al., 2024).

DDC represents a specific form of organizational culture that is realized through data orientation. A DDC emphasizes that organizational decisions are grounded on insights from data, which fosters continuous knowledge and skills acquisition within the organization. Organizational culture encompasses a collection of values, beliefs, and attitudes held by organizational members. As it is based on organizational cultures, DDC's complexity lies in its need for consistency with decision-making principles (Fattah, 2024).

Becoming data-driven is about building capabilities, tools and most important a culture that is acting on data (Anderson, 2015).

In DDC data must be shared across the organization. DDC focus on setting goals, measuring success, interaction, feedback, learning and recognition of Data Literacy (DL) (Anderson, 2015).

Becoming data-driven is stated as one of the top priorities for organizations for the last 10 years. Numbers show clearly benefits of being data-driven. Companies that base their decisions on evidence are on average 5% more productive and 6% more profitable than their competitors (Storm and Borgman, 2020).

DDC influences business models and provides ways through which organizations develop their operations to secure higher profits (Chaudhuri et al., 2024).

Efficient and fast analysis of huge volumes of data has helped organizations make accurate decisions which could help influence innovative activities. Such an approach has also helped organizations revamp their business processes and develop smart products in relation to customer needs, eventually leading to increased profitability. Thus, the improvement of the organizational data-driven culture could lead employees to be more creative and gen-

[a] https://orcid.org/0009-0001-2524-7478
[b] https://orcid.org/0000-0002-6112-3536
[c] https://orcid.org/0000-0002-7855-0235

Bassi, C., Alves-Souza, S. and de Souza, L.
Data Governance to Be a Data-Driven Organization.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 175-186
ISBN: 978-989-758-716-0; ISSN: 2184-3228

erate novel ideas that could lead to the creation of new products to cater to the needs of dynamic markets (Chaudhuri et al., 2024).

Data Governance (DG) is the comprehensive management of usability, availability, Data Privacy and Security (DPS) and Data Quality (DQ) inside and outside the organization (Abraham et al., 2019).

It includes establishing policies, standards, processes, and structures to ensure the correct use and effective protection of data (Al-Dossari and Sumaili, 2021).

DG requires policy specification dynamics that can deal with problems related to the collection, storage, processing, sharing and use, reuse, and disposal of data throughout its life cycle (Filgueiras and Lui, 2022).

Many businesses are currently adapting digital strategies and new business models. Studies have shown that organizations of all sizes recognize the need for DG (Lis and Otto, 2020). Increasing importance has been given to incorporating DG as a means of encouraging the strategic use of data, thus promoting data-driven innovation (Lis et al., 2022).

The data-driven innovative capabilities have been enhanced by the applications of advanced Information and Communication Technology (ICT), strong analytic capabilities, effective data management and governance mechanisms (Chatterjee et al., 2024).

Beyond this, DG serves as a structured framework for organizations to manage data, recognizing data as a valuable corporate asset, thus promoting the optimal use of data. In practical terms, by organizing and mapping data-related processes within organizations, DG enables organizations to identify which data should be analyzed, unlock potential value, and overcome obstacles in developing DDC (Fattah, 2024).

DG plays a crucial role in encouraging and empowering the use of data analysis, such as DL, aligned with DDC. In this context, DG establishes decision making rights and accountability to ensure appropriate behavior in assessing, creating, using, and controlling organizational data, analytics, and information assets. Integrating DG with the overall business strategy and aligning it with data and analytical assets is considered critical for the organization's stakeholders (Fattah, 2024).

The findings indicate that DG and DL originally precede DDC. Consequently, it delivers a relevant theoretical contribution by providing empirical evidence and delving into the role of DG as a precursor to DDC. Interestingly, despite DG's crucial role in overseeing data pertinent to organizational decision-making, existing literature lacks clarity concerning the nomological network between DG constructs and other analytical capacities (Fattah, 2024).

This article consists of identifying that an organization will become data-driven when it creates a real DDC, and this will be possible through the implementation of effective DG.

2 CHALLENGES

2.1 Data-Driven Culture

DDC helps organizations to support process and product innovation eventually impacting the performance of the organizations (Chaudhuri et al., 2024).

Culture permeates every aspect of organizational actions, influencing decisions related to products, employees, customers, measurements, resource allocation, etc. While some leaders seek to embrace cultural norms using advanced technology, others may resist cultural change (Fattah, 2024).

There is necessity for shift the executive mindset by stimulating DDC and formulating strategies and mechanisms for governance (DG), as well as renewing the skills of analysts (DL) (Fattah, 2024).

Study conducted to identify the enabling factors of a DDC identified that (Berndtsson et al., 2018):

- established DG and access to data of good quality are mandatory for any type of analysis. If these features are not in place, then trust in business insights generated by various tools will deteriorate and undermine the move towards a DDC.
- top level management is important and needs to be actively involved in developing a strategy for establishing a DDC.

A survey conducted to identify challenges to creating a DDC pointed out the following (Storm and Borgman, 2020):

- a big challenge is taking away resistance to adopt a new technology.
- the struggle of creating insights derived from analysis.
- the complicated organizational structure.
- the lack of time forms a barrier among employees.
- the lack of knowledge occurs in departments where data is not originally embedded.

2.2 Data Governance

Organizations face challenges and problems in implementing a comprehensive and efficient DG program. In many cases, there is a lack of knowledge on the part of the professionals involved in conducting DG implementation projects, regarding which activities are necessary, who should be responsible for carrying out these activities, what the relationship and dependence are between these activities, as well as the impacts generated by not performing such activities properly (Bassi and Alves-Souza, 2023).

Several studies have found, among other reasons, that many of the bad decisions are due to the poor quality of the information generated from dirty, erroneous and incomplete data. This has led important companies worldwide to lose many thousands of dollars by managing information of low quality in their organization (Castillo et al., 2017).

Companies spend an average of 30% of their time on non-value-added tasks due to poor DQ and availability (Zhang et al., 2022). Key trends in DG indicate that by 2025, 80% of organizations seeking to scale digital businesses may fail if they do not adopt a modern approach to DG and analytics (Fattah, 2024).

Implementing DG is a complex project that requires long-term commitment and continuous engagement and, as such, organizations usually need to formulate a series of actions towards these goals (Zhang et al., 2022).

Mobilizing an organization to adopt DG has proven to be a challenge in practice. Taking stock of data inventory remains tedious, the potential for value creation seems abstract, and the importance of investing in DG is understood only if the company has already suffered major regulatory pressure or data breaches (Benfeldt et al., 2020).

There is a general lack for a clear understanding of what DG is and how it is currently implemented in companies (Krumay and Rueckel, 2020).

A comprehensive review of the scientific and practice-oriented literature shows a lack of understanding the activities required for introducing a DG program (Alhassan et al., 2019).

The implementation of a DG program implies a set of actions as presented by the Data Management Framework (Association, 2017) and Data Governance Institute (DGI) DG Framework (Thomas, 2024), highlighting the need to carry out:

- an inventory of assets, business processes, regulations regarding data processed by the organization.
- the constitution of a governance management committee.
- creating consistent and effective policies.
- the training of all employees in DG concepts, technologies and best practices and the creation of a data culture.

3 RESEARCH

Within the scope of a research to specify a guide with practical and coordinated actions that help organizations in the implementation of DG, a survey was carried out among professionals with experience in DG in two stages, the first through a web form and the second through an interview.

3.1 First Stage

A web form was created based on the challenges, impacts and solutions in the implementation of DG obtained from a literature review [reference - (Bassi and Alves-Souza, 2023)] to help evaluate the following points:

- validate that organizations carry out a series of actions in the context of implementing DG, among them, the mapping / survey / inventory of data, processes, regulations, and infrastructure; the creation and performance of a Management Committee; the development and application of DG, DPS and DQ policies, in addition to presentation / training in the concepts of DG and the policies.
- identify the importance, level of complexity and priority of such actions in DG.
- identify the main challenges that prevent organizations from carrying out such actions.
- identify the main challenges faced by organizations when carrying out such actions.

Table 1 in the *APPENDIX* displays the web form used in this stage.

It was estimated that a population of 81 research participants could represent at least organizations of different sizes (small, medium and large), in different stages of implementation (initial, advanced and planning) and in different market segments.

It is important to determine the size of a sample to represent the population of a research. According to Krejcie and Morgan, the recommended sample size would be around 82% of this population (Krejcie, 1970).

200 professionals who work with DG were invited to fill out a web form. 67 professionals answered the form and signed the consent form to participate in the research.

They have been working on the implementation of DG in organizations from the most different market segments and of the most varied sizes as show in Figure 1, Figure 2, and Figure 3.

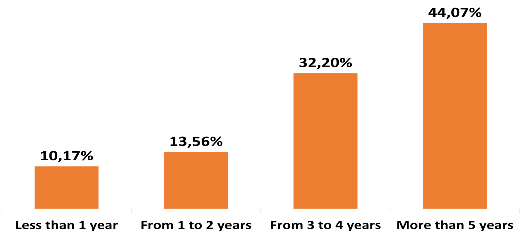


Figure 1: Experience in DG.

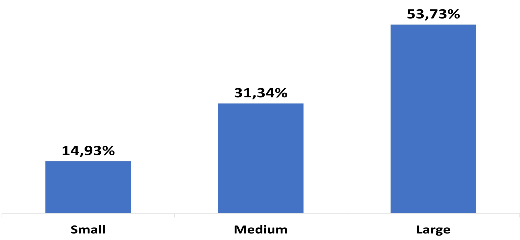


Figure 2: Size of the organization.

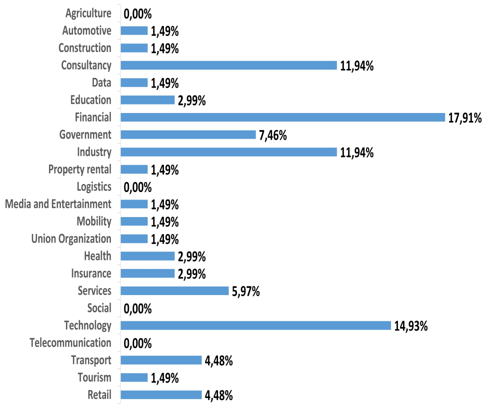


Figure 3: Organization's market segment.

3.2 Second Stage

38 professionals from the group who responded to the web form and who had more than two years of experience in DG were invited to participate in the interview phase (75% of the participating professionals have more than 3 years of experience in DG).

19 professionals agreed to participate in this second phase. A web form based on the results collected in the previous stage was designed to conduct an online interview, whose objective was to obtain a classification / prioritization of the main challenges identified at the first stage to that prevent from carrying out the actions and faced when carrying out the actions in order to help in prioritizing these actions in the implementation of the DG.

Table 2 in the *APPENDIX* displays the web form used in this stage.

4 ANALYSIS

Among the various challenges highlighted by professionals during the first phase of the research, it was possible to identify that the challenges listed below are directly related to cultural and organizational aspects that impact the implementation of effective DG and alignment with the organization's strategic objectives:

Resistance (RST) - professionals' reluctance to adopt new processes, policies, procedures, and tools due to concerns about changes in organizational culture, the perception of loss of control or impact on the activities performed or, the lack of awareness / training.
Employee Engagement (EEG) – active and motivated participation of employees in the process of understanding, adopting, executing, and effectively using practices, policies, procedures, and tools in their activities.
Culture / Knowledge / Empowerment (CKE) – forming a data-driven mindset, promoting understanding and effective use of practices and procedures through training and ongoing education.
Experience (EXP) - practical knowledge and in-depth understanding of practices, procedures, and tools by the organization's professionals.
Alignment / Communication (ALC) - ensure all stakeholders are coordinated and informed about policies, procedures, and objectives to promote effective collaboration and successful implementation.
Perception of Value / Benefits (PVB) – everyone involved recognizes the potential positive impacts, encouraging support, adherence and commitment to the initiatives are being implemented.
Management Support (MSP) – demonstration of leadership, commitment, and strategic direction to ensure adequate resources, organizational alignment and prioritization of initiatives are being implemented by the organization's executives.

4.1 Mapping / Survey / Inventory Challenges

The implementation of a DG program implies the need to know the data from the organization, the business processes that manipulate this data, the regulations, and standards that they must be followed and the infrastructure that supports the processing of the data.

To achieve this, the following actions are necessary:
Data Inventory (DI) - comprehensive identification, cataloging and documentation of all data that the or-

ganization collects, processes and stores, with the aim of understanding how that data is used, who has access to it and where the data is stored.

Process Mapping (PM) - identify, visualize, and document in detail all business processes, aiming to understand how data flows through the organization and identify points of intervention.

Regulatory Survey (RS) - identify and analyze relevant laws, regulations, internal policies, and external standards that impact data management, security, and privacy, ensuring legal compliance and risk mitigation.

Infrastructure Mapping (IM) - identify and document the structure of hardware, software, and data storage systems, ensuring a comprehensive understanding of the technological infrastructure that supports the management and manipulation of organizational data.

Figure 4 presents the *'Challenges that Prevent from Carrying Out'* and the *'Challenges Faced when Carrying Out'* the Mapping / Survey / Inventory (MSI). The x-axis of the figures shows the number of citations of the challenge by participants.

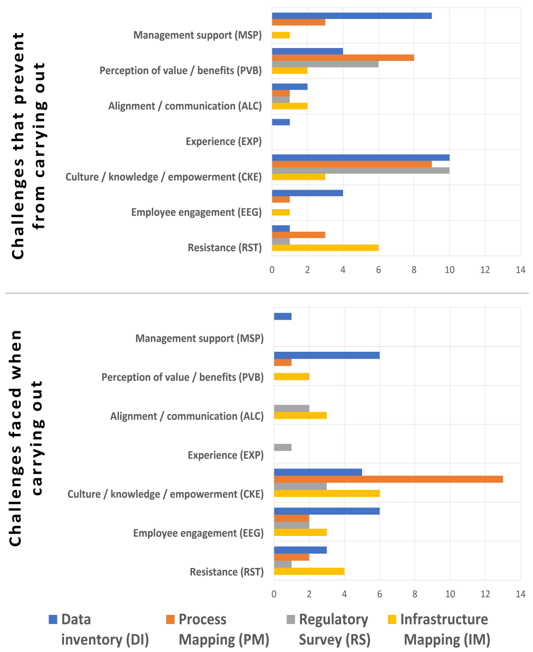


Figure 4: Challenges to carry out the mapping / survey / inventory (MSI).

The MSP, PVB and CKE are challenges that practically impact the carrying out of these actions as indicated by some research participants.

[P54] - *"Culture of the organization, people trained to implement the process and mainly lack of support from the organization's leadership."*

[P66] - *"The lack of technical knowledge about process mapping is one of the main challenges, due to the complexity of the processes and the level of maturity of the processes. Another point is not knowing what benefits this mapping will bring to the company objectively."*

The challenges CKE, EEG and RST impact the execution of these actions.

[P75] – *"Engagement of the organization to map concepts and metadata, size of the current Data Governance structure and understanding of the importance of the subject by board members."*

[P01] – *"Resistance and commitment of those responsible."*

When professionals were asked to classify / prioritize the most relevant *'Challenges that Prevent from Carrying Out'* and the *'Challenges Faced when Carrying Out'* for these actions, the following ranking was obtained as shown in the Figure 5.

The challenges MSP and the PVB are the main *'Challenges that Prevent from Carrying Out'* and *'Challenges Faced when Carrying Out'* in the execution of an MSI.

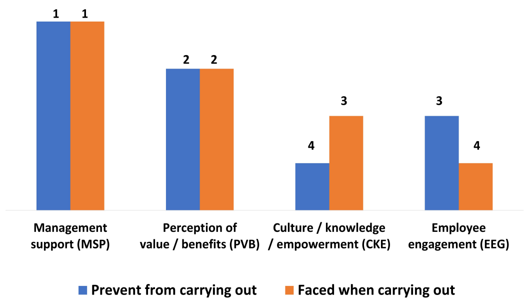


Figure 5: Ranking of challenges to carry out the mapping / survey / inventory (MSI).

4.2 Management Committee Challenges

The creation and effective performance of a management committee will have a significant impact on the implementation of DG.

The Data Office establishes vision, strategy and governance of initiatives, projects, actions of the council data, ensures the smooth functioning and management of data systems, guarantees ethic, responsibility, and lawful use of data. It consists of a multidisciplinary team, as well as an information and records manager and experts in efficiency, transparency, accountability, and life cycle management (Cerrillo-Martínez and Casadesús-De-mingo, 2021).

Figure 6 presents the *'Challenges that Prevent in the Creation'* and the *'Challenges Faced when in the*

Performance' of the Management Committee (MC). The x-axis of the figures shows the number of citations of the challenge by participants.

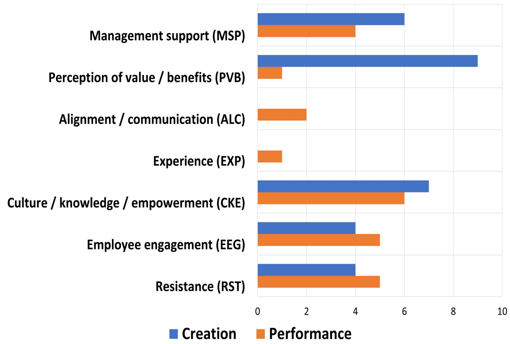


Figure 6: Challenges that prevent in the creation and faced when in the performance of the Management Committee (MC).

The MSP, PVB and CKE are challenges that practically impact in the creation of the MC while the challenges CKE, EEG and RST impact the performance of the MC.

[P68] – *"The lack of a data governance culture in the organization, which recognizes the strategic importance of data as an asset and a resource for decision making."*

[P55] – *"Management of conflicts of interest related to corporate vs departmental prioritization, lack of knowledge and experience, as well as resistance to change."*

Figure 7 presents the classification / prioritization carried out by professionals of the *'Challenges that Prevent in the Creation'* and the *'Challenges Faced when in the Performance'* of the MC.

The creation of the MC is impacted by the challenges MSP and PVB while your performance is impacted by the challenges PVB and CKE.

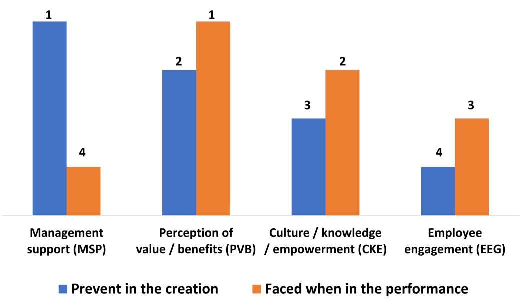


Figure 7: Ranking of challenges in the creation and faced when in the performance of the Management Committee (MC).

4.3 Policy Challenges

Companies specify guidelines and rules for the creation, acquisition, storage, security, quality and permitted use of data, developing standard processes and policies for data, and establishing employee organizations that support dedicated data governance activities (Zhang et al., 2022).

The following policies are relevant within the DG program:

Data Governance Policy (DGP) – set of principles, guidelines and procedures formally established to guide the management, quality, security, privacy, and ethical use of data, aligned with the organization strategic and regulatory objectives.

Privacy and Data Security Policy (PSP) – set of formal guidelines that establish the requirements and procedures to protect the confidentiality, integrity, and availability of data, ensuring compliance with privacy and security regulations, such as the General Data Protection Regulation (GDPR) - European Community and the Lei Geral de Proteção de Dados Pessoais (LGPD) - Brazil, and mitigating risks related to data breaches.

Data Quality Policy (DQP) – set of guidelines that defines standards, processes, responsibilities to ensure the reliability, accuracy, integrity of data throughout its life cycle, aiming to effectively support operations and decision making.

Figure 8 presents the *'Challenges that Prevent the Developing and Application'* and the *'Challenges Faced when Applying'* these policies. The x-axis of the figures shows the number of citations of the challenge by participants.

The challenges MSP, CKE and EEG are CPDA of the policies while the application of the policies is impacted by the challenges PVB, ALC, CKE and EEG.

[P27] – *"The preparation of data owners and data stewards to carry out their role within data quality."*

[P31] – *"The company's culture does not favor the practical application of a Data Governance policy."*

[P35] – *"Changing behavior and adopting a policy is always delicate and if it is not done correctly, all the work may have been in vain."*

Figure 9 presents the classification / prioritization carried out by professionals of the *'Challenges that Prevent the Developing and Application'* and *'Challenges Faced when Applying'* of the policies.

The challenges MSP, PVB and CKE impact on the developing and application of the policies.

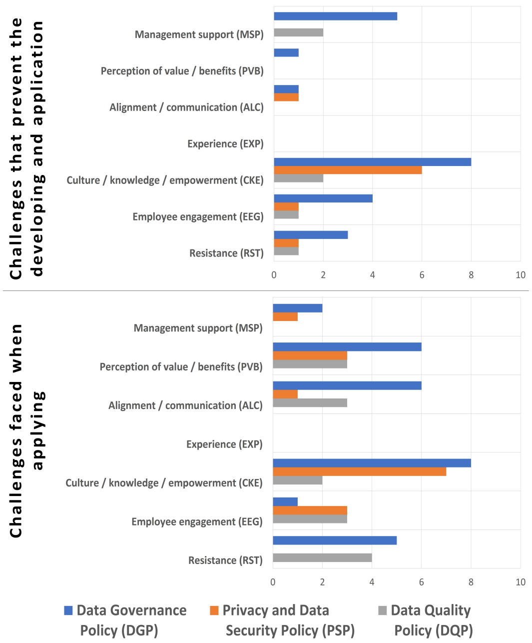


Figure 8: Challenges to develop and apply the policies.

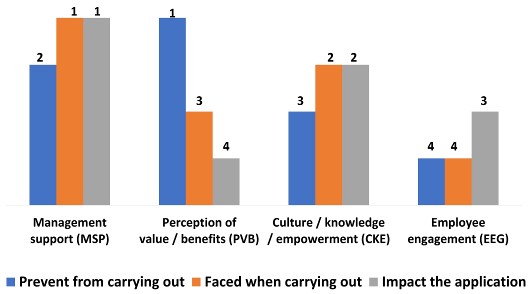


Figure 9: Ranking of challenges that prevent the developing and applying and in the application of the policies.

4.4 Presentation / Training Challenges

Different actions / interactions are recommended to ensure appropriate employee data competencies. The most important action / interaction is 'training', such as continuous training in dealing with and implementing data policies as well as data processes and procedures and includes internal and external training (Alhassan et al., 2019)

Figure 10 presents the *'Challenges that Prevent from Carrying Out'* and the *'Challenges Faced when Carrying Out'* the presentation / training (PT) on the concepts involved in DG and the elaborate policies. The x-axis of the figures shows the number of citations of the challenge by participants.

The MSP, PSB and CKE and employee engagement are challenges that impact the carrying out of these actions.

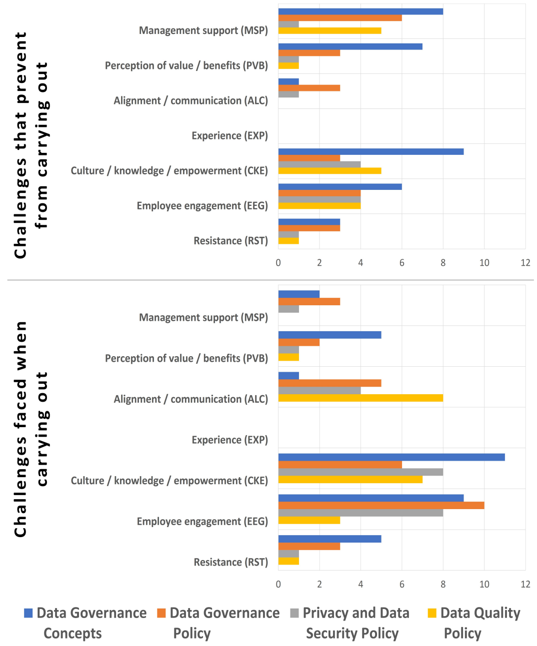


Figure 10: Challenges to carry out the presentation / training (PT).

[P31] – *"Include the training agenda in a broader literacy program with a focus on governance themes associated with business challenges."*

[P46] – *"The clear definition of a data strategy for the entire institution from the organization's management team and not from the information technology area."*

The execution of these actions is impacted by the challenges ALC, CKE and RST.

[P45] – *"Employees who do not participate in training, not all managers are concerned with Data Governance."*

[P68] – *"Measuring and demonstrating the benefits and results of the Data Governance Policy, which demonstrate improvements in the quality, security, privacy and value of data."*

Figure 11 presents the classification / prioritization carried out by professionals of the *'Challenges that Prevent from Carrying Out'* and the *'Challenges Faced when Carrying Out'* for the PT.

The challenges MSP and CKE are the main *'Challenges that Prevent from Carrying Out'* while the EEG and PVB are the *'Challenges Faced when Carrying Out'* for the PT.

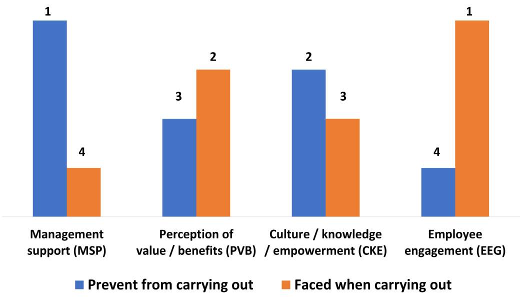


Figure 11: Ranking of challenges to carry out the presentation / training (PT).

5 DISCUSSION

It can be identified that cultural aspects are extremely relevant to the successful implementation of a DG program in organizations. These aspects comprehend a set of values, beliefs and behaviors that define how the organization conducts its business and how it treats its customers and partners (Bassi and Alves-Souza, 2023).

A Scoping Review (SR) of publications that present case studies (CS) on DG projects identified that the challenges related to cultural aspects like the lack of perception of the value of data as an asset and the lack of understanding and training of those involved in the concepts, technologies and best practices have significantly impact on the implementation of DG in organizations from different market segments and countries (Bassi and Alves-Souza, 2023).

The importance of incorporating DG as means of encouraging the strategic utilization of data, therefore promoting innovation and culture, is increasing (Lis et al., 2022).

One of the solutions identified in the literature for implementing DG is train and improve DL for all staff across organizations that participate in the project (Kawtrakul et al., 2021).

Implementing DG is a complex project that requires long-term commitment and continuous engagement and, as such, organizations usually need to formulate a series of actions towards these goals (Zhang et al., 2022).

A DDC should be developed, engaging the entire business, and sparking employee interest and motivation. DDC must be extended in an integrated way throughout the organization, instead of being segregated (Anton et al., 2023).

Data-driven organizations require strong, top-down data leadership. They need a leadership that inspires, promotes a data-driven culture, and actively drives and supports all aspects of the analytics value chain, from data collection through to data-driven decision making and institutional learning (Anderson, 2015).

Establishing clear policies for DG helps ensure the responsible and ethical use of data within the organization (Anton et al., 2023).

The implementation of a DG will effectively contribute to making an organization data-driven. It turns out that implementing a DG program involves overcoming a series of cultural and organizational challenges.

In the research performed with professionals who work in DG, it was possible to identify the four main challenges directly related to cultural and organizational aspects that prevent or impact the carrying out of the actions necessary to implement the DG program as shown in the Figure 12.

Previous knowledge of these challenges helps organizations that desire to implement effective governance of their data to better plan and take the necessary actions to mitigate the impacts that these challenges generate (Bassi and Alves-Souza, 2023)

The successful implementation of a DG program will create a DDC and, consequently, make the organization data-driven.

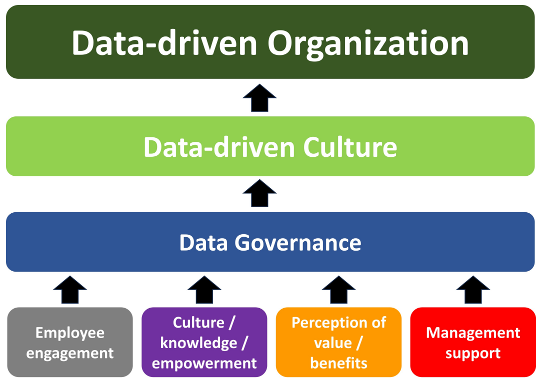


Figure 12: Cultural challenges of a Data Governance.

6 CONCLUSIONS

Many organizations are on the path to becoming data-driven. It turns out that there are a series of challenges that must be overcome in how they manage their data so that they can enjoy the benefits that data orientation offers.

The main challenges to be overcome are related to data culture and the culture of the organization itself. Management support will be fundamental in all actions involved in implementing a DG program.

The perception of value and benefits that data has and that its governance offers enhances the creation of this data-driven culture.

Management support and the perception of value / benefits will have a direct impact on employee engagement in all actions necessary to implement data governance and culture.

Carrying out training, awareness, and engagement of all employees in the concepts and activities involved with data governance, the policies, the procedures, and methodologies necessary to perform all necessary actions will facilitate the perception of value / benefits for everyone in the organization, facilitating the allocating of financial resources and human needed.

The research conducted with professionals allowed us to obtain details related to the challenges in implementing DG. In addition to the cultural and organizational challenges discussed in this paper, there is information related to data, organizational structure, technology support, policies and DG implementation projects that will be explored in other papers.

Many of the results presented here were used to validated the specification of a guide with practical and coordinated actions that help organizations in the implementation of DG, overcoming the main challenges, evolving in their maturity and, thus, creating a culture to obtain the desired benefits to be a data-driven organization. This guide will be presented in another article as soon as the research is completed.

Implementing the actions specified in this guide in some organizations will allow us to assess how much the implementation of DG contributes to creating a data-driven culture and, consequently, making these organizations data-driven in contrast to organizations that have not implemented DG.

REFERENCES

Abraham, R., Schneider, J., and vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda.

Al-Dossari, H. and Sumaili, A. A. (2021). A data governance maturity assessment: A case study of saudi arabia. *International Journal of Managing Public Sector Information and Communication Technologies*, 12:19–30.

Alhassan, I., Sammon, D., and Daly, M. (2019). Critical success factors for data governance: A theory building approach. *Information Systems Management*, 36:98–110.

Anderson, C. (2015). *Creating a data-driven organization : practical advice from the trenches*. O'Reilly Media, Inc.

Anton, E., Oesterreich, T. D., Aptyka, M., and Teuteberg, F. (2023). Beyond digital data and information technology: Conceptualizing data-driven culture. *Pacific Asia Journal of the Association for Information Systems*, 15.

Association, D. M. (2017). *DAMA-DMBOK - Data Management Body of Knowledge*. Technics Publications, 2nd edition edition.

Bassi, C. A. and Alves-Souza, S. N. (2023). Challenges to implementing effective data governance: A literature review. In *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K - Proceedings*, volume 3, pages 17–28. Science and Technology Publications, Lda.

Benfeldt, O., Persson, J. S., and Madsen, S. (2020). Data governance as a collective action problem. *Information Systems Frontiers*, 22:299–313.

Berndtsson, M., Forsberg, D., Stein, D., and Svahn, T. (2018). Becoming a data-driven organisation.

Castillo, L. F., Raymundo, C., and Mateos, F. D. (2017). Information architecture model for data governance initiatives in peruvian universities. In *9th International Conference on Management of Digital EcoSystems*, volume 2017-January, pages 104–107. Association for Computing Machinery, Inc.

Cerrillo-Martínez, A. and Casadesús-De-mingo, A. (2021). Data governance for public transparency. *Profesional de la Informacion*, 30.

Chatterjee, S., Chaudhuri, R., and Vrontis, D. (2024). Does data-driven culture impact innovation and performance of a firm? an empirical examination. *Annals of Operations Research*, 333:601–626.

Chaudhuri, R., Chatterjee, S., Mariani, M. M., and Wamba, S. F. (2024). Assessing the influence of emerging technologies on organizational data driven culture and innovation capabilities: A sustainability performance perspective. *Technological Forecasting and Social Change*, 200.

Fattah, I. A. (2024). The mediating effect of data literacy competence in the relationship between data governance and data-driven culture. *Industrial Management and Data Systems*.

Filgueiras, F. and Lui, L. (2022). Designing data governance in brazil: an institutional analysis. *Policy Design and Practice*.

Kawtrakul, A., Chanlekha, H., Waiyamai, K., Kangkachit, T., D'Orazio, L., Kotzinos, D., Laurent, D., and Spyratos, N. (2021). Towards data-and-innovation driven sustainable and productive agriculture: Bio-agri-watch as a use case study. In *International Conference on Big Data*, pages 3407–3415. Institute of Electrical and Electronics Engineers Inc.

Krejcie, R. V. (1970). Determining sample size for research activities.

Krumay, B. and Rueckel, D. (2020). Data governance and digitalization – a case study in a manufacturing company. In *Pacific Asia Conference on Information Systems Proceedings*, pages 6–22. AIS Electronic Library.

Lis, D., Arbter, M., Spindler, M., and Otto, B. (2022). An investigation of antecedents for data governance adoption in the rail industry - findings from a case study at thales. *IEEE Transactions on Engineering Management*.

Lis, D. and Otto, B. (2020). Data governance in data ecosystems - insights from organizations. In *Ameri-*

cas Conference on Information Systems Proceedings. AIS Electronic Library.

Storm, M. and Borgman, H. P. (2020). Understanding challenges and success factors in creating a data-driven culture. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Thomas, G. (2024). The data governance institute (dgi) data governance framework.

Zhang, Q., Sun, X., and Zhang, M. (2022). Data matters: A strategic action framework for data governance. *Information and Management*, 59.

APPENDIX

Table 1: Web form used at the First Stage of research.

Begin of the Table 1

	Question	Alternatives
1	**OBJECTIVES**	
2	**IDENTIFICATION OF THE RESEARCH PARTICIPANT**	
2.1	Do you have experience in DG?	Yes No
2.2	If you have experience in DG, indicate the approximate time.	Less than 01 year 01 to 02 years 03 to 04 years More than 05 years
3	**ORGANIZATION IDENTIFICATION**	
3.1	Size	Small Medium Large
3.2	Market segment	Agriculture Consulting Education Financial Government Industry Logistics Health Insurance Services Social Technology Telecommunications Transportation Other
4	**STAGE IN DATA GOVERNANCE**	
4.1	How would you classify the organization's stage of DG?	Planning Under implementation Implemented Not planning to implement
4.2	How long has the organization been implementing DG?	Not applicable 06 months 01 year 02 years More than 02 years
4.3	What are the main factors that motivate or motivated the organization to implement DG?	Assist in decision-making Improve data quality Gain competitive advantage Reduce data management costs Comply with legislation/regulations Increase data security Reduce data volume Improve access to data Other
5	**DATA INVENTORY (DI)**	
5.1	Regarding the organization's carrying out of DI?	Performed Is currently being performed Intends to perform Will not perform
5.2	How important is it to conduct a DI for DG?	Not at all important Slightly important Moderately important Very important Extremely important
5.3	What level of priority should be given to DI?	No priority Low priority Neutral High priority Maximum priority
5.4	What is the level of complexity for performing DI?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
5.5	List the main challenges that prevent the organization from carrying out the DI	
5.6	List the main challenges faced by the organization when carrying out the DI	
6	**PROCESS MAPPING (PM)**	
6.1	Regarding the organization's carrying out of PM?	Performed Is currently being performed Intends to perform Will not perform

Continuation of the Table 1

	Question	Alternatives
6.2	How important is it for DG to PM?	Not at all important Slightly important Moderately important Very important Extremely important
6.3	What level of priority should be given to PM?	No priority Low priority Neutral High priority Maximum priority
6.4	What is the level of complexity involved in PM?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
6.5	List the main challenges that prevent the organization from carrying out PM	
6.6	List the main challenges faced by the organization when carrying out PM	
7	**REGULATION SURVEY (RS)**	
7.1	Regarding the organization's carrying out of RS?	Performed Is currently being performed Intends to perform Will not perform
7.2	How important is it to conduct RS for DG?	Not at all important Slightly important Moderately important Very important Extremely important
7.3	What level of priority should be given to the RS?	No priority Low priority Neutral High priority Maximum priority
7.4	What is the level of complexity in conducting a RS?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
7.5	List the main challenges that prevent the organization from conducting a RS	
7.6	List the main challenges faced by the organization when conducting a RS	
8	**INFRASTRUCTURE MAPPING (IM)**	
8.1	Regarding the organization's carrying out of IM?	Performed Is currently being performed Intends to perform Will not perform
8.2	How important is it to perform IM for DG?	Not at all important Slightly important Moderately important Very important Extremely important
8.3	What level of priority should be given to IM?	No priority Low priority Neutral High priority Maximum priority
8.4	What is the level of complexity in performing IM?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
8.5	List the main challenges that prevent the organization from performing IM	
8.6	List the main challenges faced by the organization when performing IM	
9	**MANAGEMENT COMMITTEE (MC)**	
9.1	Regarding the definition / performance of a DG MC in the organization?	Existing and active committee Committee in the process of being created Intends to establish a Committee Does not intend to establish a Committee
9.2	How important is it to establish a DG MC?	Not at all important Slightly important Moderately important Very important Extremely important
9.3	What level of priority should be given to the creation of a DG MC?	No priority Low priority Neutral High priority Maximum priority
9.4	What is the level of complexity in defining a DG MC?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
9.5	What is the level of complexity in the performance of a DG MC?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
9.6	List the main challenges that prevent the creation of a DG MC	
9.7	List the main challenges encountered in the performance of the DG MC	

Continuation of the Table 1

	Question	Alternatives
10	**DATA GOVERNANCE TRAINING**	
10.1	Regarding the presentation of DG concepts to the organization's employees?	Performed Is currently being performed Intends to perform Will not perform
10.2	How important is it to present DG concepts to the organization's employees?	Not at all important Slightly important Moderately important Very important Extremely important
10.3	What level of priority should be given to presenting DG concepts to the organization's employees?	No priority Low priority Neutral High priority Maximum priority
10.4	What is the level of complexity in training employees in DG concepts?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
10.5	List the main challenges that prevent employees from being trained in DG concepts	
10.5	List the main challenges encountered during employee training in DG concepts	
11	**DATA GOVERNANCE POLICY (DGP)**	
11.1	Regarding the definition of a DGP for the organization?	Prepared and applied It is in the development phase Intends to prepare Won't elaborate
11.2	How important is it to develop and implement a DGP?	Not at all important Slightly important Moderately important Very important Extremely important
11.3	What level of priority should be given to developing and implementing a DGP?	No priority Low priority Neutral High priority Maximum priority
11.4	What is the level of complexity in developing a DGP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
11.5	What is the level of complexity in implementing a DGP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
11.6	List the main challenges that prevent the development and implementation of a DGP	
11.7	List the main challenges encountered during the implementation of the DGP	
11.8	Regarding the presentation of DGP to the organization's employees?	Performed Is currently being performed Intends to perform Will not perform
11.9	How important is it to present DGP to the organization's employees?	Not at all important Slightly important Moderately important Very important Extremely important
11.10	What level of priority should be given to presenting DGP to the organization's employees?	No priority Low priority Neutral High priority Maximum priority
11.11	What is the level of complexity in presenting the DGP to the organization's employees?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
11.12	List the main challenges that prevent the presentation of the DGP to the organization's employees	
11.13	List the main challenges encountered when presenting the DGP to the organization's employees	
12	**PRIVACY AND DATA SECURITY POLICY (PDSP)**	
12.1	Regarding the definition of a PDSP for the organization?	Prepared and applied It is in the development phase Intends to prepare Won't elaborate
12.2	How important is it to develop and implement a PDSP?	Not at all important Slightly important Moderately important Very important Extremely important
12.3	What level of priority should be given to developing and implementing a PDSP?	No priority Low priority Neutral High priority Maximum priority

Continuation of the Table 1

	Question	Alternatives
12.4	What is the level of complexity in developing a PDSP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
12.5	What is the level of complexity in implementing a PDSP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
12.6	List the main challenges that prevent the development and implementation of a PDSP	
12.7	List the main challenges encountered during the implementation of the PDSP	
12.8	Regarding the presentation of PDSP to the organization's employees?	Performed Is currently being performed Intends to perform Will not perform
12.9	How important is it to present the PDSP to the organization's employees?	Not at all important Slightly important Moderately important Very important Extremely important
12.10	What level of priority should be given to presenting the PDSP to the organization's employees?	No priority Low priority Neutral High priority Maximum priority
12.11	What is the level of complexity in presenting the PDSP to the organization's employees?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
12.12	List the main challenges that prevent the presentation of the PDSP to the organization's employees	
12.13	List the main challenges encountered when presenting the PDSP to the organization's employees	
13	**DATA QUALITY POLICY (DQP)**	
13.1	Regarding the definition of a DQP for the organization?	Prepared and applied It is in the development phase Intends to prepare Won't elaborate
13.2	How important is it to develop and implement a DQP?	Not at all important Slightly important Moderately important Very important Extremely important
13.3	What level of priority should be given to developing and implementing a DQP?	No priority Low priority Neutral High priority Maximum priority
13.4	What is the level of complexity in developing a DQP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
13.5	What is the level of complexity in implementing a DQP?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
13.6	List the main challenges that prevent the development and implementation of a DQP	
13.7	List the main challenges encountered during the implementation of the DQP	
13.8	Regarding the presentation of DQP to the organization's employees?	Performed Is currently being performed Intends to perform Will not perform
13.9	How important is it to present DQP to the organization's employees?	Not at all important Slightly important Moderately important Very important Extremely important
13.10	What level of priority should be given to presenting DQP to the organization's employees?	No priority Low priority Neutral High priority Maximum priority
13.11	What is the level of complexity in presenting the DQP to the organization's employees?	No complexity Low complexity Moderate complexity High complexity Extreme complexity
13.12	List the main challenges that prevent the presentation of the DQP to the organization's employees	
13.13	List the main challenges encountered when presenting the DQP to the organization's employees	

End of the Table 1

Table 2: Web form used at the Second Stage of research.

	Question	Alternatives to be ranked
I	**ACTIONS RELATED TO INVENTORY / INFORMATION MAPPING**	
01	How would you rate the IMPLEMENTATION of these Inventories / Information mappings?	- Data Inventory - Process Mapping - Regulatory Survey - Infrastructure Mapping
02	How would you rate these challenges that PREVENT the PERFORMANCE of Inventory / Information Mapping?	- Planning / prioritization - Diversity of technologies / environments systems - Support of tools - Existing / updated documentation - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
03	How would you rate these challenges that IMPACT the EXECUTION of Inventory / Information Mapping?	- Planning / prioritization - Diversity of technologies / environments systems - Support of tools - Existing / updated documentation - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
II	**ACTIONS RELATED TO POLICIES**	
04	How would you rate the PREPARATION of these Policies?	- Data Governance Policy - Data Privacy and Security Policy - Data Quality Policy
05	How would you rate these challenges that PREVENT the DEVELOPMENT of Policies?	- Planning / prioritization - Diversity of technologies / environments systems - Support of tools - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear policies - Well-defined / clear processes
06	How would you rate these challenges that IMPACT on POLICY MAKING?	- Planning / prioritization - Diversity of technologies / environments systems - Support of tools - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear policies - Well-defined / clear processes
07	How would you rate the IMPLEMENTATION of these Policies?	- Data Governance Policy - Data Privacy and Security Policy - Data Quality Policy
08	How would you rate these challenges that IMPACT the IMPLEMENTATION of Policies?	- Planning / prioritization - Diversity of technologies / environments systems - Support of tools - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear policies - Well-defined / clear processes
III	**ACTIONS RELATED TO TRAINING / TRAINING / PRESENTATION**	
09	How would you rate the PRESENTATION of these Training / Presentations?	- Data Governance Concepts - Data Governance Policy - Data Privacy and Security Policy - Data Quality Policy
10	How would you rate these challenges that PREVENT the PRESENTATION of Training / Presentations?	- Planning / prioritization - Clear / defined policies - Alignment / communication - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
11	How would you rate these challenges that IMPACT the PRESENTATION of Training / Presentations?	- Planning / prioritization - Clear / defined policies - Alignment / communication - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
IV	**ACTIONS RELATED TO THE MANAGEMENT COMMITTEE**	
12	How would you rate these challenges that PREVENT the CREATION of a Management Committee?	- Planning / prioritization - Clear / defined policies - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
13	How would you rate these challenges that IMPACT the PERFORMANCE of the Management Committee?	- Planning / prioritization - Clear / defined policies - Employee engagement - Culture / knowledge / training - Perception of value / benefits - Management support - Well-defined / clear strategy
V	**GROUP OF ACTIONS**	
14	How would you rate the IMPLEMENTATION of these Group of Actions?	- Carrying out the Inventory / Mapping information - Policy Development - Implementation of elaborated Policies - Carrying out the Training / Training / Presentation - Creation of a Management Committee - Performance of established Management Committee

Smart Data Stewardship: Innovating Governance and Quality with AI

Otmane Azeroual
German Centre for Higher Education Research and Science Studies (DZHW), 10117 Berlin, Germany
azeroual@dzhw.eu

Keywords: Data Governance, Data Quality, Artificial Intelligence (AI), AI-Powered Framework, Data Integration, Quality Assurance, Data Protection Monitoring, Compliance Management, Digital Transformation, Organizational Efficiency, Case Studies, Practical Applications.

Abstract: In the modern digital landscape, data plays a crucial role in the competitiveness and efficiency of organizations. Data governance, which involves managing and ensuring data quality, faces increasing challenges due to the growing volumes and complexities of data. This paper examines how artificial intelligence (AI) offers innovative solutions for optimizing data governance and data quality. We present an AI-powered framework that includes components such as data integration, quality assurance, data protection monitoring, and compliance management. Through case studies and practical examples, we demonstrate how this framework can be implemented in real-world environments and the benefits it offers.

1 INTRODUCTION

In today's digitalized world, data has become an essential asset that forms the basis for business decisions, innovations, and strategic planning (Schildt, 2020), (Kolasani, 2023). Data governance refers to the comprehensive management of the availability, usability, integrity, and security of data within an organization (Solà-Morales et al., 2023). It is an overarching concept that defines the policies, processes, roles, standards, and metrics necessary to ensure that data can be used effectively and efficiently (Hatanaka et al., 2022). The importance of data governance cannot be overstated, as it helps to minimize risks, ensure compliance with legal and regulatory requirements, and guarantee data quality (Mahanti, 2021).

Traditional approaches to data governance face significant challenges, which are exacerbated by the exponentially growing volumes of data and the increasing complexity of data landscapes (Caparini & Gogolewska, 2021). One of the main difficulties lies in the manual management of data, which is prone to errors and time-consuming. Additionally, traditional data governance models are often not flexible enough to quickly adapt to changes in the data landscape or new regulatory requirements (Gong et al., 2020). The fragmentation of data across various systems and silos makes it difficult to ensure consistent data quality and to implement a holistic data governance strategy (Strengholt, 2020), (Janssen et al., 2020). Moreover, many organizations are confronted with limited resources and expertise, which further complicates the effective implementation and maintenance of data governance programs (Plotkin, 2020).

Artificial intelligence (AI) offers innovative solutions to address the challenges of traditional data governance (Janssen et al., 2020). By leveraging AI technologies such as machine learning (ML), natural language processing (NLP), and robotic process automation (RPA), many of the manual and time-consuming processes can be automated, significantly improving the efficiency and accuracy of Data Governance. AI-powered systems can analyze large volumes of data in real-time, identify patterns and anomalies, and take proactive measures to ensure data quality and security (Yandrapalli, 2024). Additionally, AI enables dynamic adaptation to changing regulatory requirements and business environments, ensuring flexible and future-proof data governance (Stanciu et al., 2021). These technologies can also help integrate and harmonize data from various sources, reducing data fragmentation and enabling a holistic view of the data landscape.

This paper aims to present a comprehensive framework that demonstrates how AI can be used to

Azeroual, O.
Smart Data Stewardship: Innovating Governance and Quality with AI .

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 187-196
ISBN: 978-989-758-716-0; ISSN: 2184-3228

optimize data governance and data quality in organizations. The following sections will first explain the theoretical framework for data governance and the relevance of AI technologies. Next, the various components of the proposed AI-driven framework will be described in detail, including data integration, data quality assurance, data privacy monitoring, and compliance management. The paper will also discuss the steps for implementing the framework, as well as the associated technological and organizational requirements and challenges. Additionally, case studies and practical examples will be presented to illustrate the practical applicability of the framework and analyze the benefits achieved. Finally, there will be a discussion of the effectiveness of the framework compared to traditional approaches, followed by a summary of the key findings and recommendations for practice.

2 THEORETICAL FRAMEWORK

Data governance refers to the set of measures, rules, processes, and technologies that ensure data is managed efficiently, effectively, and securely within an organization (Azeroual et al., 2023). The main goal of data governance is to ensure data quality and integrity, protect data privacy and security, and comply with legal and regulatory requirements (Brous et al., 2020). This also includes establishing clear responsibilities and accountabilities for data management and defining policies and standards for data handling.

The essential objectives of data governance include (Georgiadis & Poels, 2021), (Duggineni, 2023), (Ren, 2022), (Al-Surmi et al., 2022):

- **Ensuring Data Quality**: Avoiding data inconsistencies, duplications, and errors to provide reliable and accurate data for business decisions.
- **Compliance with Regulations**: Ensuring that data processing and storage comply with legal and regulatory requirements, such as GDPR.
- **Data Protection**: Ensuring the confidentiality, integrity, and availability of data to prevent data misuse and breaches.
- **Efficient Data Management**: Optimizing processes for data integration, processing, and utilization to enhance the efficiency and effectiveness of data management.
- **Supporting Strategic Decision-Making**: Providing high-quality and up-to-date data for strategic planning and operational decisions.

AI encompasses a variety of technologies that enable machines to mimic human intelligence and perform tasks autonomously (Jiang et al., 2022). The key AI technologies relevant to data governance include machine learning (ML), natural language processing (NLP), and robotic process automation (RPA) (Ansari et al., 2019), (Serey et al., 2021), (Sarker, 2021), (Sharma et al., 2022), (Rane et al., 2024):

- **ML** is a subset of AI that uses algorithms and statistical models to learn from data and make predictions or decisions without being explicitly programmed. ML models can be used to detect patterns and anomalies in large datasets to identify and address data quality issues. For example, ML algorithms can be used for automatic duplicate detection, error correction, and data classification.
- **NLP** enables machines to understand, interpret, and generate human language. This technology can be used to analyze and process unstructured data such as text documents, emails, and reports. In data governance, NLP can be employed to extract relevant information from unstructured data sources, categorize data, and generate metadata, thereby improving data quality and availability.
- **RPA** uses software robots or "bots" to automate repetitive and rule-based tasks. RPA can be used in data governance to automate routine tasks such as data cleaning, validation, and updating. This reduces manual intervention and minimizes error rates, thereby increasing the efficiency and accuracy of data management.

In recent years, numerous studies and approaches have been developed to integrate AI into data governance (Janssen et al., 2020), (Wirtz et al., 2020), (Zuiderwijk et al., 2021), (Taeihagh, 2021), (Khan et al., 2024). These approaches aim to enhance the efficiency and effectiveness of data management through the use of AI technologies.

- **Automated Data Quality Monitoring**: Several studies have demonstrated that using ML algorithms for continuous monitoring and improvement of data quality offers significant benefits (Lee & Shin, 2020). For instance, ML models can detect data anomalies in real time and suggest corrections, thereby sustainably improving data quality.

- **NLP-Based Data Analysis:** Research has shown that NLP techniques can be effectively used to analyze and process large volumes of unstructured data (Aladakatti & Senthil Kumar, 2023). This enables better data categorization and indexing, enhancing data availability and usability.
- **RPA for Data Management Processes:** The use of RPA to automate data management tasks has proven successful in many organizations. Studies have shown that RPA increases the efficiency of data management by automating repetitive tasks and reducing human errors (Radke et al., 2020).

An example of a successful implementation of an AI-powered data governance framework is a project by a large financial services provider that employed ML and NLP technologies to improve data quality and security (Li et al., 2021), (Xu, 2022), (Mishra et al., 2024). By integrating these technologies, the company was able to significantly enhance data integrity and compliance while reducing operational costs.

3 COMPONENTS OF THE AI-POWERED FRAMEWORK

The AI-powered framework presented for optimizing data governance and data quality encompasses four essential components: data integration, data quality assurance, data protection monitoring, and compliance management.

3.1 Data Integration

Data integration refers to the process of combining data from various sources to provide a consolidated and unified view. This is particularly important as organizations often work with a multitude of data sources and formats. Effective data integration enables harmonizing data across different systems, which in turn improves the quality and availability of data (Rangineni et al., 2023). The significance of data integration lies in its role as the foundation for reliable data analysis and decision-making. Without efficient integration, data can be fragmented and inconsistent, leading to faulty analyses and suboptimal decisions.

Artificial intelligence can significantly enhance and automate the data integration process (Aldoseri et al., 2023). ML and NLP can be used to identify complex data patterns and automatically transform and harmonize data from various sources. AI-powered data integration systems can:

- **Automatic Schema Matching and Mapping:** ML algorithms can recognize and automatically map data fields from different sources, reducing the time and effort required for manual data reconciliation.
- **Anomaly Detection:** AI can detect anomalies and inconsistencies in data sources and suggest solutions to address these issues.
- **Real-Time Data Integration:** AI-powered systems can continuously integrate and update data in real-time, enhancing data timeliness and accuracy.

3.2 Data Quality Assurance

Data quality assurance refers to the processes and measures that ensure data is accurate, complete, consistent, and up-to-date. High data quality is crucial for the reliability and credibility of data analysis and decision-making. Poor data quality can lead to incorrect conclusions and ineffective business strategies. Therefore, ensuring data quality is a fundamental aspect of data governance.

AI offers powerful tools to improve data quality through the use of advanced algorithms:

- **Anomaly Detection:** ML algorithms can be used to identify unusual patterns and outliers in the data that may indicate errors or inaccuracies.
- **Automated Data Cleansing:** AI systems can automatically detect and clean duplicates, missing values, and inconsistencies in the data.
- **Data Validation:** AI can continuously monitor data quality and apply validation rules to ensure data meets established quality standards.

3.3 Data Protection Monitoring

Data protection monitoring encompasses the measures and technologies used to ensure the confidentiality, integrity, and availability of data (Farayola et al., 2024). This is particularly important given the increasing threats from cyberattacks and data breaches. An effective data protection monitoring system helps safeguard sensitive data and

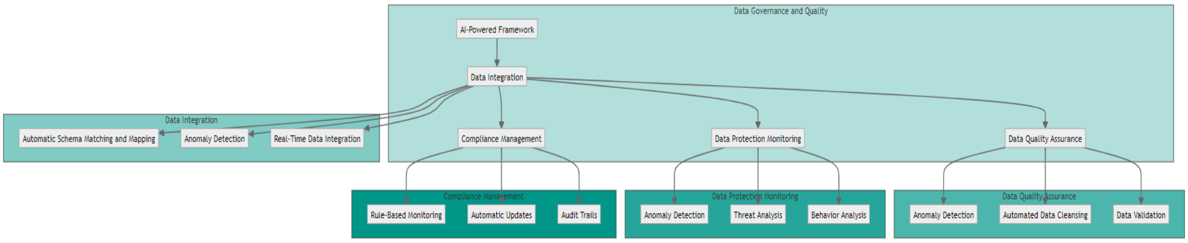


Figure 1: AI-Powered Framework for Optimizing Data Governance and Data Quality.

ensures that only authorized users have access to it (Duggineni, 2023).

AI can significantly enhance data protection monitoring by employing advanced techniques to detect security vulnerabilities and data breaches:

- **Anomaly Detection:** ML algorithms can identify unusual activities and anomalies in network traffic that may indicate potential security breaches.
- **Threat Analysis:** AI can analyze threats in real-time and proactively adjust security measures to prevent attacks.
- **Behavior Analysis:** AI-powered systems can analyze user behavior and identify abnormal activities that may indicate potential insider threats.

3.4 Compliance Management

Compliance management refers to ensuring that an organization meets all relevant legal and regulatory requirements (Biggeri et al., 2023). This is particularly crucial in highly regulated industries such as healthcare, finance, and insurance. An effective compliance management system helps minimize legal risks and build trust with customers and stakeholders (Olawale et al., 2024).

AI can significantly improve compliance management by employing advanced techniques to monitor and ensure adherence to regulations:

- **Rule-Based Monitoring:** AI-powered systems can continuously monitor compliance with regulations and automatically issue alerts when violations are detected.
- **Automatic Updates:** AI can be used to update compliance rules and regulations in real-time, ensuring the organization is always up-to-date with legal requirements.
- **Audit Trails:** AI can generate detailed logs and reports for audits and inspections to demonstrate compliance with regulations.

Summarizing the components in Figure 1, each plays a crucial role in ensuring effective and efficient data management. By leveraging AI technologies such as machine learning, natural language processing, and robotic process automation, organizations can overcome the challenges of traditional data governance and achieve higher data quality, security, and compliance. This framework offers a comprehensive approach to modernizing and enhancing data governance in the digital era.

4 IMPLEMENTATION OF THE AI FRAMEWORK: STEPS, REQUIREMENTS, AND CHALLENGES

To optimize data governance and data quality using an AI-powered framework, a systematic approach is required. Here, we detail the essential implementation steps and discuss the technological and organizational requirements necessary for organizations to effectively deploy the framework.

4.1 Steps for Implementing the Framework

The implementation of an AI-powered framework for optimizing data governance and data quality requires a systematically grounded approach. First, a comprehensive requirements analysis is essential to identify the specific needs and goals of the organization. This step involves evaluating current data governance practices, identifying weaknesses, and defining the main objectives and requirements. The analysis is supported by qualitative methods such as stakeholder interviews and quantitative methods such as data analyses. For example, a healthcare organization might conduct interviews with doctors and administrative staff to identify requirements for data quality and data protection.

Following the requirements analysis is the selection of appropriate AI technologies. This step involves evaluating ML algorithms, NLP tools, and

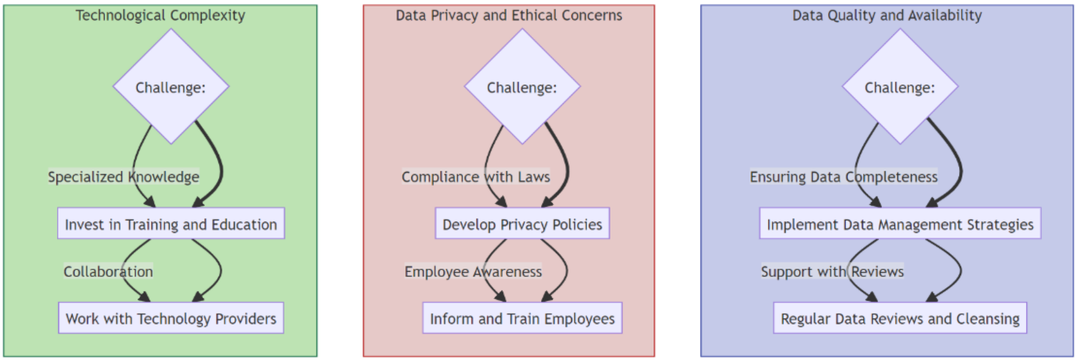


Figure 2: Automation and Efficiency Improvements through AI Technologies.

RPA software that meet the needs of data integration, data quality assurance, data protection monitoring, and compliance management. Criteria for selection include scalability, compatibility with existing IT infrastructure, user-friendliness, and cost. For instance, a financial institution might choose specific ML algorithms developed for fraud detection in transaction data and evaluate them based on their ability to process large volumes of data in real time and reliably detect anomalies.

4.2 Technological and Organizational Requirements

Implementing an AI-powered framework involves addressing both the technological and organizational requirements of the organization.

➢ **Technological Requirements**

Powerful servers and network infrastructures are required to support real-time data processing and the execution of AI algorithms. Additionally, specialized AI platforms and tools are necessary to implement ML, NLP, and RPA. Robust databases and data warehouses are also essential for efficiently storing and managing large volumes of data.

➢ **Organizational Requirements**

Training and continuous education of employees are crucial to ensure they acquire the necessary skills and knowledge to handle the new AI technologies. Furthermore, clear roles and responsibilities for implementing and managing the framework need to be defined, including appointing data stewards and AI specialists. A comprehensive change management plan is also required to promote acceptance and engagement among employees and to ensure a smooth introduction of new technologies. Challenges in implementation include ensuring data quality and availability, as well as addressing data protection and ethical concerns through robust data management strategies and clear data protection policies.

4.3 Challenges and Solutions

Adopting an AI-powered framework for data governance offers numerous benefits but also presents significant challenges. One of the greatest strengths of the framework is the automation and efficiency improvements enabled by AI technologies. These advancements lead to a significant enhancement of data quality, increased security, and improved compliance (see Figure 2).

- *Data Quality and Availability*

Challenge: Ensuring that data is complete and accurate is one of the biggest challenges. Missing or incomplete data can significantly impair the effectiveness of AI algorithms.

Solution: Implement robust data management strategies, including clear guidelines and standards to ensure data quality. This can be supported by regular data reviews and cleansing processes.

- *Data Privacy and Ethical Concerns*

Challenge: The use of AI must comply with data protection laws and consider ethical standards to maintain stakeholder trust.

Solution: Develop and implement clear data privacy policies and practices. Organizations should ensure that all employees are informed about the importance of data protection and receive appropriate training.

- *Technological Complexity*

Challenge: The implementation and maintenance of AI systems require specialized technical knowledge and resources, which are not always readily available.

Solution: Invest in training and continuing education for employees to build the necessary technical skills and knowledge. Collaboration with specialized technology providers and consultants can also be beneficial.

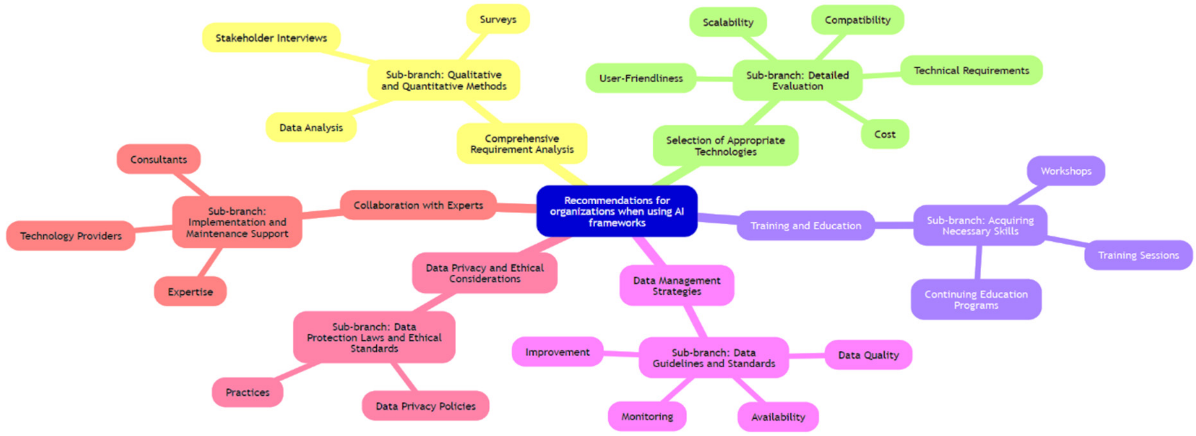


Figure 3: Recommendations for Implementing an AI-Powered Data Governance Framework.

4.4 Recommendations for Organizations

This section presents key recommendations for organizations to successfully implement an AI-powered framework for optimizing data governance and data quality. By following these guidelines, organizations can effectively address challenges and maximize the benefits of AI integration. See Figure 3 for a visual representation of these recommendations.

- *Comprehensive Requirement Analysis*

Organizations should conduct a thorough requirement analysis to identify specific needs and goals. This can be achieved through qualitative and quantitative methods such as stakeholder interviews, surveys, and data analyses.

- *Selection of Appropriate Technologies*

The selection of the right AI technologies should be based on a detailed evaluation of technical requirements and available solutions. Organizations should consider factors such as scalability, compatibility, user-friendliness, and cost.

- *Training and Education*

Organizations should invest in the training and education of their employees to ensure they acquire the necessary skills and knowledge to handle new AI technologies. This can be facilitated through training sessions, workshops, and continuing education programs.

- *Development of Robust Data Management Strategies*

Organizations should develop and implement clear data guidelines and standards to ensure data quality and availability. This includes the establishment of processes for continuous monitoring and improvement of data quality.

- *Data Privacy and Ethical Considerations*

Organizations should ensure that their AI-powered systems comply with data protection laws and consider ethical standards. This requires the development and implementation of clear data privacy policies and practices.

- *Collaboration with Experts*

Organizations should consider collaborating with specialized technology providers and consultants to support the implementation and maintenance of AI systems. This can facilitate access to expertise and resources, thereby enhancing the efficiency of the implementation process.

By following these structured recommendations, organizations can maximize the benefits of an AI-powered framework for optimizing data governance and data quality, while effectively addressing the associated challenges.

5 CASE STUDIES AND PRACTICAL EXAMPLES

To illustrate the practical applicability of the AI-powered framework for optimizing data governance and data quality, detailed case studies and practical examples are presented below. These examples demonstrate how the framework has been successfully implemented in various organizations and the resulting benefits.

5.1 Case Study 1: Healthcare Organization

Challenge: A large healthcare organization faced the challenge of significantly improving the quality and security of patient data. Due to the variety and volume of data, errors and inconsistencies frequently occurred, impacting medical care. Manual data reviews and corrections were time-consuming and inefficient, leading to delays and increased error rates.

Solution: To address these challenges, the organization implemented an AI-powered system for automatic data cleansing and anomaly detection. Machine learning (ML) was used to systematically identify and correct data errors. The AI system continuously analyzed incoming patient data for inconsistencies and deviations from established standards. Once an error was detected, the system could automatically initiate corrective actions or notify the responsible staff.

Approach: Data Integration: All relevant patient data from various sources were first integrated into a central database. ML algorithms helped harmonize and structure different data formats.

Anomaly Detection: An advanced ML model was trained to detect anomalies and inconsistencies in the data. These algorithms were capable of identifying both simple errors like missing values and complex anomalies like unusual diagnoses.

Automatic Data Cleansing: After anomaly detection, the system performed automatic data cleansing processes, removing duplicates and correcting faulty entries. For more complex issues, the staff was notified to conduct manual reviews.

Outcome: Implementing the AI system significantly improved data quality and security. Errors and inconsistencies were drastically reduced, leading to higher accuracy of medical data. This not only increased the efficiency of medical care but also improved patient satisfaction, as treatment decisions were based on more reliable data. Overall, the organization reduced the error rate in patient data by 60% and shortened data processing time by 40%.

5.2 Case Study 2: Financial Institution

Challenge: A large financial institution faced the challenge of early detection and prevention of fraudulent activities in transaction data. Traditional methods for fraud detection were inefficient and often resulted in false alarms or delayed responses, causing financial losses and erosion of customer trust.

Solution: The institution opted to implement an AI-powered system for real-time detection of fraudulent activities. ML algorithms and natural language processing (NLP) tools were used to analyze transaction patterns and identify anomalies. The system was designed to continuously monitor transaction data and immediately respond to suspicious activities.

Approach: Data Collection and Integration: Transaction data from various sources were collected and integrated into a central system. Historical data were also included to train the ML models.

Model Development: Various ML models, including supervised and unsupervised learning, were developed to detect fraudulent patterns. These models were continuously trained and refined with new data.

NLP Analysis: In addition to the ML models, NLP tools were used to analyze text data from transaction descriptions and identify semantic patterns indicative of fraudulent activities.

Real-Time Monitoring: The system continuously monitored all transactions in real-time. Upon detecting an anomaly or potential fraud, immediate actions were taken, such as freezing the affected account and notifying the customer and the security team.

Outcome: The implementation led to a significant reduction in fraud cases. Real-time monitoring and analysis of transaction data increased the security of financial transactions and enabled quick responses to suspicious activities. The number of fraud cases was reduced by 70%, and the accuracy of fraud detection increased to over 90%. Additionally, customer trust in the security of their transactions improved significantly, leading to stronger customer retention.

5.3 Analysis of Benefits and Lessons Learned

The analysis of the benefits shows that the adoption of the AI-powered framework led to significant efficiency gains. Routine tasks were automated, reducing manual effort and improving data processing accuracy. The improvement in data quality was achieved through AI algorithms capable of identifying and correcting data errors in real-time. Moreover, data security was enhanced by proactive detection and prevention of security breaches.

Lessons Learned: Importance of High-Quality Data: High-quality and complete data are essential for the success of AI systems. Organizations must invest in robust data management strategies to ensure data integrity and quality.

Training and Education: Successful implementation of AI systems requires well-trained and educated employees. Organizations should continuously invest in training their staff to equip them with the necessary technical skills and knowledge.

Flexible IT Infrastructure: A flexible and adaptable IT infrastructure is crucial to meet changing requirements. A scalable infrastructure facilitates the implementation and use of AI technologies, enabling organizations to respond quickly to new challenges.

These case studies highlight the diverse applications and practical benefits of the framework in various contexts. By applying the lessons learned, organizations can maximize the advantages of AI technologies while effectively addressing the challenges.

6 DISCUSSION

Although a direct evaluation was not conducted in this article, the assumed benefits of the AI-powered framework are based on a comprehensive analysis of theoretical and practical case studies from the literature. Implementing such a framework has the potential to achieve significant improvements in data governance. By automating routine tasks and continuously monitoring data quality, the efficiency and accuracy of data processing can be enhanced. Specifically, the use of ML algorithms for data cleansing and anomaly detection offers the possibility to identify and correct data errors in real-time, leading to improved data quality. Additionally, proactive monitoring and detection of security breaches can significantly enhance data security.

Compared to traditional approaches, the AI-powered framework offers several key advantages. Traditional data governance methods often rely on manual processes that are time-consuming and prone to errors. The AI-powered framework automates many of these processes, thereby increasing efficiency and minimizing human errors. Real-time analysis and monitoring of data enable faster responses to anomalies and security threats, which is often not possible to the same extent with traditional approaches. Furthermore, the ability to process large volumes of data in real-time and recognize patterns provides a distinct advantage over traditional methods.

Despite the theoretical and practical advantages, there are still areas for improvement and future research. A central area is the continuous development and refinement of algorithms to further enhance their accuracy and efficiency. Moreover, integrating AI technologies into existing IT infrastructures is often complex and requires further research to optimize these processes. Another important research field concerns the ethical and legal implications of using AI in data governance, particularly regarding data privacy and data integrity. Future studies should also focus on developing more user-friendly AI tools to promote their acceptance and use in non-technical domains.

7 CONCLUSIONS

The theoretical analysis and case studies indicate that AI technologies can significantly enhance data governance. Automation and real-time analysis can greatly improve data quality and security. The case studies demonstrated practical applications and the benefits achieved in various organizational contexts.

AI will play an increasingly important role in the future of data governance. AI's ability to efficiently process large volumes of data, recognize patterns, and proactively respond to anomalies will be crucial in addressing the challenges of the modern data landscape. AI technologies will enable organizations to continuously improve their data governance practices and meet growing demands.

Deploying an AI-powered framework requires careful planning and execution. Organizations should conduct a comprehensive requirements analysis, select appropriate technologies, and invest in training and educating their employees. Clear data management strategies and data privacy policies are essential to ensure data integrity and security. Collaboration with experts and continuous monitoring and optimization of AI systems are also critical success factors.

By following a structured implementation approach and considering these recommendations, organizations can maximize the benefits of an AI-powered framework for optimizing data governance and data quality, while effectively addressing the associated challenges.

REFERENCES

Aladakatti, S. S., & Senthil Kumar, S. (2023). Exploring natural language processing techniques to extract semantics from unstructured dataset which will aid in effective semantic interlinking. *International Journal of Modeling, Simulation, and Scientific Computing*, 14(01), 2243004. https://doi.org/10.1142/S1793962322430048

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. *Applied Sciences*, 13(12), 7082. https://doi.org/10.3390/app13127082

Al-Surmi, A., Bashiri, M., & Koliousis, I. (2022). AI based decision making: combining strategies to improve operational performance. *International Journal of*

Production Research, 60(14), 4464-4486. https://doi.org/10.1080/00207543.2021.1966540

Ansari, W. A., Diya, P., Patil, S., & Patil, S. (2019). A review on robotic process automation-the future of business organizations. In *2nd International conference on advances in science & technology (ICAST)*. http://dx.doi.org/10.2139/ssrn.3372171

Azeroual, O., Nikiforova, A., & Sha, K. (2023). Overlooked Aspects of Data Governance: Workflow Framework For Enterprise Data Deduplication. In 2023 *International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)* (pp. 65-73). IEEE. https://doi.org/10.1109/ICCNS58795.2023.10193478

Biggeri, M., Borsacchi, L., Braito, L., & Ferrannini, A. (2023). Measuring the compliance of management system in manufacturing SMEs: an integrated model. *Journal of Cleaner Production*, 382, 135297. https://doi.org/10.1016/j.jclepro.2022.135297

Brous, P., Janssen, M., & Krans, R. (2020). Data governance as success factor for data science. In *Conference on e-Business, e-Services and e-Society* (pp. 431-442). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-44999-5_36

Caparini, M., & Gogolewska, A. (2021). Governance challenges of transformative technologies. *Connections: The Quarterly Journal*, 20(1), 91-100. https://doi.org/10.11610/Connections.20.1.06

Duggineni, S. (2023). Impact of controls on data integrity and information systems. *Science and Technology*, 13(2), 29-35. https://doi.org/10.5923/j.scit.20231302.04

Farayola, O. A., Olorunfemi, O. L., & Shoetan, P. O. (2024). Data privacy and security in it: a review of techniques and challenges. *Computer Science & IT Research Journal*, 5(3), 606-615. https://doi.org/10.51594/csitrj.v5i3.909

Georgiadis, G., & Poels, G. (2021). Enterprise architecture management as a solution for addressing general data protection regulation requirements in a big data context: a systematic mapping study. *Information Systems and e-Business Management*, 19, 313-362. https://doi.org/10.1007/s10257-020-00500-5

Gong, Y., Yang, J., & Shi, X. (2020). Towards a comprehensive understanding of digital transformation in government: Analysis of flexibility and enterprise architecture. *Government Information Quarterly*, 37(3), 101487. https://doi.org/10.1016/j.giq.2020.101487

Hatanaka, M., Konefal, J., Strube, J., Glenna, L., & Conner, D. (2022). Data□driven sustainability: Metrics, digital technologies, and governance in food and agriculture. *Rural Sociology*, 87(1), 206-230. https://doi.org/10.1111/ruso.12415

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022). Quo vadis artificial intelligence?. *Discover Artificial Intelligence*, 2(1), 4. https://doi.org/10.1007/s44163-022-00022-8

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government information quarterly*, 37(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Khan, A. (2024). Data Quality and Governance in Healthcare: Leveraging AI and ML for Master Data Management. *International Meridian Journal*, 6(6), 1-8.https://meridianjournal.in/index.php/IMJ/article/view/33

Kolasani, S. (2023). Innovations in digital, enterprise, cloud, data transformation, and organizational change management using agile, lean, and data-driven methodologies. *International Journal of Machine Learning and Artificial Intelligence*, 4(4), 1-18. https://jmlai.in/index.php/ijmlai/article/view/35

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157-170. https://doi.org/10.1016/j.bushor.2019.10.005

Li, Y., Yi, J., Chen, H., & Peng, D. (2021). Theory and application of artificial intelligence in financial industry. *Data Science in Finance and Economics*, 1(2), 96-116. https://doi.org/10.3934/DSFE.2021006

Mahanti, R. (2021). Data governance and data management. Springer Singapore. https://doi.org/10.1007/978-981-16-3583-0

Mishra, A. K., Tyagi, A. K., & Arowolo, M. O. (2024). Future Trends and Opportunities in Machine Learning and Artificial Intelligence for Banking and Finance. In *Applications of Block Chain technology and Artificial Intelligence: Lead-ins in Banking, Finance, and Capital Market* (pp. 211-238). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-47324-1_13

Olawale, O., Ajayi, F. A., Udeh, C. A., & Odejide, O. A. (2024). RegTech innovations streamlining compliance, reducing costs in the financial sector. *GSC Advanced Research and Reviews*, 19(1), 114-131. https://doi.org/10.30574/gscarr.2024.19.1.0146

Plotkin, D. (2020). Data stewardship: An actionable guide to effective data management and data governance. Academic press. https://doi.org/10.1016/C2012-0-07057-3

Radke, A. M., Dang, M. T., & Tan, A. (2020). Using robotic process automation (RPA) to enhance item master data maintenance process. *LogForum*, 16(1). http://doi.org/10.17270/J.LOG.2020.380

Rane, N., Choudhary, S., & Rane, J. (2024). Artificial Intelligence-Driven Corporate Finance: Enhancing Efficiency and Decision-Making Through Machine Learning, Natural Language Processing, and Robotic Process Automation in Corporate Governance and Sustainability. Natural Language Processing, and Robotic Process Automation in Corporate Governance and Sustainability. http://dx.doi.org/10.2139/ssrn.4720591

Rangineni, S., Bhanushali, A., Suryadevara, M., Venkata, S., & Peddireddy, K. (2023). A Review on enhancing data quality for optimal data analytics performance.

International Journal of Computer Sciences and Engineering, 11(10), 51-58. https://doi.org/10.26438/ijcse/v11i10.5158

Ren, S. (2022). Optimization of Enterprise Financial Management and Decision□Making Systems Based on Big Data. *Journal of Mathematics*, 2022(1), 1708506. https://doi.org/10.1155/2022/1708506

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science,* 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Schildt, H. (2020). The data imperative: How digitalization is reshaping management, organizing, and work. Oxford University Press, USA. https://doi.org/10.1093/oso/9780198840817.001.0001

Serey, J., Quezada, L., Alfaro, M., Fuertes, G., Vargas, M., Ternero, R., ... & Gutierrez, S. (2021). Artificial intelligence methodologies for data management. *Symmetry*, 13(11), 2040. https://doi.org/10.3390/sym13112040

Sharma, R., Agarwal, P., & Arya, A. (2022). Natural language processing and big data: a strapping combination. In *New Trends and Applications in Internet of Things (IoT) and Big Data Analytics* (pp. 255-271). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-99329-0_16

Solà-Morales, O., Sigurðardóttir, K., Akehurst, R., Murphy, L. A., Mestre-Ferrandiz, J., Cunningham, D., & de Pouvourville, G. (2023). Data governance for real-world data management: a proposal for a checklist to support decision making. *Value in Health*, 26(4), 32-42. https://doi.org/10.1016/j.jval.2023.02.012

Stanciu, A., Titu, A. M., & Deac-Şuteu, D. V. (2021). Driving digital transformation of knowledge-based organizations through artificial intelligence enabled data centric, consumption based, As-a-service models. In 2021 *13th international conference on electronics, computers and artificial intelligence (ECAI)* (pp. 1-8). IEEE. https://doi.org/10.1109/ECAI52376.2021.9515172

Strengholt, P. (2020). Data Management at scale. " O'Reilly Media, Inc.". https://www.oreilly.com/library/view/data-management-at/9781098138851/

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and society*, 40(2), 137-157. https://doi.org/10.1080/14494035.2021.1928377

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, 43(9), 818-829. https://doi.org/10.1080/01900692.2020.1749851

Xu, J. (2022). AI Theory and Applications in the Financial Industry. Future And Fintech, The: Abcdi And Beyond, 74. https://doi.org/10.1142/9789811250903_0003

Yandrapalli, V. (2024). AI-Powered Data Governance: A Cutting-Edge Method for Ensuring Data Quality for Machine Learning Applications. In 2024 *Second International Conference on Emerging Trends in Information Technology and Engineering* (ICETITE) (pp. 1-6). IEEE. https://doi.org/10.1109/ic-ETITE58242.2024.10493601

Zuiderwijk, A., Chen, Y. C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government information quarterly*, 38(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

EX-DSS: An Explorative Decision Support System for Designing and Deploying Smart Plug Forecasting Pipelines

Giulia Rinaldi, Lola Botman, Oscar Mauricio Agudelo and Bart De Moor
KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Leuven, Belgium
giulia.rinaldi@kuleuven.be

Keywords: Decision Support System, Smart Plug Forecasting, Artificial Intelligence Pipeline.

Abstract: Artificial Intelligence pipelines are increasingly used to address specific challenges, such as forecasting smart plug loads. Smart plugs, which remotely control various appliances, can significantly reduce energy consumption in commercial buildings by about 20% when effectively scheduled using AI techniques. Designing these AI pipelines involves numerous steps and variables, requiring collaboration and shared knowledge among designers. A Decision Support System (DSS) can facilitate this process. This paper introduces the Explorative Decision Support System (EX-DSS), which extends the classical DSS framework. The EX-DSS integrates an Explorative Management Subsystem to provide project-specific recommendations and a Data Quality (DQ) module to validate user inputs, ensuring clarity and enhancing information sharing. The EX-DSS architecture framework was tested through a software prototype designed to create AI pipelines for forecasting smart plug loads. The study found that using the EX-DSS improves the quality of suggestions, making them more problem-specific and resulting in a more personalized and meaningful user experience, with a significant potential to reduce energy consumption in commercial buildings.

1 INTRODUCTION

Designing an Artificial Intelligence (AI) pipeline involves multiple steps, from data cleaning to implementing machine learning methods. This process can be complex and time-consuming, especially when trying to find the most efficient combination. In the context of smart plug forecasting, this challenge is significant, as plug loads account for over 40% of total energy consumption in commercial buildings, excluding lighting, HVAC, and water heating (Chia et al., 2023). Smart plugs, which remotely monitor and control electrical appliances, can save up to 20% of electricity through effective scheduling, making AI forecasting and scheduling methods promising solutions (Botman et al., 2024).

A Decision Support System (DSS) can expedite the design of these AI pipelines. A DSS is a flexible software tool that assists in decision-making processes and allows for shareability and reproducibility among users, necessitating collaboration functionalities and ways to evaluate user input.

This paper introduces an experimental Explorative Decision Support System (EX-DSS) aimed at designing and deploying industrial smart plug pipelines to solve forecasting problems. The novel contributions of this research include:

- Extending the conventional DSS with the Explorative Management Subsystem, offering project-specific insights and recommendations.
- Integrating a Data Quality (DQ) module to assess the quality of user inputs, ensuring reliable AI pipeline design.
- Maintaining a human-in-the-loop approach, giving users maximal control over every feature in the EX-DSS.
- Providing a practical demonstration of the EX-DSS through a software prototype for smart plug forecasting, validating the system's capabilities and showcasing its effectiveness in real-world applications, thereby highlighting its potential impact on reducing energy consumption in commercial buildings.

This paper is structured as follows. Section 2 presents related works. In Section 3, the EX-DSS architecture framework is introduced. Section 4 describes the application of this framework in the design of smart plug forecasting AI pipelines via a software prototype. In Section 5, results are discussed, and fi-

Rinaldi, G., Botman, L., Agudelo, O. and De Moor, B.
EX-DSS: An Explorative Decision Support System for Designing and Deploying Smart Plug Forecasting Pipelines.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 197-205
ISBN: 978-989-758-716-0; ISSN: 2184-3228

nally, in Section 6, conclusions are drawn, and future research lines are presented.

2 BACKGROUND

A Decision Support System (DSS) is software that helps users analyze data and make decisions. It generates insights and suggestions through a structured framework (Gonzalez-Andujar, 2020). The classical DSS includes a **Data Management Subsystem**, which stores and handles data; a **Model Management Subsystem**, which manages models for DSS tasks; a **Knowledge Management Subsystem**, which provides information to users; and a **User Interface**, which connects users with DSS subsystems (Turban et al., 2010).

DSS supports semi-structured or unstructured decisions, requiring evaluations beyond mathematical modeling (Duan and Xu, 2009). For example, buying a new tool (Jacquet-Lagreze and Shakun, 1984) or deciding on stock market entry/exit (Chandra et al., 2007). These systems use rule-based models and AI to predict scenarios and derive insights (Phillips-Wren, 2013).

DSS applications are used in various fields like business, healthcare, and agriculture. They can help medical staff plan and monitor medications (Sloane and J. Silva, 2020) and manage production costs in agriculture (Rupnik et al., 2019). Tools like Rapidminer[1] and Knime[2] support the design of AI pipelines but lack assistance during the configuration phases.

The problem considered in the EX-DSS software prototype involves forecasting the load of electrical appliances, monitored and controlled by smart plugs to optimize energy consumption (Chia et al., 2023). Smart plugs can significantly reduce energy usage in commercial buildings by scheduling appliance operations. Accurate load forecasting is crucial for creating effective schedules. Current methods, including time series analysis and AI-based techniques like neural networks and ensemble methods (Botman et al., 2024), are time-consuming and complex to implement. This underscores the need for a robust DSS framework like EX-DSS to streamline the process and improve accuracy.

This paper introduces a knowledge-driven DSS framework that incorporates detailed domain knowledge. This makes it especially useful for complex decision-making processes that require specific expertise. Specifically tailored for time series forecasting, this framework focuses on input quality assessment and fosters a collaborative research community by enabling users to share insights, models, and results.

[1]https://altair.com/altair-rapidminer

[2]https://www.knime.com/

3 EX-DSS FRAMEWORK

As mentioned in Section 2, a traditional Decision Support System (DSS) has three subsystems and a User Interface (UI). The new Explorative Decision Support System (EX-DSS) adds a Data Quality (DQ) Module (Fig.1C) and an Explorative Management Subsystem (Fig.1D.4). The DQ Module (Fig.1C) improves data quality and promotes information sharing. The Explorative Management Subsystem (Fig.1D.4) provides project-specific insights for better analysis.

The EX-DSS architecture framework has two levels: the User (Fig.1A) and the EX-DSS (Fig.1B). The User interacts with the system to solve problems using semi-automated steps. The EX-DSS analyzes problems, provides extra information, and helps create the AI pipeline. A Graphical User Interface (Fig.1E) facilitates these interactions.

The EX-DSS has two main parts: the DQ Module (Fig.1C) and the Pipeline Design Module (Fig.1D). The DQ Module includes the Intake Phase (Fig.1C.1) for uploading data and the Assessment Phase (Fig.1C.2) for evaluating data quality by generating a report. The Pipeline Design Module (Fig.1D) has four subsystems: Data Management (Fig.1D.1), which stores essential information; Model Management (Fig.1D.2), which manages the pipeline structure (Fig.1D.2.a), configuration (Fig.1D.2.b) and training (Fig.1D.2.c); Knowledge Management (Fig.1D.3), which provides general information; and Explorative Management (Fig.1D.4), which derives project-specific insights from the keywords extracted from the DQ report (Fig.1D.4.a).

The general knowledge in the Knowledge Management Subsystem (Fig.1D.3) is not tied to specific problems. For example, it answers questions like “What is forecasting?”. Project-specific knowledge in the Explorative Management Subsystem (Fig.1D.4) is detailed and tailored to specific projects, like “What are the best methods for forecasting smart plugs?”. The DSS is called “explorative” because it helps users dive into and explore specific problems, such as forecasting smart plugs for energy savings in buildings.

By adding an extra subsystem, the EX-DSS improves modularity, maintainability, and reusability. It separates general knowledge from project-specific information, allowing better customization and system performance.

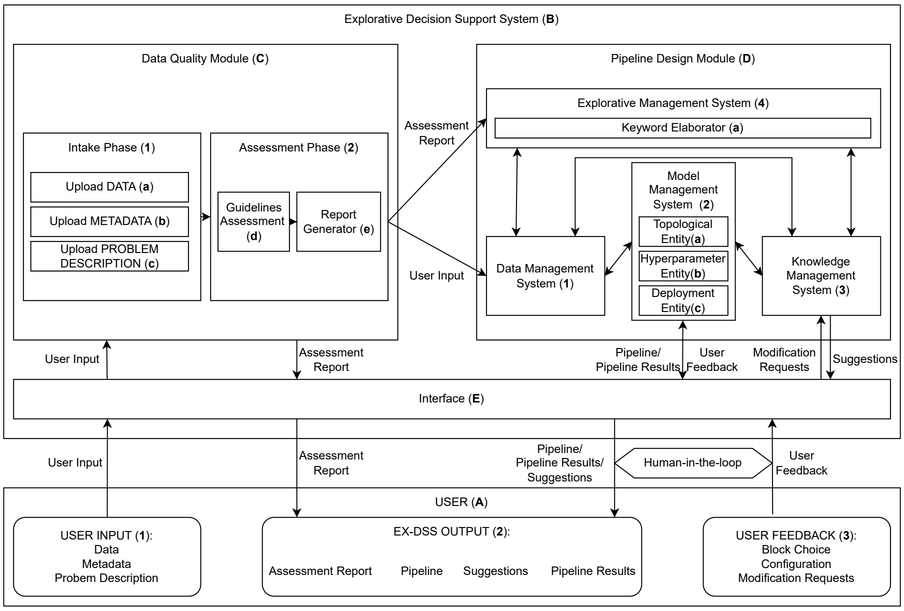


Figure 1: **EX-DSS architecture framework.** This illustrates the proposed EX-DSS framework. (A) User-level input and outputs. (B) Blocks composing the EX-DSS: (C) Block to assess the quality of the user's input and (D) Block to aid the users in designing, configuring, and deploying the pipeline. (E) User Interface that allows the user to dialogue with the EX-DSS. The arrows represent the flow of information inside the systems.

4 EX-DSS FOR SMART PLUG FORECASTING PIPELINES

This Section presents the application of the Explorative Decision Support System (EX-DSS) framework in the design of a smart plug forecasting AI pipeline via a software prototype. Section 4.1 describes the use case and the context to implement such EX-DSS, while Section 4.2 discusses in detail the EX-DSS implementation.

4.1 The Context

The EX-DSS framework described in Section 3 addresses the design of Smart Plug Forecasting pipelines. Smart plugs are devices that fit between appliances' power cords and wall sockets, enabling remote control. Smart plugs convert standard appliances into smart ones. Various appliances like printers, copiers, and TVs are monitored. Studies show that plug loads (excluding lighting, HVAC, and water heating) account for over 40% of total energy consumption in commercial buildings, and that this consumption is increasing over time (Chia et al., 2023; Tuttle et al., 2020). Smart plugs can significantly reduce energy consumption by automatically determining if a plug is idle or active and forecasting its usage. This allows scheduling devices to turn on or off accordingly, leading to potential electricity savings of up to 20%, optimizing building energy efficiency.

AI is crucial for determining plug usage, forecasting consumption, and scheduling devices (Botman et al., 2024). However, finding the best-performing method is time-consuming. The software prototype proposed in this paper aims to guide researchers in designing pipelines for smart plug usage forecasting and scheduling, thereby optimizing building energy consumption. The trained model from the design phase can also be downloaded for further use.

4.2 Description of the Software Prototype

The EX-DSS software prototype is a web application written in Python, implemented using the Dash framework[3]. It assists users by providing descriptions and suggestions for configuring each block of the AI pipeline, based on issue-specific keywords for the Smart Plug problem described in Section 4.1 and listed in the Appendix. Additionally, a chatbox allows users to ask general questions. It is integrated with the Natural Language Processing (NLP) model developed by Cohere[4].

[3]https://dash.plotly.com/
[4]https://cohere.com/

The functionalities of the EX-DSS software prototype are illustrated through an experiment using methods and datasets collected by Botman et al. (2024). While Botman et al. (2024) proposed many alternatives[5], a subset of these methods is included in this prototype. After configuring the pipeline, the EX-DSS connects to an external server for workflow management using the Airflow platform[6].

4.2.1 New Dataset

The EX-DSS software prototype's experiment began with uploading a new dataset (Fig.1C.1). The user provided a title, data source, and a short description (Fig.1C.1.b) to facilitate understanding and improve support in subsequent phases. Next, the user selected keywords from a set proposed by the EX-DSS (see Appendix for the full list). These keywords are crucial for the Explorative Management Subsystem to generate insights and target specific pipeline characteristics. Finally, the user submitted the main data file (Fig.1C.1.a) and could also upload supplementary files such as metadata or additional information (Fig.1C.1.b).

For this experiment, the input was based on the work proposed by Botman et al. (2024)::

- **Title:** Smart Plug Data,
- **Data Source:** "https://gitlab.esat.kuleuven.be/Lola.Botman/smart-plug-pipeline/-/tree/main/Dataset"
- **Data Description:** "The dataset is collected through smart plug sockets between the wall plugs and the electric appliance as detailed in Chia et al. (2023). Smart plugs from Best Energy Reduction Technologies (BERT) are deployed in fifteen buildings on the campus of the University of California, San Diego (UCSD). The power level of each appliance is recorded in mW at fifteen-minute intervals. The dataset consists of 169 high-quality smart plug time series spanning 498 days, from November 18th, 2021, to March 31st, 2023. The 169 plug loads include 146 printers, 16 copiers, 4 TVs, and 3 fax machines. This dataset is openly accessible; see Botman et al. (2024) for more details."
- **Keywords Selected:** "Smart Plugs, High Performance, Non-intrusive"
- **Main File:** "SmartPlugPreprocessedData.csv"
- **Additional Metadata Files:** ["SmartPlugMetaData.csv" "SmartPlugHolidayData.csv"]

[5]https://gitlab.esat.kuleuven.be/Lola.Botman/smart-plug-pipeline

[6]https://airflow.apache.org/

Table 1 shows a snapshot of the main dataset, "SmartPlugPreprocessedData.csv." Rows indicate the time at which power values are recorded, and columns represent monitored electrical appliances with power values in mW. The goal is to predict these power values. In the illustrated subset, recordings start at 11:45 on the 18th of November 2021 until 17:00 on the 31st, 2023, capturing the power load of 169 smart plugs at fifteen-minute intervals.

4.2.2 Assessment Report

The Assessment Phase (Fig.1C.2) begins once the system receives the data and the complementary input. In this phase, the Data Quality module in EX-DSS runs an internal analysis to assess the quality of the inputs used in subsequent steps to generate suggestions based on the work by Rinaldi. et al. (2023).

Although the Assessment Phase can be set up to run different guidelines, for the software prototype, three guidelines were chosen (Fig.1C.2.d):

- **FAIR Paradigms** (Wilkinson et al., 2016): The system checks dataset uniqueness and corrects saving (findability), attempts to open the dataset in a pandas dataframe[7] (reusability), evaluates data encoding (interoperability), and audited security rules (accessibility).
- **Data Quality Analysis:** This included checking for missing values (completeness), assessing value types (consistency), measuring distance between input and reference datasets (accuracy), and verifying if the data is up-to-date (timeliness).
- **Data Cleaning Automation:** This involved routines like evaluating missing values, identifying columns with a single value, and checking for row duplication.

The clarity of the dataset description is also evaluated. At the end of this phase, the system provides two scores: one for data quality analysis and another for data cleaning assessment, as shown in Rinaldi. et al. (2023).

Although this analysis is done in the background, users could configure it manually. They can choose whether to analyze only the main file or include metadata files, select the weight of each quality metric, and decide if any analysis should be skipped. All this information is compiled into a report (Fig.1C.2.b), which users can download.

Listing 1 shows a summary of the report that contains the analysis with the above-mentioned guidelines.

[7]https://pandas.pydata.org/

Table 1: **Smart Plugs Preprocessed Dataset snapshot.** The dataset contains power loads of 169 smart plugs.

Timestamp	**Plug_0**	**Plug_1**	**Plug_2**	**...**	**Plug_167**	**Plug_168**
2021-11-18 11:45:00	NaN	NaN	NaN	...	NaN	NaN
...	...	...	...	...	...	...
2021-11-18 13:15:00	NaN	10.413	19.656	...	NaN	NaN
...	...	...	...	...	...	...
2023-03-31 17:00:00	55.809	10.03	19.44	...	2.64	4.176

Listing 1: Extract of the JSON report containing the findings of the DSS after analyzing the information uploaded by the user.

```
{”SCORES”  {
    ”Data Quality”: 0.56,
    ”Data Cleaning”: 0.99},
”FAIR”  {
    ”Reusability”: ”The dataset was
         rightly converted to a
         dataframe.”
    ”Data_description_clarity_score”:
          ”Input text clarity score:
         85%”,  ...},
”DATA QUALITY”  {
    ”Timeliness”: ”The dataset is not
          up to date -> parameter is
         0. The user set up the
         threshold to 0 years.”  ...},
”DATA CLEANING” {
    ”Time Column”: ”Identified.”
    ”Single Columns”: ”There are no
         columns with only one value.”
           ...}
```

4.2.3 New Project

The user needed to define the project's title and description, upload any additional documents that could help understand the project's final goal, and select the most pertinent keywords from a subset proposed by the EX-DSS (see Appendix for the full keywords list). These keywords assist in formulating proper suggestions and clarifying the research intention. An example of a new project record is as follows:

- **Title:** Smart Plug Project,
- **Project's Description:** “This project implements smart plug active operating mode detection, plug-level load forecasting, and a plug scheduling methodology. A pipeline integrates the detected operating modes with forecasting and scheduling, aiming at reducing building energy consumption (Botman et al., 2024)”
- **Additional Documents:** “No additional description documents”
- **Keywords Selected:** “High Performance, Non-intrusive”

4.2.4 Pipeline Configuration and the EX-DSS Guidance

When a project is initialized, the user can start the design and configuration of the pipeline. Figure 2 shows a screenshot of the configuration page of the EX-DSS prototype software. Figure 2a lists the available blocks developed in the EX-DSS software prototype, while Figure 2b displays the suggestions generated by the Knowledge Management Subsystem with support from the Explorative Management Subsystem.

A suggestion consists of a block's description, insight generated by the Explorative Management for the specific project, and a chatbox for communication with the Knowledge Management Subsystem. The Explorative Management Subsystem (Fig.1D.4) uses the project's keywords to find the closest matching dataset and generate suggestions based on stored expert knowledge. These suggestions include a list of blocks and methods, such as the appropriate forecasting method or technique to preprocess the dataset. This modular approach provides accurate and specialized knowledge, which can be modified to enhance specific subject knowledge or reused for other purposes.

The Knowledge Management Subsystem (Fig.1D.3) maintains a general knowledge view. In the EX-DSS software prototype, it uses a Generative AI model developed by Cohere[8]. The system connects to the Cohere platform via an API, sends a request with a question, and displays the AI-generated response to the user. Additionally, it ensures the correct interconnections between blocks, such as preventing the “Forecast” block from being followed by the “Dataset” block.

Additionally, the Knowledge Management Subsystem ensures that the interconnections between blocks are correct. For example, in the EX-DSS software prototype, the “Forecast” block could not be followed by the “Dataset” block. Figure 2c represents the configuration area, where users choose methods and configure parameters. For example, in the “Forecast” block, users can select the forecasting method and the prediction horizon. In the example shown

[8] https://cohere.com/

(Fig. 2c), "Global XGBoost" was chosen with a prediction horizon of "1" day.

Figure 2d shows the area where the user can organize and connect the chosen blocks. In the example, the tasks were organized as follows: "Dataset" for selecting the dataset, "Operating Mode" for determining the appliance's operating mode, "Forecast" for predicting the appliance loads, "Schedule" for planning when the appliances should be turned on or off, and "Evaluate" for assessing the performance of the other blocks. "Data Cleaning" and "Preprocessing" were excluded since the dataset was already preprocessed

Once the configuration for each block is saved, the pipeline is ready for training.

4.2.5 Pipeline Training and Results

Once the pipeline is ready, the user initiates the Model Management Subsystem (Fig.1B.2), specifically the Deployment Entity (Fig.1B.2.C), to launch the deployment process. In the EX-DSS software prototype, this was implemented by an external server running Airflow[9], a platform designed for executing and monitoring a chain of tasks. The Deployment Entity transferred the necessary files to the Airflow server. Figure 3 displays the pipeline's runtime, which took 9 minutes and 24 seconds to train and generate results.

Upon completion, the Airflow server communicates the results back to the EX-DSS, and users can download these results through the EX-DSS interface. Additionally, users can download the forecasting model, operating mode model, predictions, and final schedule. Table 2 provides an example of the final schedule. Additionally, Table 3 displays the numerical values of the metrics used to assess smart plug scheduling following the application of the "Global XGBoost" method.

Listing 2: Example of static suggestion stored inside the EX-DSS software prototype.

```
{"Pipeline Steps" {
    "Data Cleaning" "Imputation"
    "Data Cleaning" "Normalization"
        ...},
"Operating Mode Detection" {
    "Method" "GMM"
    "Method" "Ensemble"
"Forecasting" {
    "Method" "Global XGBoost"
    "Method" "Global FNN"}
```

[9]https://airflow.apache.org/

Table 2: **Schedule Overview.** The table provides an example of the schedule for plugs 0 and 168. Both plugs are off at night and turn on in the morning, with plug 0 turning on earlier. Plug 0 turns off around lunch while plug 168 turns off later in the afternoon. The schedule can be summarized with the turn-on and turn-off times: Plug 0: {2023-01-13 05:15:00: "Turn on"; 2023-01-13 12:30:00: "Turn off"}, Plug 168: {2023-01-13 09:30:00: "Turn on"; 2023-01-13 16:45:00: "Turn off"}

Timestamp	Plug_0	...	Plug _168
2023-01-12 00:00:00	OFF	...	OFF
...	...	...	...
2023-01-13 05:15:00	ON	...	OFF
...	...	...	...
2023-01-13 09:30:00	ON	...	ON
...	...	...	...
2023-01-13 12:30:00	OFF	...	ON
...	...	...	...
2023-01-13 16:45:00	OFF	...	OFF
...	...	...	...

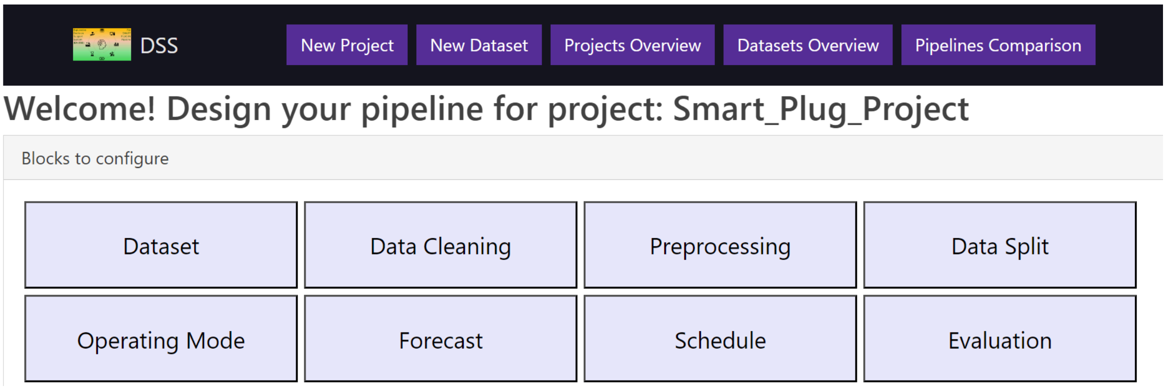
Table 3: **Forecasting Evaluation.** The table illustrates how the chosen forecasting method performed in the pipeline.

	Global XGBoost Schedule
Number of violations (%)	3.27
Missed chances (%)	27.67
Energy saved (%)	25.84
Number of turn on/off commands per plug per day	3.33
Energy Efficiency (%)	47.92

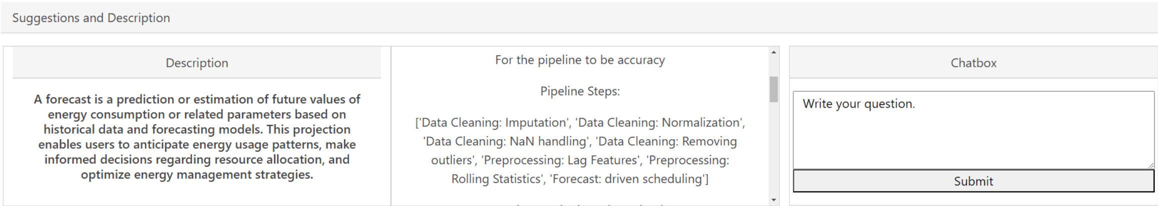
5 RESULTS AND DISCUSSION

The study's findings indicate that the Explorative Decision Support System (EX-DSS) architecture framework enhances the design and implementation of DSS for designing smart plug pipelines, optimizing data forecasting, and, consequently, aiding in reducing energy consumption. By extending the classical Decision Support System (DSS) framework, the EX-DSS incorporates a module for assessing user input quality, facilitating collaboration among users, and promoting reproducibility. This prevents the repetition of previous errors and accelerates the pipeline creation process.

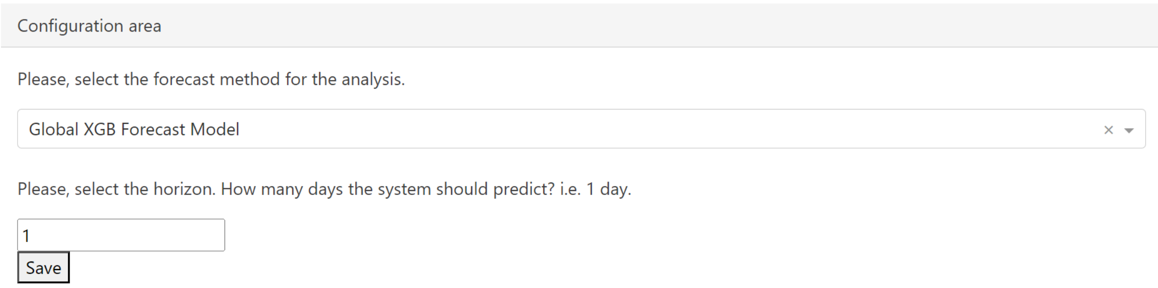
The EX-DSS allows users to upload new datasets, along with descriptions and additional documents, providing a comprehensive overview. The system internally evaluates these inputs and generates a report



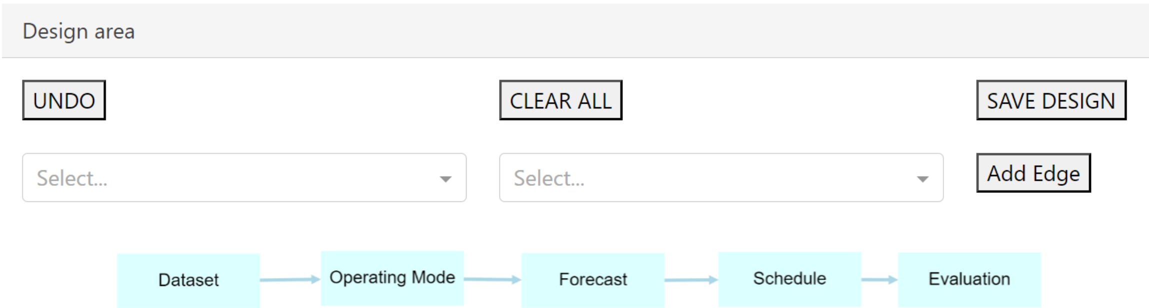
(a)



(b)



(c)



(d)

Figure 2: **Configuration page screenshots.** They were taken from the EX-DSS software prototype. (a) The list of the blocks developed inside the EX-DSS software prototype. (b) The structure of the suggestions provided to the user by the system. (c) This part of the page is dedicated to configuring the blocks. In the image, it is possible to visualize the configuration for the Forecast block. (d) This is the design area where the user can add and connect the blocks.

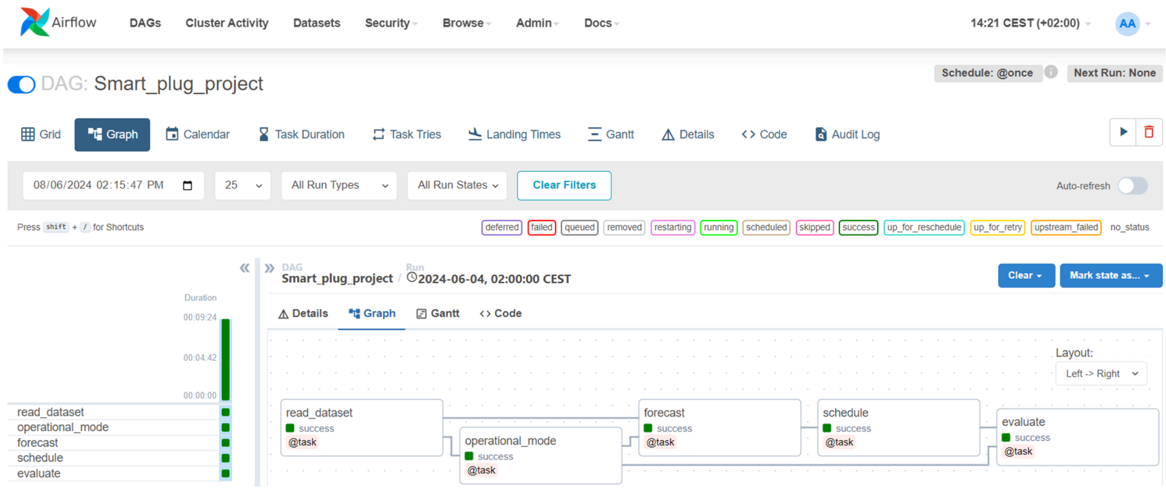


Figure 3: **Screenshot of the Airflow server.** The image shows the service used by the EX-DSS to run the pipeline.

on data quality and description clarity. By maintaining the essential human-in-the-loop characteristic of a DSS, users retain control over the types of analyses the system should perform, ensuring that the information received is of high quality and increasing the likelihood of high-performance outcomes from the deployed pipeline.

A key innovation in the EX-DSS is the inclusion of the Explorative Management Subsystem, which supports the Knowledge Management Subsystem by analyzing user inputs and providing relevant insights. By using a system of keywords tailored to the smart plug forecasting problem, the EX-DSS offers specialized support while maintaining general intelligence within the Knowledge Management Subsystem, thanks to the integration of the Cohere generative AI model. This dual support (general and specific) significantly aids users in designing and configuring forecasting pipelines, allowing for customization and catering to various needs and requirements.

The study has shown that the pipeline design, configuration, and deployment process can be carried out without coding, making it accessible to a broader range of users. However, future work should include extensive testing with diverse datasets to further validate the system's robustness and flexibility. Additionally, user feedback should be collected to refine the interface and improve the overall user experience. Integrating real-time data processing and adaptive learning capabilities could also enhance the EX-DSS, making it more responsive to changing conditions and user needs.

6 CONCLUSIONS

The Explorative Decision Support System (EX-DSS) framework, implemented through a software prototype, has proven effective in designing pipelines for scheduling smart plugs to reduce building energy consumption. By extending the classical DSS with the Explorative Management Subsystem, the EX-DSS provides specialized suggestions alongside general support, enhancing the design process.

The introduction of the data quality module has enabled the EX-DSS to assess the quality of input information, providing users with a clear overview of data status and ensuring higher performance when training models. The system also promotes the shareability of pipeline information and dataset knowledge by requiring users to describe new projects and datasets, thus promoting reproducibility and avoiding duplication.

Despite its benefits, the study faced limitations. The software is not production-ready and was developed for demonstration purposes only. Additionally, the study included only the methods proposed by Botman et al. (2024), and the development relied on input from a single expert due to time constraints and limited availability of specific profiles.

Future work should focus on investigating the interaction between the Knowledge Management Subsystem and the Explorative Management Subsystem to optimize their integration. Extending the software prototype to include result visualization directly within the EX-DSS would improve the user experience. Additionally, developing functionality to compare the results of similar pipelines would help identify the best-performing ones. Transitioning the software from a prototype to a production-ready system is essential, as is conducting extensive testing with diverse datasets and gathering user feedback to refine the interface and enhance overall usability. These steps will help validate the EX-DSS's robustness, improve its flexibility, and ensure it meets the needs of various users.

ACKNOWLEDGEMENTS

This research received funding from • KU Leuven: Research Fund (projects iBOF/23/064, C3/20/117, C3I-21-00316), Industrial Research Fund and several Leuven Research and Development bilateral industrial projects; • Flemish Government Agencies: ○ FWO: SBO project S005319N, TBM Project T001919N; FWO PhD Grant 11K5623N; ○ EWI: the Flanders AI Research Program; ○ VLAIO: CSBO (HBC.2021.0076) Baekeland PhD (HBC.20192204); • EU: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885682). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or ERC. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

Botman, L., Lago, J., Fu, X., Chia, K., Wolf, J., Kleissl, J., and De Moor, B. (2024). Building plug load mode detection, forecasting and scheduling. *Applied Energy*.

Chandra, A., Krovi, R., and Rajagopalan, B. (2007). Risk visualization: A mechanism for supporting unstructured decision making processes. *The International Journal of Applied Management and Technology*.

Chia, K., Lebar, A., Agarwal, V., Lee, M., Ikedo, J., Wolf, J., Trenbath, K., and Kleissl, J. (2023). Integration of a Smart Outlet-Based Plug Load Management System with a Building Automation System. *2023 IEEE PES Grid Edge Technologies Conference and Exposition, Grid Edge 2023*.

Duan, Y. and Xu, M. (2009). Decision support systems in small businesses. *IGI Global - Business and Information Systems Research Centre (BISC)*.

Gonzalez-Andujar, J. (2020). *Introduction to Decision Support Systems*, pages 25–38.

Jacquet-Lagreze, E. and Shakun, M. F. (1984). Decision support systems for semi-structured buying decisions. *European Journal of Operational Research*.

Phillips-Wren, G. (2013). *Intelligent Decision Support Systems*.

Rinaldi., G., Crema Garcia., F., Agudelo., O. M., Becker., T., Vanthournout., K., Mestdagh., W., and De Moor., B. (2023). A framework for a data quality module in decision support systems: An application with smart grid time series. In *Proceedings of the 25th International Conference on Enterprise Information Systems - Volume 1: ICEIS*. INSTICC, SciTePress.

Rupnik, R., Kukar, M., Vračar, P., Košir, D., Pevec, D., and Bosnić, Z. (2019). Agrodss: A decision support system for agriculture and farming. *Computers and Electronics in Agriculture*.

Sloane, E. B. and J. Silva, R. (2020). Artificial intelligence in medical devices and clinical decision support systems. In Iadanza, E., editor, *Clinical Engineering Handbook (Second Edition)*. Academic Press, second edition edition.

Turban, E., Sharda, R., and Delen, D. (2010). *Decision Support and Business Intelligence Systems*. Prentice Hall Press, 9th edition.

Tuttle, R., Trenbath, K., Maisha, K., and LeBar, A. (2020). Assessing and reducing plug and process loads in office buildings. *Technical Report*.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Santos, L. O. B. d. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T. W., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., González-Beltrán, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J. N., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B. L., Rocca-Serra, P., Roos, M., Schaik, R. v., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Lei, J. v. d., Mulligen, E. M. v., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*.

APPENDIX

Keywords used to define datasets and projects: time series, energy, smart plugs, fast, high speed, less computation, accuracy, high performance, precise, reliable, user convenient, minimal disruption, user-friendly, non-intrusive.

Assessing the Use of Online Platforms in Sharing Tacit Knowledge in Innovation Networks

Nathalya Guruge and Jyri Vilko
Department of Industrial Engineering and Management, LUT University, kauppalankatu 13, 45100 Kouvola, Finland
nathalya.guruge@lut.fi, jyri.vilko@lut.fi

Keywords: Tacit Knowledge, Knowledge Sharing, Innovation Networks.

Abstract: In an increasingly digitalized society, sharing tacit knowledge has emerged as a critical activity for driving innovation, especially within innovation networks. This paper presents a systematic literature review to assess the role of online platforms in facilitating tacit knowledge sharing. It explores how digital tools impact the exchange of tacit knowledge, offering a conceptual framework to understand this process. The findings provide strategies for leveraging online platforms to foster innovation within diverse knowledge-driven ecosystems.

1 INTRODUCTION

Tacit knowledge, which is uncodified and personal, has become a crucial asset for innovation in the modern knowledge society (Polanyi, 1966; Nonaka, 1994). As organizations increasingly prioritize knowledge sharing to maintain competitive advantages, particularly within the digital space, understanding how tacit knowledge can be effectively shared online is critical (Ichijo & Nonaka, 2007). This paper ex-amines the practices and tools used in tacit knowledge sharing, motivating and inhibiting factors that affect tacit knowledge sharing and the outcomes of this sharing particularly within innovation networks, through a systematic review of literature.

2 THEORETICAL BACKGROUND

This section explores the foundational theories of tacit knowledge and innovation networks, setting the stage for examining how digital platforms support this knowledge exchange.

2.1 Tacit Knowledge

Tacit knowledge, first explored by Michael Polanyi in 1958, refers to the inexpressible knowledge that resides in the human subconscious, which is difficult to articulate in written or spoken form (Polanyi, 1966). Nonaka (1994) later expanded on this concept, asserting that the knowledge expressible in words is only a small part of the broader body of knowledge. Tacit knowledge, as further developed by Nonaka and Takeuchi (1995), includes personal experiences, beliefs, and values, and is harder to express explicitly, but can be articulated in certain situations. In management studies, Nonaka's definition of tacit knowledge, emphasizing its implicit and experiential nature, is more widely accepted.

The initial view of tacit knowledge suggested by Polanyi sees a minor change during Nonaka's time in such a way that it becomes known as the part of knowledge that is possible to be articulated and expressed in certain situations. In management studies these philosophical definitions of tacit knowledge by Polanyi are not commonly used but Nonaka's definitions are accepted as more relevant. Hence, tacit knowledge is the implicit knowledge possessed by individuals including skills, insights, intuitions, and experiences, which are often difficult to express explicitly.

2.2 Tacit Knowledge Sharing Online

Tacit knowledge sharing is defined as the exchange of intuitive and unarticulated knowledge that takes the form of personal skills, know-how, experience, and expertise in knowledge management literature. This type of knowledge, which is intangible and personal, occurs primarily between peers or co-workers in a workplace and it involves direct

Guruge, N. and Vilko, J.
Assessing the Use of Online Platforms in Sharing Tacit Knowledge in Innovation Networks.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 206-213
ISBN: 978-989-758-716-0; ISSN: 2184-3228

exchange of knowledge between employees (Majewska and Szulczynska, 2014). However, the importance of the need of support to foster trust and overcome resistance to knowledge transfer is highlighted by Majewska and Szulczynska (2014) signalling of an establishment of a new argument related to factors that can catalyse tacit knowledge transfer which is addressed in this present study.

Tacit knowledge sharing is viewed as tacit-to-tacit conversations (Socialization) and as tacit-to-explicit conversations (Internalization) according to the knowledge creation theory of Nonaka and Takeuchi (1995) (Marwick, 2001; Sarkiunaite and Kriksciuniene, 2005; Lopez-Nicolas and Soto-Acosta, 2005). This is because of the interactive and dynamic relationship between tacit and explicit knowledge (Panahi et al., 2015). The scholars explained this further using the spiral model of knowledge management and stated that Nonaka and Takeuchi's theory on knowledge creation which adopts a spiral movement indicates it is unavoidable to externalize and internalize tacit knowledge when communicating. With the introduction of web initiatives, arguments on new tools that can drive sharing tacit knowledge through interactive and collaborative technologies was discussed (Panahi et al., 2012). ICT has been one of the main enablers of knowledge sharing activities (Panahi et al., 2016). Building on this, online platforms, particularly those leveraging Web 2.0 technologies, have enabled more dynamic and interactive forms of knowledge sharing.

2.3 Innovation Networks

Innovation builds upon previous discoveries and inventions, much like Newton's famous analogy of "standing on the shoulders of giants."(Acemoglu et al., 2016). A consistent and reliable network of innovators facilitates the ongoing accumulation of technological and scientific advancements leading to economic growth and transformation. This is well explained through Neo-Schumpeterian economics that places technological innovation at the core of economic advancement. According to the theory accumulation of knowledge and technological advancement is fueled by research and development funding, intellectual property protection (Law), Support for entrepreneurship and policymaking (Government intervention). Hence, this is an innovation system (Beije and Groenewegen, 1992) and in modern literature known as 'Innovation Networks'. The term refers to the actors in an innovation system and their relationships.

Difficulties in knowledge creation and learning significantly impact the overall results and achievements of a network of collaborating organizations (Lampela, 2009). One practical solution put forward by Lampela (2009) to overcome the challenges in innovation networks is implementing virtual innovation teams. Sharing knowledge within virtual teams can be complicated since it is unlike face-to-face communication. Knowledge, especially tacit knowledge, is sensitive information that is embedded in a person's sub conscious mind. The sharing of sensitive information in innovation networks necessitates a high degree of trust. Simultaneously, the operational processes demand speed and flexibility, despite facing uncertainty, complexity, and ambiguity in the available information and operating environment (Lampela, 2009). Hence, it is worth assessing the factors that enable and moderate virtual tacit knowledge sharing in innovation networks.

3 METHODOLOGY

This study utilizes a systematic literature review approach, analysing peer-reviewed journal articles and conference papers published between 2003 and 2023. The review spans various academic databases, including Scopus and Web of Science, and focuses on studies that explore the use of online platforms in tacit knowledge sharing.

Main research question:

What is the use of online platforms in sharing tacit knowledge in innovation networks?

This question is split into five questions to in a way that they cover the wide span of the above research question.

1. What are the practices used to share tacit knowledge online?

2. What are the tools used to share tacit knowledge online?

3. Which factors enhance tacit knowledge sharing online?

4. Which factors inhibit tacit knowledge sharing online?

5. What are the possible outcomes of tacit knowledge sharing online?

Once formulating the research questions, a comprehensive search strategy was developed to find articles and conference papers. The strategy includes which scientific databases are used, the search terms and inclusion exclusion criteria of data. This review spans academic databases such as Scopus, Web of Science and Ebsco. The descriptor contained synonyms and related terms from each category of key words to ensure the comprehensive coverage of the topic. These words were either present in the Title of the article which was optional and were present in both the abstract and the key words section. This bibliographic search was conducted from June 2023 to July 2024. The search was refined and repeated several times until the optimum results were yielded. The search for relevant literature was restricted to peer-reviewed journal articles and conference papers published in English between 2003 and 2023. Textbook chapters were not included, and while the search was limited to English language publications, there were no restrictions based on geographic location. The initial search yielded 80 articles, and after filtering, 24 journal articles were selected for review.

In search of conference papers, the same search strategy was applied which resulted in 18 papers from 16 conferences. They were subjected to researcher triangulation and 11 papers were considered suitable to study the research questions. To ensure the quality of studies conferences with a h-Index ranking were selected. H-index was introduced by Jorge E. Hirsch in 2005 to measure a researcher's scientific productivity and impact based on the number of publications and citations (Hirsch, 2005; Meho and Yang, 2007). This metric, now widely adopted in academia, considers an H-Index of 30 or above as a benchmark for selecting conferences in this study.

Once applying the H-Index seven articles were initially chosen for full-text assessment. However, two additional conference papers were included due to their relevance to the research objectives and quality: the Wireless Telecommunications Symposium paper, a good H-Index of 20, and "Potentials of social media for tacit knowledge sharing amongst physicians: Preliminary findings" (Panahi et al., 2012), which was deemed exceptionally useful despite not meeting the strict criteria. This resulted in a total of eight conference papers being selected for review. Hence, a total of 30 journal articles and conference papers were analysed in this review.

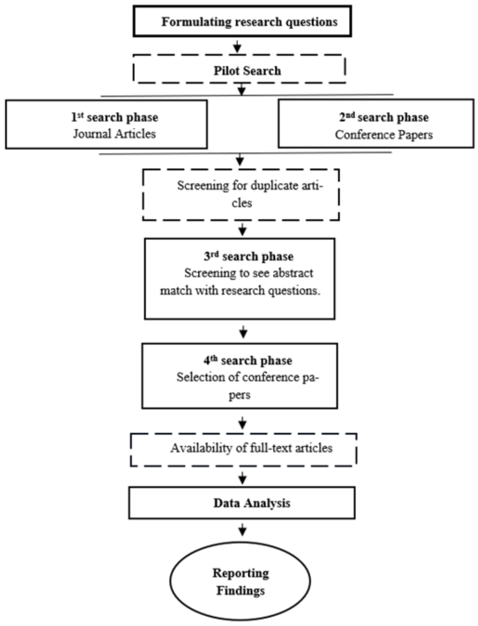


Figure 1: Research Design.

4 DATA ANALYSIS

This study employed both content and thematic analysis to analyse data from 30 articles, using NVivo and Microsoft Excel for qualitative data analysis. Content analysis, following Gaur and Kumar (2018), quantified specific elements within the data, while thematic analysis, adhering to Braun and Clarke's (2017,2022) guidelines, identified broader themes and patterns. Data from the articles were recorded in an Excel sheet with key columns such as author, DOI, and findings, which were analysed to identify descriptive and thematic results. A coding framework was developed based on Gaur and Kumar's (2018) categories, focusing on practices, tools, factors enhancing or inhibiting tacit knowledge sharing, and outcomes. Iterative adjustments were made to account for emergent themes that did not fit the initial categories.

5 ANALYSIS

The line graph depicts the selected number of journal articles, conference papers published each year and the total number of data used in this review. The highest number of articles has been published in 2019, and there are no conference papers this year. This total drastically decreases in the following year, the year when the covid-19 pandemic started spreading outside of China. The lowest number of journal articles and conference papers were seen in 2015 with zero data from the year. There are 8 years in which only one article or conference paper was found. The overall trend is an increase which indicated promising future for research interests in tacit knowledge sharing online.

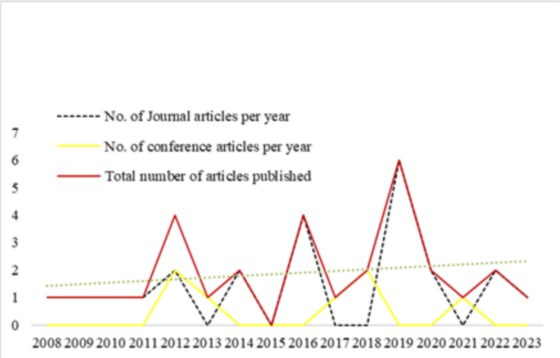


Figure 2: Number of journal articles and conference papers according to the year published.

Of the 30 articles analysed, 22 mentioned practices or/and tools used in TKSO. In the present context practices are the processes and interactions that facilitate TKSO. Hence, explaining how a tool, or an online platform enables individual to share expertise in online environments. The identified practices are; Structuring information (Kogl and Gilbert, 2019), Virtual communities (Chi-cheng et al., 2022), Virtual learning activities (Haag and Duan, 2012; Hildrum, 2009), case analyses and sharing of experiences (Deng et al., 2023) broadcasting information to wider audiences, faster dissemination, personalized feeds, staying up-to-date, documenting experiences, and improved retrievability (Panahi et al., 2016).

Tools in the present research context are specific technologies and platforms that assist the execution of the practices. Such tools identified in the reviewed studies are; Web 2.0 technologies (E-ling and Xiaoxia, 2019; Jarrahi et al., 2019; Fang and Gong, 2012; Panahi et al., 2012), Social media platforms (Panahi et al., 2016; Panahi et al., 2016 Mladenovoic and Krajina, 2020; Liana et al., 2014; Amidi et al., 2018; Amidi et al.,2017; Mislan et al.,2016), Online community forums (Rachaneewan, 2011; Yuan et al., 2016; Ku, 2014), Social networking applications (SNAs) (Diptee and Diptee, 2013),Videos, Text, interactive elements (drag and drop exercises (Hon et al., 2008; Mislan et al.,2016), IoT, Digital twins, Siemens tecnomatrix (Guerra-Zubiage et al., 2021).

16 studies presented factors affecting tacit knowledge sharing online. These factors either belonged to motivating factors or hindering factors. The following tables (table 1 and 2) depict each factor identified along with the relevant references.

Table 1: Factors motivating tacit knowledge sharing online.

Motivating factors	References
Closeness of the network	(Chi-Cheng et al.,2022; Cao et al.,2012; Mislan et al., 2016)
Frequency of interaction	(Cao et al.,2012; Chi-Cheng et al.,2022; Tee and Karney,2010; Diptee and Diptee,2013)
Stability of member interaction	(Cao et al.,2012; Chi-Cheng et al.,2022; Ku, 2014; Großer et al.,2018; Panahi et al., 2016)
Intrinsic motivations (enjoyment & self-efficacy)	(Wang et al.,2022; Mislan et al., 2016; Hildrum, 2009; Amidi et al.,2017)
Personal traits	(Yuan et al., 2016; Diptee and Diptee,2013)
Trust in the informant	(Diptee and Diptee,2013; Großer et al.,2018; Panahi et al., 2016)
Communication style	(Diptee and Diptee,2013; Großer et al.,2018)
Appropriate technology use	(Großer et al.,2018; Amidi et al.,2017; E.-Ling and Xiaoxia, 2019)

Factors that act as barriers to online tacit knowledge sharing have not been studied by many researchers hence the few refences.

Table 2: Factors hindering tacit knowledge sharing online.

Hindering factors	References
Lack of communication	(Metin, 2019)
Conflicting perspectives	(Metin, 2019; Jarrahi et al., 2019)
Misaligned priorities	(Metin, 2019)
Lack of trust	(Cao et al.,2012; Mislan et al., 2016)
Virtual rewards	(Wang et al.,2022)
Value dissimilarity	(Jarrahi et al., 2019)
Economic factors	(Deng et al.,2023)
Network reciprocity	(Deng et al.,2023)

Regarding the outcomes of Tacit knowledge sharing online (TKSO), nine studies directly or indirectly discussed the impact of TKSO. These outcomes were concentrating on Organizational (2 Studies), Interpersonal (2 Studies), Both individual and Organizational (1 study), and Individual (3 studies). The sector with the most outcomes is the healthcare sector.

6 FINDINGS

The systematic review of 30 articles reveals a diverse landscape of tools and practices employed for TKSO. 22 of such articles addressed tools or practices used in different contexts. Hence, this paper identifies practices used for TKSO as; information structing, memberships in virtual communities, virtual learning activities, virtual case discussions, and the use of social media features such as broadcasting and personalized feeds. These practices collectively enhance the accessibility and co-creation of tacit knowledge among innovation networks.

The tools identified in this study largely falls into two categories; 1. communicating and collaborating tools, 2. Social media platforms. Skype, slack, e-learning platforms, blogs, and wikis fall under communicating and collaborating tools while Twitter, Facebook and LinkedIn can be categorized under social media platforms. However, both these categories of tools enable real-time interaction, collaborative content creation and knowledge disseminations among innovation networks.

Real-time communication, collaborative features, and personalization/ filtering capabilities within these tools are essential for effective TKSO in innovation networks. They promote engagement, facilitate interaction, and help individuals find relevant up to date information. In addition to the established tools that are discussed above, this review identified emerging technologies such as digital twins and IoT being explored to capture tacit knowledge transfer between humans and collaborative robots (Guerra-Zubiage et al.,2021) highlighting the evolving landscape of TKSO.

Out of 30 revied studies, 16 identified factors influencing tacit knowledge sharing (TKSO), categorized as moderating or inhibiting. The four key moderating factors are network characteristics, intrinsic motivation, trust, and communication technology. Strong relationships and frequent interactions within a network promote effective TKSO, with trust in the knowledge provider—based on benevolence, competence, and integrity—being crucial. Organizations should encourage trust through transparent communication and recognition of expertise. Additionally, clear communication and appropriate technology, such as real-time communication tools, are vital for successful knowledge exchange, with the choice of technology and communication style being tailored to specific contexts for optimal results.

Several factors inhibiting tacit knowledge sharing (TKSO) were identified, including lack of trust and poor communication, which are interconnected barriers in virtual platforms. Building trust through open communication and clear expectations is crucial. Differences in values or negative attitudes within innovation networks also hinder knowledge sharing. Additionally, virtual rewards, if poorly designed, can demotivate TKSO. Economic disparities between regions influence knowledge flow, with wealthier regions typically sharing more knowledge with less successful ones. Lastly, network reciprocity where knowledge is shared only when something valuable is received can limit the willingness to share tacit knowledge.

The findings highlight the complexity of tacit knowledge sharing (TKSO), influenced by both individual and contextual factors. Positive influences include network characteristics and intrinsic motivation, while the effects of external rewards and social dynamics are more nuanced. TKSO is crucial for innovation networks, promoting trust, collaboration, and the sharing of expertise to drive growth. The outcomes of TKSO extend beyond the act of sharing, producing tangible benefits such as increased knowledge and skills, and intangible ones such as enhanced collaboration and innovation.

Nine studies, primarily focused on healthcare, examined TKSO outcomes at individual and organizational levels. Panahi et al. (2016) and Muhammed Kashif et al. (2019) found that sharing medical knowledge through social media enhances both professional and personal development. At the individual level, TKSO improves skills, motivation, and innovation capabilities. Hildrum (2009) and Tee and Karney (2010) emphasized that TKSO fosters learning and problem-solving skills. Organizational-level outcomes, as noted by Buunk et al. (2018) and Panahi et al. (2016), include improved problem-solving and innovation, showing TKSO as a catalyst for individual and organizational growth and innovation.

7 CONCLUSION

This systematic review highlights the crucial role of online platforms in facilitating tacit knowledge sharing (TKSO) within innovation networks. It identifies a range of tools, from social media to emerging technologies like digital twins, which promote TKSO. Both individual-level and organizational-level benefits, such as professional development, innovation, and problem-solving, are noted. However, TKSO's effectiveness is influenced by factors like network characteristics, motivation, trust, and appropriate technology use.

The review offers insights for fostering a knowledge-sharing culture by emphasizing trust, collaboration, and the selection of suitable tools. A conceptual framework is developed to illustrate the interplay between individual characteristics (moderating factors) and contextual factors (enabling factors) in the TKSO process. The framework underscores the recurring nature of knowledge sharing and provides a foundation for future research and practice.

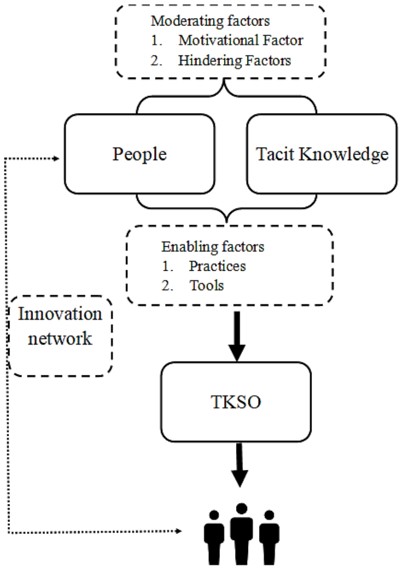


Figure 3: Conceptual framework for TKSO in innovation networks.

Overall, this review highlights the multifaceted impact of TKSO, extending beyond the mere exchange of information. By understanding these outcomes, organizations can develop strategies to leverage online platforms and create environments that encourage and reward the sharing of tacit knowledge, ultimately driving individual growth and organizational success.

REFERENCES

Acemoglu, D., Akcigit, U., Kerr, W. R. (2016). Networks and the Macroeconomy: An Empirical Exploration. NBER Macroeconomics Annual, 30, 273-335. DOI: 10.1086/685961.

Amidi, A., Jabar, M., Jusoh, Y.Y. and Abdullah, R. (2017). Appropriation of social media for fostering effective tacit knowledge sharing: developing conceptual model. Journal of Physics: Conference Series, [online] 892, p.012012. doi: https://doi.org/10.1088/1742-6596/892/1/012012.

Amidi, S., Yusmadi Yah Jusoh, Mar Yah Said, Jabar, M.A. and Abdullah, R. (2018). Application of Social Media Among Medical Practitioner for Sharing Tacit Knowledge: A Pilot Study. [online] doi: https://doi.org/10.1007/978-981-10-8276-4_32.

Beije, P. R. & Groenewegen, J. (1992). A Network Analysis of Markets. Journal of economic issues. [Online] 26 (1), 87–114. DOI: 10.1080/00213624.1992.11505263.

Braun, V. & Clarke, V. (2022). Thematic analysis: a practical guide. SAGE Publications, London.

Buunk, I., Smith, C.F. and Hall, H. (2018). Tacit knowledge sharing in online environments: Locating 'Ba' within a platform for public sector professionals. Journal of Librarianship and Information Science, [online] 51(4), pp.1134–1145. doi: https://doi.org/10.1177/0961000618769982.

Cao, W., Xu, L., Liang, L. and Chaudhry, S.S. (2012). The impact of team task and job engagement on the transfer of tacit knowledge in e-business virtual teams. Information Technology and Management, [online] 13(4), pp.333–340. doi: https://doi.org/10.1007/s10799-012-0129-6

Clarke, V. & Braun, V. (2017). Thematic analysis. The journal of positive psychology. [Online] 12 (3), 297–298.

Deng, Z., Deng, Z., Liu, S. and Evans, R. (2023). Knowledge transfer between physicians from different geographical regions in China's online health communities. Information Technology & Management. [online] doi: https://doi.org/10.1007/s10799-023-00400-3.

Diptee, D. and Diptee, J. (2013). Tacit knowledge acquisition in virtual teams. 2013 Wireless Telecommunications Symposium (WTS). [online] doi: https://doi.org/10.1109/wts.2013.6 566228

Duan, Y. and Haag, M. (2011). Understanding personal knowledge development in online learning environments: an instrument for measuring externalization, combination, and internalization. Electronic Journal of Knowledge Management. [online] Available at: https://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=78296196&site=ehost-live

Fang, J. and Gong, L. (2011). Web2.0 Environment of Personal Knowledge Management Applications. Lecture notes in electrical engineering, [online] pp.1639–1646. doi: https://doi.org/10.1007/978-94-007-1839-5_177

Gaur, A. and Kumar, M. (2018). A systematic approach to conducting review studies: An assessment of content analysis in 25 years of IB research. Journal of World Business, 53(2), pp.280-289.

Großer, B., Kepplinger, S., Vogel, C. and Baumöl, U. (2018). Knowledge Processes in Virtual Teams - Tacit Knowledge. [online] 20th International Conference on Enterprise Information Systems. doi: https://doi.org/10.5220/0006674602470254

Guerra-Zubiaga, D.A., Navid Nasajpour-Esfahani, Phan, N.Q., Gupta, S. and Block, L. (2021). Tacit Knowledge

Capture Using Digital Tools in a Human-Robot Interaction: A Case Study. Volume 2B: Advanced Manufacturing. [online] doi: https://doi.org/10.1115/imece2021-66084.

Haag, M. & Duan, Y. (2012). Understanding Personal Knowledge Development in Online Learning Environments: An Instrument for Measuring Externalization, Combination, and Internalization. The Electronic Journal of Knowledge Management, 10(1), pp. 39-47.

Hildrum, J. M. (2009). Sharing Tacit Knowledge Online: A Case Study of e-Learning in Cisco's Network of System Integrator Partner Firms. Industry & Innovation, [online] 16(2), pp.197–218. doi: https://doi.org/10.1080/13662710902764360.

Hirsch, J. E., 2005. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, 102(46), pp.16569-16572.

Huang, C.-Y., Gamage, B., Bryce, E.A., LoChang, J., Yassi, A., Maultsaid, D. and Yu, S. (2008). Personal protective equipment in health care: Can online infection control courses transfer knowledge and improve proper selection and use? American Journal of Infection Control, [online] 36(10), pp. e33–e37. doi: https://doi.org/10.1016/j.ajic.2008.07.007

Hsiao, E-Ling. and Huang, X. (2019). Strategies to Support Personal Knowledge Management Using a Wiki Site in Online Courses. The Journal of Educators Online, [online] 86 16(1). doi: https://doi.org/10.9743/jeo.2019.16.1.5

Jarrahi, M. H., Philips, G., Sutherland, W., Sawyer, S. and Erickson, I. (2019). Personalization of knowledge, personal knowledge ecology, and digital nomadism. Journal of the Association for Information Science and Technology, [online] 70(4), pp.313–324. doi: https://doi.org/10.1002/asi.24134

Kashif Imran, M., Fatima, T., Aslam, U. and Muhammad Javed Iqbal, S. (2019). Exploring the Benefits of Social Media Towards Knowledge Sharing Among Doctors. Pakistan Journal of Psychological Research, [online] 34(2), pp.331–352. doi: https://doi.org/10.33824/pjpr.2019.34.2.18.

Ku, E. C. S. (2012). Distributed Fascinating Knowledge Over an Online Travel Community. International Journal of Tourism Research, [online] 16(1), pp.33–43. doi: https://doi.org /10.1002/jtr.1895

Kögl, S. and Silvius, G. (2019). Using Patterns to Capture and Transfer Tacit Knowledge in Virtual Project Teams. Semantic Scholar, [online] 7(20). doi: https://doi.org/ 10.19255/JMPM02002

Lampela, H.: Inter organizational learning within and by innovation networks. Lappeenranta University of Technology. http://lutpub.lut.fi/handle/10024/45412 (2009)

Lee, C.-C., Lee, L.-C. and Kao, R.-H. (2022). How do Enterprises promote innovation performance? A study on the relationship between online communities and innovation 87 performance-exploring the mediating effect of tacit knowledge. Current Psychology. [online] doi: https://doi.org/10.1007/s12144-022-02890-

Liu, E. and Song, M.: Innovation Networks and R&D Allocation *. [online] Available at: https://ernestliu.scholar.princeton.edu/sites/g/files/toruqf4426/files/documents/paper.pdf (2024)

López-Nicolás, C. and Soto-Acosta, P., 2010. Analyzing ICT adoption and use effects on knowledge creation: An empirical investigation in SMEs. *International Journal of Information Management*, 30(6), pp.521-528.

Marwick, A., 2001. *The New Nature of History: Knowledge, Evidence, Language*. Oxford: Oxford University Press.

Majewska, M. & Szulczyńska, U. (2014). Methods and Practices of Tacit Knowledge Sharing Within an Enterprise: An Empirical Investigation. Oeconomia Copernicana. [Online] 5 (2), 35–48.

Meho, L.I. and Yang, K., 2007. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), pp.2105-2125.

Metin, F. (2019). 'The factors influence online tacit knowledge sharing in public organisations: A qualitative case study from Turkey', *Gaziantep University Journal of Social Sciences*, 18(3), pp. 1121-1142. doi: 10.21547/jss.442144.

Mislan, A., Sahibi, M.S., Jamaluddin, A., Sufi Alawi Idris, A. and Hartina Ahmad, N. (n.d.). The arts of Knowledge Transfer in Virtual Organization: Tacit Knowledge. International Journal of Computer Science and Network Security, [online] 16(4). Available at: http://paper.ijcsns.org/07_book/201604/20160408.pdf

Mladenović, D. and Krajina, A. (2019). KNOWLEDGE SHARING ON SOCIAL MEDIA: STATE OF THE ART IN 2018. Journal of Business Economics and Management, [online] 21(1), pp.44–63. doi: https://doi.org/10.3846/jbem.2019.11407

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. New York: Oxford University Press.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization science (Providence, R.I.). [Online] 5 (1), 14–37. DOI: 10.1287/orsc.5.1.14.

Nilmanat, R. (2011). Investigating image usage and tacit knowledge sharing in online communities.

Nonaka, I. & Teece, D. J.: Managing industrial knowledge: creation, transfer, and utilization. London: Sage.ISBN: 0-7619-5498-8 (2001)

International Journal of Innovation and Learning, [online] 10(4), p.350. doi: https://doi.org/10.1504/ijil.2011.043095

Panahi, S., Watson, J. and Partridge, H. (2012). Potentials of social media for tacit knowledge sharing amongst physicians: preliminary findings. Science & Engineering Faculty. [online] Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84878301263&partnerID=40&md5=c5aabe4d94691a551f8ab7d717846694.

Panahi, S., Watson, J. and Partridge, H. (2015). Information encountering on social media and tacit knowledge sharing. Journal of Information Science, [online] 42(4), pp.539–550. doi: https://doi.org/10.1177/0165551515598883

Panahi, S., Watson, J. and Partridge, H. (2016). Conceptualizing social media support for tacit knowledge sharing: physicians’ perspectives and experiences. Journal of Knowledge Management, [online] 20(2), pp.344–363. doi: https://doi.org/10.1108/jkm-06-2015-0229.

Polanyi, M. (1958). Personal knowledge; towards a post-critical philosophy. University of Chicago Press, Chicago.

Polanyi, M. (1966). The Tacit Dimension, University of Chicago Press, London (1966).

Rachaneewan, M. (2011). Investigating image usage and tacit knowledge sharing in online communities," International Journal of Innovation and Learning, Inderscience Enterprises Ltd, vol. 10(4), pages 350-364.

Razmerita, L., Kirchner, K. and Nabeth, T. (2014). Social Media in Organizations: Leveraging Personal and Collective Knowledge Processes. Journal of Organizational Computing and Electronic Commerce, [online] 24(1), pp.74–93. doi: https://doi.org/10.1080/10919392.2014.866504.

Sarkiunaite, I. and Kriksciuniene, D., 2005. Impacts of Information Technologies on Tacit Knowledge Sharing: Empirical Approach. Informacijos Mokslai, 35, pp.69-79.

Ichijō, K. & Nonaka, I. (2007). Knowledge creation and management: new challenges for managers. Oxford: Oxford University Press.ISBN: 9780195159622.

Tee, M. Y. and Karney, D. (2010). Sharing and cultivating tacit knowledge in an online learning environment. International Journal of Computer-Supported Collaborative Learning, [online] 5(4), pp.385–413. doi: https://doi.org/10.1007/s11412-010-9095-3.

Tidd, J., Bessant, J., & Pavitt, K.: "Managing innovation; integrating technological, market and organizational change, 2d ed", 2001/11//, Reference and Research Book News, vol. 16, no. 4 (2001).

Wang, N., Yin, J., Ma, Z. and Liao, M. (2021). The influence Practice of rewards on knowledge sharing behaviors in virtual communities. Journal of Knowledge Management, [online] 26(3), pp.485–505. doi: https://doi.org/10.1108/jkm-07-2020-0530

Yuan, D., Lin, Z. and Zhuo, R. (2016). What drives consumer knowledge sharing in online travel communities? Personal attributes or e-service factors? Computers in Human Behavior, [online] 63, pp.68–74. doi: https://doi.org/10.1016/j.chb.2016.05.019

Yudhistira, D.S. and Sushandoyo, D. (2020). Does political self-disclosure in social media hamper tacit knowledge sharing in the workplace? VINE Journal of Information and Knowledge Management Systems, [online] 50(3), pp.513–530. doi: https://doi.org/10.1108/vjikms-08-2019-0128

Development of a Concept Map Evaluation Support System for Social Studies Learning

Keitaro Tokutake[1][a], Dai Sakuma[2][b] and Masao Murota[3][c]
[1]School of Environment and Society, Tokyo Institute of Technology, Japan, 2-12-1, Ookayama, Meguro-ku, Tokyo, Japan
[2]Faculty of Teacher Education, Shumei University, 1-1 Daigaku-cho, Yachiyo-shi, Chiba, Japan
[3]Institute for Liberal Arts , Tokyo Institute of Technology, Japan, 2-12-1, Ookayama, Meguro-ku, Tokyo, Japan
tokutake.k1128@gmail.com, sakuma@mailg.shumei-u.ac.jp, murota@ila.titech.ac.jp

Keywords: Educational Measurement, Educational Evaluation, Instructional Material Structural Analysis, Social Studies Education.

Abstract: In social studies learning, it is crucial for students to develop a "structural awareness" that systematically organizes the connections between social phenomena. One approach to achieving this is concept mapping, and Tokutake et al. (2019) developed the S-R Score Table as a method for teachers to evaluate students' concept maps. However, the procedure for utilizing this method is complex, and interpreting the results requires specialized knowledge and insight. Therefore, in this study, we developed an evaluation support system that automates the creation of the S-R Score Table and displays the comparison results of the concept maps created by teachers and students in a comprehensive view. This system is designed to make it easier for teachers to evaluate the overall trends in students' structural awareness. The application of this system in actual classroom settings suggested that it could enhance teachers' ability to evaluate the structural awareness trends of the entire class.

1 INTRODUCTION

In social studies learning, it is crucial for students to grasp the meaning, significance, characteristics, and interrelationships of social phenomena. Brahami and Nada (2019) found that the process of extracting expert knowledge and mapping relationships improves creativity and innovation efficiency. Based on this, we believe that for students to form a structural awareness of social matters, it is first necessary for them to be able to grasp the "structural awareness" formed by the teacher, who uses a "structuring perspective" as an expert.

Methods to visualize students' structural awareness include the concept mapping method developed by Novak et al. (1984) and the hierarchical directed graphs by Sato (1987), both of which students can draw. In this study, Sato's hierarchical directed graphs are considered one method of drawing concept maps.

Research evaluating students' concept maps includes scoring comparisons between learners' and experts' concept maps (Aliya et al.,2021) and link comparisons (Kato et al., 1988, Jaruwat,2016). These studies compared individual learners' concept maps with those of experts, making it difficult to grasp the recognition trends of all learners. Therefore, Tokutake et al. (2019) developed the S-R Score Table to evaluate the structural awareness of individual students and the entire student body by comparing the connections in structural graphs drawn by teachers and students. In the S-R Score Table, each connection in the teacher's structural diagram is categorized based on the perspectives and ways of thinking required for the connection, giving meaning to the connections. This helps teachers evaluate individual students' structural perspectives and structural awareness based on the presence or absence of these connections in students' concept maps.

Next, by displaying the connection information of students' concept maps in a list, the S-R Score Table

[a] https://orcid.org/0009-0002-1099-5518
[b] https://orcid.org/0009-0007-3638-8229
[c] https://orcid.org/0009-0000-9727-3096

Tokutake, K., Sakuma, D. and Murota, M.
Development of a Concept Map Evaluation Support System for Social Studies Learning.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 214-221
ISBN: 978-989-758-716-0; ISSN: 2184-3228

supports teachers in evaluating the structural perspectives and recognition trends of the entire student body. Also, in the S-R Score Table, attention coefficients are indicated as a metric to identify students with unique recognition patterns and links where the entire student body may have unique recognition patterns.

However, these methods are difficult to use directly in schools because of the complexity of the procedures for analysis and the specialized knowledge and insight required to read the indicators. In order to solve these problems, it is necessary to consider ways to facilitate their interpretation and reading.

Therefore, the purpose of this study is to develop an evaluation support system using the S-R Score Table to make it easier for teachers to understand the structural recognition trends of individual students and the entire student body.

2 OVERVIEWS OF THE S-R SCORE TABLE

The S-R Score Table, which lists the connection information from the concept maps of the entire student body, is developed with reference to the S-P table by Sato (1998), a method for graphically interpreting students' learning achievement. An overview of the S-R Score table is shown in Figure 1.

To create the S-R Score Table, each node in the teacher's concept map is assigned a sequential number. If the teacher connects node 1 to node 2, the link is labeled "1→2". Next, to explicitly show the structural perspectives required to draw the connections between nodes, the connections are categorized using the items in Table 1. These three item combinations define the structural perspectives in the S-R Score Table.

For instance, if the teacher connects nodes 1 and 2, which relate to "politics" and "culture" respectively, and this relationship is neither explicitly nor abstractly described in the textbook, requiring students to analyze and infer the causal relationship between the events, the connection between nodes 1 and 2 is categorized as "different fields, no description, analysis". The connections in the teacher's concept map, classified according to the items and elements in Table 1, are placed in the connection items of Figure 1.

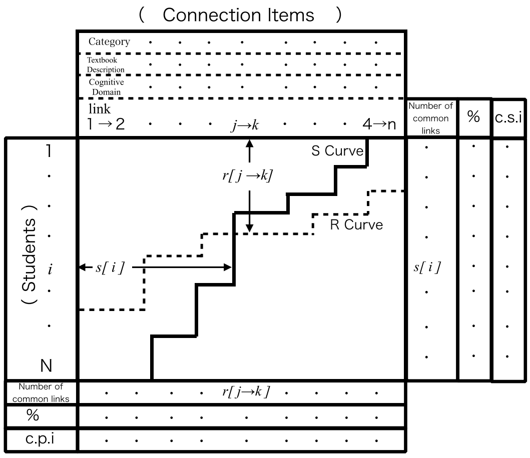


Figure 1: Basic structure of the S-R Score Table.

When comparing the teacher's and students' concept maps, common links are marked as "1" and unique teacher links, which are not drawn by students, are marked as "0" in the table. Students are then ranked in descending order based on the number of common links, and each link item is similarly ranked.

Based on the number of common links for each student, an S (Student) curve (solid line in Figure 1) is drawn. In Figure 1, *i* represents the total number of common links for student *i*.

Next, for each link between nodes, an R (Recognition) curve (dotted line in Figure 1) is drawn according to the number of students who made the common link. In Figure 1, *r[j→k]* represents the total number of students who recognized the relationship and made the common link between nodes *j* and *k*.

Table 1: Classification items of links.

Item	Element	Content
Category	Same Category	Links drawn between events in the same category.
	Different Category	Links drawn between events in different category.
Relationship Description	Described	Links explicitly explained in the textbook.
	Undescribed	Links not explicitly explained in the textbook.
Cognitive Domain	Knowledge	Links inferred from relationships explicitly stated in the textbook.
	Interpretation	Links inferred from abstract descriptions or observations.
	Analysis	Links inferred from causal relationships arising from events.

In the S-R Score Table, the overall structural awareness of students is evaluated using the S curve and R curve. The S curve, drawn in descending order according to the total number of common links for each student, indicates that if the curve leans to the right side of the table, a higher number of common links are present, suggesting a well-formed structural awareness of social phenomena. Conversely, by examining the R curve and identifying links with high and low numbers of common links, teachers can discern which social phenomena were easily recognized by many students and which were challenging in terms of forming structural awareness.

3 METHOD OF SUPPORT FOR TEACHERS' EVALUATION

In the S-R Score Table, an attention coefficient is calculated to identify individual students and link items with unusual recognition patterns compared to the overall trend. These are denoted as C.S.i for individual students and C.P.i for link items. The calculation of the attention coefficient follows the method proposed by Sato (1998) for the S-P table. By examining the attention coefficient, interpretations can be made as shown in Figures 2 and 3.

As seen in Figure 2, C.S.i values exceeding 0.5 indicate unusual recognition patterns, while students with a correct response rate below 30% may be interpreted as having insufficient learning or unique response patterns. Similarly, Figure 3 shows that C.P.i values exceeding 0.5 indicate unusual recognition for specific link items.

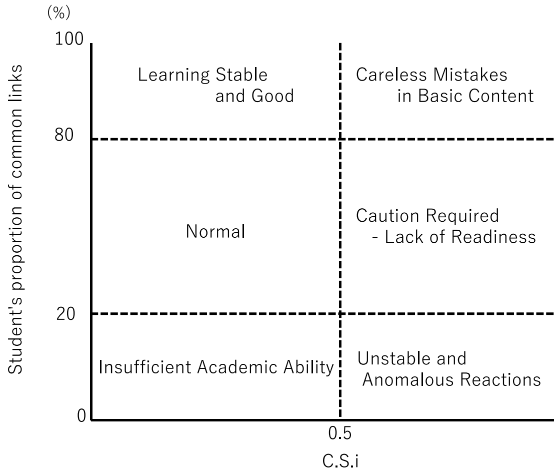


Figure 2: Interpretation of C.S.i in S-P Score Table.

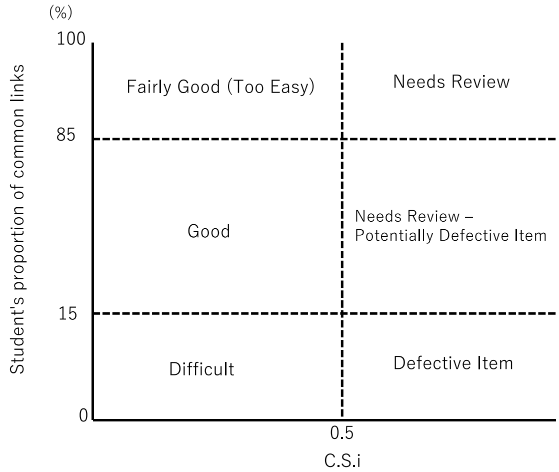


Figure 3: Interpretation of C.P.i in S-P Score Table.

Based on these findings, we established the following two requirements to support evaluation activities using the S-R Score Table:

1. By plotting the C.S.i values of each student in a scatter plot, teachers can understand the overall structural recognition trends of the students.
2. By plotting the C.P.i values of each link item in a scatter plot, teachers can understand the relationships between phenomena that students found difficult to understand.

Based on the above requirements, we developed the evaluation support system.

4 SYSTEM FOR CONCEPT MAP EVALUATION SUPPORT

The connection information of the students' concept maps is written into a CSV file in a predefined format and uploaded to the system, which then generates the S-R Score Table. Figure 4 shows the screen displaying the S-R Score Table generated by the system, where the S curve is shown in blue and the R curve is shown in red.

Category	Different	Same	Different	Different	Same	Different	Same			
text	Described	Undescribed	Undescribed	Described	Undescribed	Described	Undescribed			
cog	Knowledge	Interpretation	Analysis	Knowledge	Analysis	Knowledge	Knowledge			
student/item	6→10	5→7	8→11	1→2	7→8	3→4	9→10	S-SUM	%	C.S.i
StudentA	1	0	0	1	1	1	1	5	71.4	1.75
StudentB	1	1	1	1	0	0	0	4	57.1	0.0
StudentC	0	1	1	0	1	0	0	3	42.9	0.67
StudentE	1	1	1	0	0	0	0	3	42.9	0.0
StudentD	1	0	1	0	0	0	0	2	28.6	0.2
StudentF	1	1	0	0	0	0	0	2	28.6	0.0
StudentG	1	1	0	0	0	0	0	2	28.6	0.0
R-SUM	6	5	4	2	2	1	1	21		
%	86.0	71.0	57.0	29.0	29.0	14.0	14.0	42.9		
C.P.i	1.0	1.5	1.0	0.0	0.33	0.0	0.0			

Figure 4:S-R Score Table created by system.

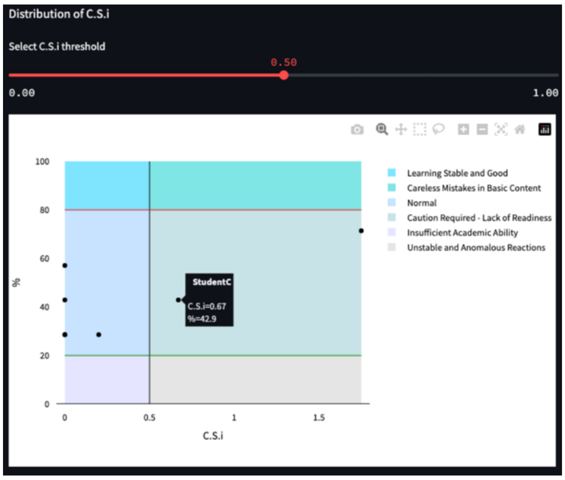


Figure 5: Scatter diagram of C.S.i created by system.

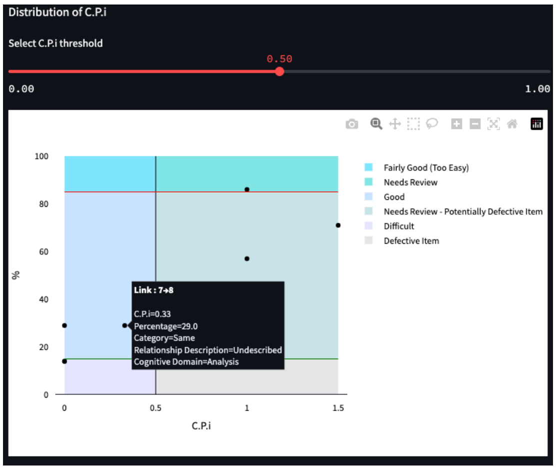


Figure 6: Scatter diagram of C.P.i created by system.

The interface displays not only the S-R Score Table but also scatter plots of C.S.i (Figure 5) and C.P.i (Figure 6).

A slider labeled "Select C.S.i threshold" is provided on the C.S.i scatter plot, allowing the threshold value for the attention coefficient to be adjusted. This feature supports teachers in reinterpreting the threshold value based on the scatter plot, making it easier to interpret students' structural awareness even if there are students with attention coefficients slightly below the standard threshold value, such as 0.45. Hovering over a point in the scatter plot displays the student's name, C.S.i, and the percentage of common links. By examining the scatter plot of students' C.S.i values, teachers can easily determine whether there are more students in a stable group with well-formed structural awareness or in a deficient group with insufficient learning. For students with a percentage of common links below 30% and an attention coefficient exceeding the threshold value, it can be interpreted that they may have made inappropriate connections or formed unique historical perspectives.

The same functionality is implemented for C.P.i. By examining the scatter plot of C.P.i values for the concept map links, teachers can visually interpret the proportion of links that were easy for students to understand and those that were difficult. Additionally, for links where 15% to 85% of the students have made the common links and the C.P.i value exceeds the threshold, it can be interpreted that the content of the nodes or links created by the teacher might not have been appropriate.

5 EXPERIMENTAL TRIAL

5.1 Experimental Setting

The developed system was applied to evaluate concept maps drawn by 21 second-year high school students enrolled in a history class. These students participated in lessons on drawing concept maps over a six-month period, ensuring they understood the method and were deemed suitable subjects for this study.

To avoid the influence of the teacher's instruction on the content of the students' concept maps, no direct instruction on the study material was provided. Instead, students were instructed to read the textbook and create their concept maps based on their understanding. Figure 7 shows the concept map created by the teacher.

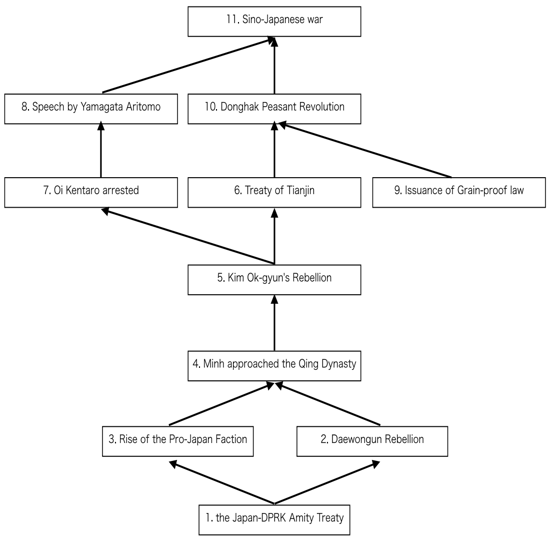


Figure 7: Concept map drawn by teacher.

Students were given the nodes from the teacher's concept map and instructed to independently arrange the nodes and draw the links.

5.2 Analysis and Result

5.2.1 Understanding Structural Perspectives and Trends in Structural Awareness

Based on the link information from the students' concept maps, the S-R Score Table generated by the developed system is shown in Figure 8.

Regarding Figure 8, focusing on the R curve, it can be observed that the number of common links decreases beyond "6→10." When examining the links from "2→4" to "6→10" in terms of structural perspectives, these links are classified as "textbook described, knowledge," regardless of whether they belong to the same field or different fields. Additionally, focusing on the S curve, it can be seen that the S curve to the right of "6→10" includes about half of the students, indicating that the percentage of common links exceeds 50%.

From these observations, it can be evaluated that approximately half of the students in the history course tend to develop structural awareness by utilizing structural perspectives to interpret the relationships between social phenomena described in the textbook, regardless of whether they are in the same or different fields. However, for the links classified as "described, knowledge" such as "1→2," "3→4," and "9→10," the percentage of common links falls below 30%. Therefore, it is necessary to further investigate the reasons for the decrease in the percentage of common links while reviewing the textbook and the structural diagrams drawn by the students.

On the other hand, for the links beyond "8→11" where the number of common links decreases, it can be seen that many of these links are classified under "analysis" or "interpretation" when focusing on the structural perspectives.

From these observations, it can be evaluated that students in the history course tend to find it difficult to develop structural awareness using structural perspectives for relationships between social

Category	Different	Different	Different	Different	Different	Different	Different	Different	Different	Different	Same	Different			
text	Described	Described	Described	Described	Described	Described	Undescribed	Undescribed	Described	Described	Undescribed	Undescribed			
cog	Knowledge	Knowledge	Knowledge	Knowledge	Knowledge	Knowledge	Analysis	Interpretation	Knowledge	Knowledge	Analysis	Knowledge			
student/item	2→4	4→5	5→6	1→3	10→11	6→10	8→11	5→7	1→2	3→4	7→8	9→10	S-SUM	%	C.S.i
StudentA	0	1	1	0	1	1	0	0	1	1	1	1	8	66.7	1.35
StudentB	1	1	1	1	0	1	1	1	1	0	0	0	8	66.7	0.29
StudentC	1	1	1	1	0	0	1	1	0	0	1	0	7	58.3	0.52
StudentD	1	1	1	1	1	1	1	0	0	0	0	0	7	58.3	0.0
StudentE	1	0	1	1	1	1	1	1	0	0	0	0	7	58.3	0.32
StudentF	1	1	1	1	1	1	0	1	0	0	0	0	7	58.3	0.06
StudentG	1	1	1	1	1	1	0	1	0	0	0	0	7	58.3	0.06
StudentH	1	1	1	1	1	0	0	0	0	1	0	0	6	50.0	0.28
StudentI	1	1	1	1	0	1	1	0	0	0	0	0	6	50.0	0.15
StudentJ	1	1	0	1	1	1	1	0	0	0	0	0	6	50.0	0.25
StudentN	1	0	1	1	1	1	0	0	0	0	0	0	5	41.7	0.16
StudentO	1	1	0	1	1	0	1	0	0	0	0	0	5	41.7	0.26
StudentK	1	1	0	1	1	0	0	0	1	0	0	0	5	41.7	0.35
StudentM	1	1	0	1	0	1	0	1	0	0	0	0	5	41.7	0.39
StudentL	1	1	1	0	0	0	1	0	0	0	0	1	5	41.7	0.58
StudentP	0	1	1	0	1	0	0	0	1	0	0	0	4	33.3	0.58
StudentQ	1	0	1	1	1	0	0	0	0	0	0	0	4	33.3	0.11
StudentR	1	1	1	0	0	0	0	0	1	0	0	0	4	33.3	0.36
StudentS	1	0	0	1	1	0	0	0	0	0	0	0	3	25.0	0.18
StudentT	1	0	1	0	0	1	0	0	0	0	0	0	3	25.0	0.22
StudentU	1	1	1	0	0	0	0	0	0	0	0	0	3	25.0	0.0
R-SUM	19	16	16	15	13	11	8	6	5	2	2	2	115		
%	90.0	76.0	76.0	71.0	62.0	52.0	38.0	29.0	24.0	10.0	10.0	10.0	45.7		
C.P.i	1.23	0.48	0.68	0.51	0.78	0.36	0.46	0.27	0.83	0.4	0.2	0.6			

Figure 8: S-P Score table created by the system

phenomena that require interpreting abstract descriptions in the textbook or inferring causal relationships arising from the phenomena.

5.2.2 Understanding Trends Using Scatter Plots of Attention Coefficients

The focus is on the students' C.S.i and the links' C.P.i. Based on the concept maps obtained from the experiment in this study, scatter plots of C.S.i and C.P.i created by the system are shown in Figures 9 and 10, respectively.

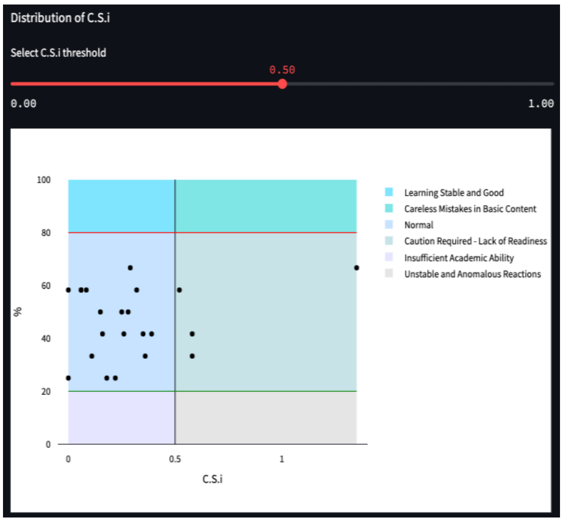


Figure 9: Scatter diagram of C.S.i in Experimental trial.

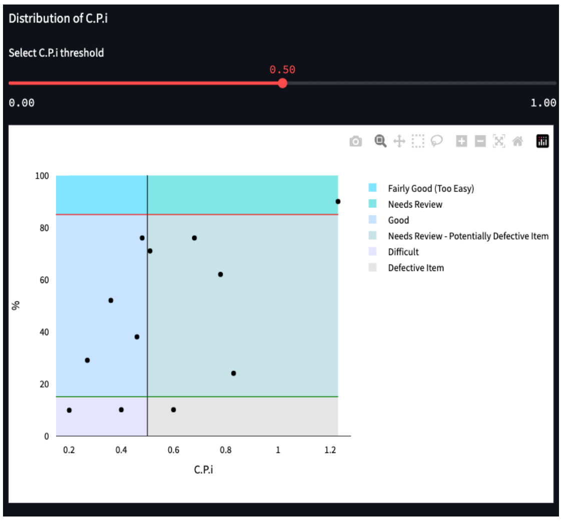


Figure 10: Scatter diagram of C.P.i in Experimental trial.

In Figure 9, it can be seen that 17 out of 21 students belong to the normal group. Therefore, it can be evaluated that the students in this class may have insufficient formation of structural awareness. Additionally, it is observed that four students have attention coefficients exceeding 0.5. These students are "studentA," "studentC," "studentL," and "studentP." According to the S-R Score Table, studentA and studentC have a common link percentage of about 60%. This suggests that while these two students have formed some degree of structural awareness, their structural awareness may be insufficient in certain areas.

On the other hand, studentL and studentP have a common link percentage of about 40% or lower. This suggests that these two students may be forming a unique historical understanding different from that of the teacher. In this study, the structural awareness held by the teacher is used as the correct model, but it cannot be said that the students' unique historical understanding is necessarily incorrect. Therefore, when evaluating these two students in a real classroom setting, it is necessary to review their concept maps to understand their structural awareness.

Regarding C.P.i, focusing on Figure 10, it can be seen that two links are classified as "Difficult." These links are "7→8 (Same fields, Undescribed, Analysis)" and "3→4 (Different fields, Described, Knowledge)." The link 7→8 is considered difficult for students to grasp the relationship between the phenomena as it is not described in the textbook, reflecting the teacher's professional perspective. On the other hand, the link 3→4, although described in the textbook, has the content on different pages, making it difficult for students to grasp the relationship.

For the link "9→10 (Different fields, Described, Knowledge)" classified as "Defective item" although the relationship is described in the textbook, it is on different pages similar to "3→4." Moreover, the content of node 6, which also influenced node 10, is described in detail, suggesting that the influence of node 9 on node 10 is minimal and thus not suitable to be included in the map.

On the other hand, the link "2→4 (Different field, Described, Knowledge)" classified as "Too Easy" with an attention coefficient exceeding 0.5 is explicitly described in the textbook, making it easy for students to recognize the relationship between the phenomena. However, the C.P.i is high because some students, despite the high common link percentage, missed this description and did not draw the link.

In the S-R table, C.S.i is calculated for each student and C.P.i is calculated for each link, but it is difficult to discern the tendencies of students and links from this data alone. However, by viewing the scatter plots of C.P.i and C.S.i implemented in this system, it has become easier for teachers to grasp these tendencies.

If teachers can grasp the overall tendencies of students and links, they can adjust the difficulty level of the lessons and design better instructional content. Additionally, using the scatter plots in this system makes it easier to identify students and links that deviate from these tendencies. If teachers can identify students who deviate from the norm, they can analyze those students' individual learning situations in more detail and consider optimal instructional strategies. Similarly, if teachers can identify links that deviate from the norm, they can determine which parts require more explicit teaching, thus aiding in the design of their lessons.

6 DISCUSSION

6.1 Practical and Managerial Implications of Plotting Results

In this study, we developed a system to assist teachers in evaluating students' structural awareness by comparing concept maps created by both teachers and students, and visualizing the differences. The S-R Score table proposed by Tokutake et al. (2019) is highly effective as a method for comparing concept maps between teachers and students. However, when teachers use this tool for lesson planning, it is necessary to focus not on the results of individual students, but on the overall trends among all students.

Therefore, the system developed in this study, which plots the results of each student and allows them to be viewed at a glance, is considered to be highly effective in helping teachers understand the overall trends in students' structural awareness and in considering the level and content of the lessons.

Additionally, plotting the correctness information and the attention coefficient (C.P.i) for each link is considered to be highly effective in helping teachers review the accuracy of their knowledge structure as experts and in understanding the relationships between phenomena that are difficult for students to grasp.

From these points, we believe that the system developed in this study sufficiently supports teachers in evaluating students' structural awareness.

6.2 Generalization of Methods and Feasibility in the Field

In this study, the developed system has been used in the context of history education and its effectiveness has been discussed. The use of concept maps to form structural awareness is also practiced in geography, politics and economics, which are different areas of social studies, and in science classes. In order to apply the S-R table and the system developed in this study to these subjects, we believe that it is necessary to change the classification items of the link. For example, in history education, historical events in the political field are sometimes related to historical events related to culture. To be able to capture the relationship between these events is very important in forming a structural awareness. For this reason, the classification of the "Same Category" and the "Different Category" are used to categorize the connections. However, in geography classes, not only causality and influence among events, but also inclusive relationships among events are sometimes considered important. Therefore, it may be necessary to reflect items such as "preconditions" and "inclusions" as elements of "Category".

In the system developed in this study, the S-R Score table reflects the elements written by the teacher in the csv file. Therefore, the system is expected to be able to handle such changes adequately. In addition, the creation of the S-R Score table is automatic, so there is no need for teachers to follow complicated procedures.

Therefore, we believe that the system developed in this study is applicable to other fields and can be easily introduced to schools.

7 CONCLUSIONS

The purpose of this study is to develop an evaluation support system using the S-R Score Table to make it easier for teachers to understand the structural recognition trends of individual students and the entire student body.

As a result, using the S curve and R curve, it was possible to understand the students' perspectives on structurally capturing the relationships between social phenomena and the trends in their structural awareness.

The scatter plot of the attention coefficient C.S.i, which indicates the heterogeneity of students' structural awareness, revealed the proportion of students with unique recognition. Additionally, by examining the percentage of common links with the teacher's concept map, it became possible to make detailed interpretations of the students' recognition.

The scatter plot of the attention coefficient C.P.i, which indicates the heterogeneity of recognition for each link in the concept map, allowed for a visual understanding of the relationships between social phenomena that students likely have insufficient understanding of. Furthermore, it enabled the identification of nodes in the teacher's concept map that may be considered unnecessary for organizing the relationships between phenomena. From this, it was suggested that the function of the concept map evaluation support system developed in this study has the potential to assist teachers in easily understanding the structural recognition trends of individual students and the entire class.

Future task include the following:

1) The interpretation of C.S.i and C.P.i used in this study is based on the content of the S-P table, which measures students' attainment of test questions. Future tasks include improving the interpretation of C.P.i and C.S.i to be unique to concept map.
2) The developed system was introduced in a school setting and its effectiveness was verified, it was only done in one case. Hence, it is necessary to have multiple teachers use the system and evaluate its usefulness.
3) S-R Score Table evaluates students' structural awareness by comparing it with the concept map created by the teacher. However, in social studies learning, students' independently formed understandings cannot always be deemed incorrect. Therefore, a separate method needs to be considered to evaluate the validity of such unique structural awareness formed by students.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS KAKENHI Grant Number 24H02486. This work utilized OpenAI's ChatGPT for initial drafting, which was thoroughly reviewed, edited, and supplemented by the authors. We therefore assume full responsibility for the final content of this publication.

REFERENCES

Aliya Nugumanova, Yerzhan Baiburin, (2021). Evaluation metrics for automatically constructed concept maps. *In Proceedings of the 21st International Conference on Control, Automation and Systems (ICCAS),* 365-370. https://doi.org/10.23919/ICCAS52745.2021.9649911

Brahami Menaouer, and Nada. Matta, (2019). The Relationship between the Knowledge Mapping and Open Innovation Process: The Case of Education System, Artificial Intelligence for Engineering Design, *Analysis and Manufacturing (AIEDAM)*, 33, 1－13. DOI: https://doi.org/10.1017/S0890060419000325.

Jaruwat Pailai, Warunya Wunnasri, Yusuke Hayashi, Tsukasa Hirashima. (2016). Automatic Concept Map Assessment in Formative Assessment Approach. *The 76th Advanced Learning Science and Engineering Workshop.* The Japanese Society for Artificial intelligence. https://doi.org/10.11517/jsaialst.76.0_09

Kato, H., Kurata, M., Sato, T., & Ozawa, S. (1988). Measurement and analysis of structural understanding state using hierarchical directed graphs drawn by learners. Gakushusha ga egaita gakushu naiyo no kaisoteki yukō gurafu ni yoru kōzō teki rikai jōtai no sokutei to bunseki (in Japanese). *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,* A J-71(10): 1955-1965.

Keitaro Tokutake, Hiromichi Mori, Masao Murota. (2019). Consideration of a Method to Grasp the Trend of How to Understand the Causal Relationship in Historical Learning. *Japan Journal of Educational Technology.* 43(Suppl.),133-136. https://doi.org/10.15077/jjet.S43083

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.

Takahiro Sato. (1987). *[ISM structural learning method]ISM kozo gakushu h*o(in Japanese). Tokyo: Meiji Tosho.

Takahiro Sato. (1998). *[Explanation of the S-P Score Table Analysis Method] S-P Hyo Bunsekiho no Kaisetsu*(in Japanese). Hiroshima. Chuo Kyoiku Kenkyujo.

Design and Evaluation of Microteaching: Emergent Learning for Acquiring Classroom Management Skill in Teacher Education

Dai Sakuma[1][a], Keitaro Tokutake[2][b] and Masao Murota[3][c]
[1]*Faculty of Teacher Education, Shumei University, 1-1 Daigaku-cho, Yachiyo-shi, Chiba, Japan*
[2]*School of Environment and Society, Tokyo Institute of Technology, Japan, 2-12-1, Ookayama, Meguro-ku, Tokyo, Japan*
[3]*Institute for Liberal Arts, Tokyo Institute of Technology, Japan, 2-12-1, Ookayama, Meguro-ku, Tokyo, Japan*
sakuma@mailg.shumei-u.ac.jp, tokutake.k1128@gmail.com, murota@ila.titech.ac.jp

Keywords: Emergent Learning, Classroom Management, Microteaching, Teacher's Knowledge, Adaptive Teaching.

Abstract: Schools in countries struggling with academic achievement gaps need to improve the teaching and support skills for students who facilitate classes in these gaps. This study focused on methods for acquiring complex classroom management skills for pre-service teachers. The aim of the study was to design and validate a method for teacher candidates to learn these behaviors. To achieve this, microteaching sessions in which unexpected behaviors occur were designed and carried out. A video recording of the microteaching was used in the evaluation experiment. Evaluators were five expert teachers in Japan. Statistical tests using the results of the questionnaire responses revealed that the simulated situations by student roles were close to actual situations with real students. It was also confirmed that the teaching candidates experienced situations that required various management behaviors. These results indicate that the microteaching sessions designed by the authors are useful as a method for emergent learning to achieve management skills in the classroom to control unexpected behaviors.

1 INTRODUCTION

Many countries have recognized inclusive education, but its definitions and implementations vary widely (Haug, 2017). In countries with large achievement gaps, students who struggle to follow teachers' instructions are often labeled as exhibiting "unexpected behavior".

Emergent learning, where preservice teachers respond flexibly to students' needs, is crucial for effective classroom management and enhances educational quality. It requires not only planned lessons but also improvisation to engage students.

Microteaching has been a method to train prospective teachers by allowing them to practice teaching skills in a controlled environment (Allen 1966; Sakamoto 1981; Sakuma et al. 2019). This technique helps improve skills such as attention management, questioning, and class control (Gower et al., 1995; Capel et al., 1998; Kilic, 2010). The Learner-Centred Microteaching (LCMT) model involves decision-making, planning, application, evaluation, and reflection, and is used to help teachers learn emergent management skills (Kilic, 2010).

However, there are limited examples of emergent learning in microteaching, especially regarding management procedures and behaviors for unexpected behavior. Microteaching can teach emergent behaviors that prevent unexpected classroom disruptions. Research on teachers' decision-making and information processing is important in this context (Pittman, 1985; Yoshizaki, 1988).

To design learning in a simulated classroom, it's important to consider teachers' information processing and decision-making models. Pittman (1985) identified three strategies teachers use for management: training, corrective, and push-in. Yoshizaki (1988) viewed teachers as information processors who explore routines and adapt to classroom situations. Some studies have attempted to approximate the environment of microteaching to that

[a] https://orcid.org/0009-0007-3638-8229
[b] https://orcid.org/0009-0002-1099-5518
[c] https://orcid.org/0009-0000-9727-3096

Sakuma, D., Tokutake, K. and Murota, M.
Design and Evaluation of Microteaching: Emergent Learning for Acquiring Classroom Management Skill in Teacher Education.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 222-229
ISBN: 978-989-758-716-0; ISSN: 2184-3228

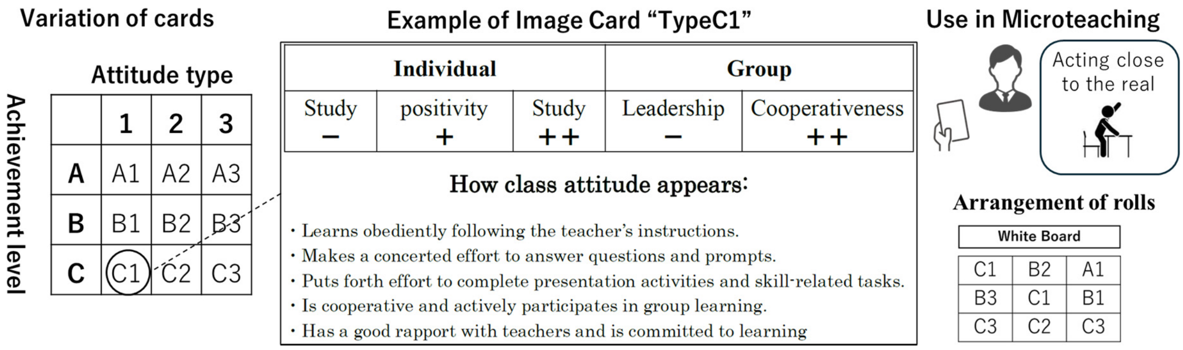


Figure 1: Overview of Microteaching Design and Simulation.

of an actual class, enhancing its effectiveness. Sakuma et al. (2019) developed image cards to assist pupil roles' acts in microteaching for this purpose.

This study provides insights into the design and implementation of microteaching sessions that incorporate unexpected student behaviors.

2 PURPOSE OF STUDY

2.1 Design of Microteaching

In this study, we developed and evaluated a microteaching method for prospective teachers to manage unexpected student behaviors in classroom settings using Image-cards (Sakuma et al. 2019).

In the microteaching which we designed, teacher-role extracts the situations of attitudes and behaviors of the student-role from the classroom situation and recognizes and discriminates between expected and non-expected behaviors. The teacher roles experienced situations by the student role's behaviors include delay in learning, interruption of learning, withdrawal from learning, disturbance to others and disturbance to the teacher. These categories were developed by classifying examples of attitudes and behaviors from Sakuma et al. (2019) image cards. It is assumed that, depending on the perceived attitude/behaviorsituation. The teacher roles then recognise the discrepancies between the lesson plan and the actual situation and the factors.

In addition, the teacher role has a learning opportunity to invoke or create management actions to resolve the discrepancy between the plan and the actual situation. Through this learning opportunity, the teacher role learns management behaviors to control the unexpected behavior.

Thus, we assume emergent learning, in which the number of perceived unexpected behaviors, including disruptive behaviors to the lesson, decision-making activities related to management are activated, and the teacher role invokes management behaviors and creates alternative solutions.

Figure 1 shows that the overview of the student image cards used in designing the microteaching (Sakuma et al. 2019). They consisted of three types based on the degree of 'learning achievement' (A: high - C: low) and three types based on the degree of 'difficulty in following instruction' (1: easy to follow - 3: difficult to follow): A1, A2, A3, B1, B2, B3, C1, C2, C3. Figure 1 shows the types of image cards used and the seating arrangements of the student roles when a microteaching is conducted with nine student roles.

The microteaching which we designed has system to make it easy for situations to occur in which student roles in C1, C2, and C3 are more than half of the total number of students in the class, and in which they become noisy and cannot follow instructions. Specifically, the roles and their number were set up with reference to the characteristics of a disrupted class in which more than half of the student roles are dissatisfied with class life or are unable to comply with class rules, and in which the roles of student C1, C2 and C3 are more than half of the total number of student roles in the class, which can lead to a situation of withdrawal from class and a situation of general noisiness and lack of following instructions.

Additionally, place student roles who are difficult to reach for instruction in seats that are closest to the teacher and within the teacher's sight. The student roles whose guidance is easy to follow are put in the role of supporting the learning of the student roles. Therefore, to create unexpected behavior, C3 was placed in the seat farthest away from the teacher, contrary to the considerations. In addition, A1 was placed in the seat farthest away from C3.

Table 1: Examples of situations which confirmed in Microteaching authors designed.

Design1

No.	Situation
1	Situations where instructions are not followed and the class is held up
2	Messing with other children
3	Playing with stationery or teaching aids
4	Throwing things
5	Threatening or provoking other children
6	Going outside without permission
7	Inviting other children to play
8	Shouting or shouting
9	Playing with stationery or teaching aids
10	Pointing out minor mistakes by the teacher
11	Situations where children start to play
12	Talking about topics unrelated to the lesson
13	Drawing on the blackboard
14	Singing a song when bored with learning
15	Situations where the pupil's gaze is not looking in the direction of anyone other than the teacher

Design2

No.	Situation
1	Sleeping situation from the beginning of the class
2	Situations where they turn their back
3	Playing with stationery or misbehaving
4	Hitting another friend to wake them up
5	Messing with a friend
6	Dropping things
7	Walking around situation
8	Going outside
9	Hitting another friend to wake them up
10	Turns his/her back
11	A situation where the child starts reading a book
12	Situations where the whole place becomes noisy
13	Situations where the child does not want to cooperate in a cooperative learning situation
14	Throwing erasers or scraps of paper
15	A situation in which the whole class becomes increasingly noisy while writing on the board
16	A situation in which greetings are not coordinated

2.2 Application

Microteachi ng designs were incorporated into a practice-based class for second-year university students aspiring to become primary school teachers. The implementation date was December 12, 2015.

The participants in each group consisted of one teacher, nine students, and an extra observer. The microteaching took approximately 30 minutes. The scope of study covered the third grade of p rimary schools. The learning objectives for both Design1 and Design2 were "to understand how to add and subtract decimals to and from one decimal place, and to be able to perform these calculations".

The learning task for the class, designed by the teacher role in Design1, was to line up A4-sized sheets of colored paper, which were regarded as 1, with thin sheets of colored paper, which were regarded as 0.1, and ask students to think about how many 0.1s could be placed in the sheet ”1”. One set of these materials was prepared for each group of three. The lesson learning task designed by the teacher of Design2 was to ask the students to think about how many 0.5L and 0.3L together would be, through the juxtaposition of 'colored paper with a picture of two beakers' and 'thin colored paper that is 0.1L.

The learning task for the lesson designed by the Design1 teacher was to line up A4-sized colored paper regarded as 1 with thin colored paper regarded as 0.1 and ask the students to think about how many 0.1s could be put into 1. One set of these materials was prepared for each group of three.

The learning task of the lesson designed by the teacher in Design2 is to have the students think about how many pieces of paper 0.5L and 0.3L add up to through the arrangement of "colored paper with a picture of two beakers" and "thin colored paper that is considered to be 0.1L".

In addition, the teacher role of Design1 facilitated the class with the student roles sitting on the floor, without using a desk. On the other hand, the teacher of the Design2 conducted the class with the student roles sitting on chairs, in the seating order shown in Figure.1 (right side). The reasons for the differences in the teaching materials prepared by the teachers of the Design1 discussed below. The characteristics of the students played by the pre-service teachers in the microteaching were known from the lesson design stage for Design1. Therefore, the teacher had the original idea of having the students sit on the floor for the class and tried to attract the students' interest. In addition, it can be said that the teacher tried to avoid unexpected behavior by giving the student a specific task to do in groups of three, namely 'laying out the cards (0.1) on A4 paper (1)', so that the student role could concentrate on their learning.

2.3 Management Behavior

Hereafter, the unexpected behavior that occurred in the microteaching designed in this study will be referred to as the simulated situation. The events and situations that occur in the actual classroom are referred to as actual situations.

Examples of simulated situations that occurred in the microteaching designed by the authors are shown, based on a series of utterances obtained by transcription from the video recordings. For reasons of space limitation, approximately one minute of each lesson is shown for both Design1 and Design2. The simulated situations are single underlined. The management behaviors created by the respective teacher roles are underlined with a chain line.

The following are the classroom situations observed at the beginning of the Design2 practice and the simulated situations.

[Teacher]: Good morning, everyone, I'd like to start the first period, Tom (C1), please wake up.
[C3 role]: *wake up - Tom (C1 role)!*
[Any roles]: *'Wake up, first period is starting'.*
[Teacher]: *Tom, you must be sleepy.*
[C1 role] : *I want to sleep, and going to home.*
[Teacher]: *George, what are you doing? Well, It's time to start first period, Bob, what are you doing?*
[C3 role]: (turning back, looking restless)
[Teacher]: *Bob, I'm going to start, but I thought I'd check one rule, Bob, face forward.*
[C3 role]: *eh.*
[Teacher]: *Look forward, yes, please. And everyone, I'm going to go over the rules that we've been going over all this time, and if, while we're teaching, you get an itch and you feel like you want to stand up, please raise your hand immediately. The reason is that if you stand up, the teacher will be worried, so if you make a signal before that, you can do it, so do you remember the rules? Bob, are you okay?*

From the utterance, it can be said that the teacher role created the management behavior of 'taking up stationery'. Similarly, it can be said that the teacher role of the Design2 implemented the management behavior of 'checking the rules of the class during the lesson.

2.4 Teacher Role's Reflections

Immediately after the practice of the microteaching, the teacher roles were asked to reflect on their own classes. The parts of transcripts of their utterances are shown below. The transcripts of the teacher role of Design2's utterances are shown below.

[Teacher]: *Well, for the time being, since I was in the lower ranks...[omission]...my goal was to do 1.0, 1 and up to what we did today on the assumption that I couldn't go to the problem areas, but it still took a lot of time to deal with the student roles who were doing something or not doing something. I realized that it really takes a lot of time to deal with every student role who is doing something or not doing something, or who is standing up and walking around, and I thought I still don't know where to switch and ignore them. ...[omission]...I also thought that it was very difficult to know where to switch from caution to scolding, and I was thinking about this as I tried to deal with the student roles who were moving around. I was also thinking about how to deal with the student roles who didn't write a lot, and there were a lot of student roles who didn't write a lot this time, and whether to adapt to the role of the student roles who wrote a lot or the majority who didn't write a lot, so I adapted to that role.*

The teacher role of the Design2 was searching for a teaching method that could achieve the goal in a simulated situation where unexpected student role behavior was observed. As a result, it can be said that, as the authors intended, they created their own management behavior of 'changing the form of the learning activity' during the lesson.

3 EXPERT ASSESSMENT

3.1 Method of Evaluation

An Expert evaluation experiment was conducted on the environment and learning of a microteaching designed by the authors. The following three evaluations of the simulated classroom environment were obtained.

・Whether the simulated situation as a whole is close to the actual situation
・Whether each simulated situation is close to the actual situation
・Whether each simulated situation is an opportunity to learn management behaviors

Is it close to the individual situation that causes it (assessment of similarity)?

・Whether it is an opportunity for the teacher role to learn management behaviors (evaluation as a learning opportunity).

For evaluating learning effects, the teacher roles experienced creating management strategies that mitigate unexpected behaviors during microteaching sessions

The implementation periods were 16, 23, and 30 April and 9 and 14 May 2016. On each date, one evaluator was invited to the laboratory to carry out the evaluation experiment. The total experimental time spent per person was around 120 minutes. Evaluators were expert teachers with an average of 29.8 years of experience (S.D: 10.8 years).

The procedure for the evaluation experiment is as follows.

1. Experimental teaching
2. Viewing of video recordings of the microteaching and interviews
3. Evaluation of the simulated situation from five perspectives through a questionnaire survey.

For (1), the purpose and flow of the experiment were explained to the evaluators. For (2), the evaluators were asked to randomly select and watch one of the two designs and one of the designed children that were practiced in this study. During the viewing of the video, the stop-and-motion method of Fujioka (1991) was used to obtain the evaluators' learning of the microteachings designed by the authors and their evaluative utterances of the situations that the students' roles caused.

The following questions were set for the class evaluation and semi-structured interviews were conducted.

・What teacher skills and knowledge were learned through experiencing the focused event?

The stop-and-motion method of Fujioka (1991) was used when the evaluator spoke about the simulated situation, to implement the same format as in a classroom review meeting in a school setting.

In addition, the evaluators were asked to watch a simulated situation randomly selected by the authors beforehand, and to rate whether the simulated situation was close to the actual situation or not using a five-point scale (1: does not apply - 5: applies). At the same time, using Asada and Sako's (1991) classification of eight types of management behaviors, the teachers were asked to choose which of the simulated situations they were asked to watch corresponded to a learning opportunity for creating management behavior. Multiple answers were allowed. The eight options were A. Inserting teaching materials, B. Changing the form of children's activities, C. Changing the order of nomination, D. Changing the sequence of questions, E. Changing the nomination-response rule, F. Changi ng the form of communication, G. Changing the response method, and H. Attention and instruction.

Table 1 shows the extracted simulated situations. Regarding (3), the authors clarified from five perspectives whether the situations in the student role that occurred within the microteaching designed by the authors were closer to the actual classes compared to the traditional microteaching.

In the present study, this questionnaire item was also used, and the evaluators were asked to answer the questions using a five-point scale (1: does not apply - 5: applies). The five question items used by Sakuma et al. (2019) were:

(i). Diverse situations,
(ii). Individual child situations
(iii). Overall child situations
(iv). Impact and change on other children
(v). Events that test the trust relationship with the teacher.

3.2 Results of Analysis

To determine whether the five simulated situations - (1) various situations, (2) individual child situations, (3) overall child situations, (4) effects and changes on other children, and (5) events that test the trust relationship with the teacher - approximated the actual situations, a one-sample t-test was conducted with the population mean considered to be 3. The results of the analysis showed a significant trend and a significant difference in the results of all the responses of the rater groups. The test results are shown in Table 2.

Table 2: Results of t-tests for similarity (N=5).

	Design1			Design2		
	Degree of approximation			Degree of approximation		
Quetion	**Mean**	**S.D.**	***P-value***	**Mean**	**S.D.**	***P-value***
(□) Diverse situations	4.60	0.55	*0.032* *	4.60	0.55	*0.024* *
(□) Individual pupil situation	4.20	0.84	*0.002* **	4.20	0.45	*0.003* **
(□)Overall situation of the pupil	4.40	0.55	*0.004* **	4.20	0.45	*0.002* **
(□)Impact and change on other pupils.	4.40	0.89	*0.004* **	4.20	0.45	*0.003* **
(□)Test the trust relationship with teachers.	4.20	0.45	*0.032* *	4.00	1.00	*0.099* +

*not significant: n.s. p<.10: + p<.05: * p<.01: ** p<.001: ****

Table 3: Results of t-tests for degree of similarity.

Design 1

No.	Mean	S.D.	*P-value*
1	4.8	0.45	***
2	4.8	0.45	***
3	4.6	0.55	**
4	4.8	0.45	***
5	4.2	1.3	*n.s.*
6	3.6	1.67	*n.s.*
7	3.6	1.79	*n.s.*
8	4.2	0.84	**
9	4.2	0.45	**
10	4.2	0.84	*
11	4.6	0.55	**
12	4.8	0.45	***
13	3.6	1.95	*n.s.*
14	3.6	1.52	*n.s.*
15	4.6	0.89	*

Design2

No.	Mean	S.D.	*P-value*
1	3.8	1.3	*n.s.*
2	4.8	0.45	***
3	4.5	0.58	**
4	3.0	0.71	*n.s.*
5	4.2	0.45	**
6	3.8	0.45	*
7	4.4	0.55	**
8	4.2	1.3	*n.s.*
9	3.4	0.55	*n.s.*
10	4.2	0.45	**
11	4.0	1.22	*n.s.*
12	4.8	0.45	***
13	4.2	0.84	*
14	4.0	1	*n.s.*
15	4.4	0.55	**
16	3.8	1.3	*n.s.*

*not significant: n.s. p<.10: + p<.05: * p<.01: ** p<.001: ****

Table 4: Percentage of similar simulated situations.

	Event(X)	No similarity (Y)	Similarity (X-Y)	Percentage
Design1	15	5	10	0.67
Design2	16	6	10	0.63

Table 5: Frequency of simulated situations that create the management skills confirmed in the experiment.

		A	B	C	D	E	F	G	H
		Add new tasks for pupil	**Change how to teach**	**Change notable pupil**	**Change how to learn**	**Encourage pupil to learn together**	**Change how to communicate with pupils**	**Give formative feedback to pupils**	**Give cautions and warnings for pupil**
Design1	**Measured value**	22	12	1	15	1	10	5	31
Design2	**Measured value**	16	17	4	5	1	13	6	30
Total	**Measured value** **Expected value**	38 (23.63)	29 (23.63)	5 (23.63)	20 (23.63)	2 (23.63)	23 (23.63)	11 (23.63)	61 (23.63)

The results presented in Tables 3 highlight the significant differences in the degree of similarity and frequency of management behaviors that were confirmed in the experiment. A one-sample t-test was conducted to determine whether the mean similarity scores for the simulated situations significantly differed from the expected mean score of 3. To determine the proportion of simulated situations that occur in the microteaching designed by the authors that are close to the actual situations experienced by the group of evaluators, a one-sample t-test was conducted using the results of five responses to a total of 31 simulated situations, 15 from Design1 and 16 from the Design2.

Table 4 presents the percentage of similar simulated situations by using scores of Tables 3.

From these analyses, key findings from the statistical analysis indicate that the microteaching sessions designed in this study approximate real classroom scenarios in approximately 60% of cases, providing useful insights for future study.

In addition, a chi-square test was used to compare the frequency of different types of management behaviors observed in the two designs, revealing significant differences in specific behaviors from Table 5. A chi-square test revealed no significant difference between the two designs in terms of overall learning opportunities for management behaviors. However, a subsequent analysis using the data in Table 5 found significant differences in the specific types of management behavior opportunities experienced by the teacher role (χ^2 (7) = 110.894, p < .01).

Multiple comparisons using Ryan's nominal levels revealed significant differences among management behaviors. Inserting teaching materials was more effective than changing the nomination order, nomination-response rules, or response methods. Changing the children's activities was more effective than changing the nomination order or rules but less effective than attention and instruction. Additionally, changing the order of nomination was less effective than changing the sequence of questions, communication, or attention and instruction.

Additionally, changing the order of nomination was less effective than changing the sequence of questions, communication, or attention and instruction. Changing the sequence of questions was less effective than attention and instruction. Finally, changing the nomination-response rule was less effective than changing the form of communication or attention and instruction. Furthermore, it was observed that F: changing the form of communication <H: the way of attention/direction *(critical ratio = 4.0, p = 0.0002)*. Finally, it was found that G: changing the method of response < H: the way of attention/direction *(critical ratio = 5.8, p = 0.0002)*.

These results indicate that through the practice of the microteaching designed in this study, the teacher roles had the opportunity to learn the management behavior of inserting teaching materials rather than changing the order of nomination, nomination-response rules and response methods to establish a lesson. It was evident that the students had the experience. Similarly, it can be said that the teacher role experienced the opportunity to learn the management behavior of changing the form of the children's activity rather than changing the order of nomination or changing the nomination-response rule.

4 DISCUSSIONS

4.1 Environmental Assessment

It was found that the microteaching designed by the authors could be practiced to enable learning in a situation close to the actual unexpected behavior. Furthermore, the similarity between the simulated situation and the actual situation was approximately 60%, which means that more than half of all situations in the microteaching designed by the authors were inevitable situations in which the teacher role had to invoke and create management behaviors.

In other words, a certain quality is guaranteed as a method for learning management behaviors to control unexpected behavior.

4.2 Causes of Low Similarity

The following reasons can be given as to why a total of 11 simulated situations that did not show statistically significant differences were not close to the actual situations.

One possible reason for the low similarity between simulated and actual situations is the over-exaggeration of certain student roles based on the image cards. For example, a student role labeled C3 may have been overly disruptive due to a lack of nuanced understanding of the behavior expectations, leading to a deviation from realistic classroom dynamics. Future research could involve more detailed role-playing instructions to mitigate such discrepancies. This indicated that, to have the opportunity to learn in a context close to the actual classroom, the children acting out C3 needed to be given prior instruction to avoid overacting. However, no differences were found in the proportion of simulated situations that occurred within each microteaching. In other words, there was no difference in the type and number of simulated situations occurring between the different designs.

4.3 Evaluation of Learning Effectiveness

The teacher roles in both Design1 and Design2 experienced various management behaviors, including attention, instruction, material insertion, question sequencing, communication, and activity changes. They found that unexpected behavior triggered management behaviors like cautioning, scolding, and interrupting the class. While students were confused, the teacher roles learned to identify key student roles, adjust their lesson plans, and implement appropriate management actions. This suggests the method's effectiveness as an emergent learning approach for invoking and creating management behaviors.

4.4 Comparison with Previous Studies

This study proposed a method to cultivate classroom management skills in pre-service teachers using microteaching.

Previous studies have highlighted the potential importance of emergent learning in classroom management. Haug (2017) noted the varying definitions and implementations of inclusive education across countries, emphasizing the need for flexible responses to students' needs. This aligns with our findings, suggesting that emergent learning, where preservice teachers adapt to unexpected behaviors, may be beneficial for effective classroom management.

Additionally, this study supports the potential effectiveness of microteaching in improving teaching skills such as attention management, questioning, and class control (Gower et al., 1995; Capel et al., 1998; Kilic, 2010). Our findings indicate that microteaching could help in teaching emergent behaviors to handle unexpected classroom disruptions.

Furthermore, Sakuma et al. (2019) developed image cards to assist student roles in microteaching, enhancing its effectiveness. Our study builds on this by incorporating unexpected behaviors into microteaching sessions, suggesting that this approach may provide a more realistic and comprehensive training experience for preservice teachers.

5 CONCLUSIONS

This study provides insights into the design and implementation of microteaching sessions that incorporate unexpected student behaviors. While the findings suggest potential benefits, further research is needed to establish more robust scientific validation.

The microteaching sessions provided valuable insights into the challenges faced by teachers in managing unexpected behaviors. However, more comprehensive studies are needed to validate these findings across different contexts and sample sizes. By refining the pre-teaching preparation of student roles and considering more diverse simulated situations, future studies can better assess the effectiveness of emergent learning strategies in classroom management training.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP24K00423. This work utilized OpenAI's ChatGPT for initial drafting, which was thoroughly reviewed, edited, and supplemented by the authors. We therefore assume full responsibility for the final content of this publication.

REFERENCES

Allen, D. W. (1967). *Microteaching: A description*. Stanford Teacher Education Program.

Asada, T., & Sako, H. (1991). Identification of and a model for classroom management behaviour in instructional situations: Introducing a managerial point of view in analysis of classroom instruction. *Japan Journal of Educational Technology, 9*(1), 12-26. https://doi.org/10.15077/jmet.15.3_105

Capel, S., Leask, M., & Turner, T. (1998). *Learning to teach in the secondary school*. Routledge.

Gower, R., Phillips, D., & Walters, S. (1995). *Teaching practice handbook*. Heinemann.

Fujioka, N. (1991). Methods of classroom research using stop-motion technique. *Japan Journal of Classroom Research, 4*(2), 102-115.

Haug, P. (2017). Understanding inclusive education: Ideals and reality. *Scandinavian Journal of Disability Research, 19*(3), 206-217. https://doi.org/10.1080/15017419.2016.1224778

Kilic, A. (2010). Learner-centered microteaching in teacher education. *International Journal of Instruction, 3*(1).

Pittman, S. I. (1985). A cognitive ethnography and quantification of a first-grade teacher's selection routines for classroom management. *The School Journal, 85*(4), 541-557.

Sakamoto, T. (1981). The effects of simplified microteaching for pre-service teacher training. *Educational Technology Research, 5*, 1-13.

Sakuma, D., Takaishi, T., Imai, T., Hasegawa, K., & Murota, M. (2019). Designing microteaching using imagination-cards of pupil's actual condition. *Japan Journal of Educational Technology, 43*(2), 91-103. https://doi.org/10.15077/jjet.42163

Yoshizaki, S. (1988). Development of a model for teachers' decision making. *Japan Journal of Educational Technology, 12*(2), 51-59. https://doi.org/10.15077/etr.KJ00003899067

A Comprehensive Approach for Graph Data Warehouse Design: A Case Study for Learning Path Recommendation Based on Career Goals

Viet Nguyen[1,2], Vu Bui[1,2] and Thu Nguyen[1,2][a,b]
[1]*Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam*
[2]*Viet Nam National University, Ho Chi Minh City, Vietnam*
{20127664, 20127668}@student.hcmus.edu.vn, ntmthu@fit.hcmus.edu.vn

Keywords: Data Warehouse, NoSQL Databases, Graph Data Warehouse, NoSQL-Based Data Warehouse Modeling, Methodology.

Abstract: Today, organizations leverage big data analytics for insights and decision-making, handling vast amounts of structured and unstructured data. Traditional data warehouses (TDW) are suboptimal for such analytics, creating a demand for NoSQL-based modern data warehouses (DWs) that offer improved storage, scalability, and unstructured data processing. Graph-based data models (GDMs), a common NoSQL data model, are considered the next frontier in big data modeling. They organize complex data points based on relationships, enabling analysts to see connections between entities and draw new conclusions. This paper provides a comprehensive methodology for graph-based data warehouse (GDW) design, encompassing conceptual, logical, and physical phases. In the conceptual stage, we propose a high-abstraction data model for NoSQL DW, suitable for GDM and other NoSQL models. During the logical phase, GDM is used as the logical DW model, with a solution for mapping the conceptual DW model to GDW. We illustrate the GDW design phases with a use case for learning path recommendations based on career goals. Finally, we carried out the physical implementation of the logical DW model on the Neo4j platform to demonstrate its efficiency in managing complex queries and relationships, and showcase the applicability of the proposed model.

1 INTRODUCTION

The proliferation of social networks, cloud computing, and IoT devices has precipitated the generation of vast amounts of data, heralding the era of 'big data.' This surge in data volume presents significant challenges for TDW systems, which increasingly struggle with scalability and the management of unstructured data Oussous et al. (2017). Although once efficient, TDWs are now inadequate in addressing the growing demands for speed and complexity in data processing Bhogal and Choksi (2015). In response to these limitations, NoSQL databases have emerged as a critical solution, offering enhanced flexibility, scalability, schemalessness, high availability, and cost-effectiveness in data storage and analysis Bhogal and Choksi (2015).

Among the various NoSQL data models, GDMs stand out for their capability to manage large-scale data and represent complex relationships, making them particularly suited for applications such as recommendation systems and social networks Akid and Ayed (2016); Sellami et al. (2019). Consequently, a growing number of researchers and organizations are transitioning from Relational Online Analytical Processing (ROLAP) to NoSQL DW, despite the technical challenges that accompany this shift Banerjee et al. (2021). This transition necessitates not only technological adaptations but also a thorough design process that spans from conceptual to physical models Abdelhédi et al. (2017); Banerjee et al. (2021).

Nevertheless, a significant research gap persists in the absence of a unified conceptual model specifically designed for NoSQL DW Banerjee et al. (2021). Addressing this gap is crucial for ensuring the effective design and efficient implementation of NoSQL DW across various organizational contexts. This study contributes to the field by proposing:

- A novel symbolic system for conceptual modeling in NoSQL DW, building on prior research Banerjee et al. (2021).
- A comprehensive NoSQL DW design methodol-

[a] https://orcid.org/0009-0009-6961-3976
[b] Corresponding author

Nguyen, V., Bui, V. and Nguyen, T.
A Comprehensive Approach for Graph Data Warehouse Design: A Case Study for Learning Path Recommendation Based on Career Goals.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 230-237
ISBN: 978-989-758-716-0; ISSN: 2184-3228

ogy, with a particular focus on GDM.

- A case study that demonstrates the application of the proposed system within GDW tailored for learning path recommendations.

The paper is structured as follows: **Section 1** introduces the context and objectives of the study. **Section 2** reviews the relevant literature. **Section 3** presents the symbolic system and the proposed conceptual DW model. **Section 4** elaborates on the conversion process to GDM, supported by a case study. **Section 5** discusses deployment results and performance analysis. Finally, **Section 6** concludes the paper and explores future research directions.

2 RELATED WORKS

In NoSQL DW research, implementations have predominantly focused on the logical level, targeting specific NoSQL types. Yangui et al. (2016) proposed a method for transforming multidimensional conceptual models into column- and document-oriented NoSQL models. Chevalier et al. (2015) introduced three conversion strategies for document-oriented models, and their study rigorously evaluated the performance and memory utilization of these strategies.

Recent studies highlight the superior performance of GDW compared to traditional DWs. For instance, Akid et al. (2022) demonstrated GDW's efficiency in handling OLAP queries, while Nguyen et al. (2022) explored the practical applications of GDM in complex scenarios, including learning path consultancy.

However, as Banerjee et al. (2021) points out, there remains a significant gap due to the absence of a unified conceptual model for NoSQL DW. Their proposed symbolic system and model, despite its novel approach, has constraints, particularly in depicting many-to-many relationships and defining dimension table attributes.

Building on these studies, our research proposes a novel symbolic system for constructing a conceptual DW model for NoSQL DW, applicable across various NoSQL databases, with a focus on GDM. We also offer transformation rules and a case study to illustrate the practical application, culminating in a physical implementation on the Neo4j platform.

3 A SYMBOL SYSTEM FOR CONCEPTUAL DW MODELING

This section introduces a set of symbols and rules for conceptual NoSQL DW modeling, using Crow's Foot notation for relationships and tabular symbols for facts and dimensions, ensuring clarity in representing attributes of fact and dimension tables.

Before presenting the symbol system, fundamental concepts from the multidimensional model, as detailed by Sellami et al. (2019), are adapted. This model focuses on a business-centric view of data, emphasizing facts, dimensions, measures, and their interrelationships.

3.1 Conceptual Multidimensional Model

A conceptual multidimensional model denoted MDM is defined by the triplet $(FTs, DTs, Star^{MDM})$, where:

- $FTs = \{FT_1, FT_2, \ldots, FT_n\}$ is a set of fact tables.
- $DTs = \{DT_1, DT_2, \ldots, DT_m\}$ is a set of dimension tables.
- $Star^{MDM} : F^{MDM} \rightarrow 2^{DTs}$ is a function that associates each fact table $FT_i \in FTs$ with the set of dimension tables $DT_i \in DTs$, and 2^{DTs} is a set of all combinations of the dimension tables.

A fact table, denoted $FT_i \in FTs$, is defined by the pair $(F_{name}, F_{measures})$, where:

- F_{name} represents the name of the fact table.
- $F_{measures} = \{m_1, m_2, \ldots, m_k\}$ is the set of measures of the fact table.

A dimension table, denoted $DT_i \in DTs$, is defined by the pair $(D_{name}, D_{attributes})$, where:

- D_{name} represents the name of the dimension table.
- $D_{attributes} = \{a_1, a_2, \ldots, a_k\}$ is a collection of dimension table attributes.

3.2 The Set of Symbols for Conceptual NoSQL DW Representation

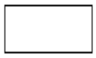
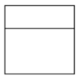
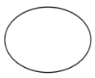
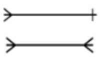
In this section, we propose a refined set of symbols, building upon previous research of Banerjee et al. (2021), to effectively represent the conceptual model of a NoSQL DW. This enhanced set of symbols aim to address and overcome the limitations identified in earlier studies, providing a robust framework for modeling DW at the conceptual level.

3.2.1 The Refined Set of Symbols

The Crow's Foot notation is widely recognized for its simplicity and clarity in Entity-Relationship (ER) modeling, making it particularly effective for conceptual database representation Puja et al. (2019). This notation intuitively depicts relationships, including many-to-many connections, using Crow's Foot symbols, while tabular symbols represent facts and dimensions along with their attributes and measures. This dual-notation approach ensures clear expression of fact and dimension table attributes, as summarized in **Table 1**.

The conceptual model for a NoSQL DW will be constructed using these symbols, with definitions of model components and relationships detailed in the next section.

Table 1: Proposed symbol system based on Banerjee et al. (2021).

Shape	Symbol	Significance
Rectangle		Components within the Collection layer
Tabular		Table of facts or table of dimensions
Ellipse		The optional attribute for a dimension table
Arrow		Indicates a relationship in which a component of one layer encapsulates a component of another layer
Crow's Foot Notation		Indicates the relationship between a dimension table and a fact table, or a dimension table and another dimension table

3.2.2 Conceptual NoSQL DW Model

Consistent with the symbol set presented in Table 1, this section presents a conceptual DW model specifically designed for NoSQL DW that can be applied to different kinds of NoSQL data models. Particular guidelines for conceptual information representation are included in this model. In-depth information about the elements and how they interact with one another in the model will also be supplied. The suggested conceptual DW model is shown in Figure 1, which provides a clear and organized representation of these elements and their relationships. The model is explained in depth in the following:

Components of the Model

This section presents comprehensively information regarding the components of the model, which is structured into three distinct layers, namely the Collection layer, Family layer, and Optional attribute layer.

a. **Collection Layer (CL):** The topmost layer of the model is defined by its primary component, col. CL comprises one or more col:

$$CL = \{col_1, \ldots, col_q\}$$

b. **Family Layer (FL):** The middle layer of the conceptual DW model is represented by this. The field of FL encompasses two distinct categories, namely Fact Family (FF) and Dimension Family (DF). Fact tables and dimension tables are the primary components of FF and DF, respectively.

- **Fact Family (FF):** The components of FF consist of fact tables (FT), denoted as:

$$FF = \{FT_1, \ldots, FT_n\}$$

- **Dimension Family (DF):** The components of DF consist of dimension tables (DT), denoted as:

$$DF = \{DT_1, \ldots, DT_n\}$$

c. **Optional Attribute Layer (OAL):** This layer is the last layer of the model and includes optional attributes that can be present or absent between instances of dimension tables $DT \in DF$. The key component of this class is OAT, which is defined by several pairs (D_{name}, OA), where:

- D_{name} represents the name of DT.
- OA is the optional attribute name of the DT

$$OAT = \{(D_{name_1}, OA_1), \ldots, (D_{name_k}, OA_k)\}$$

Relationship

This section will deal with the relationships between the components identified in the above section. Relationships within the model can be classified into two fundamental types: internal and external relationships. Internal relationships indicate relationships inside the same component, whereas external links illustrate interactions between distinct components.

External Relationship: According to Figure 1, these following relationships are considered external relationships:

a) **Component FF – Component DF (FactDimRel)**: the FactDimRel relationship is a mapping $f : FF \to \mathcal{P}(DF) \setminus \{\emptyset\}$ where $\mathcal{P}(DF)$ is the power set of DF. In this relationship, each $FT \in FF$ is

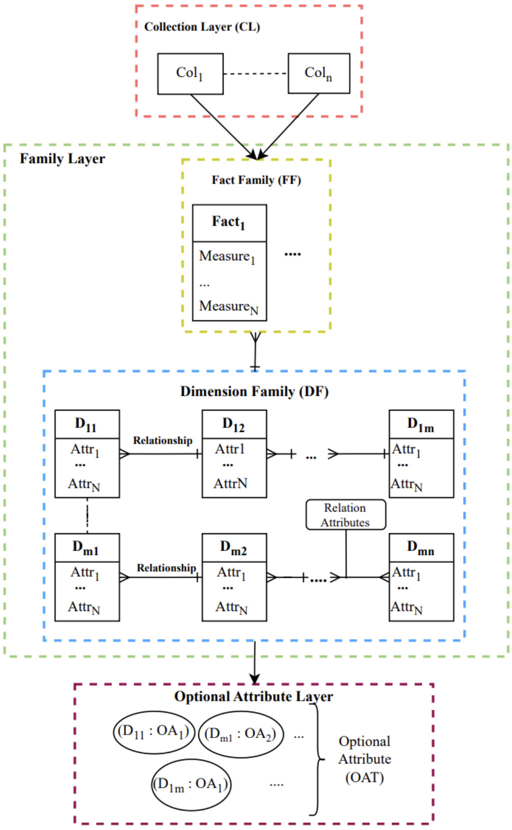


Figure 1: Proposed conceptual DW model for NoSQL DW.

mapped to a non-empty collection of dimension tables $DTs \subseteq DF$. This can be represented by the function f as follows:

$$f(FT) = DTs \text{ with } DTs = \{DT_1, ..., DT_m\}$$

b) **Component DF – Component OAT (DimOAttrRel)**: The DimOAttrRel relationship is a mapping $f : DF \rightarrow \mathcal{P}(OAT)$ where $\mathcal{P}(OAT)$ is the power set of OAT. In the DimOAttrRel relationship, each $DT(D_{name}, D_{attributes}) \in DF$ is mapped to a set of optional attributes $oat \subseteq OAT$. This can be represented by the function f as follows:

$$f(DT) = oat, \text{ where } oat = \{(D_{name}, OA_1), (D_{name}, OA_2), ..., (D_{name}, OA_m)\}$$

c) **Component Col – Component FF (CLFFRel)**: The CLFFRel relationship is a mapping $f : Col \rightarrow \mathcal{P}(FF) \setminus \{\emptyset\}$ where $\mathcal{P}(FF)$ is the power set of FF. In the CLFFRel relationship, each $Col \in CL$ is mapped to a non-empty set of fact tables $FT_s \subseteq FF$. This can be represented by the function f as follows:

$$f(Col) = FT_s \text{ with } FT_s = \{FT_1, ..., FT_m\}$$

Internal Relationships: According to Figure 3.1, internal relationships occur within the DF component. There are two types: hierarchical relationships (many-to-one) and semantic relationships (many-to-many).

a) **Hierarchical Relationships (HierarRel)**: Representing many-to-one relationships between two dimensions DT_1 and DT_2 of DF, this is a mapping: $f : DT_1 \rightarrow DT_2$ such that each element in DT_1 (child) is mapped to a unique element in DT_2 (parent).

$$f : DT_1 \rightarrow DT_2$$
$$\forall y \in DT_1, \exists!\ x \in DT_2 \text{ such that } f(y) = x.$$

b) **Semantic Relationships (SemanRel)**: Representing many-to-many relationships between two dimensions DT_1 and DT_2 of DF. To represent this SemanRel, an intermediate structure M is required. In this structure:

- M is the intermediate set containing pairs of mappings between elements in DT_1 and DT_2 along with the set of attributes RA derived from their relationship.
- $M \subseteq DT_1 \times DT_2 \times \mathcal{P}(RA)$ where $\mathcal{P}(RA)$ is the power set of the set of attributes derived from RA.
- Each element of M can be represented as a tuple (d_1, d_2, ra) where $d_1 \in DT_1, d_2 \in DT_2$ and $ra \subseteq RA$.

4 CONVERTING CONCEPTUAL NOSQL DW MODEL TO GRAPH-BASED LOGICAL MODEL RULES

In this section, we propose rules to convert the conceptual DW model to a logical DW model, which uses a graph-based data model. Our rules are stated based on Sellami et al. (2019), a study about rules for mapping a multidimensional model to a graph-based logical DW model. Before delving into the conversion rules, we will mention the foundation of a graph-based data model, laying the groundwork for the conversion.

4.1 Graph-Based Data Model Foundation

Information about the concept of the GDM is inherited from reference Sellami et al. (2019), with some adjustments for higher clarity and consistency. Specifically:

A graph-oriented database G is defined by (V^G, E^G, P^G), where:

- $V^G = \{V_1, V_2, \ldots, V_n\}$ is the set of nodes.
- $E^G = \{E_1, E_2, \ldots, E_m\}$ is the set of edges denoting relationships between nodes.
- P^G is the set of nodes' properties in the graph.

An attribute is defined by a key-value pair.

Node: A node V consists of properties and labels, identified by (id^V, P^V, L^V), where:

- id^V is the identifier of the node.
- P^V is the set of properties of the node.
- L^V is the set of labels of the node, each label l shows meaning to the node, helping in identifying the roles for each node.

Relationship: A relationship R is defined as a link connecting two nodes, which represent two entities that interact or relate to one another. Each relationship might contain more information through additional properties. A relationship can be identified by $(id^R, sourceNodeID, targetNodeID, T^R, P^R)$, where:

- id^R is the identifier of the relationship.
- $sourceNodeID$ is the identifier of the source node.
- $targetNodeID$ is the identifier of the target node.
- T^R denotes the relationship type, helping determine which relationship is linking the two nodes.
- P^R is a set of properties representing additional information about it.

4.2 Rules for Mapping Components

We will offer the conversion rules in the respective components, taking into account the attributes of the components given in the suggested conceptual DW model and the symbols of the graph-based data model described above. There are two main groups into which these rules fall: relationships and entities. Here, we will offer greater detail about these guidelines.

4.2.1 Rules for Mapping Entities

Rule 1: Mapping col and FT. Each fact table $FT(F_{name}, F_{measures})$ of the specified conceptual DW (via *CLFRFel* relationship) will become a node $V(id^V, P^V, L^V)$ where:

- Label l_1 represents the table type: $l_1 = \{fact\}/L^V = \{l_1\}$
- Label l_2 represents the table name: $l_2 = \mathrm{F_{name}}/L^V = L^V \cup l_2$
- Label l_3 represents the related collection name: $l_3 = \mathrm{col}/L^V = L^V \cup l_3$ (optional)
- Each measure $m_i \in \mathrm{F_{measures}}$ will be converted into a property $p \leftarrow m_i/P^V = P^V \cup p$

Rule 2: Mapping DT. Each dimension table $DT(D_{name}, D_{attributes})$ of the DF will become a node $V(id^V, P^V, L^V)$ where:

- Label l_1 represents the table type, which can be *dimension* or *parameter* depending on the following cases:
  - If DT is related to a fact table via the FactDimRel relationship or is related to another dimension via the SemanRel relationship, then $l_1 = \{dimension\}/L^V = \{l_1\}$
  - If DT is related to a dimension table via the HierarRel relationship and is a parent, then $l_1 = \{parameter\}/L^V = \{l_1\}$
- Label l_2 represents the table name: $l_2 = \mathrm{D_{name}}/L^V = L^V \cup l_2$
- Each identifier a_w (if any) will be converted into a property $p \leftarrow a_w/P^V = P^V \cup p$
- Each attribute $a_i \in D_{attributes}$ will be converted into a property $p \leftarrow a_i/P^V = P^V \cup p$

Rule 3: Mapping OA. Each optional attribute $a_k \in oa^{Att}(oa^{Att} \subseteq OA^{Att})$ of DT via DimOAttrRel relationship will be converted into a properties $p \leftarrow a_i/P^V = P^V \cup p$ in dimension node.

4.2.2 Rules for Mapping Relationships

Rule 4: Mapping FactDimRel Relationship. Each FactDimRel relationship will be transformed into a relationship $R(id^R, sourceNode, targetNode, T^R, P^R)$:

- sourceNode represents the fact;
- targetNode represents the dimension.
- t_1 denotes the relationship type, $t_1 = \{linkfact - dimension\}/TR = \{t_1\}$

Rule 5: Mapping HierarRel Relationship. Each HierarRel relationship will be transformed into a relationship $R(id^R, sourceNode, targetNode, T^R, P^R)$:

- sourceNode represents the child.
- targetNode represents the parent.
- t_1 denotes the relationship type, $t_1 = \{Precede\}/TR = \{t_1\}$

Rule 6: Mapping SemanRel Relationship. Each SemanRel relationship will be transformed into a relationship $R(id^R, sourceNode, targetNode, T^R, P^R)$:

- sourceNode represents the source node.
- targetNode represents the target node.
- For each attribute $a_i \in ra^m \subseteq RA^M$ it will be transformed into a property $p \leftarrow a_i / P^R = P^R \cup p$
- t_1 denotes the relationship type, $t_1 = \{dim-to-dim\} / TR = \{t_1\}$

5 CASE STUDY

This section presents the development of a personalized learning path recommendation system (LPRS) based on career goals, showcasing our graph-based NoSQL DW. We provide a practical example of applying the conceptual, logical, and physical design phases for NoSQL DW to validate our conceptual DW model and highlight graph databases' strengths in handling complex relationships and queries in recommendation systems.

Graph NoSQL databases outperform in complex relationship modeling Fernandes and Bernardino (2018); Sellami et al. (2019), making them ideal for LPRS that analyze relationships between courses, learning objectives (LO), and job positions. While previous studies have focused on graph databases Beutling and Spahic-Bogdanovic (2024); Nguyen et al. (2022), we emphasize GDW's scalable data integration and advanced analytics for personalized recommendations. Based on our previous Knowledge Graph (KG) architecture Nguyen et al. (2022) links educational resources with career objectives across three layers: Career Concept, Course Concept, and Competency Concept. We then create a GDW using this KG to enhance data-driven recommendations for LPRS.

The following subsections detail the next three phases of NoSQL DW design, involving conceptual, logical, and physical design.

5.1 Conceptual DW Model Design

In the conceptual DW model design process, we use a top-down approach to identify essential data components and their relationships, ensuring alignment with problem requirements. For this case, we focus on addressing learning path recommendation issues based on users' existing skills and career goals. We identify the main entity groups: Courses, Careers, Users, and Skills. Skills encompass various entities such as Programming Languages, Frameworks, Platforms, Knowledge Areas, and Tools, forming the ”Family Layer.” Within this layer, we define two types of tables: ”Fact Family” and ”Dimension Family.” The ”Fact Family” includes fact tables like job postings and course enrollments, along with related measures. The ”Dimension Family” supports each fact table with detailed data. For example, the job posting fact table has dimensions like posting location, time, career, website, and organization, while the course enrollment fact table includes course name, instructor, website, and training organization.

We also identify optional attributes for dimensions, such as skill requirements, proficiency levels for courses, certifications for instructors, and phone numbers for users, to optimize data querying and support advising. The complete conceptual DW model is presented in Figure 2.

5.2 Mapping Logical DW Model Design

After designing the conceptual NoSQL DW model, transitioning to a logical GDW model is crucial for simplifying deployment and enhancing efficiency on specific database types. This section outlines how to map the conceptual NoSQL DW model to the logical GDW model using the mapping rules from section 4. We'll apply these rules to transform the LPRS problem's conceptual NoSQL DW model into a complete logical GDW model, focusing on mapping entities and relationships to corresponding components, thereby streamlining deployment on graph databases. The process involves two main steps:

Entity Mapping: Each entity identified in the conceptual DW model, such as courses and websites, is mapped to a separate node in the graph database. The names and types of these entities are transformed into corresponding labels, ensuring clear node identification. Attributes are also mapped to the properties of corresponding nodes.

Relationship Conversion: The relationships between entities, depicted by crow's foot symbols in the conceptual DW model, are converted into edges connecting the corresponding nodes in the graph. This step elucidates relationships, enabling a comprehensive and interconnected representation of nodes in the graph database. Figure 3 illustrates the logical graph model after conversion.

5.3 Building LPRS Solution Based on GDW

Based on the proposed logical GDW model, we deployed it using the Neo4j platform, incorporating data from 4,000 course records and 2,000 job postings sourced from public job websites. In this GDW, data entities are represented as nodes, and relationships are effectively modeled.

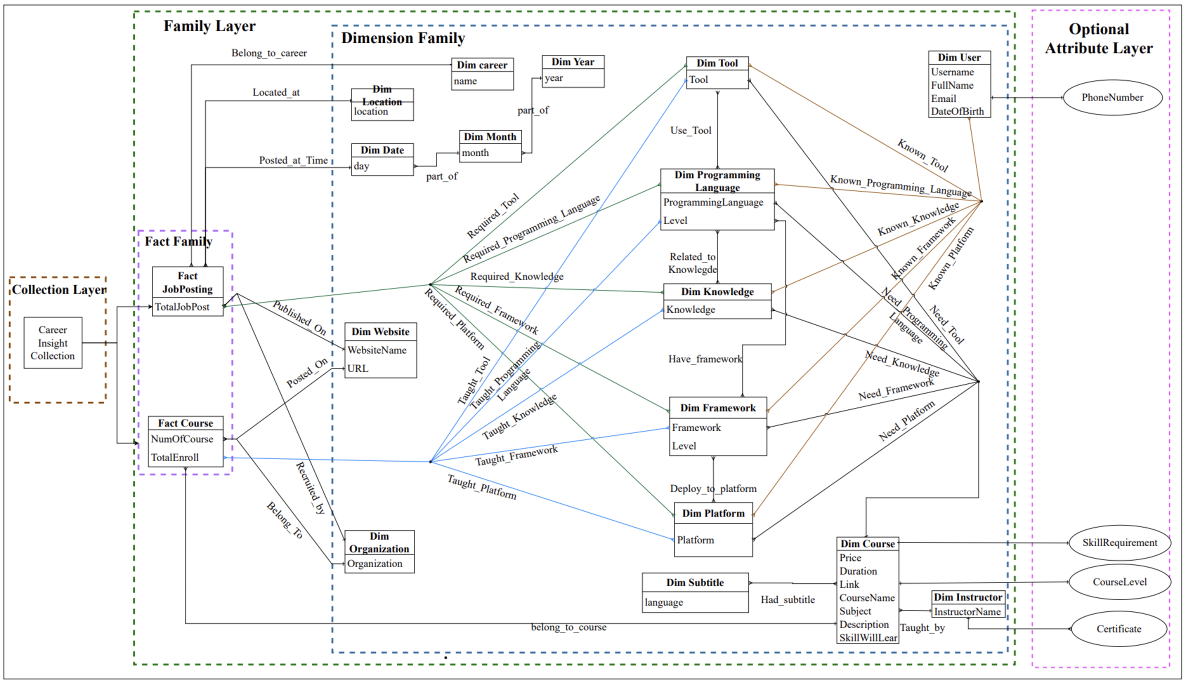


Figure 2: Proposed conceptual DW model for the LPRS problem.

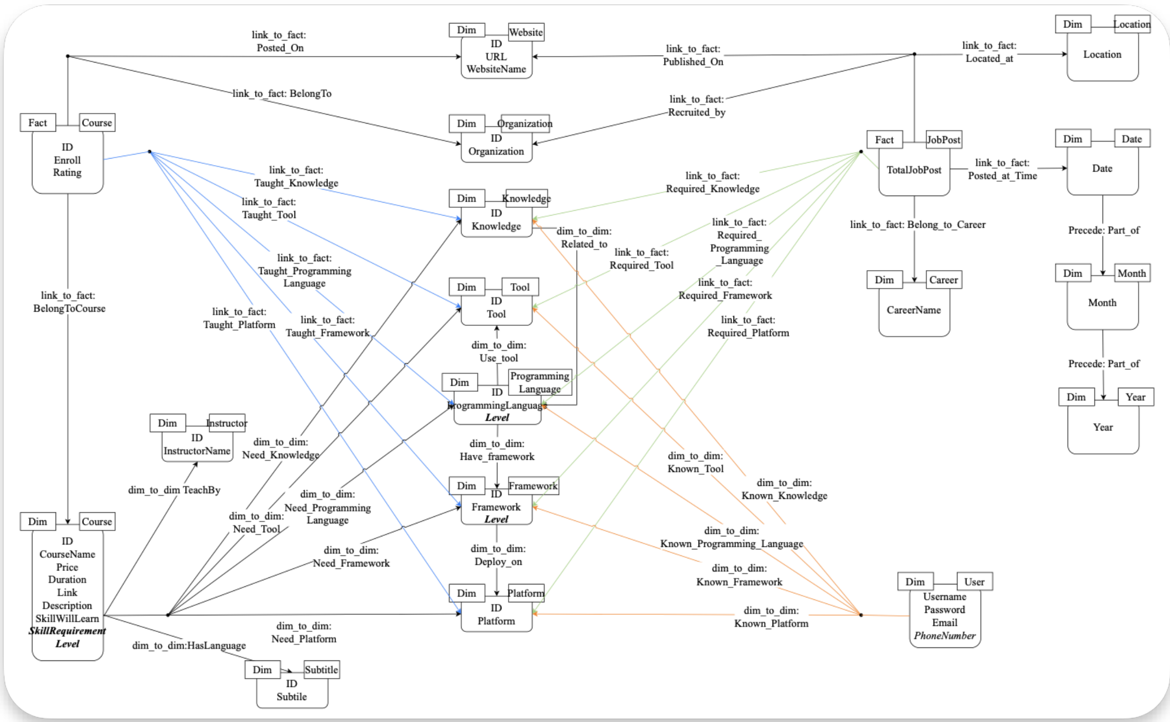
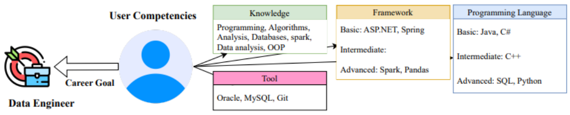


Figure 3: Logic model for the LPRS problem.

Two graph mining algorithms were developed for the Learning Path Recommendation System (LPRS). The first algorithm identifies critical competencies for a career and filters them based on the user's existing skills to recommend appropriate competencies. The second algorithm uses these competencies to find relevant courses, building a personalized learning pathway for the user based on course prerequisites and levels. Figure 4 provides an example of the input and output for these algorithms.



Learning Path	Prerequisite Course	Basic Course	Medium Course	Advanced Course
1	-	Foundations for Big Data Analysis with SQL	Scalable Machine Learning on Big Data using Ap...	Big Data Analysis: Hive, Spark SQL, DataFrames...
2	-	Process Data from Dirty to Clean	Big Data Analysis Deep Dive	Big Data Services: Capstone Project
3	-	Exploring and Preparing your Data with Big...	Business intelligence and data warehousing	How to Win a Data Science Competition: Learn f...
4	-	Introduction to Data Engineering	Scalable Machine Learning on Big Data using Ap...	Big Data Analysis: Hive, Spark SQL, DataFrames...
5	-	Data Processing Using Python	Big Data Analysis Deep Dive	How to Win a Data Science Competition: Learn f...
6	-	Python for Data Science, AI & Development	AI Workflow: Business Priorities and Data Inge...	AI Workflow: Data Analysis and Hypothesis Testing

Figure 4: An example of the results of LPRS.

The implementation of LPRS on Neo4j underscores its efficiency and scalability, particularly in managing complex queries with intricate relationships. The GDW enables rapid and precise course recommendations tailored to user profiles, showcasing its capability to handle complex data models.

6 CONCLUSIONS

This study presents a comprehensive methodology for GDW design and implementation, introducing new symbols system and transformation rules that enhance the efficiency of model conversion. Through the Neo4j-based LPRS case study, we demonstrated GDW's ability to effectively manage complex data relationships, offering enhanced flexibility and performance in analyzing intricate data structures. GDW's strengths make it particularly valuable in Business Intelligence applications, such as recommendation systems and real-time analytics, where accurate and insightful decision-making is critical. Looking ahead, our future work will involve experimenting with large datasets (big data), exploring other NoSQL types like document, key-value, and column stores, and further refining GDW's performance and scalability. These efforts will help to validate GDW's benefits and broaden its applicability across diverse data management scenarios.

ACKNOWLEDGEMENTS

This research is partially supported by research funding from the Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam. This research is funded by the Faculty of Information Technology, University of Science, VNU-HCM, Vietnam, Grant number CNTT 2023-15.

REFERENCES

Abdelhédi, F. et al. (2017). Logical unified modeling for nosql databases. In *2017*.

Akid, H. and Ayed, M. B. (2016). Towards nosql graph data warehouse for big social data analysis. In *2016*.

Akid, H. et al. (2022). Performance of nosql graph implementations of star vs. snowflake schemas. *IEEE Access*, 10:48603–48614.

Banerjee, S. et al. (2021). A unified conceptual model for data warehouses. *Annals of Emerging Technologies in Computing (AETiC)*, 5(5):162–169.

Beutling, N. and Spahic-Bogdanovic, M. (2024). Personalised course recommender: Linking learning objectives and career goals through competencies. In *Proceedings of the AAAI Symposium Series*, volume 3, page 72–81.

Bhogal, J. and Choksi, I. (2015). Handling big data using nosql. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshop*.

Chevalier, M. et al. (2015). Implementation of multidimensional databases with document-oriented nosql. In *Big Data Analytics and Knowledge Discovery: 17th International Conference, DaWaK 2015, Valencia, Spain, September 1-4, 2015, Proceedings*, volume 17. Springer International Publishing.

Fernandes, D. L. and Bernardino, J. (2018). Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb. *Information Sciences*.

Nguyen, T., Vu, N., and Ly, B. (2022). An approach to constructing a graph data repository for course recommendation based on it career goals in the context of big data. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE.

Oussous, A. et al. (2017). Nosql databases for big data. In *2017 IEEE World Congress on Services*.

Puja, I., Poscic, P., and Jaksic, D. (2019). Overview and comparison of several relational database modelling methodologies and notations. In *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*.

Sellami, A., Nabli, A., and Gargouri, F. (2019). Transformation of data warehouse schema to nosql graph database. In *Advances in Intelligent Systems and Computing*, page 410–420.

Yangui, R., Nabli, A., and Gargouri, F. (2016). Automatic transformation of data warehouse schema to nosql database: comparative study. *Procedia Computer Science*, 96:255–264.

Exploring Centralized, Decentralized, and Hybrid Approaches to Micro-Credential Issuance in HEI Alliances

Padmasheela Kiiskilä[a] and Henri Pirkkalainen[b]
Tampere University, Tampere, Finland
{padmasheela.kiiskila, henri.pirkkalainen}@tuni.fi

Keywords: Micro-Credentials, Digital Credentials, MC Governance, Knowledge Management, Data Governance.

Abstract: Micro-Credentials (MCs) are seen as a way by Higher Education Institutions (HEIs) to equip learners with the essential skills for their careers or professional development. In Europe, HEIs are joining forces to form alliances to offer a broad range of MCs and make them tamper-proof, verifiable, and shareable. Although extensive research is being done on MCs, there is a major research gap in identifying and comparing different ways alliances can manage issuance of MCs. We identified two approaches in practice and through a case study, identified a third approach alliances can utilize. This paper also addresses this gap by using the data governance contingency model to provide a comparison of all three approaches that alliances can utilize in selecting the most suitable one for their business strategy. To achieve this, a qualitative case study is conducted with in-depth interviews with administrators from HEIs that are partners of an alliance. This study contributes to the governance of MCs through identification and comparison of the three approaches - centralized, decentralized, and hybrid in the context of MC issuance by HEI alliances.

1 INTRODUCTION

In recent years, interest in micro-credentials (MCs) has been on the rise for multiple stakeholders, including higher education institutions (HEIs). Multiple alliances have formed in Europe, consisting of HEIs and other institutions to pilot MCs at different levels (ECIU, 2023; ENHANCE 2022; Una Europa, 2022). MC governance depends on management of data, processes and technology within an alliance (Doering et al., 2022). Current research focuses mostly on the implementation (Abdullah & Ghazali, 2024), security and recognition (McGreal, 2024; Saad et al., 2024), and affordances (Reed et al., 2024) of MCs but discussion on issuing MCs is limited to technology used such as blockchain (EBSI, 2022; Kiiskilä et al., 2023) and platforms (Saad et al., 2024). Two approaches, as referred in this study as hybrid and decentralized, can be found from practical examples (ENHANCE, 2022; SDG Campus, 2024; Una Europa, 2022) for MC governance by alliances. In the hybrid approach, one or more partners in the alliance issue on behalf of all partners and in decentralized, each partner issue on their own. However, empirical studies evaluating different approaches for MC governance in alliances including how to issue MCs that satisfy the MC recommendations are missing.

Issuing MCs require knowledge of data such as student data, MC data, institutional data and processes to use specific technologies such as EDCL platform. Through knowledge management, alliances can develop the required processes to offer and issue MCs that might not be possible as individual institutions (Jiang and Li, 2009). The complexities of HEI alliances, combined with the need to use platforms and the data residing in multiple systems in multiple HEIs need to be taken into consideration for MC governance. MC governance in an alliance can also depend on the scope of the alliance (Oxley and Sampson, 2004).

Traditional data governance involves structured data managed locally (Al-Ruithe, 2018) and presume that the same model can fit any organization (Weber et al., 2009). While research on using contingency factors to find the right data governance approach is available (Chelliah et al., 2016; Lee et al., 2017), empirical studies evaluating contingency factors

[a] https://orcid.org/ 0000-0002-6347-9595
[b] https://orcid.org/ 0000-0002-5389-7363

Kiiskilä, P. and Pirkkalainen, H.
Exploring Centralized, Decentralized, and Hybrid Approaches to Micro-Credential Issuance in HEI Alliances.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 238-245
ISBN: 978-989-758-716-0; ISSN: 2184-3228

relevant to alliances issuing MCs is missing. Moreover, current studies do not address how these contingency factors can be used in choosing the right approach for MC governance in an alliance setting. To address these research gaps, the following two questions are studied:

RQ1: *What are the different approaches alliances can issue MCs to their students?*

RQ2: *Based on the data governance contingency factors, how do different approaches compare for an alliance to issue MCs?*

Through the qualitative case study, we present a new centralized approach that is currently not utilized in the issuance of MCs by alliances. Taking inspiration from Weber et al.'s (2009) contingency model, the factors for the governance of MCs are identified and used for comparison of all three approaches in issuing MCs by alliances. Because issuing MCs, especially by alliances, is a rapidly growing practice-based area with limited theoretical understanding, we chose qualitative approach to gain a deeper understanding of this important area for HEIs.

The present paper contributes to MC literature by identifying a third approach that alliances can use and also to compare the three approaches for issuing MCs by alliances using data governance contingency factors. As a practical contribution, this can help HEI alliances understand the pros and cons of the three approaches to issuing MCs using the data governance contingency model. The rest of the current paper is organized as follows: Section 2 provides a theoretical background related to MCs and data governance models. Section 3 presents the data collection and analysis, which is followed by the findings in section 4. Section 5 covers the discussion and conclusion.

2 BACKGROUND

2.1 Micro-Credentials

After an extensive effort that was coordinated by the European Commission, MCs have been defined as *"the record of the learning outcomes that a learner has ac-quired following a small volume of learning*" (European Commission, 2022). Along with these efforts, the frameworks, and related technologies of MCs (DCC, 2023; Digivisio, 2023; MicroHE, 2021) started developing, and the metadata that they should contain have been described (European Commission, 2022). Within the context of HEIs, most of the literature focuses on the content, importance, and implementation of the MCs (Flynn et al., 2023; Oliver, 2019). Although some studies have explained the features needed in MC platforms, (Kiiskilä et al., 2022), and new technologies are emerging to facilitate some of those features such as tamper-proofing and verifiability (EBSI, 2022; EDCL, 2022), there is a limited understanding of how they can be used in an alliance setting for MC governance.

Most of the literature is focused on what type of MCs are offered by the alliances (Shanahan and Organ, 2022), why alliances offering them benefit learners and HEIs (Romiță et al., 2021) and recognition of MCs (Ashizawa et al., 2024). From knowledge management perspective, literature also covers how alliances can be a differentiating factor in offering unique products such as MCs (Parise and Sasson, 2002), how alliances can leverage strength from the partners to create new opportunities (Inkpen, 1998) and governance adaptation at alliance level (Reuer and Zollo, 2000). Literature on how alliances can issue MCs and their strategy for MC governance is, however, limited to a few practical examples (Una Europa, 2022; ENHANCE, 2022). There are two approaches that can be seen in these practical examples for alliances to issue MCs.

(a) Hybrid approach: One or more of the alliance partners issue MCs on behalf of all the partners. This requires one or more of the partners to develop a system internally to issue MCs for all the partners (Una Europa, 2022). It includes a system in place for alliance partners to offer the MCs in a mutually agreed platform or their own internal platform (SDG Campus, 2024) and clear path for students how to apply for the MC (credential portion) when they successfully complete it.

(b) Decentralized approach: Each partner issues their MCs. In the decentralized approach, the MCs might be offered through the alliance, but each partner need to develop the MC issuance system internally or use commercial systems such as BC Diploma (BCDiploma, 2023). The decentralized approach is similar to existing practices in HEIs in issuing study records to students coming to their institution to study a single course or a fixed amount of time for mobility (Cuzzocrea and Krzaklewska, 2023). While in normal mobility programs, regular study records are issued to students from other HEIs, for MCs partners need to either enhance their existing systems or partner with external systems to issue MCs.

In both cases, issuing MCs should include the requirements such as making them tamper-proof, for example, by using an electronic seal to sign the credentials (eSealing). Depending on a variety of reasons such as institutional or national strategy for MCs, institutional bandwidth to enhance existing systems or partner with external systems (Kiiskilä et al., 2023), these options might not work for all alliances.

2.2 Data Governance

Data governance can be described as a framework of policies, processes, and guidelines for managing data as a strategic enterprise asset (Abraham et al., 2019). It specifies rights and accountabilities for an organization's decision-making about its data (Khatri & Brown, 2010). Different frameworks have been proposed with a focus on traditional IT assets (Weill and Ross, 2004), and domains relevant to data (Khatri and Brown, 2010; Otto, 2011).

Possible data governance approaches include centralized, decentralized and hybrid (Asgarinia et al., 2023; Colona & Jaffe, 2016; Xie et al., 2024). Although most of the data governance models presume the same model can fit any organization, Weber et.al. (2009) proposed a model that takes contingency factors into account for data governance. By using these contingency factors, organizations can determine the right data governance approach (Otto, 2011). The data governance model for any organization depends on the external and internal organizational factors that determine the organizational context related to data governance (Otto, 2011). According to the Weber's contingency model (Weber et al., 2009), a specific data governance configuration needs to be designed so that it fits the organization's context factors. It consists of two distinct parameters (a) organizational placement of data quality management (DQM) activities and (b) coordination of decision making for DQM.

However, the factors proposed by Weber do not consider external contexts such as platforms used by organizations. For institutions and alliances to issue MCs, platform(s) are a key component and so the governance factors related to platforms should also be considered. Lee et al. platform governance model (Lee et al., 2018) is an extension to the Weber model that includes factors specific to organizations using platforms. For this case study, we adapted Weber's contingency factors and two of the factors from the platform governance model namely, Open strategy and Platform maturity that contribute to the MC governance. For alliances offering and issuing MCs, open strategy can open the platform for multiple use cases and platform maturity ensures strict data governance.

3 DATA COLLECTION AND ANALYSIS

A case study was conducted with an alliance of multiple HEIs offering MCs to students of the alliance partners. At the time of data collection, there were 12 HEIs in the alliance from 12 different countries throughout Europe. All the partners were already offering courses through the alliance and issued a certificate of completion, along with the required academic transcripts to the students.

Multiple sources have been used for data collection, here following the recommendations of Yin (Yin, 2009). These include internal documents, project documents, white papers, and external documents such as platform documentation. However, data from the interviews are the major source. 19 individuals in administrative roles were interviewed. The interviews lasted between 28 and 63 minutes, with an average of 52 minutes. The interviews were conducted remotely using Microsoft Teams and were recorded with the permission of the participants. Furthermore, all the institutions participated in an early adopter program with EDCL.

All the interviews were transcribed. The research approach can be considered as an interpretive case study (Elliott & Timulak, 2005). In the present study, the three coding procedures proposed by Strauss and Corbin (1990) were used: open, axial, and selective coding. NVivo software was used to identify concepts from the interview transcripts. The incident(s), action(s), or event(s) from the raw data were given conceptual labels (Strauss & Corbin, 1990). Then the identified concepts were compared with the emerging theories and the most relevant and acceptable concepts were categorized. These categories were further developed into subcategories, ensuring the existence of linkage between them.

The two approaches, named in the present study as hybrid and decentralized, were derived from the practical examples as mentioned in section 2, and the third approach is derived from the empirical analysis. The contingency factors were used as influencing factors in providing the comparison of all three approaches for MC governance.

4 FINDINGS AND ANALYSIS

4.1 Approaches for Credential Governance

As described in section 2, the issuance of MCs can happen in two ways for alliances. (a) Hybrid approach, where one or more of the alliance partners issue MCs on behalf of all the partners or each partner issues their MCs. (b) Decentralized approach which requires all the partners to build the capability to issue MCs. *In our case study, during the EDCL early adopter program and our interviews it became evident that no one institution from the alliance would like to take the responsibility to issue for other institutions nor have the capability to issue MCs.*

It became clear that neither hybrid nor decentralized approaches would work for this alliance. This resulted in the need to identify a different approach to offer and issue MCs for alliances such as the one in the present study. Based on the empirical analysis, a third approach is identified:

(c) Centralized approach: The alliance seals and issues the MCs on behalf of the partners.

The centralized approach requires the alliance to have a central platform to offer and issue the MCs. This platform becomes the central management and governance entity for all the MCs offered by the alliance. The central platform becomes the repository of all the MCs offered by the partners, ensuring all the information required to issue a MC is captured. This approach requires the central platform to provide a way for the partners to indicate when students complete the requirements to issue the MCs. The central platform also needs to have a mechanism to re-issue the same MCs if the need arises in the future.

4.2 Contingency Factors for the Governance of Credentials

4.2.1 Comparison of Approaches

To understand and evaluate the best approach for the governance of MCs by alliances, the contingency model was applied. As discussed in section 3, the contingency factors include both organizational as well as platform context factors. Table 1 shows the comprehensive contingency factors and the comparison of all three approaches.

Certain contingency factors such as the degree of market regulation stay the same for all three approaches because, irrespective of how the MCs are issued, all the data regulation policies need to be followed. The centralized approach provides an open strategy because the platform is specifically designed to offer and issue MCs and customizing to a different market is easier. Similarly, it is easier to expand the portfolio of MCs offered in the centralized approach because it is a cumulative effort of all partners. The quality of the portfolio can be higher in both hybrid and centralized approaches because interoperability and uniformity are inherent.

Table 1: Contingency factors and comparison of the three approaches for MC governance.

Contingency factor	Hybrid (A)	Decentralized (B)	Centralized (C)
Performance strategy *(High<->Low)* *Quality* *Growth*	 High Low	 Low Low	 High High
Diversification breadth *(High<->Low)* Portfolio/Market	Low	Medium	High
Organizational structure *(Centralized<->Decentralized)*	Centralized	Decentralized	Centralized
Competitive strategy *Branding* *(High<->Low)*	High	Low	High
Process harmonization *(Global<->Local)*	Semi-global	Local	Global
Degree of market regulation *(High<->Low)*	High	High	High
Decision making style *(Hierarchical<->Co-operative)*	Hierarchical	Segmented	Co-operative
Platform maturity *(Robust<->Scalable)*	Robust	Scalable	Robust & scalable
Open strategy *(Open<->Closed)*	Closed	Closed	Open

4.2.2 Case Study

The contingency factor matrix shows the comparison of the three approaches to issuing MCs. The alliance in the present case study chose centralized issuance as the best approach because it solved the issues identified earlier and opens other opportunities. As mentioned in the previous section, none of the partners in this alliance were ready to take up the responsibility of issuing MCs on behalf of all the partners, and not all the partners were ready to issue MCs on their own. When asked for preference in the method of issuing MCs, administrators from all the partner institutions responded with a unanimous vote for the centralized issuance of MCs. The reasons included administrative ease and control over what is issued.

"Administratively, it makes much more sense to have a centralized process and to have one place that issues all the credentials and has the control of accreditations and all the parameters that you need to issue the credentials". - National co-ordinator.

Administrators felt issuing from the central platform would ensure all the MCs would be consistent. Each institution can provide the same information for all the MCs that can be stored in the central platform.

"It would make it equal and not different from each university". – Chief consultant

Institutional strategy and the work needed to convince and organize is another factor mentioned by administrators. The strategic vision for the alliance and making it a good user experience for the students were at the top of the list for all institutions.

"If you look at the administrative process, then centralized would be better because it is much more work if you have to organize this within your own university". – Educational consultant

Another factor mentioned by administrators is branding. It helps with the strategic vision and branding of the alliance as the individual institutions can focus on offering high-quality MCs under the alliance brand.

"Centralized marketing, centralized administration, enrolment, record, and storage, and so on". – Teacher and work package contact for MCs.

The alliance being a legally registered entity, acquired a Qualified eSeal and acted as the "issuer" of MCs by eSealing them, and the institutions offering the MCs acted as the "awarding body". In every MC, an evidence statement confirming the issuer and awarding body roles and responsibilities was included by EDC.

To facilitate this, the alliance added an addendum to its existing accord for recognition of credits. stating that all partner institutions agreed on the alliance to eSeal and issue all the MCs on behalf of the partners. This addendum was saved on the central platform server and a link to that document was included in every MC. This new ability to let an alliance issue MCs on behalf of all the partners also led to identifying steps in the MC process and responsibilities at the institutional and central alliance levels.

4.2.3 Potential Barriers for Centralized MC Issuance

Although numerous advantages were cited for issuing MCs centrally by the administrators, certain limitations also exist in the approach. The platform is central but there are no IT pipeline integrations done from partner institutions to the central platform to streamline data transfer. Therefore, all the relevant information for MCs including content related information, and admissions requirements for each MC, need to be entered into the central system manually. This includes the mandatory information to be included into a MC such as learning outcomes, description of the course, and workload. This can be time-intensive and requires dedicated resources from each institution to manage the information.

Because most of the MCs offered by this alliance can be formal and credit-based, the grade was an important factor to include in MCs. Due to lack of IT integration, they need to be entered manually by each institution into the central platform. Grades are sensitive information for a student and so a decision was made not to include grades in the MCs as it is considered too risky and error-prone to enter them manually. Since the students need a proof of completion including the grade to receive credit in their own institution, all the institutions still need to provide proof of grade, such as transcript of records. This renders MCs complementary to the existing process, a digital proof of participation and not replacement of current study record

Even though students register to a MC in the central platform, the admission process is done by the institution offering the MC and students receive access to their learning management platform. Since the MCs are created and issued from the central platform, there is a need to establish the identity of the students to ensure the MCs are issued to the right students. A process was established to obtain the level of security for the students either from their home institution or the host institution who is offering the MC by conducting an identity check. In cases where host institutions do not conduct an identity check if partner students register and home institution doesn't have an established process to set the security level for the students, the level of security remains low which results in MCs not issued to those students. This requires either an additional workload for IT team in the institutions or an additional process to be followed in the central platform to do manual identity checks for those students with low level of security through video chats. This adds additional burden either on the institutions to ensure the level of security is added or on the central staff to perform manual checks. This may also increase privacy and security related issues for the students with additional identity checks.

Centralized issuance of MCs can add additional privacy and security related governance issues and need to be carefully considered including discussions with data protection officers and designing comprehensive policies.

5 DISCUSSION AND CONCLUSION

5.1 Theoretical Contributions

The first key contribution is the identification of a new approach in the context of MC issuance by alliances. Studies have examined MCs and their potential in various fields for career development (Vordenberg et al., 2024) and possible hurdles in implementing them from HEI perspective (Raj et al., 2024; Saad et al., 2024). Issuing MCs is discussed only as an end to means rather than as an important influencing factor (Alsobhi et al., 2023; Halim et al., 2024) especially in the case of alliances. Few studies that are available on alliances offering MCs (Ipsilandis et al., 2024) focus on dynamics of alliances themselves. MC governance required to issue MCs is critical and needs thoughtful consideration for the success of alliances in offering MCs. The two approaches we identified through practical examples work for certain alliances but not for all. Factors such as an individual institution's ability to issue MCs and willingness to take the burden of issuing MCs on behalf of all the partners determine whether either of the approaches can be used. Alliances where neither option is viable are left with no solution. With the centralized approach that we identified, alliances can consider building a central system and using the alliance legal entity to eseal and issue MCs for all the partners. This gives the institutions the flexibility to build the capability of issuing MCs in step with their institutional strategy. This centralized approach for the alliances to issue MCs has not been used in practice or studied to our knowledge. This contributes to the MC literature to include different approaches for MC governance.

The second key contribution is the comparison of the three approaches in the context of MC issuance. Although the three approaches namely: hybrid, decentralized and centralized can be found in existing literature in the data governance context (Coche et al., 2024; Lemieux et al., 2020), the same cannot be applied to MC governance. Existing studies offer knowledge about technology to use for MC governance (Subramanian et al., 2024). Studies involving governance in alliances also discuss about governance mechanism that promote knowledge sharing (Eden et al., 2011) and how governance form can enable partners to develop, transfer and protect knowledge (McGill, 2007). However, our findings illustrate that in an alliance setting, who and how MCs can be issued using the technology is still needed. To meet requirements such as making the MCs tamper proof and include certain information mandatory in each MC, every alliance needs to look at how management and governance can be done and choose the right approach for them. This paper provides an understanding how the contingency factors from data governance can be used to find the right approach for any given alliance. To the best of our knowledge, this is one of the first studies to provide a comparison of all three approaches in the context of MC governance which includes issuing MCs. This contribution extends extends the MC literature to compare different approaches alliances can take. This also contributes to the data governance literature to use contingency factors in determining the right approach for alliances to take.

5.2 Practical Contributions

The findings make practical contributions by explaining 1) reasons why certain approaches such as hybrid and decentralized, might not meet the needs of some alliances for MC governance and 2) provide comparison of approaches from data governance perspective so alliances can make informed decisions. MC research has lacked empirical studies that could inform alliances different approaches feasible for MC governance or how to choose the right approach. Only a handful of practical examples are available and the approaches from those examples, don't necessarily fit for all alliances as was evident from the findings. The present case study presents a centralized approach for MC governance and a comparison of all three approaches provided in this study from data governance perspective can help alliances when considering the right approach for their MC governance.

5.3 Limitations and Future Research Topics

The present study has certain limitations that warrants further research. First, the interviews were done during spring 2022, when MCs were still new for the administrators who were interviewed. However, all the administrators had considerable knowledge of the early adopter program in which the alliance participated in and helped in understanding the institution's ability to issue MCs at that time. Second, the comparison of the three approaches was done specifically, with a focus on the issuance of MCs by the alliances. Upcoming studies should also explore the validation of these approaches for other business strategies. Third, the contingency factors considered for the case study were specifically for an alliance of HEIs. Further research is needed to better understand the influence of contingency factors on data governance for other types of alliances as well. Fourth, the alliance in this case study is an established entity with an existing history. Further research is needed to understand whether a similar approach would work for alliances that form just for a single project.

ACKNOWLEDGEMENTS

This research was supported by ECIUn+ (101089422) European Universities funding.

REFERENCES

Abdullah, F., & Ghazali, N. (2024, January). MCs at Higher Education Institutions: Towards Smooth Sailing Ahead. In *4th International Conference on Communication, Language, Education and Social Sciences (CLESS 2023)* (pp. 118-126). Atlantis Press.

Abraham, R., Schneider, J. and Vom Brocke, J., (2019). Data governance: A conceptual framework, structured review, and research agenda. International journal of information management, 49, pp.424-438.

Alsobhi, H. A., Alakhtar, R. A., Ubaid, A., Hussain, O. K., & Hussain, F. K. (2023). Blockchain-based micro-credentialing system in higher education institutions: Systematic literature review. *Knowledge-Based Systems*, *265*, 110238.

Al-Ruithe, M. S. (2018). Development and evaluation of a holistic framework and maturity assessment tools for data governance in cloud computing environments (Doctoral dissertation, Staffordshire University).

Asgarinia, H., Chomczyk Penedo, A., Esteves, B., & Lewis, D. (2023). "Who Should I Trust with My Data?" Ethical and Legal Challenges for Innovation in New Decentralized Data Management Technologies. *Information*, *14*(7), 351.

Ashizawa, S., Ziguras, C., & Yonezawa, A. (2024). Convergence or fragmentation? Recent developments in recognition of microcredentials and their impact on higher education in Asia and the Pacific. Journal of International Cooperation in Education, 26(1), 116-130.

BCDiploma, (2023). https://www.bcdiploma.com/en/micro Certification.

Chelliah, J., Boersma, M., & Klettner, A. (2015, January). Governance challenges for not-for-profit organisations: Empirical evidence in support of a contingency approach. In *Australasian Conference on Business and Social Sciences 2015, Sydney*.

Coche, E., Kolk, A., & Dekker, M. (2024). Navigating the EU data governance labyrinth: A business perspective on data sharing in the financial sector. *Internet Policy Review*, *13*(1).

Colona, F., & Jaffe, R. (2016). Hybrid governance arrangements. *The European Journal of Development Research*, *28*, 175-183.

Cuzzocrea, V., & Krzaklewska, E. (2023). Erasmus students' motivations in motion: understanding super-mobility in higher education. Higher Education, 85(3), 571-585.

DAMA. International, D. (2017). *DAMA-DMBOK: data management body of knowledge*. Technics Publications, LLC.

DCC, (2023). https://digitalcredentials.mit.edu/

Digivisio2030, (2023). https://digivisio2030.fi/en/publications/modularity-and-MCs-preliminary-study/.

Doering, C., Reiche, F., & Timinger, H. (2022). Cross-university Platforms as an Enabler for Knowledge Management and Transfer. In KDIR (pp. 185-191).

EBSI, (2022). https://ec.europa.eu/digital-building-blocks/wikis/display/EBSI/MCs.

ECIU, (2023). https://www.eciu.eu/news/presentation-of-the-third-eciu-university-MCs-paper-a-vision-for-european-learners-values-and-priorities.

EDCL, (2022). https://europa.eu/europass/en/stakeholders/european-digital-credentials.

Eden, L., Hitt, M. A., Ireland, R. D., Garrett, R. P., & Li, D. (2011). Governance in Multilateral R&D Alliances. Organization Science, 1-20.

Elliott, Robert, and Ladislav Timulak. "Descriptive and interpretive approaches to qualitative research." A handbook of research methods for clinical and health psychology 1, no. 7 (2005): 147-159.

ENHANCE, (2022). https://enhanceuniversity.eu/wp-content/uploads/2023/06/ENHANCE-Template-for-Micro-credentials.pdf.

ENHANCE, (2023). https://enhanceuniversity.eu/wp-content/uploads/2023/07/Final_Policy-Paper_EEA-Implementation_ENHANCE-Alliance.pdf

European Commission (2022). Annex 1.b to the Proposal for Council Recommendation on a European Approach to MCs for Lifelong Learning and Employability. Luxembourg: Publications Office of the European Union https://data.consilium.europa.eu/doc/document/ST9237-2022-INIT/en/pdf.

Flynn, S., Cullinane, E., Murphy, H., & Wylie, N. (2023). MCs & Digital Badges: Definitions, Affordances and Design Considerations for Application in Higher Education Institutions. AISHE-J: The All Ireland Journal of Teaching & Learning in Higher Education, 15(1).

Halim, F. S. A., Luaran, J. E., & Jill, L. S. S. (2024). Unravelling Challenges of Higher Education Institutions in Implementing Effective MCs: A Multi-Stakeholder Qualitative Study. *Asian Journal of University Education*, *20*(1), 114-126.

Inkpen, A. (1998). Learning, knowledge acquisition, and strategic alliances. European Management Journal, 16(2), 223-229.

Ipsilandis, P., Kokkinos, K., & Samaras, N. (2024). A Roadmap for European Universities Policy Recommendation to Achieve HEI Sustainability: The Case of INVEST EU Alliance. *ESI Preprints*, *25*, 170-170.

Jiang, X., & Li, Y. (2009). An empirical investigation of knowledge management and innovative performance: The case of alliances. Research Policy, 38(2), 358-368.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148-152.

Kiiskilä, P., A. Hanafy, and H. Pirkkalainen. 2022. "Features of Micro-Credential Platforms in Higher Education." *CSEDU* 1: 81–91.

Kiiskilä, P., Hylli, O., & Pirkkalainen, H. (2023). How Can European Blockchain Services Infrastructure Be Used For Managing Educational Digital Credentials?. In *The proceedings of the 14th Scandinavian Conference on Information Systems (SCIS)*. Association for Information Systems.

Lee, S. U., Zhu, L., & Jeffery, R. (2018). Designing data governance in platform ecosystems. https://scholarspace.manoa.hawaii.edu/items/cc309dc4-7e72-4a89-afa5-e95 9e91be508. Queried 04/04/2004

Lemieux, V. L., Rowell, C., Seidel, M. D. L., & Woo, C. C. (2020). Caught in the middle? Strategic information governance disruptions in the era of blockchain and distributed trust. *Records Management Journal*, *30*(3), 301-324.

McGill, J. P. (2007). Technological knowledge and governance in alliances among competitors. International Journal of Technology Management, 38(1-2), 69-89.

McGreal, R. (2024). Empowering MCs Using Blockchain and Artificial Intelligence. In *Global Perspectives on Micro-Learning and MCs in Higher Education* (pp. 75-90). IGI Global.

MicroHe, (2023). https://microhe.microcredentials.eu/

Oliver, B. (2019). Making MCs work for learners, employers and providers. Retrieved from dteach.deakin.edu.au/microcredentials.

Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1), 3.

Oxley, J. E., & Sampson, R. C. (2004). The scope and governance of international R&D alliances. Strategic Management Journal, 25(8‐9), 723-749.

Parise, S., & Sasson, L. (2002). Leveraging knowledge management across strategic alliances. Ivey Business Journal, 66(4), 41-47.

Raj, R., Singh, A., Kumar, V., & Verma, P. (2024). Achieving professional qualifications using micro-credentials: a case of small packages and big challenges in higher education. *International Journal of Educational Management*.

Reed, A., Kong, Y., & Abramovich, S. (2024). Assessment, credential, or both? higher education faculty's design principles for MCs. *Discover Education*, *3*(1), 16.

Reuer, J., & Zollo, M. (2000). Managing governance adaptations in strategic alliances. European management journal, 18(2), 164-172.

Romiță, B., Ciolan, L., Nedelcu, A., & Cartis, A. (2021). Why micro-credentials should become educational "Macro-policies" for defining the future European study programmes.

Saad, A., bin Abdul Jamal, M. H., & bin Amran, A. R. (2024, January). Before the Badge: Tackling Initial Hurdles in Integrating MCs at Institutes of Higher Learning. In 2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-8). IEEE.

SDG Campus. (2024). https://sdg-campus.de/course/view.php?id=65

Shanahan, B. W., & Organ, J. (2022). Harnessing the benefits of micro credentials for industry 4.0 and 5.0: Skills training and lifelong learning. IFAC-PapersOnLine, 55(39), 82-87.

Subramanian, S. S., Krishnan, A. S., & Seetharaman, A. (2024). Blockchain Revolution in Education. In *Frameworks for Blockchain Standards, Tools, Testbeds, and Platforms* (pp. 96-130). IGI Global.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Sage publications.

Una Europa, (2022). https://www.una-europa.eu/study/micro credential-sustainability.

Vordenberg, S. E., Fusco, N. M., Ward, K. E., Darley, A., Brady, J. H., Culhane, N. S., ... & Matsumoto, R. R. (2024). An Integrative Review of Micro-credentials and Digital Badges for Pharmacy Educators. *American Journal of Pharmaceutical Education*, *88*(3), 100660.

Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. Journal of Data and Information Quality (JDIQ), 1(1), 1-27.

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results*. Harvard Business Press.

Xie, S., Luo, N., & Yarime, M. (2024). Data governance for smart cities in China: the case of Shenzhen. *Policy Design and Practice*, *7*(1), 66-86.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.

Evaluating Healthcare Automation: A Multi-Case Study on the Utilization of Automation Initiatives in Healthcare Operations

Jani Kaitosalmi[a] and Milla Ratia[b]

Department of Industrial Engineering and Management, Aalto University, Otakaari 1B, Espoo, Finland

Keywords: Healthcare Automation, Impact Assessment, Robotic Process Automation, Intelligent Automation.

Abstract: Automation technologies such as robotic process automation (RPA) and intelligent automation (IA) are essential for managing rising healthcare costs and ensuring sustainable health services. Although these solutions have been implemented in several Finnish healthcare organizations, their overall impact has not been systematically evaluated. This research investigates the impact and evaluation of healthcare automation through a multi-case study conducted in two Finnish healthcare organizations. While automation has improved resource utilization, process efficiency, and standardization across units, the findings highlight the need for a more comprehensive evaluation and continuous monitoring of automation benefits. Future research should focus on developing a specific evaluation framework tailored to healthcare automation technologies. The adoption of holistic evaluation methods could allow healthcare organizations to better understand the impact of automation and further enhance operational efficiency and patient care.

1 INTRODUCTION

Mestres (2017) posits that healthcare faces three key challenges: rising costs, a decrease in available physicians, and an increase in patients. Similar trends are noted in Finland, with increasing healthcare costs and declining availability of personnel (Kirkonpelto, Mäntyranta, et al., 2023). Digitalization offers solutions by improving productivity, transforming care delivery, and simplifying administration (Sony, Antony & Tortorella, 2023). Although digitalization can help healthcare organizations meet their objectives, the outcomes of digital interventions often take time to materialize and can be challenging to measure (Cresswell, 2023).

This paper focuses on one specific avenue of digitalization in healthcare – the impact of automating digital workflows. Automation initiatives can deliver a wide range of benefits to healthcare organizations through cost savings, standardized and always-available processes, and freeing employees from repetitive manual tasks (Kedziora and Smolander, 2022; Ratia et al., 2021; Kedziora and Kiviranta, 2018).

Although automation solutions are recognized for their potential, it remains unclear how these back-office solutions generate value in the public healthcare sector (Ratia et al., 2021). In particular, there is a lack of relevant research on the impact assessment of automation (Meironke and Kuehnel, 2022), with discussions being "ad-hoc and scattered, with minimal empirical and theoretical support" (Denagama Vitharanage et al., 2020). The research on automation solutions often focuses on the technical dimensions, giving less emphasis on the benefits and value creation (Ratia et al., 2021). Various patient-centric attempts have been made to create an overall impact assessment framework for digitalization efforts (e.g. WHO, 2016; Lillrank et al., 2019; Parviainen et al., 2017; Karunasena & Deng, 2012). However, research on automated back-office solutions that do not directly affect patient outcomes has been more limited. Especially in the Finnish public sector, systematic evaluations of automation technologies are limited (Kääriäinen et al., 2018).

This research aims to examine the outcomes of automation solutions in healthcare and provide information on how their impact can be measured. The study focuses on two methods: *robotic process automation* (RPA) and *intelligent automation* (IA), applied within two Finnish public healthcare districts: The Wellbeing Services County of Pirkanmaa

[a] https://orcid.org/0009-0009-1800-5137
[b] https://orcid.org/0000-0002-3360-9555

Kaitosalmi, J. and Ratia, M.
Evaluating Healthcare Automation: A Multi-Case Study on the Utilization of Automation Initiatives in Healthcare Operations.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 246-255
ISBN: 978-989-758-716-0; ISSN: 2184-3228

(PIRHA) and the Helsinki and Uusimaa Hospital District (HUS). Although both healthcare organizations had implemented various automation tools, they had not fully assessed the overall impact of these solutions. The objective of this paper is two-fold: to identify key automation benefits and potential risks and to provide information on what should be considered when evaluating automation outcomes. The research questions are the following:

RQ1. *What are the benefits and risks of utilizing automation solutions in healthcare?*
RQ2. *What should be considered when evaluating automation in healthcare?*

2 BACKGROUND LITERATURE

2.1 Utilizing Automation to Enhance Healthcare Operational Processes

Workflow automation, which involves identifying sequences of tasks that can be streamlined with digital tools, provides opportunities to address process inefficiencies in healthcare (Zayas-Cabán, Okubo, & Posnack, 2023). Automating simple workflows is needed in healthcare, where professionals are burdened by repetitive administrative duties like data entry, documentation, and scheduling – tasks that often distract from patient care. Automation technologies can take over these routine tasks, allowing healthcare providers to dedicate more time to meaningful patient interactions. (Mohamed & Frank, 2022.)

Often automating business processes require the use of application programming interfaces (API) to connect with legacy systems (Herm et al., 2023; Syed et al., 2020). However, such solutions are not always feasible in healthcare due to technical constraints, legacy systems, or vendor limitations. In these cases, lightweight and non-invasive automation solutions, such as RPA and IA, can present an opportunity to improve operational processes across care delivery and administration (Ferris, Ackers & Borhani, 2022; Zayas-Caban, Okubo & Posnack, 2023).

RPA automates repetitive digital tasks typically performed by humans (Ivančić, Suša Vugec, and Bosilj Vukšić, 2019; Kääriäinen et al., 2018; Willcocks, Lacity, and Craig, 2015; Ratia et al., 2021). These robots mimic human actions on system interfaces (Herm et al., 2023). Due to its low cost, quick implementation, and minimal system changes, RPA is attractive for healthcare providers struggling with system integration (Ratia et al., 2021; Osmundsen et al., 2019; Kedziora & Kiviranta, 2018).

RPA is effective in automating routine processes, such as physician credentialing, patient scheduling, and billing (Jain and Bhatnagar, 2019). For example, during COVID-19, RPA accelerated diagnoses, distributed targeted health information, and updated quarantine data (Doğuç, 2021). In Finland, RPA was used to process up to 2,000 COVID-19 vaccinations daily, a task previously done by 10-15 employees (Adolfsson, 2021). Similarly, in HUS, RPA saved over 13,000 workdays in 2021, equivalent to the work of 65 full-time employees (HUS, 2021).

While RPA is a powerful tool, its applications are primarily limited to tasks that are highly rule-based, structured, standardized, and supported by well-documented decision logic (Ng, 2021). IA enhances RPA with AI features such as machine learning and natural language processing to replicate human cognitive skills (Kedziora and Hyrynsalmi, 2023; Coombs et al., 2020). With these capabilities, IA can handle unstructured data, make real-time decisions, and perform content-aware computing, overcoming many limitations of RPA (Ng et al., 2021).

IA can assist in complex scheduling, capacity management, and process optimization in operating rooms and emergency departments by managing patient movement between diagnostics and wards (Garcia et al., 2020). Additionally, IA can coordinate patient information, issue health risk alerts, predict health outcomes, and optimize logistics processes (Secinaro et al., 2021). However, IA research is still in its early stages with challenges and potential applications yet to be fully explored (Ng, 2021).

2.2 Impact Evaluation of Automation Initiatives

Porter (2013) defines healthcare value as "maximizing value for patients," meaning achieving the best outcomes at the lowest cost. In other words, *healthcare value = patient outcomes / total costs*. Specifically, Porter (2014) argues that while process measurement and improvements are valuable tactics, they cannot replace the importance of measuring the results of care. Thus, healthcare value is the result of care, not the volume of services delivered. What is then the impact of back-office automation solutions that do not directly affect patient outcomes?

To measure the overall healthcare outcomes, the primary metric used is *effectiveness*, which refers to the impact of treatment on the health conditions of a population under routine conditions (Ikonen, 2019; Pitkänen et al., 2018). From effectiveness, several other important metrics are derived, such as *cost-effectiveness*, which evaluates the ratio of effectiveness to costs, and *productivity*, which is the ratio of output (the result of an operation) to input (the resources or costs used to achieve that output) (Sintonen et al., 2021). Productivity can be enhanced by either reducing costs while keeping the output constant or by increasing output while maintaining the same costs. Consequently, improving productivity also enhances the cost-effectiveness of healthcare processes (Sintonen et al., 2021).

Automation is often implemented to improve cost-effectiveness by enhancing productivity. Automation benefits can be tangible, like reduced costs and faster processing, or intangible, such as better customer satisfaction and employee motivation. Evaluating both types separately is crucial since not all benefits, like financial gains, are easily measurable (Axmann and Harmoko, 2021).

While the automation benefits are recognized, their impact assessment is still lacking (Meironke & Kuehnel, 2022; Kääriäinen et al., 2018). According to Kääriäinen et al. (2018), organizations tend to assess automation mainly using a narrow set of criteria, with a focus on internal savings. Many of the assessment frameworks focus on the requirements, feasibility, and readiness of a company to implement new technology, such as the 5D digital technology assessment (Axmann and Harmoko, 2021), RPA selection indicators by Kim (2023), or the method for RPA process selection proposed by Wanner et al. (2019). However, they do not consider how to conduct continuous monitoring and evaluate the impact after the implementation.

Moreover, Meironke and Kuehnel (2022) identified 62 unique metrics in the literature to evaluate the benefits of automation. Most of these metrics focus on efficiency and costs, emphasizing the number of transactions, work hours saved, and process time. Accuracy or error rates are also commonly used to measure quality and compliance benefits. On the other hand, metrics related to implementation effort, employee and customer satisfaction, availability, and interoperability are less commonly found. Thus, the authors conclude that the assessment of automation benefits “shows a tendency” to prioritize quantifiable economic metrics over qualitative and non-economic ones. (Meironke and Kuehnel, 2022.)

3 METHODOLOGY

This research aims to provide insights into evaluating automation outcomes. To address the research questions, the study has focused on the benefits, risks, and evaluation challenges of automation in two Finnish healthcare districts: PIRHA and HUS.

The research was conducted as a multi-case qualitative study. This study employed a qualitative design to explore a relatively novel topic, which is ideal for gaining an in-depth understanding of phenomena within specific contexts when little is known about a topic (Saunders et al., 2019; Antwi & Hamza, 2015). Moreover, the research employed an embedded multi-case study design; it contained more than one sub-unit of analysis to conduct an in-depth examination of a current phenomenon (the ‘case’) within its real-life setting (Yin, 2018). The method allowed to investigate stakeholders' experiences with various automation implementations across multiple units. The two case organizations were ideal for the study due to their size and previous experience with multiple automation projects. Furthermore, while both healthcare organizations had implemented various automation solutions, they had not fully assessed their overall impact.

3.1 Data Collection

The empirical data was gathered through 32 semi-structured interviews with employees, administrative staff, and stakeholders in PIRHA and HUS experienced with RPA and IA. Participants included department secretaries, nurses, doctors, pharmacists, digitalization experts, and head physicians. External IT companies providing automation solutions were also interviewed to capture their perspectives. Table 1 lists all participants.

The interview data was gathered in two phases: from PIRHA in Spring 2023 and HUS in Autumn 2023. PIRHA interviews focused on RPA solutions, while HUS interviews covered IA processes. Purposive sampling was used, meaning the participants were selected based on characteristics that matched the research objectives (Andrade, 2021; Etikan et al., 2016). Specifically, participants had prior experience with automation and represented a diverse set of roles. The research team provided preferences for roles and units but had limited control over participant selection. Due to the policies of both organizations, the final participant selection was made within a tight timeframe by a designated contact person. Consequently, convenience and availability also played a significant role in recruitment.

All interviews followed a similar structure and were conducted remotely via Microsoft Teams. Participants discussed automation benefits and risks from the perspectives of employees, patients, costs, and processes and shared their views on impact assessment. The interviews were recorded, transcribed, and pseudonymized with participants' permission before analysis.

While the interviewees discussed several automated processes, the primary focus was on two processes: referral handling and medical dosage building. Both organizations had implemented a *referral sorter* to shift manual work from nurses and department secretaries. "*Before automation or the electronic patient record system's XDS archive, the paper referrals arriving at the unit were placed on the doctor's desk,*" as one secretary from HUS describes it (Ahlskog, 2022). Now, the robot continuously processes new doctors' referral texts in a virtual referral center. It categorizes them into the correct queues, where doctors can access and review the referrals for further evaluation. In PIRHA, the RPA-based solution uses predefined logic to automate repetitive referral handling. However, if the robot misplaces a referral, an employee must correct it. Moreover, HUS has enhanced the referral sorter with AI, using machine learning to interpret symptoms and diagnoses from referral texts.

HUS has also an automation solution for filling patients' dosage information in the electronic health record system, Apotti. This *medical dosage builder*, used by nurses and pharmacists, transforms free-text medication info into the required structured format. It provides code suggestions, reducing the need for manual entry of the patient's medication list. The structured medical information is standardized data that is used to generate dosage instructions with a similar structure for all patients, reducing the need for manual input and minimizing errors.

3.2 Data Analysis

Data analysis was conducted using qualitative content analysis with the QDA software ATLAS.ti. The study followed a directed content analysis approach presented by Hsieh and Shannon (2005), starting with relevant research findings and preliminary theory-based categories for automation benefits and risks. The empirical data consisted of two datasets, one for each case, with data collection divided among the researchers. Consequently, none of the researchers participated in all interviews, resulting in a final dataset that was a mix of primary and secondary data for all researchers.

Given the mix of primary and secondary data, the study employed tactics for an abductive content analysis as outlined by Vila-Henninger et al. (2024). Firstly, a deductive theory-based codebook was created and expanded with inductive codes. For instance, in this step, the benefit dimensions provided by Meironke and Kuehnel (2022) were utilized to categorize the benefits. The "AI coding" feature of ATLAS.ti was also used for exploratory coding. Secondly, to reduce data volume, codes were combined into broader categories. Here, the 'Query Tool' was used to sort codes based on the four perspectives that were investigated. Finally, a detailed manual qualitative analysis was performed to identify emerging themes and compare differences between the two cases.

Table 1: Participant list from PIRHA and HUS.

Region	Role	Number of participants
PIRHA	Department Secretary	5
	Nurse	2
	Chief Physician / Director	2
	Medical Doctor	2
	Digitalization Specialist	2
	Head / Deputy Head Nurse	2
	Product Owner (External)	1
	Midwife	1
	Service Provider (External)	1
HUS	Pharmacist / Senior Pharmacist	5
	Digitalization Specialist	2
	Department Secretary	2
	Nurse	1
	Product Owner	1
	Deputy Chief Physician	1
	Product Owner (External)	1
	Data Scientist (External)	1
Total		**32**

4 FINDINGS

4.1 Perceived Benefits

The thematic interviews in both healthcare counties expressed that the overall perception of automation has been positive. The most mentioned direct benefits were related to process efficiency, improved resource utilization, and availability of services.

In HUS, one external expert describes that the referral processing time has decreased significantly, from 32-35 hours to approximately 3-4 hours. However, while quicker referral handling provides information to patients faster, it doesn't necessarily lead to faster treatment. More precisely, participants describe that the referral sorter has merely moved the bottleneck from the referral handling to the next step of the patient journey. Thus, speeding up one part of the process does not necessarily reduce overall treatment time. However, faster referral handling can be impactful for urgent patients. Participants describe how automation robots process urgent cases faster and thus increase the accessibility to treatment: *"For urgent referrals that take 1-7 days, it makes a significant difference if they can be processed in 2-4 hours instead of five working days."*

While interviewees often mention time savings, they struggle to pinpoint where the extra time goes. Nurses and pharmacists note a slight increase in patient care time, which enhances the overall quality of the patient experience. However, these time savings per patient inquiry are deemed small and hard to measure. In HUS, doctors have faced additional workload from the referral sorter, as incorrect referrals have been directed to them. On the contrary, in PIRHA, the effect has been more moderate. For secretaries, the possible benefits seem most direct, as they have more time for more challenging administrative tasks, such as appointment scheduling and other phone-related work.

Moreover, participants in HUS described that IA solutions has created standardization across the organization. Before automation, there were large variations in how medication lists were written. As automation requires rule-based inputs, it has standardized the medical information, leading to consistent interpretations across the organization. *"The fact that information is the same for all users. Perhaps that's the best benefit here. It also guides us in standardizing practices across different areas of healthcare,"* describes one pharmacist. *"This is a massive organization, and with automation, we have achieved greater unity across specialties,"* summarizes another participant.

Other process-related benefits were related to compliance and interoperability. The solutions have helped to meet reporting requirements and facilitated data and system linkage. *"The smoother we can make reporting, the better,"* describes one participant. "*From the perspective of organizational development, it supports knowledge management,"* continues another interviewee. Moreover, while automation robots act fast, they also handle information without biases, positively impacting equal access to treatment. *"It's not affected by whether someone is in a bad mood, had a rough morning, or is running late for work. Automation is consistent; it doesn't get tired or have biases."*

Automation solutions have also created indirect benefits in both organizations, such as enhanced employee and patient experience. The participants describe that automation can help "find meaningful job roles" and provide time for "brain-intensive work that professionals are trained for."

While cost savings are discussed, the perception of it varies across units and roles. The digitalization experts highlight the fast payback time and scalability of automation solutions. *"The time savings accumulate when creating such easily scalable processes. There are no additional costs with expansion,"* describes one digitalization expert. On the contrary, nurses and secretaries find it hard to evaluate cost savings, one reason being that some of the potential saved time goes to monitoring and correcting the results: *"Any freed-up time currently goes into fixing and monitoring the results. However, in the long term, when hopefully everything is running smoothly, I would see cost savings occurring,"*

Lastly, participants note that patients likely don't notice automation since it works in the background. However, automation has positively impacted patients in various ways, such as delivering treatment information faster and allowing doctors more time for patient care.

4.2 Recognized Risks

In both cases, the two main recognized potential risks of automation were additional work created by technical errors and decreased process quality. The risk of additional work appears to be significantly higher during what participants describe as the "infancy stage," meaning the early stages of development. PIRHA reported significantly more issues with its early-stage RPA solutions than HUS, which had both more experience and more advanced automation solutions implemented.

Table 2: An overview of the perceived main benefits and risks in HUS and PIRHA.

Direct benefits	**Description**
Process efficiency	Increase in volumes and decrease in throughput times and delays.
24/7 availability	The robots are always available to work.
Increased resource utilization	Time savings enable employees to complete more complex tasks.
Standardization	Automation provides similar outputs across units.
Faster access to treatment (urgent cases)	Automation can prioritize urgent cases.
Compliance	Helps to meet reporting requirements.
Interoperability	Easier data and system linkage.
Scalability	The solutions are fast and cheap to scale across units.
Equal access to treatment	Robots handle information without biases.
Indirect benefits	
Increased employee experience	Increased job satisfaction and meaningfulness of work.
Increased patient experience	Provides information faster to the patient and can speed up treatment visits.
Cost-effectiveness	Increase in the value of time gains (difference in cost of process by a human vs automation)
Potential risks	
Additional work	Automation solutions are not error-free. The risk of additional work is higher during the "infancy stage."
Decrease in quality	Humans perform tasks more accurately than robots.
Risks in patient safety	Automation faces challenges in interpreting complex data. Outputs involving critical patient information still need to be reviewed by a professional.

While additional work is a risk, it also matters to whom and how much of it is created. In PIRHA, most of the additional work was targeted to secretaries and nurses, some of whom felt that "it has consumed work hours and taken away time from patient care." The participants in PIRHA describe that the RPA robot should be able to handle 50 percent of the referrals to be beneficial. For example, one participant noted a success rate of only 27 percent during a two-week test period. Some other interviewed secretaries and nurses share similar experiences of automation feeling like an "additional burden." *"There is still a substantial amount of manual checking required. In the last report, more than 3/4 of cases had errors,"* describes, for example, another participant. In HUS, while the intelligent referral sorter appears to be more effective than the regular RPA solution in PIRHA, it has also created some extra work for the doctors. However, on the organizational level, the overall effect has been positive: *"The doctors don't see the advantage yet in the referral handling work. However, when I try to consider it from the perspective of the clinic's operation, I see that it creates a positive impact"*, describes one participant in HUS.

While automation increases quantity, it does not necessarily increase quality. Participants underline that humans still perform tasks with higher accuracy than robots. Thus, there appears to be a trade-off between quality and efficiency: *"I don't believe it has improved quality. But it has shifted mechanical work away from humans,"* concludes one doctor. *"The process is less precise; errors occur more frequently than with a human. However, the robot performs faster and around the clock. That's the trade-off,"* continues one external expert.

Lastly, interpreting data with changing inputs appears to be challenging. Several participants described that the dosage builder struggles with situations where the medication dosage changes within a time period. *"Often if there is a variable dosage, like one tablet in the morning and two in the evening… It cannot handle such situations properly, so it just makes a guess,"* describes one pharmacist. Thus, to maintain patient safety, the outputs need to be verified carefully by a professional. Moreover, the referral sorter in both case organizations has had issues interpreting complex data, for instance, sending the patient to a clinic outside their regular municipality.

4.3 Challenges Measuring the Impact of Automation

Measuring the impact of automation initiatives is perceived hard in both case organizations. The main assessment challenges relate to a lack of holistic data, baseline measurements, targets, and a regular evaluation process.

Active evaluation and monitoring have been missing from many implementations in both organizations. *"Not in any way, at least not in our unit,"* is how one participant describes the current state. *"I don't know if there's any monitoring at the*

PIRHA level, but at the unit level, there isn't," continues another employee.

The assessment has mainly focused on whether the technology works as intended, not on how people interact with it or how it improves outcomes: *"We have tried to gather genuine user feedback, but it has not been very successful. Essentially, what we document is the result of encountering bugs."* Evaluating the impact of automation is also challenging because it's "just a small part of it all," making it unclear which outcomes are directly caused by automation. Feedback has mostly been technical error reports: *"We get these cold reports that tell us how many times the robot has run and how many cases there are per month or week."*

Interviewees in both organizations note that evaluating progress is difficult without a baseline for comparison. *"We didn't do measurements before the start of this referral processing, such as how much time we spend now and how much we used to spend. Therefore, it's a bit challenging to assess time savings,"* describes one secretary. In some cases, potential benefits have been calculated in advance without collecting data afterward. *We haven't had any active monitoring. It's been more like we've calculated in advance the potential benefit and how quickly the process would pay for itself,"* summarizes one digitalization expert.

While participants acknowledge the time savings, it seems unclear how the saved time is utilized. *"What would interest me is somehow measuring how the saved work time is being used. Whether it means being able to serve one more patient or making more phone calls,"* describes one digitalization expert. Measuring time savings appears challenging because automation replaces only small specific tasks, not entire workflows. "*Measuring how much time various small tasks take is always challenging,"* as one digitalization expert describes. Secondly, assessment efforts would require an additional layer of monitoring, which would be difficult both technically and due to possible resistance: *"Adding extra monitoring to the busy daily work might not be the most effective solution."*

In some units, monitoring is seen as a sensitive issue. The work culture in these units seems to affect how openly the benefits of automation are discussed: *"If you have a difficult work atmosphere, you don't go tell your boss you have more bandwidth. You enjoy the fact that you have more space to do things. You don't report that you saved another 5 hours of work time this week,"* as one participant describes. However, collecting impact data could help in motivating employees to use the solutions: *It would likely turn even sceptical individuals towards a more positive outlook.*

4.4 Evaluation Metrics

Process metrics like time savings, throughput volumes, and error rates are frequently mentioned, especially those affecting treatment delivery, such as handled referrals for urgent patients. At HUS, greater emphasis is placed on qualitative metrics like worker wellbeing and satisfaction, reflecting their more mature solutions. In contrast, PIRHA focuses more on technical aspects, as their solutions are in earlier stages of development.

Back-office automation benefits patients indirectly, such as faster referral processing or clearer prescription instructions. Key metrics from the patient's perspective include patient safety, satisfaction, treatment efficiency, and service accessibility. Pharmacists and nurses using the medical dosage builder particularly emphasize patient safety as a crucial metric. Regarding the referral sorter, the main patient risk is treatment delays due to incorrect classification: *"The biggest risk we've identified is a delay in treatment due to incorrect classification. However, the risk is very low, and the consequences should not be significant."*

Measuring cost efficiency appears challenging, particularly due to difficulties in quantifying time savings. Cost savings may differ across units depending on whose tasks are automated. For instance, automating a doctor's tasks is more valuable than a secretary's. However, organizations can estimate time gains by "converting the saved time into work hours and the hourly rate," as one participant summarizes. Additionally, automation enhances organizational capabilities by increasing standardization and simplifying scalability. This organizational impact could be measured by metrics like the number of standardized processes, speed of implementation, and the number of employees trained in automation.

5 DISCUSSION

The results emphasize the need for an impact assessment model, as no proper "template or tool" exists. Despite some used metrics, employees found the value of implemented solutions unclear: *"The robot works well, but very few feel it's needed in their process."* In some cases, not seeing the benefits has also caused resistance to adoption: *"Some employees* *feel that they don't want to adopt automation because they perceive it as an additional burden."*

Based on both PIRHA and HUS interviews, evaluation has been inadequate because a process has not been built around it. Essentially, assessing digital transformation should answer two questions: *is the project doing things right* and *is the project doing the right things* (Pritchett et al., 2013)

More concretely, impact assessment includes two steps: monitoring implementation activities and evaluating the monitored outcomes. *Monitoring*, which refers to routine data collection, review, and analysis, is the most time-consuming part of the process (WHO, 2016). *Evaluation*, the systematic and objective assessment of the implemented solutions, is only the final step in determining whether objectives have been met and the impact has been achieved (WHO, 2016). In other words, evaluation aims to determine whether changes in the monitored metrics are the result of the digital intervention.

The findings from the two case organizations emphasize six key steps in impact assessment: 1) defining clear objectives, 2) choosing suitable metrics, 3) setting baselines and targets, 4) implementing monitoring mechanisms, 5) gathering feedback, and 6) conducting regular evaluations.

Firstly, when setting goals, it is essential to consider both the potential direct and indirect benefits of automation. In HUS and PIRHA, the primary advantages include operational efficiencies such as time savings and productivity boosts, while indirect benefits include reduced costs, improved transparency, and enhanced experiences for patients and employees. Automation outcomes that improve the health system's effectiveness are mostly indirect second-order effects, such as freeing up time for patient interaction or more complex administrative work.

The maturity of the solution should also be considered for in the evaluation (WHO, 2016). In other words, consider if the automation solution is developed and evaluated for the first time or if it is undergoing scale-up. For instance, PIRHA's newer RPA solutions had more errors than HUS's mature systems, which requires evaluating them differently depending on their development stage.

Moreover, not all metrics are relevant to every context. More specifically, the impact can vary based on the context in which they are implemented. Thus, a full acknowledgment of the different perspectives and boundaries is necessary (Williams 2015). More precisely, organizations must broadly consider their specific stakeholders' perspectives: what goals are relevant for whom and what is needed to measure. Automation goals may not be achieved if stakeholders are not engaged in setting the objectives (Zayas-Caban et al., 2021).

Baseline measurements are essential for setting realistic targets. Both PIRHA and HUS faced challenges due to a lack of baseline data, as no initial measurements were taken before implementation. Inconsistent monitoring was also a major obstacle. While technical data was collected, qualitative insights have been lacking. Employees stressed the need for gathering feedback to fully understand automation's impact. Collecting feedback is important, as unanswered quality or safety concerns can undermine the long-term success of automation initiatives (Zayas-Caban et al., 2021).

Lastly, evaluation should occur regularly and be based on evidence collected at across multiple time points (WHO, 2016). While interviews identified cost efficiency and resource utilization as key factors for evaluation, both organizations would benefit from considering other dimensions. Similarly, Axmann and Harmoko (2024) argue that traditional cost-benefit analysis overlooks many automation benefits, possibly leading to poor decisions. They propose a balanced scorecard (BSC) framework to evaluate RPA projects, categorizing benefits into four areas: financial, process improvement, customer satisfaction, and learning. With adjustments, the model could potentially be applied to healthcare. The BSC, originally developed by Kaplan and Norton (1992), has already been used in healthcare, both with original and modified perspectives (Amer et al., 2022; Betto et al., 2022). Further research is needed to determine whether BSC framework could effectively evaluate healthcare automation outcomes.

6 CONCLUSIONS

This research explored the impact of automation solutions in two Finnish healthcare districts, providing insights into the benefits, risks, and evaluation of automation initiatives in healthcare. However, several limitations must be acknowledged. Firstly, this study focused only on two regions with varying levels of automation maturity, which may limit the generalizability of the findings to other healthcare environments. Additionally, the research primarily gathered insights from employees familiar with automation without including patients. Although participants were asked to reflect on the impact of automation on patients, no direct feedback was collected from patients themselves.

The findings contribute to future research on developing an impact assessment framework that could help healthcare organizations better understand and enhance the use of automation. Future research should explore whether successful evaluation frameworks from other industries can be adapted for the healthcare sector.

REFERENCES

Adolfsson, F. (2021). Västerbotten sparar 100 000 timmar med RPA. *Voister*.

Ahlskog, B. (2022). Ohjelmistorobotit ja tekoäly lähetekäsittelyn apuna. *Husari*, 4/2022.

Amer, F., Hammoud, S., Khatatbeh, H., Lohner, S., Boncz, I., & Endrei, D. (2022). A systematic review: the dimensions to evaluate health care performance and an implication during the pandemic. *BMC Health Services Research*, 22(1), 621.

Antwi, S. K., & Hamza, K. (2015). Qualitative and quantitative research paradigms in business research: A philosophical reflection. E*uropean journal of business and management*, 7(3), 217-225.

Andrade C. (2021). The Inconvenient Truth About Convenience and Purposive Samples. *Indian journal of psychological medicine*, 43(1), 86–88.

Axmann, B., & Harmoko, H. (2021). The five dimensions of digital technology assessment with the focus on robotic process automation (RPA). *Tehnički glasnik*, 15(2), 267-274.

Betto, F., Sardi, A., Garengo, P., & Sorano, E. (2022). The evolution of balanced scorecard in healthcare: A systematic review of its design, implementation, use, and review. *International Journal of Environmental Research and Public Health*, 19(16), 10291.

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, 29(4), 101600.

Cresswell, K., Anderson, S., Montgomery, C., Weir, C. J., Atter, M., & Williams, R. (2023). Evaluation of Digitalisation in Healthcare and the Quantification of the "Unmeasurable." *Journal of General Internal Medicine*, 38(16), 3610-3615.

Denagama, V., Manuraji, I., Bandara, W., Syed, R., & Toman, D. (2020). An empirically supported conceptualisation of robotic process automation (RPA) benefits. *Association for Information System*, Marrakech.

Doğuç, Özge (2021). "Robotic process automation (RPA) applications in COVID-19". In: Dincer, H. & Yüksel, S. (Eds.) *Management Strategies to Survive in a Competitive Environment: How to Improve Company Performance* (pp. 233–247). Springer.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1), 1-4.

Ferris, T., Ackers, J., and Borhani, M. (2022). RPA in the NHS: Guidance for designing, delivering and sustaining RPA within the NHS. May. NHS.

Garcia, J.F, Spatharou, A., Hieronimus, S., Beck, J.P., and Jenkins, J. (2020). Transforming healthcare with AI. *McKinsey & Company: Our Insights.*

Herm, L. V., Janiesch, C., Helm, A., Imgrund, F., Hofmann, A., & Winkelmann, A. (2023). A framework for implementing robotic process automation projects. *Information Systems and e-Business Management*, 21(1), 1-35.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.

HUS (2021). Tarkastuslautakunnan arviointikertomus 2021 Helsingin ja Uudenmaan sairaanhoitopiirin valtuustolle. April. HUS.

Ikonen, T (2019). Asiakasryhmäkohtainen tieto laadusta ja vaikuttavuudesta sosiaali- ja terveydenhuollon tiedolla johtamisessa ja ohjauksessa. *Sosiaali-ja terveysministeriön raportteja ja muistioita,* 70.

Ivančić, L., Suša Vugec, D., & Bosilj Vukšić, V. (2019). Robotic process automation: systematic literature review. In *Business Process Management: Blockchain and Central and Eastern Europe Forum: BPM 2019 Blockchain and CEE Forum, Vienna, Austria, September 1–6, 2019, Proceedings 17* (pp. 280-295). Springer International Publishing.

Jain, R and R Bhatnagar (2019). Robotic process automation in healthcare- a review. *International Robotics & Automation Journal* 5(1), 12–14.

Kääriäinen, J., Aihkisalo, T., Halén, M., Holmström, H., Jurmu, P., Matinmikko, T., Seppälä, T., Tihinen, M., and Tirronen, J., (2018). Ohjelmistorobotiikka ja tekoäly–soveltamisen askelmerkkejä. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja,* 65/2018.

Kaplan, R. S. (1992). The Balanced Scorecard-Measures That Drive Performance. *Harvard Business Review*.

Karunasena, K., & Deng, H. (2012). Critical factors for evaluating the public value of e-government in Sri Lanka. *Government information quarterly*, 29(1), 76-84.

Kedziora, D., and Smolander, K., 2022: Responding to Healthcare Emergency Outbreak of COVID-19 Pandemic with Robotic Process Automation (RPA), P*roceedings of the 55th Hawaii International Conference on System Sciences*, 6207-6216

Kedziora, D., & Hyrynsalmi, S. (2023, May). Turning robotic process automation onto intelligent automation with machine learning. *Proceedings of the 11th International Conference on Communities and Technologies*, 1-5

Kedziora, D., & Kiviranta, H. (2018). Digital Business Value Creation with Robotic Process Automation (rpa) in Northern and Central Europe. *Management (18544223)*, 13(2), 64- 174.

Kim, S.H., (2023). Development of Evaluation Criteria for Robotic Process Automation (RPA) Solution Selection. *Electronics* 12(4), 986.

Lillrank, P., Tenhunen, H., Hörhammer, I., Halminen, O., Lyly, T., Linna, M., Silander, K., Laurila, R., Hiltunen, A-M, Riikonen, E., Miettinen, S., Tanila, T., Chen, A. & Vesinurm, M. (2019). Terveydenhuollon digitaalisten ratkaisujen vaikuttavuuden osoittaminen. HEMA-instituutti.

Meironke, A., & Kuehnel, S. (2022). How to measure RPA's benefits. A review on metrics, indicators, and evaluation methods of RPA benefit assessment. 1*7th International Conference on Wirtschaftsinformatik.*

Mestres, J. C. (2017). State of the art of health care: The cubism period. The digitization of healthcare: New challenges and opportunities, 3-21.

Ng, K. K., Chen, C. H., Lee, C. K., Jiao, J. R., & Yang, Z. X. (2021). A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives. *Advanced Engineering Informatics*, 47 (101246).

Osmundsen, K., Iden, O. and Bygstad, B. (2019). Organizing Robotic Process Automation: Balancing Loose and Tight Coupling. *Proceedings of the 52nd Hawaii International Conference on System Sciences 2019*. 6918-6926.

Parviainen, P., Kääriäinen, J., Honkatukia, J., and Federley, M., (2017). Julkishallinnon digitalisaatio–tuottavuus ja hyötyjen mittaaminen. *Valtioneuvoston selvitysja tutkimustoiminnan julkaisusarja,* 3/2017.

Pitkänen, L., Haavisto, I., Vähäviita, P., Torkki, P., Leskelä, L., & Komssi, V. (2018). Vaikuttavuus SOTE:ssa. Suoritteista tuloksiin. *Nordic Healthcare Group NHG.*

Pritchett, L., Samji, S., & Hammer, J. S. (2013). It's all about MeE: Using Structured Experiential Learning ('e') to crawl the design space. *Center for Global Development working paper*, (322).

Ratia, M., Myllärniemi, J. and Helander, N. (2018). Robotic process automation-creating value by digitalizing work in the private healthcare? *Proceedings of the 22nd International Academic Mindtrek Conference*, 222-227

Ratia, M., Myllärniemi, J. and Helander, N. (2021). Intellectual Capital and Robotic Process Automation (RPA)-a Capability Perspective. *International Forum on Knowledge Asset Dynamics*, 1810-1824

Saunders, M., Lewis, P., Thornhill, A., and Bristow, A., (2019). "Research Methods for Business Students" Chapter 4: Understanding research philosophy and approaches to theory development". In: *Research Methods for Business Students*. Pearson Education, pp. 128–171.

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. B*MC medical informatics and decision making*, 21, 1-23.

Sintonen, H., Blom, M., Roine, R. P., Räsänen, P., & Ryynänen, O. P. (2021). Terveystaloustiede on tärkeää lääkärille. *Lääkärilehti*, 76(39)

Sony, M., Antony, J., & Tortorella, G. L. (2023). Critical success factors for successful implementation of healthcare 4.0: A literature review and future research agenda. *International Journal of Environmental Research and Public Health*, 20(5), 4669.

Syed R, Suriadi S, Adams M, Bandara W, Leemans SJ, Ouyang C., Hofstede AH, van de Weerd I, Wynn MT, Reijers HA (2020). Robotic process automation: contemporary themes and challenges. *Computers in Industry*, 115, 103162.

Vila-Henninger, L., Dupuy, C., Van Ingelgom, V., Caprioli, M., Teuber, F., Pennetreau, D., Bussi, M., & Le Gall, C. (2024). Abductive coding: Theory building and qualitative (re) analysis. *Sociological Methods & Research*, 53(2), 968-1001.

Wanner, J., Hofmann, A., Fischer, M., Imgrund, F., Janiesch, C., & Geyer- Klingeberg, J. (2019). Process selection in RPA projects-towards a quantifiable method of decision making. *40th International Conference on Information Systems, Munich*.

WHO (2016). Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. *World Health Organization: Publications*.

Willcocks, L., Lacity, M. and Craig, A. (2015a). The IT Function and Robotic Process Automation. The Outsourcing Unit Working Research Paper Series. 15(05), 1-39.

Willcocks, L., Lacity, M. and Craig, A. (2015b). Robotic Automation at Xchanging. The Outsourcing *Unit Working Research Paper Series,* 15(03), 1-26.

Willcocks, L.P., Lacity, M.C., Craig, A. (2017). Robotic Process Automation: Strategic Transformation Lever for Global Business Services? *Journal of Information Technology Teaching Cases*. 7, 17–28

Williams, Bob (2015). "Prosaic or profound? The adoption of systems ideas by impact evaluation". *IDS Bulletin* 46(1), 7–16.

Yin, R. K. (2018). *Case study research and applications: Design and Methods*. (6th ed.). Sage Publication.

Zayas-Cabán, T., Okubo, T., and Posnack, S., (2023). Priorities to accelerate workflow automation in health care. *Journal of the American Medical Informatics Association* 30(1), 195–201.

Knowledge Management in Civil Protection at the Example of Fire Brigades

Andreas Schultz[a], Fabian Dotzki[b] and Iryna Mozgova[c]
Data Management in Mechanical Engineering, Faculty for Mechanical Engineering, Paderborn University, Germany
{*andreas.schultz, iryna.mozgova*}*@uni-paderborn.de*

Keywords: Knowledge Management, Civil Protection, Systematic Literature Review, Fire Brigade.

Abstract: Knowledge management is essential for successful disaster management. This paper conducts a Systematic Literature Review at the intersection of the knowledge management field and disaster management and examines the available body of literature. Fire departments are chosen as the focus group as they are the most prevalent emergency services. There are many publications that deal with knowledge management during the response phase of an emergency. Often, the literature focuses on the application of knowledge management in large-scale disasters to link the various organizations on-scene. What is missing in most approaches is a prior step of implementing and training the knowledge management system. Therefore, this literature review seeks to provide an overview of approaches for daily routines and small-to-medium incidents that serve as a training ground. However, literature on non-incident phases and smaller incidents is scarce. As information technologies are developing rapidly, there is no modern and recent description of the current use of knowledge management solutions in this area.

1 INTRODUCTION

Growing administrative regulations, new high-tech equipment and changing environmental circumstances constantly increase the necessary information and knowledge that civil protection personnel need to know (Schultz et al., 2024b), (Weidinger et al., 2021). This knowledge about processes, techniques and equipment has to be stored adaptively so that it is easy to learn new things, refresh familiar ones and be aware of changes (Oktari et al., 2020). Existing knowledge is rarely stored in a structured way, so new or promoted civil protection members have to retrieve information from their experienced colleagues or documents (Schultz et al., 2024b). It is rarely possible to inform oneself or look up information from one central source. These circumstances were the trigger for the examination of the existing literature body concerning knowledge management (KM) solutions within civil protection (CP) by performing a Systematic Literature Review (SLR), according to (Xiao and Watson, 2017). This paper focuses on fire departments as one of the most widespread emergency services (Brushlinsky et al., 2019). The remaining paper is structured as follows: Chapter 2 contains a brief introduction to the fields of CP and KM, followed by a discussion on the application of KM within CP and its characteristics and particularities. Chapter 3 describes the application of the SLR. The organizational aspects are outlined and then categories are derived into which the resulting publications can be classified. Chapter 4 discusses the findings from the SLR in detail and identifies the research gaps. Chapter 5 concludes the work, gives a summarizing statement and outlines subsequent steps.

2 FOUNDATION

The preface gives a brief overview of the theoretical principles of information and knowledge management in section 2.1. The heterogeneity of CP systems is presented in section 2.2 in order to illustrate differences between countries or regional authorities. Section 2.3 links the two topics KM and fire brigades and outlines the characteristics of KM applied in the structures of fire brigades.

[a] https://orcid.org/0009-0006-0213-1663
[b] https://orcid.org/0009-0000-8940-0497
[c] https://orcid.org/0000-0002-6761-0220

Schultz, A., Dotzki, F. and Mozgova, I.
Knowledge Management in Civil Protection at the Example of Fire Brigades.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 256-264
ISBN: 978-989-758-716-0; ISSN: 2184-3228

2.1 Knowledge Management

KM is a key success factor and a necessity of many businesses due to the increasing knowledge bases, the need for a competitive advantage or internationalization and managing of e.g. staff turnover (International Organization for Standardization, 2018), (Omerzel and Gulev, 2011). The ISO norm 30401 describes KM as follows: "Knowledge management is a holistic approach to improve learning and effectiveness through optimization of the use of knowledge, in order to create value for the organization. Knowledge management supports existing process and development strategies" (International Organization for Standardization, 2018). Most literature related to KM uses the building blocks of knowledge management by (Probst et al., 2012) as a basis for describing KM processes (e.g. (North, 2021), (Oliveira and Pinheiro, 2021)). The building blocks by North consist of six key components: Knowledge identification, acquisition, development, distribution, utilization, and preservation.

The Association of German Engineers (Verein Deutscher Ingenieure (VDI)) has issued a guideline for KM in the engineering sector that describes the basic concepts in VDI-5610:2009 (VDI Verein Deutscher Ingenieure e.V., 2009). The guideline aims at advising businesses on introducing a knowledge management system (KMS) and distinguishes the terms data, information and knowledge as concepts that build on each other. Two different types of knowledge are differentiated: implicit and explicit knowledge. Implicit knowledge is "bound to persons, difficult to communicate and hardly to formalize", whereas explicit knowledge "can be formalized on different levels (e.g. speech, writing); it is therefore communicable and storable in various media" (International Organization for Standardization, 2018). The process of KM in (VDI Verein Deutscher Ingenieure e.V., 2009) (depicted in Figure 1) corresponds with the building blocks described by (Probst et al., 2012).

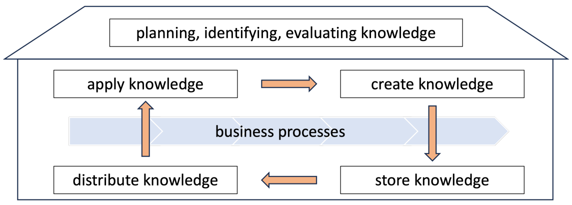


Figure 1: Knowledge management process according to (VDI Verein Deutscher Ingenieure e.V., 2009).

In summary, it can be stated that the presented scientific foundations (North, 2021), and (Probst et al., 2012) have their origin in organizational and especially business-oriented research domains. They still form the scientific basis for KM in recent literature. Also the presented norms (International Organization for Standardization, 2018) and (VDI Verein Deutscher Ingenieure e.V., 2009) focus on fostering KM in businesses. Nevertheless, these concepts are versatile and are being used in other areas as business-oriented KM as well.

2.2 Civil Protection

CP is used as a summary for all activities that aim at protecting people in a certain area from natural (e.g., earthquake, drought flood or hurricane) or man-made (e.g., technical failure, terror) hazards (Waugh and Tierney, 2007). Among others the following organizations comprise the CP: emergency services (fire brigades, emergency medical response/ ambulance, and police), public administration (municipal, region/county/provincial/national etc.) and also technical/scientific services (Waugh and Tierney, 2007). It is organized differently from country to country (cf. (Alexander, 2010) and (Zambrini et al., 2020)) and thus structures and prerequisites differ. Setting the focus on fire brigades, many countries maintain a system with mostly professional (paid) firefighters, such as Italy or Great Britain (Alexander, 2010). On the contrary, Germany, Austria or Switzerland and other Western European countries like Poland or the Netherlands, among others, maintain a firefighting system that heavily relies on volunteer firefighters (whereas professional firefighters are also present, mostly in bigger cities) (Brushlinsky et al., 2019), (Zambrini et al., 2020). The procedure of CP organizations can be categorized into different phases related to the central disaster. The disaster management circle (Inan et al., 2018) (Haddow et al., 2021) is used in multiple variations whereas the core steps are always similar: Prevention, Preparedness, (Disaster/Impact), Response and Rehabilitation. The transition between two phases is not necessarily a certain point but rather a gradual change.

The tasks of the various CP organizations are diverse. As mentioned at the beginning of this section, they can be categorized into emergency services, public administration and technical/scientific services. Each of them has its own field of responsibility, which may vary across borders. But all of them need to evolve and have to adapt to new technologies and procedures.

In recent years, various IT-solutions have been developed to support emergency operations. The examined technologies range from technical solutions such as drones and exoskeletons (Gottschalk, 2019), emergency response information systems (Weidinger et al., 2021) to the digital interconnection of units with the help of IT (Spaling et al., 2018). The introduction of new technologies is also linked to the need to adapt existing procedures to new circumstances. This may lead to challenges, that emergency responders face while using IT in incidents and are presented by Elmasllari who identifies twelve problems that arise when using IT for disaster management (Elmasllari, 2018) which are excerpted below:

- Reliability: Digital systems have to function when they are used in an emergency.
- Interoperability: IT-based tools must be able to exchange data and information with other systems and thus provide interfaces for import and export. The usage of IT restricts the user to the means of communication provided by the system.
- The price of structured data: Data, information and knowledge that should be stored in a system need to be expressible and explicit to be able to store it in a digital system. The graphical user interface limits the possibilities to those intended by the programmers.

2.3 Knowledge Management in Civil Protection

This section combines the two aforementioned topics KM and CP and outlines the particularities of KM in CP. As already mentioned in Section 2.1, all of the KM concepts and the norms ISO 30401 and VDI 5610 focus on the operational context of enterprises. The special character of (volunteer) CP is not reflected directly in the KM concepts (Lee et al., 2011). In addition, the mostly volunteer staff often does not have the time and money (as companies do) to deal with KM and implement it as a business strategy (Oliveira and Pinheiro, 2019), (Alexander, 2010). This leads to an underrepresented usage of KM in CP while the tools and techniques do exist for enterprises. Furthermore, the amount of knowledge required for successful disaster management is already enormous and vast, yet much of it is tacit and drawn as expertise from previous experience (Oliveira and Pinheiro, 2019), (Oktari et al., 2020) (Klein, 2017). CP organizations are moreover mostly run or organized by the state or public bodies, which are not run as businesses with a profit intention but are funded by the taxes allocated to them (Waugh and Tierney, 2007). This is not necessarily a downside, as there is hardly any competitive situation among two units (Omerzel and Gulev, 2011); the willingness to collaborate across unit borders is way higher than it is among companies (Schultz et al., 2024b). The particularities of KM in CP compared to the enterprise context is, among others, the following: The knowledge has to be available more or less immediately in crisis situation since the disaster and its effects do not allow any delay (Elmasllari, 2018). Furthermore, the need to share knowledge between units is greater than between companies. Therefore, specific requisites must be elaborated for KM systems which shall be used for emergency services that respond to an event.

There are already several digital solutions to support one or another dimension of knowledge. Information Management Systems (IMS) are used as administrative support systems for keeping track of incident reports, personal member data and equipment history (Schultz et al., 2024b). KMS are defined as a part of management systems related to knowledge (International Organization for Standardization, 2018) and are rarely used in CP (Schultz et al., 2024b). There are only few documentations about the implementation and usage. Two German examples are the Einsatzleiterwiki[1] (incident commander wiki) or the web-based portal BKS-Portal[2] for CP personnel in the state of Rhineland-Palatinate. The first one provides basic information sorted by keywords or alarm patterns, whereas the latter serves in the state of Rhineland-Palatinate as a central platform for exchanging information and knowledge. Another approach describes the implementation of a KM infrastructure in New South Wales, Australia (Pickles, 2004). It is implemented by a content management system for managing web-content and e.g. a document management system. The application of KM in CP has been a research topic for many years as the existing literature reviews show (Fauzi et al., 2024), (Anand et al., 2022), (Oktari et al., 2020), and (Dorasamy et al., 2013). Nevertheless, the implementation of successful KMS is still a topic of ongoing research and development. The fire brigades have an important role in the CP ecosystem as they are involved in large scale disasters as well as small and daily occurring incidents (Weidinger, 2022), (Kapalo et al., 2019) and are usually among the first organizations on-scene. Which leads to the intention of this paper to examine the literature available with a special focus on fire brigades and identify research needs and gaps in the literature body.

[1] https://einsatzleiterwiki.de/

[2] https://bks-portal.rlp.de/

3 SYSTEMATIC LITERATURE REVIEW

The process of the SLR, as proposed by (Xiao and Watson, 2017), aims to capture and evaluate the current status of a certain research field. It consists of eight stages, which are grouped into the following three phases: Planning, conducting and reporting. The literature review process can be partially performed iteratively in order to deal with unforeseen issues and to amend the search with extended keywords. The remainder of this chapter is structured following the SLR process proposal and depicted in Figure 2.

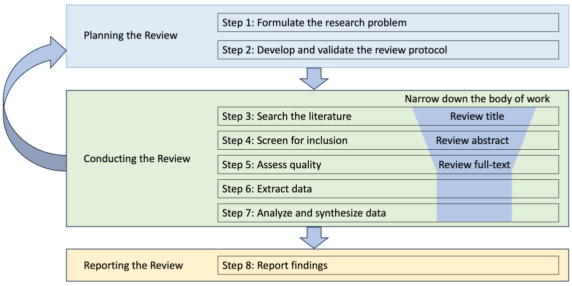


Figure 2: Systematic Literature Review process based on (Xiao and Watson, 2017).

3.1 Formulate the Research Problem

As mentioned before, the literature body concerning KM in CP, and in particular in the fire brigades, is not very mature and scattered. There are many publications on KM in business environments. A decent amount of publications focus on KM in CP but the focus is mostly on large-scale disasters like earthquakes, floods, etc. These few existing publications (e.g., (Inan et al., 2018), (Pickles, 2004) and other literature reviews (Fauzi et al., 2024), (Anand et al., 2022), (Oktari et al., 2020), and (Dorasamy et al., 2013)) mostly deal with one or several phases of the disaster management circle, examining the application of KM, e.g., in the response phase. The application and examination of KM in this article not only focuses on the typical disaster management circle phases but is extended to include supporting tasks that members of disaster management organizations perform like training, administrative work and other processes that can be supported by a KMS (Schultz et al., 2024b). This is intended to improve the attitude and capability of users of KMS to be able to handle a KMS during an incident in a timely and serious environment and not be overwhelmed by the additional IT system (Weidinger et al., 2021), (Elmasllari, 2018).

3.2 Develop and Validate the Review Protocol

Therefore, this literature review focuses on applications of KM in fire brigades. The literature should be analyzed with regards to the different disaster phases and solutions that might be applicable in daily routines as well. Differing prerequisites and requirements by professional and voluntary firefighters (c.f., (Alexander, 2010)) should be considered. For example, voluntary firefighters have a primary job and training, exercises, and administration take place in their free time. The research is driven by the following research questions (RQ):

- Which literature is available concerning knowledge management within civil protection (focusing on fire brigades) and to which phases of the disaster management circle does it refer to? (RQ1)
- How is the implementation and training of knowledge management systems performed, are there differences between professional and voluntary personnel, and is knowledge management applied in daily routines (non-incident times)? (RQ2)
- What digital knowledge management systems exist that are being used in civil protection? (RQ3)

The SLR is conducted as a desk study with the four-eyes principle by two researchers to minimize personal bias. Only online sources in English or German are included in the literature list. The search was performed during February and April 2023, with an update search shortly before submission in May 2024.

3.3 Search the Literature

Google Scholar[3] and Scopus[4] were used as the initial sources for publications by applying different combinations of search strings. The search was also performed as a backward search by examining the references of the initially found publications for other relevant literature. To find papers that already cited the relevant publications, the forward search by Google Scholar and Scopus was used. Various keywords were derived from the research questions and formulated as a boolean expression to search the databases. Following the initial review of the initial publications, an expansion of the search term was conducted to include additional relevant keywords. In order to cover a period of 20 years, publications from 2003 onward were considered further. Older literature was

[3]https://scholar.google.com/
[4]https://www.scopus.com/

excluded since the focus is on KM and its digital supporting tools. A wider scope does not make sense given the exponential evolution of technology. The second search iteration, which included both backward and forward searches, yielded 390 publications that met the search criteria. The search was terminated in April 2023 after the search pattern was fully exhausted. Repeating the search immediately before submission yielded two further results. Finally there were 392 publications given to the SLR process.

3.4 Screen for Inclusion

Subsequently, a two-step process was carried out to review the papers for inclusion in this review. The first step was to examine the title to determine whether the publication fits into the desired set. If the title was not sufficiently informative, the abstract of the work was taken into account. Each of the two researchers independently rated all of the 392 publications with the labels “include”, “discuss” and “discard”. Afterwards, both ratings were merged: papers marked by one or both with “include” were included (62), the “discuss” papers (if one or both marked a paper as “discuss”) were examined in a joint meeting to determine whether the paper was relevant or not (76), and all papers marked by both researchers with “discard” were discarded from the literature body (254). A subsequent backward and forward search yielded further publications (9), which were assessed for inclusion in the literature inventory and then added. Combined with the outcome of the ”discuss”-marked papers, this resulted in a total of 138 publications, which were used as the basis for the in-depth analysis.

3.5 Assess Quality

The second step of the quality assessment process is the review of the entire publication. There was no access to the full text of 11 publications through the provided library and other sources. In addition, 47 papers were excluded as not covering both key topics, two papers were found to be duplicate and another 36 papers were considered inappropriate for the research questions and were hence discarded. A total of 42 papers were considered for further analysis.

3.6 Extracting Data

By applying the inductive coding concept, publications were evaluated for inclusion in this paper. The publications were assigned to different categories. The results’ overview of the literature review can be found in the Open Research Knowledge Graph (Schultz et al., 2024a), where the categories for each publication are aligned to the ones of the literature review.

3.7 Analyze and Synthesize Data

Four overviews and reviews throughout the past 20 years were identified that are relevant for the research goal. Starting with the analysis by (Dorasamy et al., 2013), they already pointed out the need for unified terminologies. The stated need for a better understanding of determinants for KMS success factors is already partially addressed by other publications (e.g., (Seneviratne et al., 2010)). The research gap of what is missing in KM research towards a theoretical background is addressed by (Oktari et al., 2020). The investigation by (Anand et al., 2022) on KM in crisis gives a suitable overview of the literature according to the disaster management circle. The most recent review paper by (Fauzi et al., 2024) takes a bibliographic perspective and categorizes the literature body into different clusters, which should point out current and upcoming trends. No distinction is made between professional and voluntary personnel. Although the search queries are rather broad, the number of publications linking KM and CP is rather small and still partially related to management literature and strategies.

The response phase is covered by most publications that are related to the disaster management circle and there are only a few publications related to the other phases. The publication by (McNaughton and Rao, 2018) focuses on the application in the Caribbean region. Their approach is to implement a knowledge broker to share disaster information between Caribbean states. Mayor challenges are standardization and a coordinated production of knowledge and knowing what knowledge is where. The work is continued in (Rao and McNaughton, 2019). The proposed lessons-learned approach by (Rostis, 2007) gives input on how to integrate this source of knowledge into a KMS (Rostis, 2007). (Otim, 2006) chose the approach of case-based KM which relies on already captured knowledge from previous incidents, similar to Rostis approach. As a sole source of knowledge, case-based systems always face the problem of requiring a previous case that is comparable and stored in the system. Yet, it can provide guidance if cases occur more often, but then emergency personnel might already know the solution. The two following publications refer to mostly large-scale disasters like floods (Lee et al., 2011) or earthquakes (Cinque et al., 2015). They investigate the exchange of information during an event between different disaster management agencies and point out the need for information systems focusing on the public sector (Lee et al., 2011). Similarly, (Saoutal et al., 2015) cover the inter-organizational exchange of knowledge as well, augmented by elaboration on awareness issues and a proposed system architecture. The publication by (Seneviratne et al., 2010) covers disaster knowledge success factors which deal with a lessons-learned approach again, here in the recovery phase. The identification of technological, social, legal, environmental, economic, functional, institutional and political factors provide a good basis for developing and maintaining knowledge and information management systems.

The publications on models and meta-models seek to elaborate models that are able to represent the characteristics of this domain and provide means to describe activities or entities. This includes formal language models that are used as a foundation for human- and machine-readable descriptions. (Othman and Beydoun, 2010) apply a case-based reasoning approach, which accepts input queries and outputs a model fragment that should be suitable for the decision makers independently of the type of disaster. The results of the publication are four Unified Modeling Language (UML)-diagrams for each of the four disaster management circle phases. The UML-diagrams are validated with existing meta-models and augmented with missing items. This work is preparatory and continued in (Othman and Beydoun, 2016). (Franke et al., 2010) create a new description language, based on interviews and the Business Process Model and Notation (BPMN) to enable the depiction of temporal dependencies. It is supposed to be used within and between organizations and activities can be combined flexibly. A simplified process modeling language for disaster management is proposed by (Ziebermayr et al., 2011). It should help sharing the knowledge of experienced people. Unfortunately, no implementation is described in the publication. (Benaben et al., 2016) examine the information extraction from data flows for emergency management and the credibility of data sources. They propose a meta-model that helps transforming incoming data into information and knowledge subsequently. The core meta-model is surrounded by four packages for the domains context, partners, objectives and behavior (Benaben et al., 2016).

Only one relevant paper by (Oliveira et al., 2022) was found that deals with the sharing of tacit knowledge (the knowledge that is not explicitly written down but rather in the practitioners' heads (as experiences)). There are other publications in the business-related body of literature that deal with tacit knowledge, but in the field of emergency management, this is the only one that is relevant to the research questions. (Oliveira et al., 2022) elaborated on indicators for tacit knowledge sharing and measures to be implemented. Six excerpted indicators are: individual time management: whether practitioners have the time for sharing their knowledge; mutual confidence: firefighters need to trust each other; relationship network: who has which knowledge and is willing to share it with others; hierarchy: people with hierarchically higher positions need to allow access to their knowledge; knowledge storage: differentiation between storing explicit knowledge (in databases) and implicit knowledge in peoples' heads; power: having knowledge is perceived as a kind of dominance over others who lack knowledge.

Descriptions of practical implementations of KMS within the CP domain are rare. Only three scientific publications were found that cover KM solutions. (Pickles, 2004) describes a solution for the Australian New South Wales fire brigade to share information within and with other organizations. Unfortunately, no further publications could be retrieved for this solution. (Othman and Beydoun, 2016) is the continuation of (Othman and Beydoun, 2010) and describes a knowledge-based approach to structuring, storing and reusing disaster management knowledge. They describe a system architecture for a knowledge sharing system. Lastly, (Timm et al., 2013) cover a state-wide knowledge and collaboration platform which should be implemented by using open-source solutions to replace a commercial and costly solution. Except for the last one, the first two deal with KMS during or mainly related to an incident. Yet, the application and usage of KM systems in non-incident phases can provide users with an increased level of confidence.

4 DISCUSSION

The SLR revealed papers that are related to KM in the disaster management field. The scarce amount of recent publication shows that there is a need to research KM within disaster management to foster the usage of digital KM systems in that field. Studies on sharing knowledge in non-profit organizations are rare but do exist, although limited only to tacit knowledge (Oliveira and Pinheiro, 2021) or are focused on emergency response information system (Weidinger et al., 2021), (Alexander, 2010). One interesting approach is the development of a knowledge broker for Small Islands Developing States (SIDS) by McNaughton and Rao in (McNaughton and Rao, 2018) and (Rao and McNaughton, 2019). It deals with important aspects like avoiding knowledge silos, distributing information about who has which information and aligning organizations' vocabularies to gain a common understanding. Their use-case is on SIDS, but their assumptions are transferable to the fire brigades as well.

Concerning RQ1, this SLR revealed a decent amount of scientific literature dealing with KM in CP. The focus, although, is more on generic disaster management than especially on fire brigades (e.g. (Saoutal et al., 2015), (Oktari et al., 2020)). The response phase is covered by the majority of publications; the other phases are only covered by a few. The usage of a KM system in non-incident phases can help the users get familiar with the system; however, only a few publications could be found. None of them were suitable for the research goal (e.g., too specific on a location or circumstance). The conclusion that using a KMS in non-incident phases will improve user capabilities during a stressful and timely limited incident scenario cannot be drawn from the identified scarce literature. Also, KMS are mostly used in large-scale disasters with multiple agencies involved and many people affected. It is not described that practitioners use these systems already in small-to-medium incidents with manageable challenges.

The implementation in existing organizations and procedures as well as the training with KMS was addressed in RQ2. Unfortunately, the identified publications only refer to theoretical concepts (e.g. (Otim, 2006), (Rostis, 2007), (Franke et al., 2010)). Literature, describing a practical implementation with a validation, is scarce. One solution is presented by Cinque et al. in (Cinque et al., 2015). Their supporting platform aims at connecting various stakeholder organizations by introducing a common ontology and vocabulary. The actual implementation in existing procedures or the training is not part of their publication. Differences between voluntary and professional fire fighters are only addressed in a few publications that focus on the sharing of knowledge in volunteer or non-profit organizations (e.g. (Oliveira et al., 2022), or (Oliveira et al., 2022)). The integration of KM into daily routines to familiarize with these systems is not described by any of the identified publications. Partially, results from business sciences might be applicable for fire departments as well, up to the point where the organization switches from non-incident to an incident. There is an immense change of requirements and conditions taking place, which also affect the usage of a KMS (weather conditions, stress, limited time) (Weidinger et al., 2021), (Elmasllari, 2018).

Concerning RQ3, there are few publications describing concrete KM solutions and their applications, but these systems are either private and not accessible by the authors or in an old state and describe software that is or was used years ago. As information technologies are evolving quickly, a modern and recent description is lacking that describes the current usage of KM solutions in the field.

5 CONCLUSION

This literature review was carried out in order to analyze the intersection of KM and CP. KM in fire brigades is considered particularly relevant, as these organizations have to deal with different types of incident scenarios. Following a SLR based on (Xiao and Watson, 2017), the current state of the literature body on KM in CP was elaborated. It identified that the predominant focus is on KM during an incident to help practitioners in the response phase. There is not much literature covering non-incident phases. However, it can be argued that systems for everyday operations would be useful. This could help users to get acquainted with the system and eventually increase their confidence during stressful and time-sensitive incidents. It is not only necessary to have a KMS, but also the knowledge culture (cf. (International Organization for Standardization, 2018)) has to be embodied in an organization. As stated by Rao and McNaughton, the practitioners need to be aware of which information and knowledge exists and where it is stored. The maintenance of knowledge silos hampers the development of a shared understanding of knowledge that could be helpful for coping with incidents and disasters (Rao and McNaughton, 2019). The findings will be used in further research dealing with the development of KMS for fire brigades.

REFERENCES

Alexander, D. (2010). The voluntary sector in emergency response and civil protection: review and recommendations. *International Journal of Emergency Management*, 7(2).

Anand, A. et al. (2022). Exploring the role of knowledge management in contexts of crisis: a synthesis and way forward. *International Journal of Organizational Analysis*.

Benaben, F. et al. (2016). A metamodel for knowledge management in crisis management. volume Proceedings of the Annual Hawaii International Conference on System Sciences.

Brushlinsky, N. et al. (2019). World fire statistics. *Center for Fire Statistics of CTIF*.

Cinque, M. et al. (2015). A collaboration platform for data sharing among heterogeneous relief organizations for disaster management. volume Proceedings of the 12th ISCRAM Conference – 12th International Conference on Information Systems for Crisis Response and Management.

Dorasamy, M., Raman, M., and Kaliannan, M. (2013). Knowledge management systems in support of disasters management: A two decade review. *Technological Forecasting and Social Change*, 80(9).

Elmasllari, E. (2018). Why IT systems for emergency response get rejected: examining responders' attitude to IT. volume Proceedings of the 15th ISCRAM Conference – 15th International Conference on Information Systems for Crisis Response and Management.

Fauzi, M. A. et al. (2024). The application of knowledge management in disaster management: past, present and future trends.

Franke, J., Charoy, F., and Ulmer, C. (2010). A model for temporal coordination of disaster response activities. volume Proceedings of the 10th ISCRAM Conference – 10th International Conference on Information Systems for Crisis Response and Management.

Gottschalk, J. (2019). Digitalisierung im Einsatzdienst. *Technical report*.

Haddow, G. D., Bullock, J. A., and Coppola, D. P. (2021). *Introduction to Emergency Management*. Elsevier.

Inan, D. I., Beydoun, G., and Pradhan, B. (2018). Developing a decision support system for disaster management: Case study of an indonesia volcano eruption. *International Journal of Disaster Risk Reduction*, 31.

International Organization for Standardization (2018). ISO 30401:2018, Knowledge management systems — Requirements.

Kapalo, K. A., Wisniewski, P., and LaViola Jr., J. J. (2019). First in, left out : Current technological limitations from the perspective of fire engine companies. *Proceedings of the 16th ISCRAM Conference – 16th International Conference on Information Systems for Crisis Response and Management*.

Klein, G. A. (2017). *Sources of Power: How People Make Decisions*. The MIT Press.

Lee, J. et al. (2011). Group value and intention to use - a study of multi-agency disaster management information systems for public safety. *Decision Support Systems*, 50(2).

McNaughton, M. and Rao, L. (2018). Governing knowledge commons in caribbean disaster management: A comparative institutional analysis. *Information Services and Use*, 37(4).

North, K. (2021). *Wissensorientierte Unternehmensführung*. Springer Fachmedien Wiesbaden, 7 edition.

Oktari, R. et al. (2020). Knowledge management practices in disaster management: Systematic review. *International Journal of Disaster Risk Reduction*, 51.

Oliveira, M. et al. (2022). How to overcome barriers to sharing tacit knowledge in non-profit organizations? *Journal of the Knowledge Economy*, 13(3).

Oliveira, M. and Pinheiro, P. (2019). Factors in the sharing of tacit knowledge among volunteer firefighters. *International Journal of Business and Management*, 14(9).

Oliveira, M. and Pinheiro, P. (2021). Factors and Barriers to Tacit Knowledge Sharing in Non-Profit Organizations – a Case Study of Volunteer Firefighters in Portugal. *Journal of the Knowledge Economy*, 12(3).

Omerzel, D. G. and Gulev, R. E. (2011). Knowledge resources and competitive advantage. *Managing Global Transitions: International Research Journal*, 9(4).

Othman, S. and Beydoun, G. (2010). Metamodelling approach to support disaster management knowledge sharing.

Othman, S. and Beydoun, G. (2016). A metamodel-based knowledge sharing system for disaster management. *Expert Systems with Applications*, 63.

Otim, S. (2006). A case-based knowledge management system for disaster management: Fundamental concepts.

Pickles, A. (2004). A knowledge management infrastructure for the NSW Fire Brigades. *Australian Journal of Emergency Management*, 19(2).

Probst, G., Raub, S., and Romhardt, K. (2012). *Wissen managen*. Gabler Verlag.

Rao, L. and McNaughton, M. (2019). A knowledge broker for collaboration and sharing for sids: the case of comprehensive disaster management in the caribbean. *Information Technology for Development*, 25(1).

Rostis, A. (2007). Make no mistake: The effectiveness of the lessons-learned approach to emergency management in canada. *International Journal of Emergency Management*, 4(2).

Saoutal, A., Matta, N., and Cahier, J. (2015). How to pick up the needed information about what is going around us: Information awareness in crisis management. volume 3.

Schultz, A., Dotzki, F., and Mozgova, I. (2024a). Comparison of literature dealing with knowledge management in fire brigades.

Schultz, A., Dotzki, F., and Mozgova, I. (2024b). State of Knowledge Management among German firefighters. *Proceedings of the 21st ISCRAM Conference – 21th International Conference on Information Systems for Crisis Response and Management*.

Seneviratne, K., Baldry, D., and Pathirage, C. (2010). Disaster knowledge factors in managing disasters successfully. *International Journal of Strategic Property Management*, 14(4).

Spaling, G., Peters, R., and Wilson, F. (2018). Improving crisis response by interconnecting data worlds. *Electronic Journal of e-Government*, 16(2).

Timm, E., Alsbach, S., and Wimmer, M. A. (2013). Wissens- und kollaborationsplattform im brand- und katastrophenschutz: Open-source als kostengünstige option? *INFORMATIK 2013–Informatik angepasst an Mensch, Organisation und Umwelt*.

VDI Verein Deutscher Ingenieure e.V. (2009). VDI 5610 Blatt 1:2009-03 Wissensmanagement im Ingenieurwesen - Grundlagen, Konzepte, Vorgehen.

Waugh, W. L. and Tierney, K. J. (2007). Emergency management: Principles and practice for local government.

Weidinger, J. (2022). What is known and what remains unexplored: A review of the firefighter information technologies literature. *International Journal of Disaster Risk Reduction*.

Weidinger, J., Schlauderer, S., and Overhage, S. (2021). Information Technology to the Rescue? Explaining the Acceptance of Emergency Response Information Systems by Firefighters. *IEEE Transactions on Engineering Management*.

Xiao, Y. and Watson, M. (2017). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1).

Zambrini, F. et al. (2020). Comparing natural hazard and risk representations in a transboundary area to enhance civil protection international cooperation. *EGU General Assembly Conference Abstracts*.

Ziebermayr, T. et al. (2011). A proposal for the application of dynamic workflows in disaster management: A process model language customized for disaster management.

Digital Transformation of B2B Sales Processes

Laura Lahtinen[1], Tommi Mahlamäki[1][a] and Jussi Myllärniemi[2][b]
[1]*Unit of Industrial Engineering and Management, Tampere University, Tampere, Finland*
[2]*Unit of Information and Knowledge Management, Tampere University, Tampere, Finland*
laura.lahtinen@upm.com, tommi.mahlamaki@tuni.fi, jussi.myllarniemi@tuni.fi

Keywords: Digitalization, Digital Transformation, B2B Sales, Sales Processes.

Abstract: This paper explores the impact of digital transformation on sales practices. It emphasizes how businesses adopt new digital technologies to enhance efficiency, customer engagement, and overall business outcomes. Digital transformation involves both digitalizing analog processes and leveraging innovative technologies like artificial intelligence and automation in sales operations. Some companies have successfully integrated digital tools to improve B2B sales effectiveness and profitability. However, challenges persist, including organizational resistance and the need for updated managerial practices and performance metrics. The main contribution of this paper is the identification of changes in sales caused by digital transformation. Despite these changes and challenges, digital transformation in sales signifies a shift towards more consultative and adaptive sales approaches. This shift necessitates a re-evaluation of traditional sales practices and a proactive embrace of technological advancements for sustained competitive advantage in dynamic markets.

1 INTRODUCTION

This paper examines the profound impact of digital transformation on sales processes. Digital transformation, characterized by the integration of advanced digital technologies into business processes, has reshaped traditional sales paradigms. It encompasses both digitalization—transforming analog processes into digital—and broader business improvements facilitated by new technologies (Reis et al., 2018; Überwimmer et al., 2021).

Key findings include that while some companies actively drive digitalization for efficiency gains, others are compelled by external pressures (Überwimmer et al., 2021). Digital transformation has significantly altered sales dynamics, enhancing revenue, effectiveness, and customer understanding through digital tools (Mattila et al., 2021). However, challenges persist, with anticipated gains in productivity often unrealized (Wengler et al., 2021).

The advent of novel technologies, like artificial intelligence and machine learning, has further accelerated above mentioned changes, prompting a shift towards digital sales channels and automation of routine tasks (Syam & Sharma, 2018). This evolution necessitates a rethinking of sales strategies and roles, with a growing emphasis on consultative selling and multi-channel engagement (Singh et al., 2019; Thaichon et al., 2018).

As organizations undergo these transformations, they must overcome barriers such as technological adaptation, organizational readiness, and the evolving role of sales personnel (Alavi & Habel, 2021; Wengler et al., 2021). This paper provides a comprehensive overview of the current landscape and future prospects as businesses pursue to leverage digital transformation for sustained competitive advantage in sales processes and practices. The paper contributes also by identifying the key changes in B2B sales caused by digital transformation.

2 SALES PROCESSES

Digital transformation has a potential to change the way companies operate and do business. In sales the digitalization has been commonplace in the sales of non-complex consumer goods markets (Mahlamäki et al., 2016; Rodríguez et al., 2020) and now the B2B sales processes are following suit. This paper will

[a] https://orcid.org/0000-0003-3329-4351
[b] https://orcid.org/0000-0002-2848-0426

Lahtinen, L., Mahlamäki, T. and Myllärniemi, J.
Digital Transformation of B2B Sales Processes.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 265-272
ISBN: 978-989-758-716-0; ISSN: 2184-3228

therefore focus on the transformation of B2B sales processes.

As the focus of the paper is on B2B sales processes as a whole, it is therefore important first to understand the individual processes and tasks taking place within B2B sales. Moncrief and Marshall (2005) present "the oldest paradigm" of sales, which constitutes of seven steps: prospecting, pre-approach, approach, presentation, overcoming objections, closing and follow-up. The authors say that the model has changed very little over the years, but the orientation has altered to more customer-oriented (Moncrief & Marshall, 2005). Sales processes vary depending on the industry and even between companies within the same industry, but the principles are the same. Rodríguez et al. (2020) identified that different markets have commonalities in sales processes. In business-to-business (B2B) markets sales processes call for strong efforts by the seller in reaching out, closing deals and maintaining relationships with customers compared to another traditional market, business-to-customer (B2C). The B2B sales process is also more detailed in stages and retaining customers is a key to generating higher sales and thus long-term customer relationships are necessary. (Rodríguez et al., 2020). The next section takes a closer look at the features of the industrial sales process.

3 INDUSTRIAL SALES PROCESS

An explicit way to describe sales is to define it as an action, where value is created through an interaction between individuals who either represent themselves or an organisation. As a result, customer's problem is solved, the monetary exchange takes place and thus value is created for all parties involved. (Hänti et al., 2016) Around this core, the implementation of sales might differ not only between industries but within companies of the same kind.

In B2B context one way of organizing sales and operation processes is make-to-order (MTO) method. MTO refers to the manufacturing implementation where the products are produced partly or entirely only after the customer order has been received; the opposite approach is make-to-stock where the production and customer orders are not connected but the products are complete when the customer places an order (Kingsman et al., 1996). Sales and production are strongly connected in MTO companies; production is demand-driven and allows product customization. The products are configurable according to customers' requests but yet designed to be manufactured in an efficient way that keeps the cost at the level of mass-produced products (Custódio et al., 2018).

Dealing with customer inquiries is an essential issue for MTO companies (Kingsman et al., 1996), but before that the customer's inquiry or request for quotation (RFQ) needs to be initiated and that is the responsibility of sales. In a "traditional" sales process the salesperson spends time learning the customer's processes and then makes a lucrative offer to the customer to attract the customer's interest in the company's offering (Kotler et al., 2010). Figure 1 below presents three traditional B2B sales processes recognized in the literature: a six-step process from the 1920s in How to Increase Your Sales (1920, 17th ed. as cited in Moncrief & Marshall, 2005), the seven-step process by Moncrief & Marshall (2005) and the six-step process by Kotler et al. (2010). The content of each process is largely the same, only labelling differs.

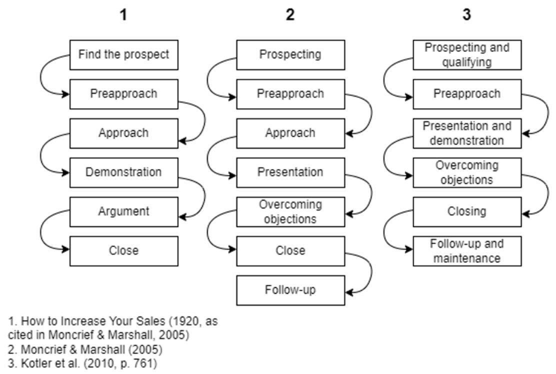


Figure 1: Traditional sales processes presented in the academic publications.

Figure 1 shows that the steps of the B2B sales process have remained almost unchanged over a century, the only remarkable change being the addition of the last step, "follow-up" (Moncrief & Marshall, 2005). According to Kotler et al. (2010), this step is vital to ensure customer satisfaction and loyalty in terms of repeat business. The approach to sales has shifted from forceful, closed means to a relationship-selling approach (Moncrief & Marshall, 2005) where long-term buyer-seller relationships lead to higher sales through rebuys and follow-up orders (Rodríguez et al., 2020). Especially in an industrial context the relationships between buying and selling companies are important, however, the nature of the relationship depends on the strategic importance of the product and process of the two companies; the importance and closeness of the relationship are different with office supplies provider and the supplier of a product which is essential to buyer's offering. (Ford, 1980) The original advantages of relational selling were cost-benefit consequences for both parties (Hohenschwert & Geiger, 2015) and for the selling company to ensure market position as the close, long-term relationships would pose a barrier for new players to enter the market (Ford, 1980). However, the perception of relational value has recently changed towards cross-organizational problem-solving, where monetary value for customers is created through gains in efficiency and improved processes, for instance (Hohenschwert & Geiger, 2015). The value is co-created with customers (Viio & Grönroos, 2014). Salespersons have a key role in value creation in B2B relationships as they have an excellent position to understand customers' value drivers and try to influence customers' value perceptions in favour of the selling company (Hohenschwert & Geiger, 2015).

The creation of exceptional customer value is a necessity for companies' long-term survival and growth (Terho et al., 2012). Customer value creation is so important that new sales approaches have emerged around it: solution, consultative and value-based selling (Hohenschwert & Geiger, 2015). Many companies are in the process of transitioning to more value-driven sales by developing their solution offering and sales. This is a wider trend in the B2B markets, where even more often the value is created through solutions selling – meaning that products and services are sold as a complete package rather than separately (Brady et al., 2005). Solution-based selling enables companies to go "down-stream" in the value chain (Salonen et al., 2021) which offers an initiative for business-to-business companies to maintain a close buyer-seller relationship (Wise & Baumgartner, 1999 as cited in Salonen et al., 2021). The sales team is a key enabler of this transition as it is up to them to craft a solution offering and communicate its value to the customer (Salonen et al., 2021). With the transition towards solution selling the position of sales changes from being an independent and somewhat isolated function to becoming an integrated, cross-functional part of customer management (Storbacka et al., 2011).

The process of selling a solution is different to selling a product (Brady et al., 2005; Salonen et al., 2021). First, selling solutions is a cyclical process where the post-project services act as an igniter to a new sales project. In traditional product sales, the handover of the product marks the end of the project, but in solution sales, the responsibility of maintenance, support and other lifecycle services remains with the seller (Brady et al., 2005). Second, the stages of solution selling are different from those of product selling. Solution selling comprises four relational stages (Salonen et al., 2021): (1) definition of customer requirements, (2) integration and customization of goods and/or services, (3) project execution and (4) post-project support. From salespeople point of view this means having to understand customers' businesses better and making more customized solutions for customers' needs (Salonen et al., 2021). Ulaga & Loveland (2014) claim that selling solutions requires different skills and attitudes than selling goods – solution sellers benefit from general intelligence and learning orientation. On the other hand, Storbacka et al. (2011) argue that instead of individual salesperson's skills the unit of analysis should be shifted to the capabilities of the sales unit as sales is becoming an important part of driving strategic initiatives.

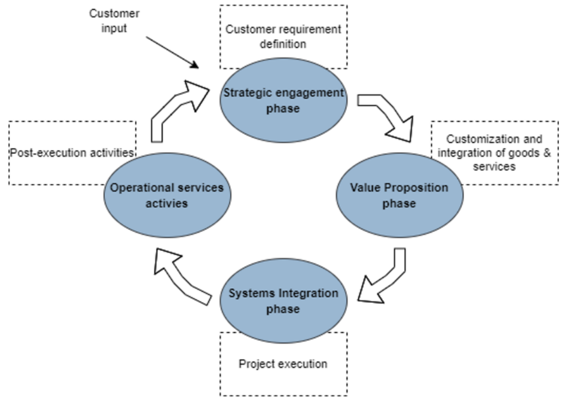


Figure 2: Sales process of selling solutions (adapted from Brady et al., 2005; Salonen et al., 2021).

Figure 2 shows the main stages of the solution sales process. The process is superficial and does not reveal the micro activities happening throughout the sales process, such as customer touchpoints, capabilities needed and so forth. Rabetino et al. (2018) have recognized that existing research tends to simplify the sales process and not go into the individual tasks and interfaces within the process. They've formed a five-step framework for selling solutions in an industrial context which is a consensus of the traditional 7-step sales paradigm (Moncrief & Marshall, 2005) and the solution selling frameworks (Brady et al., 2005; Salonen et al., 2021; Storbacka, 2011). The steps are (Rabetino et al., 2018)

1) Information acquisition
2) Initial negotiation
3) Value proposition and real negotiation
4) Offering deployment and value authentication
5) Customer operations maintenance and support

The most important purpose of the information acquisition -step is to find customers whose needs match the seller's resources (Storbacka, 2011). This is done, for instance, by regular visits to customers by a dedicated salesperson (Rabetino et al., 2018) who engages the customer in a discussion of their business problems before any official invitation to tender has been issued (Brady et al., 2005). This ensures the right contacts in the customer company and acts as an ignitor for preliminary negotiation in step 2 (Moncrief & Marshall, 2005). The aim is to make the customer interested in the seller's offering (Töytäri et al., 2011).

In value-based selling the value is determined iteratively together with the customer preferably in a cross-organizational team and this is what happens in step 3 (Storbacka et al., 2013; Töytäri et al., 2011). In the actual negotiation the quantified value should be turned into an offering which meets or even exceeds the customer's expectations (Brady et al., 2005; Töytäri et al., 2011). Steps 4 and 5 take place in parallel and are initiated when the customer and seller have reached a mutual contractual agreement (Brady et al., 2005; Rabetino et al., 2018). The solutions are usually difficult to price as the lifetime costs of the solution must be considered and a common understanding of how value will be measured in terms of pricing needs to be formed together with the customer. In addition, differing from traditional product offerings project managers in steps 4 and 5 have to pay attention to customer satisfaction besides the usual constraints of budget and schedule. (Brady et al., 2005)

Though companies' offering still rely heavily on products, they are heading more and more towards solution offering and value-based selling and their sales processes resemble the above-mentioned industrial B2B solution sales process very much. This can be explained by the nature of MTO products, which require more interaction between customers and sales than make-to-stock products which are sold from an inventory and customers cannot affect the product design (Parente et al., 2002). So, selling MTO products includes cooperative problem-solving and value-creation between seller and buyer by default. In some cases, it is noteworthy that the sales scene is determined by competitive bidding: when a customer decides to enquire about a product, the inquiry is usually sent to all suppliers simultaneously and the competing quotes will be evaluated before selecting the company with which the order will be placed (Kingsman et al., 1996). This means that even though the seller has been negotiating closely with the customer over several months the customer might end up choosing another supplier and the seller is left with empty hands; Tobin et al. (1988) argue that the percentage of quotes becoming firm orders varies from 3 % to virtually 100 %. Another determining factor of the sales process of MTO products is the cooperation of production and sales units. Production is driven by demand and on the other hand production capacity sets limits to selling products (Feng et al., 2008; Tobin et al., 1988).

Production and sales are traditionally two non-coordinated business units whose decisions have a significant impact on the company's financial performance and operational efficiency (Feng et al., 2008). Ideally, the process of bidding or quoting for orders (which is part of sales) should be used to some level to design the order book so that it can be produced profitably (Tobin et al., 1988). Sales and Operations Planning (S&OP) is a tool used to bring the two departments together strategically (Feng et al., 2008), but on a daily basis, this means that the salesperson and product manager are in close contact during the bidding process. This is illustrated in Figure 3 below, which draws a consensus on the sales processes presented in this section.

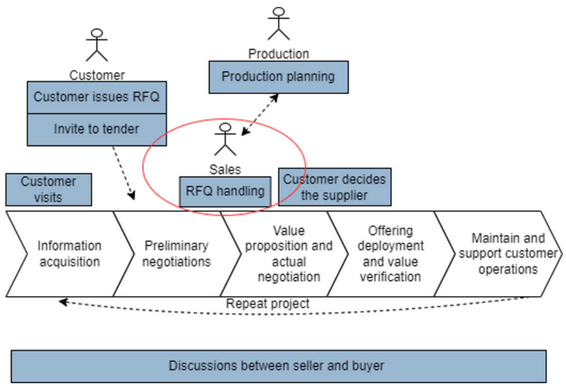


Figure 3: Solution selling process (adapted from Feng et al., 2008; Kingsman et al., 1996; Rabetino et al., 2018).

4 EFFECTS OF DIGITAL TRANSFORMATION ON SALES

This section discusses both digital transformation and digitalization and reflects towards our future research. Digital transformation means the use of new digital technologies that enable extensive business improvements and influence customers' life thoroughly (Reis et al., 2018). Digitalization means altering analogue processes into digital processes (Überwimmer et al., 2021). According to Überwimmer et al. (2021), companies either proactively drive digitalization in their operations or are forced to digitalize by external pressure. Companies driving digitalization consider improvements in efficiency as one of the primary goals to be achieved with digitalization. (Überwimmer et al., 2021) Despite the efforts made to enhance digital transformation the gains in productivity and efficiency have hardly materialized (Wengler et al., 2021). On the other hand, some studies reveal that technologies and digital tools have improved B2B sales in terms of revenue, effectiveness, profitability and understanding of customer's needs as argued by Mattila et al. (2021). Nonetheless, companies are often forced to change by the market and their customers (Überwimmer et al., 2021) and by major disruptions brought about by advancements in digital technology, such as artificial intelligence and machine learning (Singh et al., 2019). Even though the current sales practices and theories are threatened, new opportunities for sales practice and research are opening (Singh et al., 2019) and companies need to embrace this development to stay competitive (Überwimmer et al., 2021).

In sales, digital transformation can be defined as improving customer outcomes in existing business models and advancing competencies and rethinking the value proposition of the firm by applying digital technologies to current company resources. Digital transformation requires digitalization. (Singh et al., 2019) The applied digital technologies fall into three categories (Ahearne & Rapp, 2010): salesperson-oriented, customer-oriented or shared. Salesperson-oriented and customer-oriented technologies are used exclusively either by sales or customers and shared technologies are interactive, social media being an example of this. (Ahearne & Rapp, 2010) Überwimmer et al. (2021) see the shared technologies profoundly changing the way of contacting and communicating with customers: first contacts are preferred to be made in social media and meetings with customers are reduced, but simultaneously the speed of response is expected to increase. Singh et al. (2019) say that B2B companies are launching self-service platforms for customers to browse items and track or place orders; similarly, buyers are introducing online platforms for suppliers to participate in tender processes.

Sales channels are being rapidly digitized, again, to stay competitive and save on costs, improve selling efficiency and improve customer value (Thaichon et al., 2018). Überwimmer et al. (2021) consider the biggest changes brought about by digitalization in sales to be the new channels to communicate, the need for real-time data availability and the role of sales. Many studies proclaim the need for a multichannel approach to contacting customers (Ramos et al., 2023; Thaichon et al., 2018). According to Syam and Sharma (2018), the sales processes will be impacted by robotics, machine learning and artificial intelligence causing the automation of routine sales tasks. The automation is targeted to routine, non-productive processes and this will free up salespersons' time for more productive, customer-facing tasks. (Syam & Sharma, 2018) Automation of sales is called sales force automation (SFA) which together with related digital sales tools is profoundly changing the work division and dynamics between the selling and buying companies (Mahlamäki et al., 2020).

Automated sales processes enable customers to find appropriate information about products and services fast and even perform simple transactions without the involvement of a salesperson (Thaichon et al., 2018). Also, more refined tools are available, such as sales configurators – digital tools designed to guide the user through a service or product configuration process (Rogoll & Piller, 2004, as cited in Mahlamäki et al., 2020). The configuration process produces a configuration which describes the make-up of an instance of a product customized to the customer's requirements within the limits set by the product architecture (Tiihonen et al., 1996). The demand for customized products is increasing from the customer side (Hvam et al., 2008) and configurable products are the answer to that, having a pre-designed basic structure that is adaptable to customer requirements (Tiihonen et al., 1996). At first, these kinds of configurators were made to assist sales representatives in their job, but recently the configurators have been made available for buyers' independent use as well without the salesperson being present (Mahlamäki et al., 2020). Configurator is just an individual example of SFAs - some professionals estimate that even over a third of the sales processes could be automated (Mattila et al., 2021).

So, digital transformation changes the role of sales from "selling classical products with the classical tools" (Überwimmer et al., 2021). The ways of interacting change and the roles each user (buyer or supplier) plays become blurred in the digital platforms (Kumar et al., 2018; Mathmann et al., 2017). Nonetheless, salespersons still have a valuable role when customers request more complex solutions (Ahearne & Rapp, 2010) and today's buyers are increasingly asking for complex combinations of services and products (Singh et al., 2019). When the solution is complex the sales process itself tends to be more complex and the deal is not made just between two people but also the hierarchical levels of both buyer and seller companies must be considered. Selling complex solutions requires more face-to-face encounters to close deals compared to selling simple solutions and products and therefore the digitalization of a complex sales process is slower. (Rodríguez et al., 2020)

Digital transformation and sales channel digitization change many things in sales, such as the value creation of salespeople which will shift from order handling to a more consultative role, especially for complex products (Thull, 2010 as cited in Singh et al., 2019). Though, some argue that value-creation function in sales profession will diminish (Singh et al., 2019). In any case, ways of working are changing quickly in the complex digital environment which requires quick learning, especially in terms of the above-mentioned customer value creation (Hartmann et al., 2018). Therefore, the sales force needs to develop a new mindset that welcomes and fosters change (Überwimmer et al., 2021). Table 1 gathers subjects affected by digital transformation in sales.

Table 1: Changes in sales caused by digital transformation and related phenomena like digitization.

Change	Source
Selling activities take place online via digital sales channels	Alavi & Habel, 2021; Überwimmer et al., 2021
Customers can place purchase orders without the involvement of sales representatives	Mahlamäki et al., 2020; Thaichon et al., 2018
More consultative role especially in the sales of more complex products	Ahearne & Rapp, 2010; Thull, 2010 as cited in Singh et al., 2019
Ways of contacting and communicating with customers	Singh et al, 2019; Überwimmer et al., 2021
Automation of routine tasks	Mattila et al., 2021; Syam & Sharma, 2018
Unlearning old routines	Mattila et al., 2021
Reorganization of the structure and roles of the sales force	Singh et al, 2019
Hybrid sales structure	Ramos et al., 2023; Thaichon et al., 2018
Multichannel strategies	Ramos et al., 2023; Thaichon et al., 2018

Mattila et al. (2021) emphasize the importance of unlearning beliefs and practices as part of digital transformation. According to Klammer and Gueldenberg (2019), individuals refrain from relinquishing their current habits and ways of working as long as the existing methods remain successful. However, when the environment changes it is necessary to unlearn outdated practices and discard false knowledge. (Klammer & Gueldenberg, 2019) This allows organizations to change processes into more efficient ones (Becker, 2010). Things to be unlearned in sales due to digital transformation have been studied by Mattila et al. (2021) and they identified the change of mindset organization-wide as one of the main themes. Also, they speak up for actively searching and identifying the need for unlearning by critically reviewing the "old" sales and managerial processes, for instance. (Mattila et al., 2021) Digital transformation also affects the structure and roles in the sales organization. Singh et al. (2019) claim that there is little information on how to structure and organize the sales force to operate soundly and effectively in the new environment. However, they've found out that many companies have altered the roles in sales into more specialized and the number of generalist salespeople is diminishing. (Singh et al., 2019)

Also, Thaichon et al. (2018) claim that digital transformation and the rise of e-commerce are evolving the sales structures – the traditional focus on the outside sales force has shifted to valuing the inside sales force which is extended with the successful use of online channels. They consider the hybrid sales model as the sales structure of the e-commerce era. (Thaichon et al., 2018) Hybrid sales model combines outside and inside sales and online channels (Ramos et al., 2023; Thaichon et al., 2018). Inside sales refers to salespeople who are remote and not engaged with any traditional face-to-face customer interaction, but they use different communication technologies. The role of inside sales has evolved into having strategic importance in customer value creation and different inside and outside sale configurations. (Ramos et al., 2023) Hybrid sales model requires cooperation between all the discrete parties in the hybrid sales model - inside sales, outside sales, and online channel - and value co-creation with the customer through all three contact points (Thaichon et al., 2018).

Dynamic capabilities are the enablers of hybrid sales (Ramos et al., 2023; Thaichon et al., 2018). According to Wilson and Daniel (2007), the required capabilities are 1) active review of "route to market" structures, 2) aligning sales structures with products and services, 3) creating innovative (sales) channel combinations, 4) integration of processes and IT to assist multi-channel customer relationships. Implementation of sales technologies, such as SFA, enables sales forces to accomplish their tasks faster and more efficiently and supports seamless, real-time communication between inside and outside sales and sales forces and customers which is central to the hybrid sales in general (Thaichon et al., 2018).

Digitization and digital transformation act as enablers of many improvements, but as in all major changes there are obstacles hindering the development. In their article, Wengler et al. (2021) studied major barriers to digital transformation in sales and in all hierarchical levels time constraint was named the number one barrier. The top three perceived barriers differed between managerial and employee levels otherwise: managers see the company's lack of know-how and sales processes, which are not defined well enough to proceed with digital transformation as main barriers for digital transformation along with time constraints. However, employees consider the company having sufficient knowledge but see customers' know-how or use of different systems as barriers along with budget constraints. (Wengler et al., 2021) Organizational readiness in terms of capabilities and processes is important for a successful digital transformation process (Vial, 2019).

Alavi and Habel (2021) state that the human factor can hardly be overestimated when it comes to digital transformation in sales. They claim that digital transformation projects rarely fail because of technical issues but of companies' poor management of human factor. Salespeople are not afraid of new technologies as such, but they feel positive about the potential of digital sales technologies. However, they have concerns regarding job autonomy, for instance, being afraid that they could be monitored to a larger extent through the new technologies. (Alavi & Habel, 2021) Mattila et al. (2021) found that sometimes the digital transformation progress is slow due to unwilling individuals and slow changing organizational processes. This type of problem can be solved with managerial practices. (Mattila et al., 2021) Lastly, Wengler et al. (2021) claim that managers still follow "old KPIs" to measure performance and those are not applicable anymore. Suitable KPIs are needed as it is hard to manage a business properly in the digital era without those. (Wengler et al., 2021)

Unarguably digital transformation is going to change the role of sales. Some transactions take place online without the involvement of a sales representative, but offerings in B2B markets are getting growingly complex and customers more demanding – therefore, sales has a strategic role as the sales activity is changing to a more consultative, solution-selling approach and maintaining customer relationships is of great importance (Ramos et al., 2023). Digitization and digital transformation are underway and related challenges remain to be dealt with. As mentioned above, the human factor, both on an individual and organizational level, plays a big role in the success of the digital transformation.

REFERENCES

Ahearne, M., Rapp, A. (2010). The Role of Technology at the Interface Between Salespeople and Consumers. *Journal of Personal Selling Sales Management,* 30(2), 111–120.

Alavi, S., Habel, J. (2021). The human side of digital transformation in sales: Review future paths. *Journal of Personal Selling Sales Management*, 41(2), 83–86.

Becker, K. (2010). Facilitating unlearning during implementation of new technology. *Journal of Organizational Change Management,* 23(3), 251–268.

Brady, T., Davies, A., Gann, D. M. (2005). Creating value by delivering integrated solutions. *International Journal of Project Management,* 23(5), 360–365.

Custódio, D. T., Vaccaro, G. L. R., Nunes, F. L., Vidor, G., Chiwiacowsky, L. D. (2018). Variant product configuration of industrial air handling units in a MTO environment. *The International Journal of Advanced Manufacturing Technology*, 95(1–4), 1025–1037.

Feng, Y., D'Amours, S., Beauregard, R. (2008). The value of sales and operations planning in oriented strand board industry with make-to-order manufacturing system: Cross functional integration under deterministic demand and spot market recourse. *International Journal of Production Economics,* 115(1), 189–209.

Ford, D. (1980). The Development of Buyer-Supplier Relationships In Industrial Markets. *European Journal of Marketing*, 14, 339–353.

Hänti, S., Kairisto-Mertanen, L., Kock, H. (2016). *Oivaltava myyntityö—Asiakkaana organisaatio*. Edita Publishing Oy.

Hartmann, N. N., Wieland, H., Vargo, S. L. (2018). Converging on a New Theoretical Foundation for Selling. *Journal of Marketing*, 82(2), 1–18.

Hohenschwert, L., Geiger, S. (2015). Interpersonal influence strategies in complex B2B sales and the socio-cognitive construction of relationship value. Industrial Marketing Management, 49, 139–150. https://doi.org/10.1016/j.indmarman.2015.05.027

Hvam, L., Mortensen, N. H., Riis, J. (Eds.). (2008). Introduction. *In Product Customization* (pp. 1–15). Springer.

Kingsman, B., Hendry, L., Mercer, A., de Souza, A. (1996). Responding to customer enquiries in make-to-order companies Problems and solutions. *International Journal of Production Economics*, 46–47, 219–231.

Klammer, A., Gueldenberg, S. (2019). Unlearning and forgetting in organizations: A systematic review of literature. *Journal of Knowledge Management*, 23(5), 860–888.

Kotler, P., Brady, M., Goodman, M., Hansen, T., Keller, K. (2010). *Marketing Management*. Pearson Education UK.

Kumar, V., Lahiri, A., Dogan, O. B. (2018). A strategic framework for a profitable business model in the sharing economy. *Industrial Marketing Management*, 69, 147–160.

Mahlamäki, T., Ojala, M., Myllykangas, M. (2016). The drivers towards interactive digital sales processes in business-to-business markets. *Proceedings of the 20th International Academic Mindtrek Conference*, 42–48.

Mahlamäki, T., Storbacka, K., Pylkkönen, S., Ojala, M. (2020). Adoption of digital sales force automation tools in supply chain: Customers' acceptance of sales configurators. *Industrial Marketing Management*, 91, 162–173.

Mathmann, F., Chylinski, M., de Ruyter, K., Higgins, E. T. (2017). When Plentiful Platforms Pay Off: Assessment Orientation Moderates the Effect of Assortment Size on Choice Engagement and Product Valuation. *Journal of Retailing,* 93(2), 212–227.

Mattila, M., Yrjölä, M., Hautamäki, P. (2021). Digital transformation of business-to-business sales: What needs to be unlearned? *Journal of Personal Selling Sales Management,* 41(2), 113–129.

Moncrief, W. C., Marshall, G. W. (2005). The evolution of the seven steps of selling. *Industrial Marketing Management*, 34(1), 13–22.

Parente, D. H., Pegels, C. C., Suresh, N. (2002). An exploratory study of the sales-production relationship and customer satisfaction. *International Journal of Operations Production Management*, 22(9/10), 997–1013.

Rabetino, R., Ogundipe, S. J., Kohtamäki, M. (2018). Solution sales process blueprinting. *International Journal of Business Environment,* 10(2), 132.

Ramos, C., Claro, D. P., Germiniano, R. (2023). The effect of inside sales and hybrid sales structures on customer value creation. *Journal of Business Research*, 154, 113343.

Reis, J., Amorim, M., Melão, N., Matos, P. (2018). Digital Transformation: A Literature Review and Guidelines for Future Research. *In Á. Rocha, H. Adeli, L. P. Reis, S. Costanzo (Eds.), Trends and Advances in Information Systems and Technologies* (pp. 411–421). Springer International Publishing.

Rodríguez, R., Svensson, G., Mehl, E. J. (2020). Digitalization process of complex B2B sales processes – Enablers and obstacles. *Technology in Society*, 62, 101324.

Salonen, A., Terho, H., Böhm, E., Virtanen, A., Rajala, R. (2021). Engaging a product-focused sales force in solution selling: Interplay of individual- and organizational-level conditions. *Journal of the Academy of Marketing Science,* 49(1), 139–163.

Singh, J., Flaherty, K., Sohi, R. S., Deeter-Schmelz, D., Habel, J., Le Meunier-FitzHugh, K., Malshe, A., Mullins, R., Onyemah, V. (2019). Sales profession and professionals in the age of digitization and artificial intelligence technologies: Concepts, priorities, and questions. *Journal of Personal Selling Sales Management*, 39(1), 2–22.

Storbacka, K., Polsa, P., Sääksjärvi, M. (2011). Management Practices in Solution Sales—A Multilevel and Cross-Functional Framework. *Journal of Personal Selling Sales Management*, 31(1), 35–54.

Storbacka, K., Windahl, C., Nenonen, S., Salonen, A. (2013). Solution business models: Transformation along four continua. *Industrial Marketing Management*, 42(5), 705–716.

Syam, N., Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, 69, 135–146.

Terho, H., Haas, A., Eggert, A., Ulaga, W. (2012). 'It's almost like taking the sales out of selling'—Towards a conceptualization of value-based selling in business markets. *Industrial Marketing Management*, 41(1), 174–185.

Thaichon, P., Surachartkumtonkun, J., Quach, S., Weaven, S., Palmatier, R. W. (2018). Hybrid sales structures in the age of e-commerce. *Journal of Personal Selling Sales Management*, 38(3), 277–302.

Tiihonen, J., Soininen, T., Männistö, T., Sulonen, R. (1996). State-of-the-practice in product configuration—A survey of 10 cases in the Finnish industry. *In T. Tomiyama, M. Mäntylä, S. Finger (Eds.), Knowledge Intensive CAD* (pp. 95–114). Springer US.

Tobin, N. R., Mercer, A., Kingsman, B. (1988). A Study of Small Subcontract and Make‐to‐Order Firms in Relation to Quotation for Orders. *International Journal of Operations Production Management*, 8(6), 46‐59.

Töytäri, P., Alejandro, T. B., Parvinen, P., Ollila, I., Rosendahl, N. (2011). Bridging the theory to application gap in value-based selling. *The Journal of Business Industrial Marketing*, 26(7), 493–502.

Überwimmer, M., Hautamäki, P., Wengler, S., Füreder, R. (2021). Sales Organizations on the Path of Digitalization – a Reflection from Germany, Finland and Austria: Obchodné organizácie na ceste digitalizácie – pohľady z Nemecka, Fínska a Rakúska. *Marketing Science Inspirations*, 16(2), 24–29.

Ulaga, W., Loveland, J. M. (2014). Transitioning from product to service-led growth in manufacturing firms: Emergent challenges in selecting and managing the industrial sales force. *Industrial Marketing Management*, 43(1), 113–125.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144.

Viio, P., Grönroos, C. (2014). Value-based sales process adaptation in business relationships. *Industrial Marketing Management*, 43(6), 1085–1095.

Wengler, S., Hildmann, G., Vossebein, U. (2021). Digital transformation in sales as an evolving process. The *Journal of Business Industrial Marketing*, 36(4), 599–614.

Wilson, H., Daniel, E. (2007). The multi-channel challenge: A dynamic capability approach. *Industrial Marketing Management*, 36(1), 10–20.

Sales Development in Business-to-Business Markets

Tommi Mahlamäki[a] and Juho Martikainen
Unit of Industrial Engineering and Management, Tampere University, Tampere, Finland
tommi.mahlamaki@tuni.fi, juho.e.martikainen@gmail.com

Keywords: Sales Development, Digitalization, Sales, B2B.

Abstract: Sales development has emerged as a critical strategy within the business-to-business (B2B) landscape, propelled by the ongoing digitalization of sales processes. This paper explores how digital technologies, including artificial intelligence (AI), machine learning, CRM systems, and e-commerce platforms, have reshaped sales strategies and interactions between buyers and sellers. As digital tools become integral to modern sales operations, the role of sales development in generating leads and fostering customer relationships is increasingly prominent. The paper reviews existing literature on sales development, highlighting its evolution from traditional sales models to contemporary digital frameworks. Furthermore, the study proposes an initial conceptual framework for understanding sales development in B2B contexts, setting the stage for empirical research to explore its implementation and effectiveness in enhancing sales performance and customer engagement.

1 INTRODUCTION

Sales and sales processes have been extensively researched, yet they have evolved significantly over the past few decades, driven by digitalization. The internet and digital tools, initially prominent in B2C contexts, are now transforming B2B markets as well. This paper explores how digitalization impacts sales and sales processes, focusing on a new concept of sales development.

Digitalization is a megatrend in sales, driven by advancements in information technology, machine learning, robotics, and AI. As Rodriguez et al. (2020) highlight, B2B markets are increasingly complex, necessitating more efficient, cost-effective processes. Digitalization and Sales Force Automation (SFA) are crucial in this transformation, making it a mandatory consideration for management (Mahlamäki et al., 2016). This paper will delve into how digital technologies and channels have dramatically altered buyer-seller interactions, emphasizing the significant role of digital tools such as CRM systems and e-commerce platforms in enhancing sales processes.

Furthermore, the relationship between digitalization and marketing technologies is explored. Marketing technologies, including marketing automation and CRM tools, play a vital role in supporting sales development by providing the necessary digital infrastructure.

This paper also examines the concept of inside sales, which has gained prominence as digital channels reduce the need for face-to-face interactions. Inside sales, as a configuration of sales development, is critical in today's sales environment, offering numerous advantages over traditional outside sales approaches.

The paper contributes to academic discussion by providing more understanding of the successful use of sales development as part of a digitalized sales process in the B2B context. Concretely this is done by introducing a framework of sales development, focusing on its role in generating leads and fostering customer relationships within a digitalized sales process. The challenges and benefits of implementing sales development are discussed, providing a foundation for further empirical research. This initial framework will be refined in the future through personal interviews with representatives from B2B companies, ensuring its relevance and applicability in the modern sales context.

[a] https://orcid.org/0000-0003-3329-4351

Mahlamäki, T. and Martikainen, J.
Sales Development in Business-to-Business Markets.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 273-279
ISBN: 978-989-758-716-0; ISSN: 2184-3228

2 LITERATURE REVIEW

2.1 Digitalization in Sales and Sales Processes

Sales and sales processes have been a focus of research for a long time. Yet during the past decades they have evolved significantly, particularly with the use of digitalization. The internet and other digital tools, initially prominent in B2C contexts, are now transforming B2B markets as well.

Digitalization is a megatrend in sales and is often called the "fourth industrial revolution." It is driven by advancements in information technology, machine learning, robotics, and AI (Syam & Sharma, 2018). Rodriguez et al. (2020) highlight that B2B markets are increasingly complex and demanding, leading to less productive operations. Complex B2B sales organizations aim for more efficient, cost-effective processes through digitalization and Sales Force Automation (SFA) (Román & Rodríguez, 2015). Thus, digitalization transforms the operational environment and is a mandatory consideration for management.

Digital technologies and channels have dramatically altered buyer-seller interactions. Companies invest more in digital technologies to streamline and enhance sales processes (Terho et al., 2022). Also, according to Terho et al. (2022), marketing plays a crucial role in acquiring new leads through digital channels, with customers relying heavily on digital resources during the initial buying phases, even in B2B contexts.

In the context of this research, digitalization in sales is an important theme because sales development is mostly related to digitalization: it can be said to require effective digital tools to be able to run properly and to gain the benefits it is offering.

Common digital sales tools include CRM systems (Ahuja & Medury, 2010) and e-commerce platforms (Alsaad & Taamneh, 2019). These tools enhance sales processes by reducing internal costs or increasing outcomes (Rodríguez et al., 2020). CRM systems facilitate internal communication and data storage. E-commerce platforms facilitate transactions between trading partners, often visible to customers as webshops. Despite substantial investments in digitalization, the link between investment and performance enhancement remains unclear. Successful technology implementation depends on the interplay of technological elements and the commitment of salespeople (Rodríguez et al., 2020).

Rodriguez et al. (2020) identified internal and external enablers and obstacles to digitalizing sales processes, categorized into organizational, technological, cultural, and legal/security dimensions. Overcoming these challenges can lead to successful digitalization and significant benefits. Priority internal sales elements for digitalization include customer identification, sales proposals, and follow-up support—steps 1, 3, and 5 in Rodriguez et al. (2020) sales steps. In the modern world, digitalization is essential for any company's existence, impacting both B2C and B2B sectors.

Machine learning and AI represent significant trends within digitalization and digital commerce. Syam and Sharma (2018) describe ongoing digitalization, including AI and machine learning, as a transformative force in the industry, reshaping sales practices. The industry is keenly observing AI's potential and limitations.

Digitalization has fundamentally altered selling. Relational selling has undergone substantial changes. Arli et al. (2018) note that sellers must adapt to larger, more organized buyers with higher demands, reflecting the evolving nature of sales in the digital age.

2.2 Digital Marketing and Marketing Technologies

When discussing sales and especially sales development and digitalization, it is essential to recognize the significant role of digital marketing and marketing technologies. These elements are intrinsically linked to sales development. Viewing sales as an isolated segment of the funnel is unwise; instead, it should be considered part of the integrated marketing and sales funnel, as Järvinen and Taiminen (2016) suggest. In addition, sales development relies on a strong connection between marketing and sales (Sleep et al., 2020; Terho et al., 2022).

The marketing department plays a crucial role in the company's business. Its efforts directly impact the sales and marketing funnel. Marketing should not function as an isolated unit. Wiersema (2013) found that integrating marketing and sales systems, like marketing automation or CRM, enhances team efficiency. Järvinen and Taiminen (2016) also noted that an integrated marketing and sales funnel increases transparency between these units, facilitating better collaboration. Marketing must work closely with sales, aligning business objectives to focus on the right leads, close more customers, and boost sales growth.

Emphasis will be on marketing technologies, defined as tools or methods aiding marketing teams in achieving their targets (Graesch et al., 2020). These technologies include database management systems enabling CRM tools, websites facilitating online marketplaces, and social media applications supporting corporate influencers.

Digitalization has profoundly impacted marketing. Before digital solutions, marketing was entirely offline. "Marketing finds itself in the middle of a digital transformation, which IT has caused and accelerated over time" (Graesch et al., 2020, p. 124). Online tools have revolutionized marketing, enhancing performance, reach, measurement, and efficiency.

From a sales perspective, marketing technologies are vital for collaboration throughout the sales funnel. Integrating and effectively using marketing automation and CRM tools are crucial for modern commerce (Sleep et al., 2020; Terho et al., 2022). Tools like CRM are often shared between marketing and sales. A common, integrated customer database helps both teams achieve their goals (Syam & Sharma, 2018).

2.3 Inside Sales

As digitalization has driven sales more towards digital channels and reduced face-to-face interactions between buyers and sellers, sales have been divided into inside sales and outside sales (Thaichon et al., 2018; Sleep et al., 2020). This division is quite old; sales have been categorized into inside and outside sales since the 1980s when telemarketing began to emerge. The term inside sales became more prevalent around the millennium, contrasting with outside sales (Kroger, 2013). Traditionally, inside sales have been defined as sales occurring remotely, whereas outside sales involve face-to-face interactions (Kroger, 2013; Thaichon et al., 2018; Sleep et al., 2020; Terho et al., 2022).

Today, these definitions may not be entirely up-to-date. A significant portion of sales and sales communications now happen remotely, which suggests that a larger part of sales can be considered inside sales (Fready et al., 2022). While some companies and roles still maintain clear distinctions between inside and outside sales, in many cases, the line between these functions has blurred. Additionally, the need for face-to-face interactions occurs much later in the B2B sales process, if at all, because modern technology can almost entirely replace face-to-face interactions (Mantrala & Albers, 2022).

According to Thaichon et al. (2018), inside sales have experienced significant growth in the last 50 years in terms of practical use. In the 1970s, sales focused almost entirely on face-to-face contact and companies relied on outside sales personnel. Over the past 50 years, sales structures have shifted from merely providing value to buyers to creating value with them. This shift and significant technological transformations have led companies to use more inside sales to meet market demands and leverage its benefits (Thaichon et al., 2018).

As inside sales constitutes an integral part of sales structures in many contemporary companies, it inherently offers several advantages. According to Gessner and Scott (2009) and Martin (2013), organizations are increasingly reallocating resources towards developing inside sales teams. For instance, Albrecht et al. (2014) observed that approximately 40% of large technology companies are shifting resources from outside sales to inside sales, reflecting the rising prominence of inside sales within corporate sales frameworks.

Sleep et al. (2020) delineate three significant shifts that have facilitated the transition from outside sales to inside sales. Firstly, inside sales now encompasses more than traditional call center roles and order-taking; it involves activities such as order generation traditionally performed by outside sales forces. Advances in technology have enabled effective customer interaction without the need for face-to-face meetings, allowing inside sales teams to manage smaller customer accounts within organizations (Sleep et al., 2020).

Secondly, Sleep et al. (2020) highlight changes in customer behavior within the B2B context, where customers increasingly utilize various technology-based channels to gather information, thereby favoring inside sales approaches. Johnson (2005) notes the evolving nature of B2B sales processes towards being more buyer-driven. An intriguing finding by Rapp et al. (2013) suggests that whether salespersons operate in inside or outside sales roles does not significantly affect customer orientation. Moreover, customers in B2B settings have reduced their reliance on face-to-face meetings in favor of online services; as early as 2012, Mantrala and Albers (2022) reported a 37% decline in face-to-face meetings among customers. The COVID-19 pandemic has accelerated these changes, making virtual and remote communication commonplace in B2B interactions (Fready et al., 2022).

Thirdly, heightened competition has compelled companies to invest in cost-effective capabilities, with inside sales playing a pivotal role in delivering value-added services through every customer interaction, focusing on upselling and cross-selling opportunities (Sleep et al., 2020). Today, inside sales represents a critical avenue for exploring new methods to enhance lead management in B2B firms, particularly through specialized inside salesforces dedicated to generating and nurturing high-quality leads for external sales teams (Kuruzovich, 2013; Thaichon et al., 2018; Sleep et al., 2020; Terho et al., 2022). This underscores the limited research into sales development, where the definition often stems from inside sales practices that differ significantly across sales domains. Thaichon et al. (2018) characterize sales development as a specialized adaptation of inside sales, underscoring its distinct operational methodologies. Terho et al. (2022) provide a rare focus on sales development and attempt to contextualize its understanding within the realm of inside sales.

In exploring sales development, understanding the relationship between inside sales and sales development is crucial. According to Sleep et al. (2020), sales development constitutes one of the four principal archetypes within the B2B inside salesforce configurations. Sales development is specifically defined as an adaptation of inside sales, alongside other configurations such as inbound sales, sales support, hybrid teams, and discrete sales functions, all outlined in Figure 1 derived from practitioner literature.

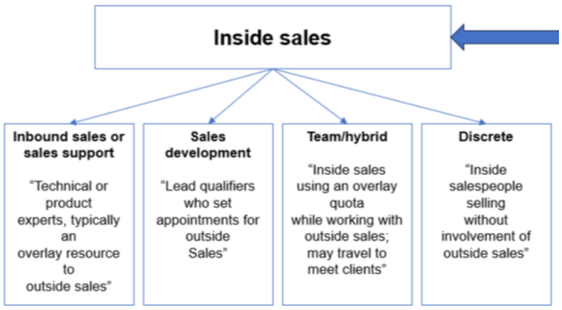


Figure 1: Major B2B inside sales force configurations (Adapted from Sleep et al., 2020).

As seen in Figure 1, sales development differs significantly from other configurations as it is not a separate function from outside sales but collaborates closely with it. For instance, discrete and hybrid configurations can operate independently from outside sales because they encompass the entire sales funnel, from prospecting to closing deals (Sleep et al., 2020). Some might argue that sales development cannot be a configuration of inside sales since Sales Development Representatives (SDRs) are not directly selling the product or service but are setting appointments for outside sales, whereas hybrid and discrete configurations manage the entire sales process. However, this distinction is not the primary concern of this study.

2.4 Sales Development

Sales development is a relatively new term in academia, with the first article focused solely on it published in 2022 (Terho et al., 2022). Although new academically, the term is well known in practice. According to non-academic sources like Salesloft (2014), Gartner (2019), Gummings (2014), Wikipedia (2024), and Arjona (2022), sales development originated in the 1980s at Oracle. The original Oracle Direct team, led by Anneke Seley, was the first to use sales development methods. Back then, sales development was quite different, with digitalization significantly transforming these methods and teams into today's modern sales development. Some argue that business development is similar to sales development, as Business Development Representatives (BDRs) may perform tasks akin to Sales Development Representatives (SDRs), but sales development is a more precise term for this specialized unit.

Although sales development is considered a configuration of inside sales, it differs significantly from the original meaning of inside sales. While sales development focuses on generating qualified leads and setting appointments for the outside salesforce, inside sales involves the entire sales process from first contact to closing. (Thaichon et al., 2018; Sleep et al., 2020; Terho et al., 2022).

As previously mentioned, Thaichon et al. (2018), Sleep et al. (2020), and Terho et al. (2022) state that sales development is primarily used in the B2B context and is a vital component of companies' sales processes. In contrast, the function is less utilized in B2C contexts, possibly due to the reasons identified by Rodriguez et al. (2020): "B2B sales processes require stronger sales efforts by the seller in reaching out, closing deals, and maintaining business relationships with customers." B2B involves more detailed stages and long-term relationships, focusing on transitioning from one-time transactions to long-term customer relationships (Rodriguez et al., 2020).

Conversely, the B2C sales process requires less effort from the sales department to reach out and achieve sales with consumers, making the process leaner (Sutanonpaiboon & Abuhamdieh, 2008). Sales development seems to fit better in B2B, but a significant obstacle is that B2B sales processes are often less digitalized than B2C, while sales development relies heavily on digital tools (Rodriguez et al., 2020; Terho et al., 2022).

Overcoming this conflict is crucial for the efficient use of sales development.

The definition of sales development is still evolving, with various references offering slightly different interpretations. However, three main publications (Thaichon et al., 2018; Sleep et al., 2020; Terho et al., 2022) are based on inside sales and define sales development similarly. Terho et al. (2022) builds on the previous definitions by Thaichon et al. (2018) and Sleep et al. (2020). Sleep et al. (2020) define sales development as an inside sales configuration focusing on lead qualification and appointment setting for outside sales. Terho et al. (2022) expand this definition, viewing the function as an integrated organizational role between marketing and sales, generating and nurturing leads, both inbound and outbound, focusing on both prospects and accounts.

As previously noted, sales development is an organizational function that bridges marketing and sales. It focuses on setting appointments for the sales force and maximizing the benefits from leads generated by marketing (Sleep et al., 2020). Terho et al. (2022) identify four distinct types of sales development processes: outbound prospect-focused, outbound account-focused, inbound prospect-focused, and inbound account-focused. In summary, prospect-focused methods have shorter turnaround cycles and benefit more from automation, while account-focused sales development requires longer cycles and more manual tasks during lead research and nurturing (Terho et al., 2022).

During their study, Terho et al. (2022) identified a few platforms besides the main topic. These findings were divided into "organizational platform decisions," "technological platform-enabling technologies," and "people platform." Each category included demands for sales development, along with benefits and challenges that can be addressed.

Organizational platform decisions refer to the seamless integration between marketing and sales, which is crucial for effective sales development. There should be jointly defined processes with clear responsibilities among units, common lead management terminology, and internal service level agreements. Shared targets, joint target setting, common goals, and aligned incentives are also essential. Additionally, a culture of open communication with regular joint meetings and active demonstration of sales development value is beneficial (Terho et al., 2022).

Technological platform-enabling technologies indicate that, as sales development is largely a back-office function building on digital sales technologies, the technological platforms facilitating this work are central enablers. For successful implementation, the process should include lead management technologies like marketing automation and customer relationship management, alongside other tools. These systems should facilitate key functionalities in sales technology, such as engagement tools for managing digital interactions, analytics and sales intelligence tools for insights, enablement tools for accessing firm content, and pipeline management tools for steering the sales process. Well-organized data, shared databases, integrated systems, and actively updated data are also important (Terho et al., 2022).

People platform refers to managing Sales Development Representatives (SDRs), who are crucial to the sales development process. Job profiles, motivation, training, evaluation, and rewards should be aligned, and SDRs should be correctly led. Typically, SDRs are junior staff in entry-level positions, which needs consideration. The job should be challenging and rewarding to motivate SDRs to perform well. Often, a fixed salary based on performance with shared targets among units is used. It is also important to differentiate between SDRs focused on outbound prospects and those focused on inbound accounts, as different profiles may be required (Terho et al., 2022).

3 FRAMEWORK OF SALES DEVELOPMENT

Researchers examining general sales processes often debate whether to focus on one-time transactions to boost immediate sales or on relationship building to enhance customer lifetime value (CLV). Based on the existing literature, sales development is not suited for one-time transactions. According to Terho et al. (2022), sales development focuses primarily on building customer relationships and generating leads to foster more relationships both within and outside the company. The appropriateness of sales development depends on the product or service a company offers. While existing literature identifies some characteristics that indicate when sales development fits into the sales process and what demands it has for successful implementation, there is still a need for deeper insight. This research aims to provide more understanding of the successful use of sales development as part of a digitalized sales process in the B2B context.

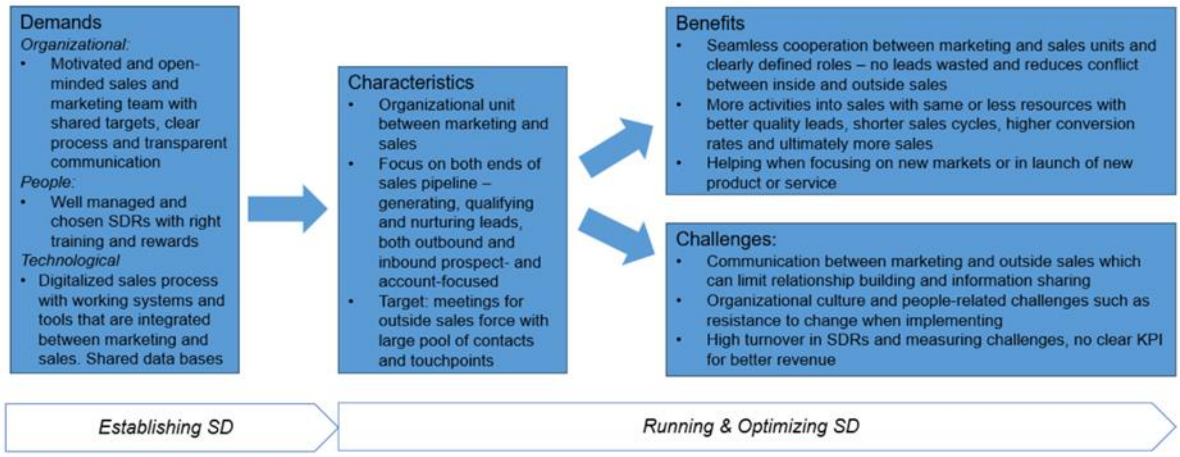


Figure 2: Initial conceptual framework for sales development.

The sales landscape has transformed significantly over recent decades, with the pace of change accelerating. Syam and Sharma (2018) describe this shift as the fourth industrial revolution, driven by digitalization in sales and business. With communication and transactions increasingly moving online, the sales environment is changing drastically. As face-to-face communication diminishes, companies are adopting or naturally transitioning to inside sales teams. This shift necessitates rethinking business strategies to achieve greater efficiency with fewer resources.

While the general sales process has evolved from Dubinsky's (1981) model, the fundamental theme of sales as a person-to-person activity remains. However, as Syam and Sharma (2018) note, AI and machine learning are becoming more integral to sales, signaling further changes in the near future. The trend toward inside sales, driven by the internet and overall digitalization, reduces the need for face-to-face meetings, potentially enhancing sales efficiency. Thaichon et al. (2018) identify this trend as the fourth evolution in sales structures, with the internet playing a crucial role and companies expanding their inside sales forces. This evolution began in the 2000s, emphasizing customer co-creation of value through various channels, supported by advancing technologies that enhance inside sales efficiency.

As inside sales becomes increasingly prevalent, whether through dedicated teams or integrated working methods, companies seek to optimize their sales processes to maximize value for both the company and the customer. Sales structures often blend outside, inside, and online sales, with inside sales playing a major role. One promising configuration for inside sales is sales development, which can significantly enhance company performance (Sleep et al., 2020).

The aforementioned changes have prompted companies to consider establishing and utilizing sales development in their processes. The need for sales development is evident, and its efficient implementation demands certain conditions. The literature suggests that sales development should be positioned between sales and marketing, with the primary goal of arranging sales meetings for the outside sales team. The benefits of implementing sales development include a more efficient and streamlined process. However, challenges related to organizational and personnel issues can arise. An initial conceptual framework for sales development, built on existing theory, is presented in figure 2.

The figure aims to present the main themes and concerns highlighted in the literature, commonly considered by companies when establishing or using sales development. This model serves as an initial framework based on B2B literature and will be refined in the subsequent steps of the research through empirical studies, including personal interviews with representatives from B2B companies.

REFERENCES

Ahuja, V., & Medury, Y. (2010). Corporate blogs as e-CRM tools – Building consumer engagement through content management. Journal of Database Marketing & Customer Strategy Management, 17(2), 91–105.

Albrecht, C., Seley, A., & Heller, K. (2014). Outside in: The rise of the inside sales team. www.zs.com/publications/whitepapers/outside-in-the-rise-of-the-insidesales-team.aspx

Alsaad, A., & Taamneh, A. (2019). The effect of international pressures on the cross-national diffusion of business-to-business e-commerce. Technology in Society, 59, 101158.

Arjona, M. (2022, October 3). The Ultimate Guide to Sales Development. https://predictablerevenue.com/guides/the-ultimate-guide-to-sales-development/

Arli, D., Bauer, C., & Palmatier, R. W. (2018). Relational selling: Past, present and future. Industrial Marketing Management, 69, 169–184.

Dubinsky, A. J. (1981). The effects of sales training. Industrial Marketing Management, 10(2), 129–137.

Fready, S., Vel, P., & Nyadzayo, M. W. (2022). Business customer virtual interaction: Enhancing value creation in B2B markets in the post-COVID-19 era – an SME perspective. Journal of Business & Industrial Marketing, 37(10), 2075–2094.

Gartner. (2019, July 27). The Sales Development Team: A Proven Framework for Success. https://www.gartner.com/en/articles/the-sales-development-team-a-proven-framework-for-success

Gessner, G., & Scott, R. A. (2009). Using business intelligence tools to help manage costs and effectiveness of business-to-business inside-sales programs. Information Systems Management, 26(2), 199–208.

Graesch, J. P., Hensel-Börner, S., & Henseler, J. (2020). Information technology and marketing: An important partnership for decades. Industrial Management & Data Sys-tems, 121(1), 123–157.

Gummings, D. (2014, June 4). Sales Development Team: Most Important Sales Process Innovation in 10 Years. https://davidcummings.org/2014/04/06/sales-development-team-most-important-sales-process-innovation-in-10-years/

Järvinen, J., & Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. Industrial Marketing Management, 54, 164–175.

Johnson, D. S. (2005). Digitization of selling activity and sales force performance: An empirical investigation. Journal of the Academy of Marketing Science, 33(1), 3–18.

Kroger, K. (2013). What Is Inside Sales? The Definition Of Inside Sales; Forbes. https://www.forbes.com/sites/kenkrogue/2013/02/26/what-is-inside-sales-the-definition-of-inside-sales/?sh=4f173fb166d8

Kuruzovich, J. (2013). Sales technologies, sales force management, and online info-mediaries. Journal of Personal Selling & Sales Management, 33(2), 211–224.

Mahlamäki, T., Ojala, M. & Myllykangas, M. 2016. “Drivers towards interactive digital sales processes in business-to-business markets”, AcademicMindtrek '16 Proceedings of the 20th International Academic Mindtrek Conference: Oct. 17th-19th, 2016. New York: ACM, 42-48.

Mantrala, M., & Albers, S. (2022). The impact of the Internet on B2B sales force size and structure. In G. Lilien, A. Petersen, & S. Wuyts (Eds.), Handbook of Business-to-Business Marketing (pp. 493–514). Edward Elgar Publishing.

Martin, S. W. (2013). The trend that is changing sales. Harvard Business Review. https://hbr.org/2013/11/the-trend-that-is-changing-sales

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. Journal of the Academy of Marketing Science, 41(5), 547–566.

Rodríguez, R., Svensson, G., & Mehl, E. J. (2020). Digitalization process of complex B2B sales processes – Enablers and obstacles. Technology in Society, 62, 101324.

Román, S., & Rodríguez, R. (2015). The influence of sales force technology use on outcome performance. Journal of Business & Industrial Marketing, 30(6), 771–783.

Salesloft. (2014, July 7). The History Of Sales Development. https://www.salesloft.com/resources/blog/brief-history-sales-development/

Sheth, J. N., & Sharma, A. (2008). The impact of the product to service shift in industrial markets and the evolution of the sales organization. Industrial Marketing Management, 37(3), 260–269.

Sleep, S., Dixon, A. L., DeCarlo, T., & Lam, S. K. (2020). The business-to-business inside sales force: Roles, configurations and research agenda. European Journal of Marketing, 54(5), 1025–1060.

Sutanonpaiboon, J., & Abuhamdieh, A. (2008). Factors Influencing Trust in Online Consumer-to-Consumer (C2C) Transactions. Journal of Internet Commerce, 7(2), 203–219.

Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Industrial Marketing Management, 69, 135–

Terho, H., Salonen, A., & Yrjänen, M. (2022). Toward a contextualized understanding of inside sales: The role of sales development in effective lead funnel management. Journal of Business & Industrial Marketing, 38(2), 337–352.

Thaichon, P., Surachartkumtonkun, J., Quach, S., Weaven, S., & Palmatier, R. W. (2018). Hybrid sales structures in the age of e-commerce. Journal of Personal Selling & Sales Management, 38(3), 277–302.

Wiersema, F. (2013). The B2B Agenda: The current state of B2B marketing and a look ahead. Industrial Marketing Management, 42(4), 470–488.

Wikipedia. (2024, August 1). Sales development. https://en.wikipedia.org/wiki/Sales_development

Importance of Context Awareness in NLP

Nour Matta, Nada Matta and Philippe Herr
LIST3N, University of Technology of Troyes, 12 Rue Marie Curie, 42060 10004 Troyes Cedex, France
Nour.matta@utt.fr, nada.matta@utt.fr, philippe.herr@utt.fr

Keywords: Context Awareness, NLP, Text Ambiguity, Knowledge.

Abstract: Context is a complex notion, that enables the understanding of happenings and concepts in an environment and the analysis of their influence (Adomavicius et al, 2011) As previously mentioned, context plays a major role in assigning meanings to words, sentences, and texts when dealing with text analysis. Multiple natural language processing approaches aim to consider "context" in analyzing the information extracted and applying a sort of word sense disambiguation (Adhikari et al, 2019). Numerous intelligence systems require knowledge of happening and are context dependent, but the definition of context and context elements used varies from one application to another based on needs. Context plays several roles in text analysis especially to reduce ambiguity and semantic extraction. In this paper, main influence of context on TextMining and NLP are shown.

1 INTRODUCTION

From a semantic perspective, if the purpose was identifying business events, the verb "fire" in the sentence "a company fired 50 employees" can be considered as a business event. In contrast, the same verb in "he fired the gun" is not a business event of interest. We can highlight in this example the necessity of the words' context in understanding their meaning. But when dealing with decisions, the impact of the information extracted must also be considered in the process for two main reasons. First, the same information can have a different impact on different entities. The information "a company fired 50 employees" is considered relatively important for the company's competitors or main clients since this information may be viewed as a sign of struggle in the company of interest but can be irrelevant for unrelated organizations. Second, the information can have a different impact based on the entities involved in the event. For instance, if the company firing employees is a small company of 60 employees in total, this event may mean that the company is more likely to be closing. But if the company originally had more than 5000 employees, firing 50 employees is more or less irrelevant to the financial state of the company.

The dependency on the context of the information while extracting knowledge, from the activity domain to the participating entities, the events mentioned, the time factor, and so on, must be considered in the analysis process. Furthermore, when extracting information from texts, knowledge representation is a required task to enable the accessibility, reuse, and learning process. When dealing with the development of strategies and decision-making based on information extracted from text, the context of this information must be considered to enable the understanding of the information along with the analysis of the importance and the impact of this information.

So, our main research question is: How to deal with the context-dependency of the words semantic in texts?

In this paper, the importance of context awareness is emphasized to consider Natural Language processing techniques.

2 CONTEXT AWARENESS

Bazire and Brézillon (Bazire et al, 2005) used 150 definitions from different domains such as computer science, philosophy, economy, and business, and tried to combine all and abstract key elements. Their research showed that context may be defined by six main components: (1) the constraints, (2) the influence, and (3) the behavior of (4) a system with specific tasks to implement, where the system can be a user or a computer. The context can also be

Matta, N., Matta, N. and Herr, P.
Importance of Context Awareness in NLP.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 280-286
ISBN: 978-989-758-716-0; ISSN: 2184-3228

categorized by its (5) nature and (6) structure. Figure 2.1 shows a representation of the context as the elements that define the context and the interactions or influences between the entities. Five elements are represented in 0 (1) the context, (2) the system, (3) the item, (4) the environment, and (5) the observer. The context is the overall group of entities and how they influence each other. The system is the machine or person having specific tasks to implement on an object in an environment. The item is the object that undergoes changes while the environment can be the organization, the location, or the time in which changes are happening. Finally, the observer is an external element that will have a different opinion on the happenings considering the context. The observer enables the consideration of different cultural backgrounds and social views that might affect the reaction, or the decision made facing an event (Matsumoto, 2007). We can also notice in 0 the interaction between the different entities and the context influence of the context (orange arrow) that each entity has.

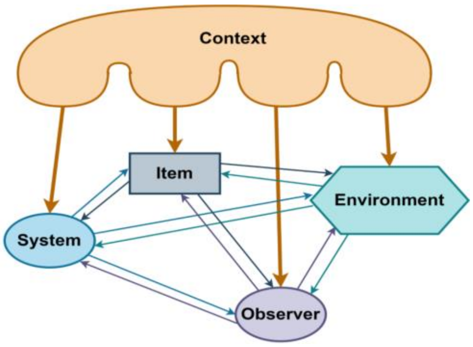


Figure 1: Context element definition and interaction.

3 CONTEXT IN TEXTMINING

3.1 TextMining and NLP

Text mining is the branch of artificial intelligence that aims to extract knowledge from structured and unstructured text (5). Text Mining enables the extraction of the knowledge available in stored textual data. Text Mining is mainly based on two components (5), the data mining and machine learning component and the computational linguistics also known as Natural Language Processing (NLP). On the data processing side, three phases can be identified (6). First, Information Retrieval allows the selection of documents of interest that are most likely related to the topic of interest. Second, data mining, machine learning algorithms, and probabilistic approaches allow the identification of patterns within the extracted data. Third, the Knowledge representation part allows the information extracted to be represented in a formal structure. On the other hand, NLP is used to simulate human's 'natural' understanding of languages to process textual data (Du, 2007). NLP researchers were first split into two divisions: stochastic and symbolic. Stochastic NLP consisted of probabilistic and statistical approaches, focusing on pattern recognition between texts. On the other hand, symbolic NLP also known as rule-based NLP was oriented on formal languages and generating syntax. NLP considers all linguistic levels (Liddy, 2001):

1. Phonetics is the study of the production of sounds
2. Phonology is the study of the arrangement of sounds
3. Morphology is the study of word structure
4. Syntactics is the study of sentence structure
5. Semantics is the study of meanings in a sentence
6. Discourse is the study of syntactic and semantics on units of text longer than a sentence
7. Pragmatics is the study of language in communication

3.2 Context Awareness in Text Analysis

Different definitions of context were provided by linguists (Lichao, 2011). Widdowson (Widdowson, 1996) presented context as a schematic construction of the circumstances of language usage relevant to the meaning. Cook worked on the relationship between literature and discourse and used the context of texts as a form of global knowledge with a (1) broad definition or a (2) narrow definition (Cook, 1994). Lichao (Lichao, 2011) divided context into three categories:

1. Linguistic context refers to the context within the text. It consists of considering the relationship between words, phrases, sentences, and paragraphs. If we consider the word "bank" we need the sentence in which the word was mentioned to be able to properly assign the corresponding meaning. The study of time, place, and people related to happenings mentioned in a text form the deictic context element. Collocation of words falls into this context categorization. Collocation is the grouping of words with their context. For example: barked and dog, born and baby, blond and hair.

2. Situational Context is also known as the context of situation where the environment, time, place, participant of the text, and their relationship form the context. The activity domain or field of text, social relationships, and the mode of text communication are also part of this type of context.
3. Cultural Context as its name indicated considers the cultural background, customs, and past history of the language and the participant (writers or speakers of a discourse). Language is influenced by social factors, social status, gender, or age.

Depending on the discipline requirements, the context of texts in text analysis may play different roles:

- Eliminate Ambiguity in multiple levels: word sentence level and groups of sentences level.
- Improving coreference resolution and indicating referents which is generally used to replace noun phrases or adverbial phrases.
- Detecting Conversational Implicature or Intentions, ie; sarcasm, irony, insults, hurting, pain, caustic, humor, vulgarity, rhetorical questions, metaphors, …

There are two different conceptualizations of context and context used in NLP. The first conceptualization evokes the context of target words in their usage in a text. The second perspective of context is relative to knowledge extraction and the use of ontologies. A popular approach that enables the application of neural networks and machine learning algorithms is the representation of words, sentences, or documents in vectors, considering the "context" which in this case is the surrounding words (Kobayashi, 2018). The research field that provided this approach is Distributional Semantics based on the distributional structure of language theory proposed by Harris (Harris, 1954). Word2vect (Mikolov et al, 2013) and BERT (Devlin et al, 2020) are two models built based on de Distribution semantics. Word2vect was the first model released in 2013, it is unidirectional. In other words, in the example "I went to the bank to sit" and "I the bank to take some money", the word bank would have the same vector because the window considered is before the targeted. While BERT is the first Bidirectional Encoder Representation from Transformers, and the two representations of the word will be different.

The use of ontologies and ontology-based technics for information retrieval purposes was popular between 2000 and 2010 (Wimalasuriya et al, 2010). The use of predefined ontologies to orient and target the information, and the domain of the ontology would play the role of the context. It is that there is a bidirectional relationship between ontologies and natural language processing (Lenci, 2010). Ontologies can be used to orient knowledge extraction from text and NLP can help build and enrich ontologies. Lenci defined four major uses of contexts for onto-lexical knowledge extraction in NLP:

1. Semantic typing is used to characterize the semantic types of linguistic expressions
2. Identify semantic similarity and relatedness in which we try to pair words with similar meanings. In this context, the aim is to identify concepts that belong to the same logical type defined by Sommers (Sommers, 1963).
3. Enable inferences and inheritance of concepts within the same type
4. Argument structure which allows combining constraints of lexical items. Using predefined relationships, using ontologies enables the extraction of concepts and relationships between concepts based on lexical dependencies.

While distributional semantic approaches, such as BERT, are highly performant in machine learning tasks such as classification and annotation, a dependence on the pretraining dataset and the conceptualization of the group built the model. The models still require a need to structure and represent the information extracted while keeping track of the context of extraction. As for ontology-based context, the limit of this approach is the relativity of the knowledge extracted from the text and maintaining the context provided by the text. When using different texts, the categorization of concepts may not identify changes in concept definitions, evolution analysis of the context, and of the elements identified in the context. A need to track temporality and limit the inferences based on their context is identified. In the following section, we will present the context-awareness field, a research field dedicated to studying context and enabling systems to be aware of the context.

4 SITUATION CONTEXT RECOGNITION

Schilit et al. (Schilit, 1995) defined context-awareness as the ability of a system to adapt to a changing environment. In their use case, they worked on a system with mobile users, and the need was for the system to be able to detect the changing location and adapt to it. In their definition context was defined as the environment, the location of users, and the users. As the definition of context changed over time, the context-awareness definition also went through changes over time as the need for context-aware processing was needed. So, situation context can be recognized by identifying the situation object, the actors that provoked changes or made actions, the occurred event, the location of event, the event happening time, the field and domain of situation (Figure 2).

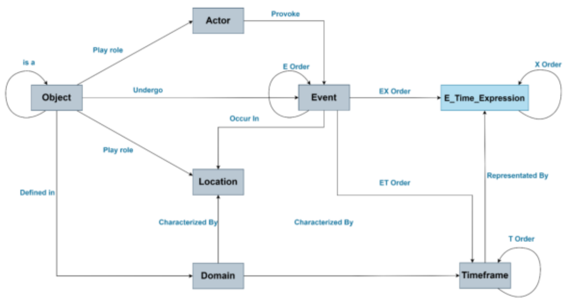


Figure 2: Main situation context elements (Matta et al, 2023).

The VerbNet parser will play a key role in identifying the agents in a sentence and their nature. VerbNet enables distinguishing between the agent 'A0' that is making the action for action verbs (Clark et al, 2021), (Brown et al, 2022). This takes into consideration the active and passive forms of the verb in a sentence. Note that the dimension actor in this approach will be represented by any entity that takes action in the text and will be the 'A0' provided by VerbNet. VerbNet also provides the part of the sentence that reflects the location and the time (Leseva et al, 2022).

Figure 3 provides an output of 4 different sentences. The 2 sentences on the top highlight the actor identification, regardless of the active or passive form of the verb, the parser detects the 'company" as the "A0", the starter of the action. As for the 2 sentences at the bottom, allows the distinction of roles the entity "France" is playing. In the first sentence, France is detected as an actor, therefore, instead of just being the country, it is also the political actor. While in the last sentence, France is the location of the event.

As for the objects, they represent any concept mentioned in the text and are extracted using hypernym/meronym relation extraction (Issa Alaa Aldine, 2022). The structural-based approach was used along with some of the Hearst Patterns (Roller et al, 2018) since they allow keeping track of the

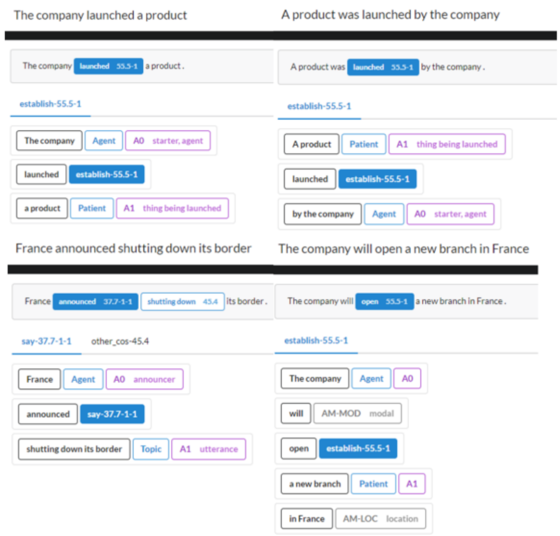


Figure 3: Example of Using VerbNet Parser in CATKoRD platform (23), (24).

semantic extraction based on how it was mentioned in the text. It also enables the discovery of new relation and is not dependent on previously defined lexicons or context independent semantic relations. For instance, in the sentence "Domestic animals such as dogs and cats" we can identify the relation "animal/domestic animal" but also "domestic animal/dog" and "domestic animal/cat". Ontologies are used to enhance inheritance among concepts within the same text and comparing objects from multiple text considering their context.

The CAToRD (Matta et al, 2023) has been developed based in these principles to identify situation context elements from text.

5 CULTURAL CONTEXT RECOGNITION

One dimension of cultural context can be related to the social evolution of a culture by analyzing the literature heritage of a civilization. When observing linguistics actors on literature text analysis, several dimensions can be identified:

- Text title type recognition that leads to text style identification.
- Text Blocs identification that emphasizes the organization of the document.
- Authors and references identification that help on literature type selection.
- Language Analysis to identify linguistics forms related to literature type.

As first step of the methodology, we want to define for cultural context recognition, a linguistic expert has been observed when analyzing two types of text from: French literature one and public scientific newspaper. These first results will help us to determine main aspects to consider in cultural context. That can be the foundations of cultural context ontology and NLP dedicated algorithms definition.

For instance, analyzing the "The Wolf and the Lamb" text leads to:

- Title of the text concerns two animals that have different characteristics.
- Text is decomposed on three blocs:
    1. An introduction that introduces the situation of two animals
    2. Discourses between two animals
    3. A conclusion on one sentence that emphasizes the text morality.
- Author: "La Fontaine", with reference, the title of the book: "Les Fables De La Fontaine" that leads to recognize the style of the literature; critics and morality documents.
- Linguistic analysis that helps to identify a conflict between a strong and weak characters.

These aspects push the linguistic analyst to isolate sentences that emphasizes this conflict (0):

- Space dimensions: The Wolf is higher than the lamb on the riverside.
- Social dimensions: social positions of the Wolf and the lamb in the society. "Magesty", "you and your family disorder my life" …
- Environmental dimensions: division of earth properties. "For you spare me little, You, your shepherds, and your dogs." …

This analysis put on a progress in conflict expression: from simple one between two animals to deeper one related to the humanity control of the environment and its impact.

In this type of text analysis, classical NLP algorithms are not sufficient to detect these types of dimensions. Analyzing sentences cannot enhance documents analysis. Cultural related to literature types must be defined. Semantic representations can be used as references that guide supervised NLP algorithms to detect such type of aspects.

Title : The wolf and the lamb **Author :** Jean de La Fontaine (1668) **Situation description:** A Lamb would drink In the current of a pure wave. A wolf comes on an empty stomach seeking adventure, And that hunger in these places attracted. **Discourse:** Who makes you so bold to disturb my drink? Says this rabid animal: **Social dimension:** You will be punished for your recklessness. 'Sire,' answers the Lamb, 'may your Majesty Do not get angry; But rather than it considers That I'm going to drink **Place dimension:** In the river, more than twenty steps below Her, And therefore, in no way, I can't disturb his drink.	You trouble her, said this cruel beast, And I know that from me you were gossiping last year. "How would I have done it if I had not been born?" The Lamb said, I will stitch my mother. **Social dimension:** "If it is not you, then it is your brother." "I have none." "So, it is someone of yours: For you spare me little, You, your shepherds, and your dogs. **Environmental dimension:** I've been told I need revenge. On it, deep in the forests The Wolf takes it, and then eats it, Without any other form of trial. **Morality:** The reason for the strongest is always the best.

Figure 4: Analysis of a Fable of La Fontaine[1].

Title : This tectonic plate is tearing under the plateau of Tibet! **Author :** Morgane Gillard (2024) **Problem Introduction:** For a long time, the question of the altitude of the Tibetan plateau has been tugging at scientists. If it is clear today that it is linked to a strong crustal thickening in connection with the Himalayan collision, the precise mechanisms remain poorly understood. A new study, however, brings a new breath to the debate: the Indian plate would be tearing in two! **Situation description:** The Himalayan chain results from the formidable collision between the Indian and Eurasian plates that began about 60 million years ago. However, the impressive rise of the Tibetan plateau has long intrigued scientists. Located north of the mountain range, this plateau regularly exceeds 5,000 metres. How can we explain the altitude maintenance of such a vast expanse of continental crust? **Problem description:** There is, of course, crustal thickening related to the collision and the passage of the Indian plate under the Eurasian plate. While it is on average 35 kilometers thick, the crust under the Tibetan plateau would thus reach 75 kilometers! Imagine an iceberg with ice added from the base: this increase in ice volume would cause the surface to rise. It's somewhat the same here. But the exact mechanisms of this thickening are still poorly understood.	**First Hypothesis:** " Plusieurs modèles existent. Certains mettent en avant la nature continentale de la plaque indienne et sa forte flottabilité. Poussée de force sous l'Eurasie, elle n'arriverait ainsi pas à plonger dans le manteau et viendrait alors se plaquer sous la croûte chevauchante, menant à un doublement de l'épaisseur crustale. D'autres modèles proposent au contraire que la plaque indienne se comprime et se déforme comme un tapis que l'on pousse contre un mur, entraînant un épaississement crustal et permettant à la partie inférieure de la plaque de s'enfoncer dans le manteau. **Second Hypothesis:** A team of researchers has come up with a new model that could reconcile these two visions. The results, presented at the international conference of the AGU, the American Geophysical Union, and available in a pre-printed version of the article, indeed suggest that the Indian subduction plate is «tearing»: the lower part, more dense, would be flowing in the mantle, thus separating from the upper part which would remain plated under Eurasia. A bit like a sole taking off from an old shoe.A hypothesis supported by seismological data and numerical modeling, and which opens new perspectives in the way of considering the evolution of the Himalayas and the Tibetan plateau.

Figure 5: Example of Analysis of a problem-solving scientific report[2].

Documents can be not only from the literature but belong to specific fields, activities, companies. For instance, technical reports are decomposed on several blocks, in which the problem is first described "*The question of the altitude of Tibetan problem…*", then observations are detailed before describing "*The Himalayan chain results from the formidable collision…*" how actors face the problem and give a solution. The title of these reports put on generally the encountered problem by presenting two hypotheses of "*floatability of the plate*" or "*separation of the Eurasian plates*". As same as, technical documents or contractual ones are presented related to reasoning schemas in which different blocks reflect actors' problem solving (0). Semantic representations and domain ontologies must be considered to emphasize

[1] https://www.poetica.fr/poeme-849/jean-de-la-fontaine-le-loup-et-agneau/

2 https://www.futura-sciences.com/planete/actualites/tectonique-plaques- cette-plaque-tectonique-train-dechirer-sous-plateau-tibet-110914/

companies' cultural context and guide the NLP analysis of these type of documents.

6 CONCLUSION

Currently, NLP techniques tend to use LLM and Generative AI to analyze texts. But this type of techniques still be hard to consider specific context of activities which are necessary to emphasize semantics of a document. In fact, they ask to define several specific prompts (Feldman et al, 2023) and don't put on global techniques for this aim. In this paper, the importance to consider context in NLP algorithms has been shown based on our first studies on this domain. Firstly, some techniques to detect situation context has been mentioned and secondly, importance of cultural context to analyze documents are emphasized.

This paper presents our first study to detect cultural context. Two types of texts have been analyzed manually to identify important parts to consider in cultural context. We aim at studying cultural works to define a dedicated ontology. Then, CAToRD platform will be augmented by extending NLP algorithms that help to consider cultural context. Global rules will be then defined to be integrated in NLP applications and LLM algorithms.

REFERENCES

Adhikari A., Ram A., Tang R. , and Lin J., 'DocBERT: BERT for Document Classification', no. *arXiv:1904.08398. arXiv, Aug. 22, 2019. doi: 10.48550/arXiv.1904.08398*.

Adomavicius G., B. Mobasher B., F. Ricci B., and Tuzhilin A., 'Context-Aware Recommender Systems', *AI Magazine, vol. 32, pp. 67–80, Sep. 2011, doi: 10.1609/aimag.v32i3.2364.*

Bazire M. and P. Brézillon P., 'Understanding Context Before Using It', in *Modeling and Using Context*, A. Dey, B. Kokinov, D. Leake, and R. Turner, Eds., *in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 29–40. doi: 10.1007/11508373_3.*

Brown S.W., Bonn J., Kazeminejad G., Zaenen A., Pustejovsky J., and Palmer M., 'Semantic representations for nlp using verbnet and the generative lexicon', *Frontiers in artificial intelligence*, vol. 5, p. 821697, 2022.

Clark P., Dalvi B., and Tandon N., 'What Happened? Leveraging VerbNet to Predict the Effects of Actions *in Procedural Text', arXiv:1804.05435 (cs), Apr. 2018, Accessed: Oct. 20, 2021. (Online). Available: http://arxiv.org/abs/1804.05435*

Cook G., 'Discourse and Literature', *Oxford University Press, 1994, p. 24.*

Devlin J., Chang M.W., Lee K., and Toutanova K., 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', arXiv:1810.04805 (cs), May 2019, Accessed: Nov. 02, 2020. (Online). Available: http://arxiv.org/abs/1810.04805

Du M., *Natural language processing system for business intelligence', p. 88, 2017.*

Feldman P., Foulds, J. R., & Pan, S. (2023). Trapping llm hallucinations using tagged context prompts. *arXiv preprint arXiv:2306.06085.*

Ghosh M., Roy M., Bandyopadhyay S., and Bandyopadhyay K., 'A tutorial review on *Text Mining Algorithms', Jun. 2012.*

Harris Z.S., 'Distributional structure', *Word*, vol. 10, pp. 146–162, 1954, doi: 10.1080/00437956.1954.11659520.

Issa Alaa Aldine A., 'Contributions to Hypernym Patterns Representation and Learning based on Dependency Parsing and Sequential Pattern Mining', *These de doctorat, Lorient, 2020. Accessed: Jun. 13, 2022. (Online). Available: http://www.theses.fr/2020LORIS575*

Kobayashi S., 'Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations', in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 452–457. doi: 10.18653/v1/N18-2072*

Lenci, A. The life cycle of knowledge. Ontology and the Lexicon. A Natural Language Processing Perspective. Cambridge University Press, Cambridge, UK, 241-257.2010

Leseva S. and Stoyanova, I. 'Linked Resources towards Enhancing the Conceptual Description of General Lexis Verbs Using Syntactic Information', in *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 2022.

Lichao S., 'The Role of Context in Discourse Analysis', *Journal of Language Teaching and Research*, vol. 1, Nov. 2010, doi: 10.4304/jltr.1.6.876-879.

Liddy E., 'Natural Language Processing', *School of Information Studies - Faculty Scholarship*, Jan. 2001, (Online). *Available: https://surface.syr.edu/istpub/63*

Mikolov T., Chen K., Corrado, G. and Dean J., 'Efficient Estimation of Word Representations *in Vector Space', no. arXiv:1301.3781. arXiv, Sep. 06, 2013. doi: 10.48550/arXiv.1301.3781.*

Matta N., Matta N., Giret E., and Declercq N., 'Enhancing Textual Knowledge Discovery using a Context-Awareness Approach', in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2021, pp. 233–237. doi: 10.1109/CSCI54926.2021.00071.

Matta N., Matta N., Marcante A., and Declercq N., 'CATKoDR: Hybrid Context-Awareness Model

Architecture for Natural Language Processing', 2023, *Accessed: Apr. 08, 2024. (Online). Available: https://ieeesmc2023.org/abstract_files/SMC23_1543_FI.pdf*

Matta N., Matta N., Declercq N., and A. Marcante, 'Semantic Patterns to Structure TimeFrames in Text', *in INTELLI 2022, The Eleventh International Conference on Intelligent Systems and Applications, May 2022, pp. 16–23. Accessed: Sep. 13, 2022. (Online). Available: https://www.thinkmind.org/index.php?view=article&articleid=intelli_2022_1_40_60016*

Matsumoto D., 'Culture, Context, and Behavior', *Journal of Personality*, vol. 75, no. 6, Art. no. 6, 2007, doi: *10.1111/j.1467-6494.2007.00476.x.*

Roller S., Kiela S. and Nickel D. (2018). Hearst patterns revisited: Automatic hypernym detection from large text corpora. *arXiv preprint arXiv:1806.03191*.

Schilit W.N., *A system architecture for context-aware mobile computing*. *Columbia University, 1995.*

Sommers F., 'Types and Ontology', *The Philosophical Review, vol. 72, no. 3, Art. no. 3, 1963, doi: 10.2307/2183167.*

Widdowson H.G., 'Linguistics', in *Linguistics*, OUP Oxford, 1996, p. 126.

Wimalasuriya C. and D. Dou D., 'Ontology-based information extraction: An Introduction and a survey of current approaches', *J. Information Science, vol. 36, pp. 306–323, May 2010, doi: 10.1177/0165551509360123.*

Exploring Knowledge Sharing Motivational Factors for Intellectual Property Lawyers: A Conceptual Framework

Nopphawat Ariyahchatraweekul[1][a], Mongkolchai Wiriyapinit[2][b] and Jintavee Khlaisang[3][c]
[1]*Technology Management and Innopreneurship Program, Graduate School, Chulalongkorn University, Bangkok, Thailand*
[2]*Department of Commerce, Chulalongkorn Business School, Chulalongkorn University, Bangkok, Thailand*
[3]*Center of Excellence in Educational Invention and Innovation, Faculty of Education, Chulalongkorn University, Bangkok, Thailand*
6481024720@student.chula.ac.th, mongkolchai@cbs.chula.ac.th, jintavee.m@chula.ac.th

Keywords: Knowledge Sharing, Lawyer, Intellectual Property, Law Firm.

Abstract: This study aims to explore the motivational factors affecting knowledge sharing among lawyers in intellectual property law firms. Through an integrative review of existing empirical and theoretical literature, the research develops a conceptual framework to understand these motivational factors. The findings identify eight key extrinsic factors and two intrinsic factors influencing knowledge sharing for lawyers in intellectual property law firms. This research provides valuable insights into motivational factors that can be used by intellectual property law firms to improve knowledge sharing practices among lawyers.

1 INTRODUCTION

Intellectual property law is a complex legal field due to the diversity of intellectual property (IP) types, such as patents, trademarks, copyrights, and trade secrets. Each type of IP is governed by specialized laws. In Thailand, patents for inventions, utility models, and design patents fall under the Patent Act B.E. 2522 (1979), which is considered the most complex IP law compared to others. Lawyers and legal counsels specializing in IP law must not only understand the legal texts but also possess technical knowledge in scientific and technological fields, such as engineering, chemistry, biotechnology, and pharmacy. This interdisciplinary expertise is essential for effective prior art searches, legal opinions, patent drafting, rights protection, infringement prevention, and litigation (Levin & Cross, 2004).

IP law firms are unique in that they require personnel with specialized knowledge in both IP law and related technical fields. Besides lawyers, these firms often employ scientists and technologists, who may be either legal professionals with additional science and technology education or experts with scientific and technological qualifications. Collaboration and knowledge sharing among these professionals are crucial for providing comprehensive legal services to clients, which distinguishes IP law firms from general law firms where lawyers typically work independently (Levin & Cross, 2004). The core operations of IP law firms are knowledge-intensive, requiring in-depth legal expertise and professional skills to advise clients on IP protection and utilization.

The rapid advancement of technology and globalization have transformed the legal profession, increasing the demand for adaptable, knowledgeable, and skilled lawyers (Flood, 2012). Efficient knowledge sharing among legal professionals within law firms has become more important to facilitate continuous learning and mutual assistance (Gardner, 2019). Knowledge sharing is the process of exchanging tacit and explicit knowledge between individuals or groups to utilize or create new knowledge (Nonaka & Takeuchi, 1995). For IP lawyers, tacit knowledge includes legal analysis, case prediction, negotiation, and understanding client industries, all of which require experience. Explicit knowledge involves documented legal texts, legal opinions, court decisions, and IP application preparation.

[a] https://orcid.org/0009-0006-7925-4055
[b] https://orcid.org/0009-0008-4959-9225
[c] https://orcid.org/0000-0002-7572-9782

Ariyahchatraweekul, N., Wiriyapinit, M. and Khlaisang, J.
Exploring Knowledge Sharing Motivational Factors for Intellectual Property Lawyers: A Conceptual Framework.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 287-294
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Client trust in law firms is built not only on the reputation of individual lawyers but also on the firm's collective expertise in specific legal fields. Effective legal services depend on the knowledge capital within the firm (Forstenlechner, Lettice, Bourne, & Webb, 2007). Studies have shown that knowledge management can support legal services, and knowledge sharing, as a part of knowledge management, is crucial for law firms (Zeide & Liebowitz, 2012; Kabene et al., 2006). Successful law firms focus on best practices for utilizing their personnel's knowledge (Fombad, Boon, & Bothma, 2009). According to Lambe (2003), law firms operate in a highly competitive and constantly changing environment, necessitating efficient knowledge sharing for maintaining expertise.

However, research by Schulz & Klugmann (2005) indicates that implementing knowledge management systems in law firms often faces challenges due to lack of acceptance. Many legal professionals may not recognize the importance of such systems or may view them as disruptive to traditional practices. This resistance can be particularly strong among senior lawyers who may see knowledge management as unnecessary, despite its potential benefits for professional practice.

Research by Meso, Bosire, & Massey (2023) highlights the importance of knowledge management in enhancing law firm performance. Their study in Nairobi, Kenya, involving 222 law firms, found that 86.5% of respondents recognized the importance of knowledge management, and 64.9% noted its positive impact on their firms' efficiency. They emphasized the need for investments in information and communication technology to support knowledge management.

Adeyemi et al. (2022) studied knowledge transfer and use in Nigerian law firms, finding significant positive correlations between these practices and firm performance. Knowledge sharing through meetings, training, seminars, and collaborative work improved creativity, financial performance, and client satisfaction.

Holinde (2015) identified intrinsic and extrinsic motivational factors influencing knowledge sharing in law firms. Intrinsic factors include altruism and self-efficacy, while extrinsic factors involve organizational rewards, reciprocity, and the use of technology.

Research by Kabene et al. (2006) identified six critical factors for successful knowledge management in law firms: culture, trust and loyalty, communities of practice, human resources roles, motivation and rewards, and the role of technology.

The absence of effective knowledge-sharing systems in law firms can lead to knowledge loss, inconsistent legal service quality, and hindered innovation (Schulz & Klugmann, 2005; Saunders, 2011; Evans et al., 2015). Despite existing research, there is limited empirical study on knowledge sharing among lawyers and the motivational factors affecting this process in IP law firms. The complexity of IP laws, the conflicts between the IP laws of different countries protecting intellectual property rights, and the challenges posed by emerging technologies create numerous IP touchpoints and uncertainties.

Given the importance and challenges mentioned above, as well as the lack of existing research on IP lawyers in IP law firms regarding the motivational factors influencing knowledge sharing, this study aims to explore the factors influencing knowledge sharing among IP lawyers and in IP law firms. The findings will contribute to identifying key motivational factors that significantly influence knowledge sharing among lawyers in law firms. This is achieved through an integrative review of existing empirical and theoretical literature, which also proposes a conceptual framework. The research question for this study is: What are the motivational factors affecting knowledge sharing among lawyers in IP law firms?

2 LITERATURE REVIEW

The authors conducted an integrative review of existing empirical and theoretical literature to provide a comprehensive understanding of knowledge sharing among lawyers in law firms. The review identifies both extrinsic and intrinsic motivational factors that influence knowledge sharing and offers insights into how these factors can foster knowledge sharing within law firms.

2.1 The Importance of Intellectual Property Law and Its Challenges in the Digital Era

The World Intellectual Property Organization (WIPO) indicates that hat in today's rapidly evolving world, IP and intangible assets are gaining increasing significance. One of the crucial challenges is to ensure that the current IP system continues to foster innovation in the era of advanced technologies, thereby driving economic growth globally. Additionally, IP has become the most valuable asset class worldwide (Ogier, 2016).

Intellectual property protection is provided by laws such as patents, copyrights, and trademarks, which allow individuals to gain recognition or financial benefits from their inventions or creations. By balancing the interests of innovators with those of the public, the IP system creates an environment conducive to creativity and innovation. IP covers a broad range of activities and is vital to both cultural and economic life, safeguarded by various national and international laws (WIPO, 2020).

The digital economy has had an impact on IP law, as evidenced by legislation addressing cybersquatting and significant advancements in legal and economic protections (Kahn & Wu, 2020). The emerging of generative artificial intelligence (AI) tools, such as ChatGPT, Midjourney, Copilot and Firefly, presents numerous IP issues and uncertainties regarding IP infringement, IP rights ownership of AI-generated outputs. To mitigate risks concerning IP rights and ownership of AI outputs, it is essential to review the terms and conditions of generative AI tools to determine who owns the intellectual property in the outputs (WIPO, 2024).

2.2 The Role and Types of Intellectual Property Law Firms

Sammon (2024) highlights the key roles of IP law firms. Firstly, they must serve as reliable business advisors, understanding each client's specific needs and business context to provide tailored legal advice, even for seemingly simple tasks like trademark registration. Beyond protecting IP rights, IP law firms should help clients plan and implement IP strategies, identify valuable IP assets, and secure the necessary protections.

Sammon categorizes IP law firms into two types: traditional and modern.

Traditional IP Firms operate as partnerships with partners, associate attorneys, solicitors, trainees, and paralegals. They offer services on fixed fees and hourly rates, including:

- Assessing IP infringement risks
- Filing applications for IP protection
- Advising on IP acquisition from state offices
- Litigating IP infringement cases
- Defending against IP infringement lawsuits
- Renewing registered IP rights
- Advising on copyright issues.

Modern IP Firms have emerged due to regulatory modernization and alternative business structures. They offer flexible fee structures, addressing concerns about hourly rates and providing budget clarity, which benefits startups and SMEs. This approach helps clients plan their budgets effectively and make informed decisions about their projects without financial uncertainty.

2.3 Knowledge Management in Law Firms

Client trust in law firms is built not only on the reputation of the lawyers but also on their knowledge, skills, and expertise in specific legal areas. Effective legal services depend on the knowledge capital that the law firm possesses (Forstenlechner, Lettice, Bourne, & Webb, 2007). According to Zeide & Liebowitz (2012), knowledge management supports lawyers in delivering legal services, with knowledge sharing being a crucial aspect, driven by lawyers' accumulated experience (Kabene et al., 2006). Therefore, law firms focus heavily on best practices for utilizing their personnel's knowledge (Fombad, Boon, & Bothma, 2009). In legal practice, knowledge management involves applying legal knowledge to specific client issues to find solutions. Lawyers represent clients in legal matters, present evidence and legal arguments, and advise on legal rights and obligations. Effective knowledge management can enhance service quality and client satisfaction, particularly in knowledge-intensive organizations like law firms (Gottschalk, 2002).

Many researchers focus on knowledge management in law firms (Rusanow, 2003; Edwards & Mahling, 1997; Gottschalk, 1999; Parsons, 2004; Du Plessis, 2004; Holinde, 2015; Meso et al., 2023). Law firms are particularly suited for knowledge management research due to their reliance on knowledge to operate and create value (Forstenlechner, 2005). Key aspects of knowledge management include improving organizational efficiency through knowledge control (Van Engers, 2001).

Effective knowledge management in law firms reduces errors, avoids duplication, solves problems faster, and enhances decision-making (Olson, 1971; Hayek, 2013; Coase, 1937). It also strengthens client relationships and improves services. However, law firms need ongoing support for knowledge management implementation and improvement. Some firms have introduced roles like Chief Knowledge Officer (CKO) to facilitate knowledge management and drive innovation (Apistola & Gottschalk, 2012).

2.4 Knowledge Sharing Among Lawyers in Law Firms

Knowledge sharing is essential for transforming law firms into learning organizations. Clients expect excellent legal services, including prompt and robust communication from lawyers. Effective knowledge sharing and collaboration within teams enhance law firm performance (Apistola & Gottschalk, 2012). Van den Brink (2003) emphasizes that knowledge sharing is a prerequisite for knowledge development in law firms, as it allows individuals to build on existing knowledge (Nahapiet & Ghoshal, 1998).

Laudon & Laudon (2011) argue that knowledge must be communicated and shared to be useful. Successful knowledge sharing involves not only intra-departmental but also inter-departmental and even inter-organizational exchange. Legal fields often overlap, and lawyers in large firms may work across multiple legal areas, necessitating sufficient knowledge in all relevant fields to provide the best advice. Effective knowledge sharing among lawyers is crucial for achieving this (Apistola, 2006).

Studies show that lawyers typically share knowledge within their departments but struggle to do so across departments. Senior lawyers often lack time to reflect on and share their experiences (Rusanow, 2002; Khandelwal & Gottschalk, 2003; Hunter et al., 2002). Apistola & Lodder (2005) identified that knowledge processes include activities related to creating, sharing, using, and preserving knowledge, enabling efficient knowledge transfer and utilization.

2.5 Motivational Factors Affecting to Knowledge Sharing Among Lawyers in Law Firms

Understanding the motivational factors that influence knowledge sharing among lawyers is crucial for fostering an environment conducive to collaboration and learning within law firms. This section explores both extrinsic and intrinsic motivational factors that encourage knowledge sharing, drawing on insights from existing empirical and theoretical literature.

2.5.1 Extrinsic Motivational Factors

Extrinsic motivational factors typically come from management and include bonuses, special awards, and other incentive methods that encourage employees to share knowledge (Disterer, 2003). From a review of the literature on knowledge sharing among lawyers and legal counsels, several extrinsic motivational factors have been identified.

Organizational culture is crucial for promoting knowledge sharing in law firms. Wong (2005) states that firms must develop a culture that values and promotes knowledge sharing, development, and use. Schulz & Klugmann (2005) highlight that a knowledge management system fosters this culture among lawyers. Key requirements include management commitment, performance evaluations based on desired behaviors, and hiring aligned with the new culture (Kabene et al., 2006). Organizational culture in law firms consists of shared norms, values, and perceptions developed through interactions, making it challenging to change (Hofstede et al., 1990).

Team interaction can motivate knowledge sharing and the development of tacit knowledge (Gore & Gore, 1999). This includes reciprocity, where mutual benefits are exchanged (Holinde, 2015). Law firms should emphasize personal relationships and clear communication to enhance knowledge sharing (Osterloh & Frey, 2000). Increasing interdependence among lawyers and fostering a collaborative environment are key strategies for improving knowledge sharing (Huysman & De Wit, 2000).

Extrinsic rewards are significant motivators. These include tangible financial incentives like salary increases, bonuses, vacations, or promotions (Hau et al., 2016; Bock & Kim, 2002; Nguyen et al., 2021). Knowledge sharing should be integral to opportunities for becoming a partner in a law firm (Apistola, 2006). Performance evaluations tied to knowledge sharing can encourage lawyers to support their colleagues (Huysman & De Wit, 2000).

Knowledge champions within the firm can stimulate knowledge sharing. Tjaden (2007) suggests leveraging the influence of these individuals, often partners, who naturally share their knowledge in daily activities, significantly impacting their peers' behavior.

Time allocation is another critical factor. In large law firms, time is often equated with money, making it challenging to allocate time for knowledge management activities (Terret, 1998; Gottschalk, 1999). Providing dedicated time for meetings and interactions can help foster a culture of knowledge sharing (Huysman & De Wit, 2000). Changing the billing structure from hourly rates to value-based billing can incentivize more efficient work and emphasize the importance of knowledge sharing (Rusanow, 2003).

Management involvement is essential for fostering a knowledge-sharing culture. Knowledge sharing relies on consistent, reliable, and credible management behavior (Disterer, 2003). Management must lead by example, actively communicating and reflecting on knowledge-sharing practices. Support from top management positively correlates with employees' perceptions of a knowledge-sharing culture and their willingness to share knowledge (Connelly & Kelloway, 2003; Lin, 2007).

Annual appraisements can effectively motivate lawyers to share knowledge. Forstenlechner (2005) suggests that incorporating knowledge sharing into annual performance evaluations ensures it is viewed as an essential part of professional development and organizational support.

Investing in technology is crucial for supporting knowledge management in law firms (Meso et al., 2023). Digital technologies enhance coordination and communication, significantly improving knowledge sharing and organizational performance (Deng et al., 2023). Technologies such as email, instant messaging, social media, blogs, wikis, discussion forums, video conferencing, document sharing tools, and web conferencing tools facilitate knowledge sharing among lawyers (Olatokun & Nneamaka, 2013). Intranets and company databases enable employees to share experiences and knowledge, promoting collaboration (Arora, 2002).

These factors illustrate the various extrinsic motivations that can encourage lawyers to share knowledge within their firms, ultimately enhancing the efficiency and effectiveness of legal services.

2.5.2 Internal Motivational Factors

Intrinsic motivational factors are non-monetary incentives that have psychological impacts and play a crucial role in encouraging individuals to share knowledge (Kabene et al., 2006). Osterloh & Frey (2000) noted that intrinsic motivation promotes tacit knowledge sharing, particularly when extrinsic motivations fail. From a literature review on knowledge sharing in both general organizations and law firms, the intrinsic motivational factors identified include intrinsic rewards and trust.

Intrinsic rewards refer to psychological incentives such as oral praise (Donnelly, 2018), recognition (Malek et al., 2020), reputation (Hung et al., 2011; Nguyen & Malik, 2020; Choi et al., 2008), and altruism (Holinde, 2015). These rewards motivate individuals to share knowledge to help others and demonstrate collegiality (Kabene et al., 2006). Law firms should acknowledge and enhance the reputation of lawyers who actively participate in knowledge sharing (Hunter et al., 2002). Forstenlechner, citing Schulz & Klugmann (2005), stresses the importance of intrinsic rewards like praise and recognition from senior management to highlight effective knowledge management.

Trust among colleagues significantly influences knowledge sharing. Evans et al. (2015) found that perceived trustworthiness mediates the relationship between social factors, like shared language and vision, and knowledge-sharing behavior. Trust enables employees to share knowledge more freely and use it effectively. It is a critical component of interpersonal interactions and can be a crucial motivator for knowledge sharing (Van den Brink, 2003). When employees recognize that their knowledge is handled carefully, they are more likely to share it (Hall, 2001; Huysman & De Wit, 2002). Trust is needed not only between individuals sharing knowledge but also within the organization as a whole (Hinds & Pfeffer, 2001).

Law firms must cultivate an attitude of trust among their members (Disterer, 2003). Mutual trust is essential for open sharing (Disterer, 2003), as it reduces the fear of opportunistic behavior. Building trust between employees and different groups facilitates more open and proactive knowledge sharing (Wong, 2005). Organizational development processes should establish a set of shared ethical standards and values for the law firm and reach consensus on acceptable practices and work habits (Disterer, 2003). These standards and values should be clearly communicated throughout the law firm to foster trust. Ensuring sufficient interaction among employees is also crucial for building trust (Kabene et al., 2006).

These intrinsic motivational factors highlight the importance of psychological incentives and trust in fostering a knowledge-sharing culture within law firms, ultimately enhancing their operational efficiency and effectiveness.

Despite existing research, there is limited empirical study on knowledge sharing among lawyers and the motivational factors affecting this process in IP law firms. The complexity of IP laws, the conflicts between the IP laws of different countries protecting intellectual property rights, and the challenges posed by emerging technologies create numerous IP touchpoints and uncertainties. Given the importance and challenges mentioned above, IP lawyers need to continuously update, discuss, and share knowledge among their teams to provide up-to-date legal advice that meets client satisfaction.

Table 1: Motivational factors affecting knowledge sharing among lawyers in IP law firms.

Factors		Authors
Extrinsic Motivation	Organizational Culture	(Wong, 2005; Kabene et al., 2006; Hofstede et al., 1990)
	Team Interaction	(Gore & Gore, 1999; Osterloh & Frey, 2000; Huysman & De Wit, 2000)
	Reward	(Hau et al., 2016; Bock & Kim, 2002; Nguyen et al., 2021; Apistola, 2006)
	Knowledge champions	Tjaden (2007)
	Time Allocation	(Terret, 1998; Gottschalk, 1999; Huysman & De Wit, 2000; Rusanow, 2003)
	Management Involvement	(Disterer, 2003; Connelly & Kelloway, 2003; Lin, 2007)
	Annual Appraisement	Forstenlechner (2005)
	Technology	(Meso et al., 2023; Deng et al., 2023; Olatokun & Nneamaka, 2013; Arora, 2002)
Intrinsic Motivation	Oral Praise	Donnelly (2018)
	Recognition	(Malek et al., 2020; Schulz & Klugmann, 2005)
	Reputation	(Hung et al., 2011; Nguyen & Malik, 2020; Choi et al., 2008; Hunter et al., 2002)
	Altruism	Holinde (2015)
	Trust	(Evans et al., 2015; Hall, 2001; Huysman & De Wit, 2002; Disterer, 2003; Wong, 2005; Kabene et al., 2006)

3 FINDING

The findings of this study, derived from an integrative review of existing empirical and theoretical literature, identify eight key extrinsic motivational factors and five intrinsic motivational factors, as summarized in Table 1, that significantly influence knowledge sharing among lawyers in law firms.

Based on the literature review and identified factors, the author proposes the conceptual framework shown in Figure 1. This framework illustrates the extrinsic and intrinsic motivational factors and their impact on knowledge sharing among lawyers in IP law firms.

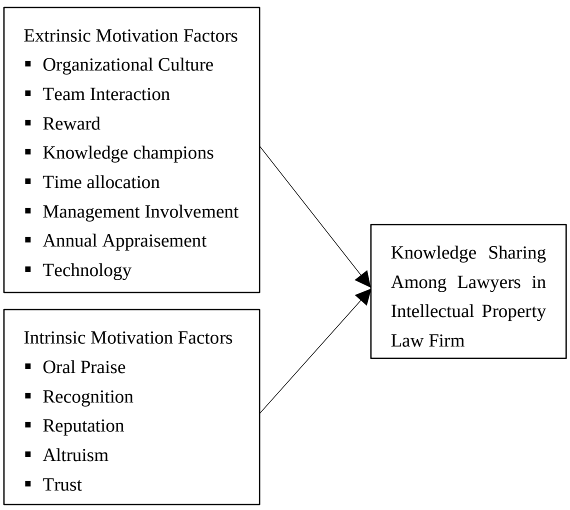


Figure 1: Conceptual Framework.

4 DISCUSSION AND CONCLUSIONS

This study emphasizes the significance of both extrinsic and intrinsic motivational factors in fostering knowledge sharing among lawyers in IP law firms. Our research contributes to the filed by identifying eight key extrinsic motivational factors and five intrinsic motivational factors, offering novel insights specific to IP law firms regarding the factors influencing knowledge sharing among lawyers in this specialized domain.

The extrinsic factors such as a supportive organizational culture, team interactions, rewards, knowledge champions, sufficient time allocation, proactive management involvement, comprehensive annual appraisals, and robust technology investments are crucial in creating a knowledge-sharing environment. These factors provide the necessary structural support to facilitate open communication and collaboration among lawyers, ensuring that knowledge can be efficiently shared within the law firm. The intrinsic factors, including oral praise, recognition, reputation, altruism and trust, are also vital in motivating lawyers to share their knowledge voluntarily and openly. By cultivating trust and offering recognition, law firms can foster an environment where lawyers feel valued and more willing to share their expertise.

By addressing these factors, lawyers in IP law firms can improve their knowledge-sharing practices, leading to enhanced efficiency, innovation, and effectiveness in providing legal services.

It is important to acknowledge that this study is based on a review of existing literature, and the proposed conceptual framework has not been empirically tested. Therefore, the findings are not generalizable to all IP law firms, and further empirical testing is recommended.

REFERENCES

Adeyemi, I. O., Uzamot, W. O., & Temim, F. M. (2022). K*nowledge Transfer and Use as Predictors of Law Firm Performance: Nigerian Lawyer's Perspectives*. International Journal of Knowledge Management (IJKM), 18(1), 1-17.

Apistola, M. (2006). *Advocaat en Kennismanagement (Lawyer and Knowledge Management)*, dissertation VU University, Amsterdam.

Apistola, M., & Gottschalk, P. (2012). *Essential knowledge and management issues in law firms*. Universal-Publishers.

Apistola, M., & Lodder, A. R. (2005). *Law Firms and IT–Towards Optimal Knowledge Management*. Journal of Information, Law and Technology, 2(3), 1-28.

Arora, R. (2002). *Implementing KM–a balanced score card approach*. Journal of knowledge management, 6(3), 240-249.

Bock, G.-W. and Kim, Y.-G. (2002). *Breaking the myths of rewards: an exploratory study of attitudes about knowledge sharing*. Information Resources Management Journal, Vol. 15 No. 2, pp. 14-21.

Choi, S. Y., Kang, Y. S., & Lee, H. (2008). *The effects of socio-technical enablers on knowledge sharing: an exploratory examination*. Journal of information science, 34(5), 742-754.

Coase, R.H. (1937). *The Nature of the Firm*, Economica (4)-16.

Connelly, C. E., & Kelloway, E. K. (2003). *Predictors of employees' perceptions of knowledge sharing cultures*. Leadership & Organization Development Journal, 24(5), 294-301.

Deng, H., Duan, S. X., & Wibowo, S. (2023). *Digital technology driven knowledge sharing for job performance*. Journal of Knowledge Management, 27(2), 404-425.

Disterer, G. (2003). *Fostering Knowledge Sharing: Why and How?*, Proceedings of the IADIS International Conference e-Society (IADIS), Lissabon, pp. 219 – 226.

Donnelly, R. (2018). *Aligning knowledge sharing interventions with the promotion of firm success: the need for SHRM to balance tensions and challenges*. Journal of Business Research, Vol. 94, pp. 344-352.

Du Plessis, T. (2004). *Information and knowledge management in support of legal research in a digital information environment* (Doctoral dissertation).

Edwards, D. L., & Mahling, D. E. (1997). *Toward knowledge management systems in the legal domain*. In Proceedings of the 1997 ACM International Conference on Supporting Group Work (pp. 158-166).

Evans, M., Wensley, A., & Frissen, I. (2015). *The Mediating Effects of Trustworthiness on Social-Cognitive Factors and Knowledge Sharing in a Large Professional Service Firm*. Electronic Journal of Knowledge Management, 13(3), pp240-253.

Flood, J. (2012). *Lawyers as Sanctifiers: The Role of Elite Law Firms in International Business Transactions*. Indiana Journal of Global Legal Studies, 19(1), 109-147.

Fombad, M. C., Boon, J. A., & Bothma, T. J. D. (2009). *Strategies for knowledge management in law firms in Botswana*. South African Journal of Information Management, 11(2), 1-16.

Forstenlechner, I. (2005). *Impact of Knowledge Management on Law Firm Performance. An Investigation of Causality Across Cultures*. (Dissertation) School of Industrial and Manufacturing Science.

Forstenlechner, I., Lettice, F., Bourne, M., & Webb, C. (2007). *Turning knowledge into value in professional service firms*. Performance Measurement and Metrics, 8(3), 146-156.

Gardner, H. K. (2019). *Smart Collaboration: How Professionals and Their Firms Succeed by Breaking Down Silos*. Harvard Business Review Press.

Gore, C., & Gore, E. (1999). *Knowledge management: the way forward*. Total quality management, 10(4-5), 554-560.

Gottschalk, P. (1999). *Use of IT for Knowledge Management in Law Firms*. The Journal of Information, Law and Technology (JILT)-3.

Gottschalk, P. (2002). *A stages of growth model for knowledge management technology in law firms*. The Journal of Information, Law and Technology (JILT), 2, 02-2.

Gottschalk, P., & Khandelwal, V. K. (2003). *Determinants of knowledge management technology projects in Australian law firms*. Journal of Knowledge Management, 7(4), 92-105.

Hau, Y.S., Kim, B. and Lee, H. (2016). *What drives employees to share their tacit knowledge in practice?*, Knowledge Management Research & Practice, Vol. 14 No. 3, pp. 295-308.

Hayek, F. A. (2013). *The use of knowledge in society*. In Modern Understandings of Liberty and Property (pp. 27-38). Routledge.

Hinds, P. J., & Pfeffer, J. (2001). *Why organizations don't "know what they know": Cognitive and motivational factors affecting the transfer of expertise*. Stanford: Graduate School of Business, Stanford University, (pp. 3-26).

Hofstede, G., Neuijen, B., Ohayv, D. D., & Sanders, G. (1990). *Measuring organizational cultures: A qualitative and quantitative study across twenty cases*. Administrative science quarterly, 286-316.

Holinde, S. (2015). *Knowledge sharing in an international law firm–A case study examining practice, barriers and motivation*. Doctoral dissertation, Aberystwyth University.

Hung, S. Y., Durcikova, A., Lai, H. M., & Lin, W. M. (2011). *The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior*. International journal of human-computer studies, 69(6), 415-427.

Hunter, L., Beaumont, P., & Lee, M. (2002). *Knowledge management practice in Scottish law firms*. Human Resource Management Journal, 12(2), 4-21.

Huysman, M., De Wit, D. (2000). *Kennis delen in de praktijk*. Vergaren, uitwisselen en ontwikkelen van kennis met ICT, Assen: Van Gorcum.

John, P Ogier (2016). *Intellectual property, finance and economic development.* WIPO Magazine.

Kabene, S. M., King, P., & Skaini, N. (2006). *Knowledge management in law firms*. Journal of Information Law and Technology (JILT)(1), 1-21.

Kahn, A., & Wu, X. (2020). *Impact of digital economy on intellectual property law*. J. Pol. & L., 13, 117.

Lambe, P (2003), What does KM mean for Law Firms?. Retrieved from http://greenchameleon.com/

Laudon, K.C., Laudon, J.P. (2010). *Management Information Systems – Managing the Digital Firm*, 12th Edition, Pearson.

Levin, D. Z., & Cross, R. (2004). *The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer*. Management science, 50(11), 1477-1490.

Lin, H. F. (2007). *Knowledge sharing and firm innovation capability: an empirical study*. International Journal of manpower, 28(3/4), 315-332.

Malek, S.L., Sarin, S. and Haon, C. (2020). *Extrinsic rewards, intrinsic motivation, and new product development performance*. Journal of Product Innovation Management, Vol. 37 No. 6, pp. 528-551.

Meso, A., Bosire, E., & Massey, E. (2023). *Knowledge management as a tool for improving work performance in selected law firms in Nairobi county, Kenya: An assessment*. Journal of Education, Society and Behavioural Science, 36(7), 85-95.

Nahapiet, J., & Ghoshal, S. (1998). *Social capital, intellectual capital, and the organizational advantage*. Academy of management review, 23(2), 242-266.

Nguyen, M., Malik, A. and Sharma, P. (2021). *How to motivate employees to engage in online knowledge sharing? Differences between posters and lurkers*. Journal of Knowledge Management, Vol. 25 No. 7, pp. 1811-1831.

Nguyen, T. M., & Malik, A. (2020). *Cognitive processes, rewards and online knowledge sharing behaviour: the moderating effect of organisational innovation*. Journal of Knowledge Management, 24(6), 1241-1261.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Olatokun, W. M., & Nneamaka, E. I. (2013). *Analysing lawyers' attitude towards knowledge sharing*. South African Journal of Information Management, 15(1), 1-11.

Olson, M. (1971). *The Logic of Collective Action. Public Goods and the Theory of Groups*. Harvard.

Osterloh, M., & Frey, B. S. (2000). *Motivation, knowledge transfer, and organizational forms*. Organization science, 11(5), 538-550.

Parsons, M. (2004). *Effective knowledge management for law firms*. Oxford University Press.

Rusanow, G. (2003). *Knowledge management and the smarter lawyer*. ALM Publishing.

Sammon (2024). How to Choose the Best IP Law Firm for Your Brand?. Retrieved from https://www.sonderandclay.com

Saunders, C. (2011). *Knowledge sharing in legal practice*. In Encyclopedia of Knowledge Management, Second Edition (pp. 935-945). IGI Global.

Schulz, M., & Klugmann, M. (2005*). Creating a Culture of Knowledge Sharing in Law Firms.*, 386-387. https://doi.org/10.1007/11590019_44.

Terrett, A. (1998). *Knowledge management and the law firm*. Journal of knowledge management, 2(1), 67-76.

Tjaden, T. (2007). The Role of Law Firm Culture in Knowledge Management. Retrieved from http://www.Slaw.Ca/2007/06/29/The-Role-OfLaw-Firm-Culture-In-Knowledge Management

Van den Brink, P. (2003). *Social, organizational, and technological conditions that enable knowledge sharing*.

Van Engers, T.M. (2001), *Knowledge Management. The Role of Mental Models in Business Systems Design.* (diss. Amsterdam VU), Belastingdienst 2001, p. 5.

Wong, K. Y. (2005). *Critical success factors for implementing knowledge management in small and medium enterprises*. Industrial management & Data systems, 105(3), 261-279.

WIPO. *Intellectual Property and Frontier Technologies.* Retrieved from https://www.wipo.int/about-ip/en/frontier_technologies/

WIPO (2020). *What is intellectual property?*. Geneva: World Intellectual Property Organization.

WIPO (2024). *Generative AI Navigating intellectual property*. Geneva: World Intellectual Property Organization.

Zeide, E., & Liebowitz, J. (2012). *Knowledge management in law: a look at cultural resistance*. Legal Information Management, 12(1), 34-38.

How Organizational Improvisational, Transformational Leadership Styles Impact Innovation Performance of Start-Up Companies in VUCA Environments

Wenxi Xiong[a] and Yongzhong Yang
Business School, Sichuan University, China

Keywords: Organizational Improvisation, Transformational Leadership, Start-Ups, Innovation Performance.

Abstract: Entering the competitive VUCA environment, the traditional management model with prediction and control as the main measures has not adapted to the needs of the times in certain situations, and it has become a hot topic for start-ups to survive and grow in the current unpredictable environment and achieve breakthrough innovation. This study constructs a correlation model between organizational improvisation, transformational leadership, and innovation performance of start-ups, and uses the VUCA environment as a moderating variable. The findings confirm that all dimensions of organisational improvisation and transformational leadership significantly affect the innovation performance of start-ups; Organizational improvisation and transformational leadership positively interact to influence innovation performance of start-ups, i.e. organizational improvisation and transformational leadership reinforce each other's influence on innovation performance of start-ups; The VUCA environment positively moderates the impact of organizational improvisation and transformational leadership on innovation performance of start-ups. This study will help start-ups to fully grasp the fleeting opportunities and respond to the changing external environment in a timely manner to further enhance the competitive advantage of start-ups, and provide practical guidance for top managers of start-ups to enhance their innovation performance.

1 INTRODUCTION

The international situation is characterized by "Volatility", "Uncertainty", "Complexity" and "Ambiguity". VUCA environment characteristics are becoming more and more obvious. Under the VUCA era environment, how start-ups can turn crisis into safety, survive and grow in the crisis has become an issue of concern for all sectors of society.

The first thing that start-ups in dynamically changing environments need to do to achieve sustainable and healthy growth is to change the decision-making paradigm away from the traditional strategic thinking of planning, execution, feedback and review, and to think and act, practice and improve, innovate and integrate with more flexible and improvised strategies (Tang & Zhou, 2017).

The concept of organizational improvisation refers to the key elements of an organization's strategic actions in a complex environment, and it differs from the traditional strategic decision-making model in that it integrates planning, decision-making, and execution. Organizational improvisation can help organizations better complete the decision-making process when unexpected events or changes in the environment deviate from expectations. Organizational improvisation also reflects the richness of a leader's experience, judgment, and decisiveness.

Weick (1993) points out that transformational leaders give their employees more autonomy and encourage them to think outside the box and promote improvisation. In today's competitive world, transformational leaders are one of the most responsive types of leaders and are valued more than ever for building strong confidence through advance learning and preparation, and for learning by experience without fear of failure, so that organizations can adjust quickly in a VUCA environment, calming immediate fluctuations while making intuitive and rational decisions to seize future opportunities.

[a] https://orcid.org/0009-0005-5551-3959

Xiong, W. and Yang, Y.
How Organizational Improvisational, Transformational Leadership Styles Impact Innovation Performance of Start-Up Companies in VUCA Environments.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 295-304
ISBN: 978-989-758-716-0; ISSN: 2184-3228

Based on this, this paper will focus on the relationship between organizational improvisation, transformational leadership and innovation performance of start-ups, and explore the mechanisms of organizational improvisation and transformational leadership on innovation performance of start-ups.

2 BACKGROUND OF THE STUDY

VUCA was first proposed by the U.S. military to describe the complex and changing military environment at that time, and with the development of information technology such as big data, artificial intelligence and cloud computing, as well as the constant changes in the international market, the external environment in which the organization operates has become more and more characteristic of VUCA. Bennett & Lemoin (2014) mentioned VUCA as "Volatility", "Uncertainty", "Complexity", "Ambiguity" in " Harvard Business Review", which will become a trend in the future, and explains its meaning. Troise et al.(2022) used digital technological capabilities, relational capabilities, and innovative capabilities premised on organizational agility - the ability to quickly anticipate or respond to external changes - is critical to surviving and competing in today's turbulent VUCA environment characterized by technological advancement and digitalization.

The study of improvisation first began with jazz metaphors, so early research on organizational improvisation was mainly based on the perspective of jazz metaphors. Through specific studies, researchers attempted to summarize the characteristics of the concept of improvisation in jazz improvisation, differentiate the degree and form of improvisation, explore the mechanisms of improvisational behaviors, and discover the role of improvisational metaphors in jazz performances in organizational practices. Weick (1998) firstly blended improvisation with organizational management by observing top managers and found that they would, like jazz musicians, identify and aim at the goals and make and follow rules, engage in directed activities, and suggested characteristics of groups with high improvisation skills. When improvisation is combined with team and environmental moderators, it has a positive impact on team innovation, as well as suggesting that organizational members can improve their improvisational skills through training. Fultzz & Hmieleski (2021) established a model linking organizational improvisation with new venture performance, demonstrating that improvisation ability is a resourceful means for startups to identify new opportunities and gain performance advantages.

For the transformational leadership research. At the individual level, the main areas involved are job performance, job satisfaction, employee happiness, individual knowledge sharing, employee creativity, employee emotions, and self-coordination. Stanescu et al. (2019) point out the positive and significant relationship between transformational leadership and employee innovative behavior and psychological empowerment by creating a greater sense of empowerment, leaders can have a higher positive impact on the employee's level of employee innovation. At the organizational level, current research focuses on the study of transformational leadership on organizational performance. Ng (2017) developed an integrative model using five theoretical driving mechanisms to explain the impact of transformational leadership on performance outcomes. Mach et al. (2021) explored how transformational leadership affects team performance through team cohesion and how this relationship is moderated by prior team performance. The results of the study indicated that transformational leadership indirectly affects team performance through the mediating role of team cohesion, and that when the level of prior performance is high, the this indirect effect is more significant.

In exploring the relationship between work environment and creativity, Amabile (1996) suggested that an innovative organizational climate promotes team support, which in turn promotes innovation. Scott & Bruce (1994) and Xie et al. (2007) argued that an innovative culture within an organization has a positive impact on performance in new tasks. The latter's empirical study found that a good innovation culture can improve the business quality and innovation level of the organization, which in turn can help firms to better improve their performance level. Leadership behavior is another important factor in influencing innovation performance. Scott (2011) supported this view by showing that innovative leadership behaviors are believed to promote an innovative and creative climate in the organization, which, in turn, improves employees' innovation ability. Wang (2021) points out that when individual employees have a higher degree of forgetfulness, it is more conducive to digital transformation to promote organizational innovation performance through dynamic capabilities.

In a VUCA environment filled with turbulence, disruption and uncertainty, start-ups are actively implementing change and innovation to cope with various changes, but encountering many difficulties and resistance in the process. These dilemmas of change and innovation have received increasing attention in management practices and theoretical studies, and have become the focus of attention for academics, researchers and managers. However, the current empirical research is still in a relatively imperfect stage, the existing research is not deep and systematic enough, and scholars and practitioners need to further integrate the research on transformational leadership and innovation performance. Meanwhile, scholars need to further integrate research on organizational change and innovation behavior and consider how to enhance transformational leadership in the spirit of organizational improvisation to promote innovation in startups; on the other hand, practitioners consider how to systematically introduce the theory of enhancing transformational leadership into the process of organizational improvisation. Overall, the relationship between organizational improvisation, transformational leadership, and innovation performance of start-ups in VUCA environments does not seem to have been systematically studied in academia.

To sum up, there are still certain black holes and deficiencies in the academic research on the related concepts, for example, the lack of integrated research on the relationship between organizational improvisation, transformational leadership and innovation performance, and the failure to organically combine the VUCA environment with organizational change and innovation management. Based on this, this paper will focus on the relationship between organizational improvisation, transformational leadership and innovation performance, and explore the mechanism of organizational improvisation and transformational leadership on the new performance of start-ups under the VUCA environment.

3 THEORETICAL MODEL AND RESEARCH HYPOTHESISSECTION HEADINGS

3.1 Theoretical Model

In the VUCA environment, it brings great impact to the survival, competition and development of start-ups. Faced with the coupling effect of external environment complexity and organizational complexity, it is especially important to create a diversified and inclusive internal environment to properly handle the crisis and turn it into a driving force for development. Responding quickly and ruling with dynamics is even more crucial for leaders of start-ups to lead their organizations to success. Based on dynamic capability theory, when an organization has the ability to react quickly to the internal and external environment, its core competitiveness will be greatly enhanced. In addition, practice has shown that in the face of new challenges and threats, the effectiveness of traditional response strategies is greatly reduced and may even fail.

A review of the literature on leadership and team performance suggests that leadership behaviors are a key factor in team innovation performance and that the right leadership behaviors can maximize team innovation performance. Bass describes transformational leaders as “leaders who make subordinates aware of the importance and responsibility of the tasks they undertake by making them leaders whose needs are met, and who also create a greater vision for their subordinates, encourage and support them to go beyond themselves, adopt new ideas and approaches, creatively solve new challenges and problems, and create an atmosphere of mutual support and harmonious innovation within the organization to promote organizational innovation” (Bass, 1985). Therefore, transformational leadership style is important for accelerating organizational innovativeness and improving organizational innovation performance.

Based on the above review of the literature and related theories, this paper argues that organizational improvisation and transformational leadership may interact to produce influencing factors on innovation performance of start-ups, while the relationship between the two and innovation performance of start-ups may be mediated by the VUCA environment, therefore, the paper proposes the following theoretical model, as shown in Fig 1.

3.2 Theoretical Model

3.2.1 Organizational Improvisation and Innovation Performance of Start-Ups

Organizational improvisation as a "time pressure" and "environmental uncertainty" triggered by the ability to respond quickly to unexpected events has a strong immediate, spontaneous, creative characteristics,

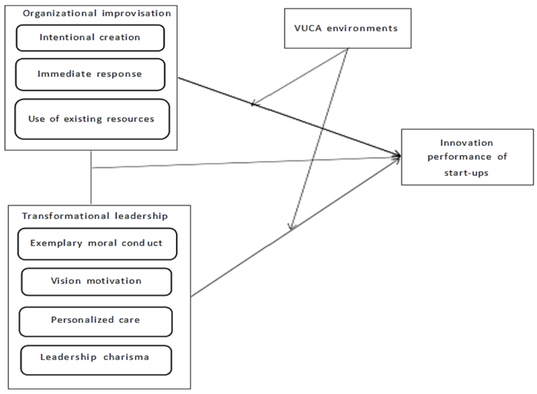


Figure 1: Conceptual model.

through the current problem facing the specificity of a timely response, and The existing resources are successfully integrated and utilized to avoid missed opportunities and problem complications. Based on the view of flexible human resource management and technological variability, Shan et al. (2021) pointed out that in an uncertain environment, technology-based firms use improvisation capabilities to enhance technology commercialization by efficiently integrating existing resources, thus positively influencing organizational innovation performance. Xiang (2021) confirmed a strong positive influence between organizational improvisational capabilities and innovation performance using the supportive incubation environment and environmental dynamics as moderating variables. Based on the moderating role of improvisational skills, Chen (2013) investigated the relationship between time-based competition and organizational performance and found that improvisational skills deepened the organization's understanding of time-based competitive cognition through rapid and immediate innovation responses, and the study found that improvisational skills effectively contributed to organizational innovation performance.

Based on this, this paper proposes the following research hypothesis: H1, H1a, H1b, as shown in Table 1.

3.2.2 Transformational Leadership and Innovation Performance of Start-Up Firms

In general, organizational development is influenced by the entrepreneurial leadership style,i.e., the leader's vision, wisdom and motivation to innovate affect the organization's performance indicators; therefore, innovation performance is closely related to the entrepreneur's values and beliefs, i.e., innovation performance is influenced by the entrepreneurial leadership style. Transformational leaders, on the other hand, tend to actively cultivate their own innovation consciousness and capabilities, inspire organizational members through their own images, attitudes, beliefs and behaviors, correct employees' work attitudes, build employees' collective cognition and perceptions of the organization, help promote mutual cognition among members, and guide the organization to accomplish innovative tasks, which is a key factor in improving organizational innovation performance. Numerous studies have shown that transformational leaders have an impact on organizational innovation performance by influencing to increase the creativity of their subordinates. For example, Nguyen (2022) showed that there is a significant correlation between transformational leadership and employee creativity. Shafi et al. (2020) stated that transformational leadership has a significant positive effect on employee creativity through idealizing influence, intellectual stimulation and motivational motivation. Zhang & Wang (2020) showed that transformational leaders improve organizational innovation performance by increasing their own work engagement. Rong et al. (2018) stated that CEO transformational leadership style has a significant positive impact on organizational innovation performance.

Based on the above analysis, the following hypotheses are proposed in this study: H2, H2a, H2b, H2c, H2d, as shown in Table 1.

3.2.3 Interactive Effects of Organizational Improvisation and Transformational Leadership on Innovation Performance of Start-up Firms

By collating the existing literature in English and Chinese, there is a gap in the current research on the interaction between the two. However, based on the existing research by scholars, it is not difficult to find that both organizational improvisation and transformational leadership have a reinforcing effect on each other to positively influence organizational innovation performance.

From the perspective of organizational improvisation, when an organization has the ability to react quickly to the internal and external environment, its core competitiveness will be greatly enhanced. In addition, the organization or its members must meet challenges with new organizational strategies through the exercise of individual creativity and improvisation, and take advantage of market opportunities to creatively solve problems, which also indirectly enhances the innovation capability of the whole organization. The transformational leader supports and facilitates the rapid growth of employees by providing them with personalized guidance, making them aware of their self-worth, and developing their ability to improve their self-leadership.

From the perspective of transformational leaders, the creative thinking, immediate responsiveness of organizational members is greatly enhanced when their behavior is unrestrained and maximizes their ability to improvise, which will help transformational leaders, through their personal charisma, build trust with employees, motivate them, and prompt them to further creatively solve new challenges and problems, creating in the organization a mutually supportive and harmonious innovation climate within the organization to achieve better organizational innovation performance. It is evident that organizational improvisation positively reinforces the role of transformational leaders on organizational innovation performance.

In summary of the analysis, the following hypothesis is proposed in this study: H3, as shown in Table 1.

3.2.4 Moderating Effect of VUCA

In the VUCA environment, various characteristics have brought a huge impact on the survival, competition and development of the company. Facing the coupling effect of external environmental complexity and organizational complexity, it is especially important to create a diverse and inclusive internal environment to properly handle the crisis and turn it into a driving force for development. Krupp & Schoemaker (2014), Bennett& Lemoine (2014), and Livingston (2014) have each made Troise et al. (2022) and others presuppose that organizational agility - the ability to quickly anticipate or respond to external changes - premised on digital technology capabilities, relational capabilities, and innovation capabilities, is essential in today's environment characterized by technological advancement and digitization as the hallmark of today's turbulent VUCA environment is critical to survive and compete. In addition to this, due to the rapidly changing domestic and international environment, start-up leaders must meet challenges with fresh organizational strategies and take advantage of market opportunities to creatively solve problems through the exercise of personal creativity and improvisation. Khan et al. (2021) showed that under the VUCA environment, knowledge workers enter the workplace more often, so leader-member exchange and leader identification play a mediating role between authentic leadership and employees' innovative work behaviors, and states that by building strong relationships with employees, managers can motivate employees to pursue innovative work behaviors.

Therefore, the following research hypotheses are proposed in this paper: H4, H5, as shown in Table 1.

Table 1: Research hypothesis.

H1	Organizational improvisation has a positive effect on innovation performance of start-ups.
H1a	Intention creation has a positive effect on the performance of start-ups.
H1b	Immediate response has a positive effect on the performance of start-ups.
H1c	Leveraging existing resources has a positive impact on start-ups.
H2	Transformational leadership has a positive effect on innovation performance of start-ups.
H2a	Virtuous exemplary behavior has a positive effect on innovation performance of start-ups.
H2b	Visionary incentives have a positive effect on innovation performance of start-ups.
H2c	Personalized care has a positive effect on innovation performance of start-ups.
H2d	Leadership charisma has a positive effect on innovation performance of start-ups.
H3	Organizational improvisation and transformational leadership reinforce each other's effects on innovation performance of start-ups.
H4	The VUCA environment positively moderates the relationship between organizational improvisation and innovation performance of start-ups, i.e., the more influenced by the VUCA environment the more innovation performance of start-ups is influenced by organizational improvisation.
H5	The VUCA environment positively moderates the relationship between transformational leadership and innovation performance of start-ups, i.e., the more influenced by the VUCA environment the more innovation performance of start-ups is influenced by transformational leadership.

4 STUDY DESIGN

4.1 Data Collection

AKumar & Das (2020) point out that enterprises that are active and founded within 8 years are new. Domestic scholars such as Cheng, Song-Song et al. (2019) and Zhang (2018) also generally define enterprises that have been established for less than 8 years as start-ups. Therefore, this study defines the investigated start-up enterprises as those whose enterprise age is less than or equal to 8 years.

In this study, 396 official questionnaires were distributed by telephone and email to the middle and senior management of a start-up enterprise in an industrial development zone in Chengdu, Sichuan Province (China) through the introduction of acquaintances, and 308 valid questionnaires were obtained after excluding invalid questionnaires such as incomplete completion or enterprises not meeting the conditions. From the viewpoint of the sample itself, the survey targets are widely distributed, covering manufacturing, service, information transmission, new energy materials, etc. From the viewpoint of the enterprise characteristics, 38 questionnaires were distributed to state-owned enterprises, 130 to private enterprises, 102 to foreign-funded enterprises, and 38 to other enterprises. 42.5% of businesses with 100 or more employees, 32.5% with 51-100 employees, 10.7% with 21-50 employees and 14.3% with less than 20 employees.

4.2 Questionnaire Design

The reliability test of the scales often uses Cronbach's α coefficient. In this paper, the reliability of each variable scale was analyzed using SPSS 27.0, and the results showed that the reliability values of each scale of organizational improvisation, transformational leadership, VUCA, and innovation performance were 0.952, 0.983, 0.940, and 0.904, respectively. The total reliability value was 0.975, thus indicating a high quality of reliability of the study data.

The validity tests were divided into exploratory factor analysis and validation factor analysis. In the exploratory factor analysis, KMO and Bartlett's spherical tests were conducted for the scales of organizational improvisation, transformational leadership, VUCA, and innovation performance, and the results showed that the KMO of all variables was greater than 0. 8, and the factor loadings of each measurement question item were greater than 0. 6, indicating good structural validity of the scales, as shown in Table 1. Validation factor analysis was conducted for each variable and the model fit was tested by indicators, in which the value of chi-square degrees of freedom was 1.143, which was less than 3, the value of RMSEA was 0.022, which was less than 0.10, the value of RMR was 0.050, which was less than 0.08, and the values of CFI, GFI, NFI, and TLI were all greater than 0.9, indicating that the absolute and value-added fitness of the model was better.

5 EMPIRICAL STUDY

5.1 Results of the Regression Ana

As can be seen from the table 2, in model M0, the adjusted R^2 is 0.004, indicating that the explained variance of firm age and firm size on individual innovation performance is 0.4%, and after adding the intention creation variable in model M1, the adjusted from R^2 increases to 0.144, and the regression coefficient P = 0.377 of virtue draping on the innovation performance of new start-up firms in model M1 is significant at the P = 0.001 level It indicates that intention creation is significantly and

Table 2: Table of regression analysis (independent variable is organizational improvisation).

Variant	M0	M1	M2	M3
Control variable				
Age of business	0.090	0.081	0.081	0.084
Enterprise size	-0.058	-0.082	-0.090	-0.100
Independent variable				
Intentional creation		0.377***		
Immediate response			0.360***	
Utilization of existing				0.370***
R^2	0.011	0.152	0.139	0.146
F-value	1.680	18.204***	16.375***	17.280***
Adjusted R^2	0.004	0.144	0.131	0.137

positively related to innovation performance of start-ups, and hypothesis H1a holds; after adding immediate response variable in model M2, the adjusted R^2 is 0.131, which has significantly improved the explanatory strength of innovation performance of start-ups compared with the basic model M0, and the regression coefficient of immediate response on innovation performance of start-ups P= 0.360, which is significant at the P = 0.001 level. It indicates that immediate response is significantly and positively related to innovation performance of start-ups, and hypothesis H2b holds; the addition of personalized care in model M3 increases the moderated R^2 to 0.137 and the regression coefficient on innovation performance is 0.370 (p=0.001), indicating that using available resources is significantly and positively related to innovation performance, and hypothesis H2c holds.

5.2 Results of Regression Analysis of Transformational Leadership and Innovation Performance

From the table 3, it can be seen that after adding the virtue pendant variable in model M1, the adjusted from R^2 increases from 0.004 in M0 to 0.213, and the regression coefficient P = 0.459 of virtue pendant on innovation performance of start-up firms in model M1 is significant at the 0.001 level, indicating that virtue pendant positively affects innovation performance of start-up firms, and hypothesis H2a holds; after adding the vision incentive variable in model M2, the adjusted from R^2 variable in model M2, the adjusted R^2 is 0.186, which has significantly improved the explanatory strength of innovation performance of start-ups compared with the basic model M0, and the regression coefficient P = 0.428 for vision incentive innovation performance of start-ups is significant at the 0.001 level, indicating that

Table 3: Results of regression analysis of transformational leadership and innovation performance).

Variant	M0	M1	M2	M3	M4
Control variable					
Age of business	0.090	0.090	0.070	0.075	0.088
Enterprise size	-0.058	-0.056	-0.071	-0.060	-0.07
Independent variable					
Set an example by virtue and behavior		0.459***			
Vision Motivation			0.428***		
Personalized Care				0.431***	
Leadership Charisma					0.445***
R^2	0.011	0.221	0.194	0.196	0.209
F-value	1.680	28.766***	24.358***	24.764***	26.698***
Adjusted R^2	0.004	0.213	0.186	0.188	0.201

vision incentive positively affects innovation performance of start-ups, and the hypothesis H2b holds; adding personalized care in model M3, the adjusted R^2 increases to 0.188 and the regression coefficient on innovation performance is 0.431 (p=0.001), indicating that personalized care is significantly and positively related to innovation performance, and hypothesis H2c holds; after adding leadership charisma in model M4, the adjusted R^2 is 0.201 and the regression coefficient on innovation performance is 0.445 (p=0.001), indicating that leadership charisma and innovation performance also show a significant positive correlation, and the hypothesis H2d holds.

5.3 Results of Interaction Effect Analysis

From the table 4, we can see that the adjusted R^2 values are 0.004,0.151,0.246,0.271, i.e., the model can explain 0.4%,15.1%,24.6%,27.1% of the variance, i.e., the model fits well; this paper constructs model M0 as the regression equation with the control variable as the independent variable, model M1 as the regression equation with the introduction of the independent variable organizational improvisation, and M2 is the regression equation with the introduction of the independent variable change-oriented leadership, and M3 is the product term with the introduction of organizational improvisation and change-oriented leadership. According to the table 4, the regression coefficient of M1 independent variable organizational improvisation is 0.386 ($p < 0.001$),

Table 4: Test table of the relationship between organizational improvisation, transformational leadership and innovation performance.

Variant	M0	M1	M2	M3
Control variable				
Age of business	0.090	0.081	0.078	0.09
Enterprise size	-0.058	-0.093	-0.078	-0.07
Independent variable				
Organizational improvisation		0.386***		-0.673**
New Leaders for Change			0.493***	-0.326
Organizational improvisation* Transformational Leadership				1.407***
R^2	0.011	0.159	0.253	0.283
F-value	1.680	19.142***	34.338***	23.845***
Adjusted R^2	0.004	0.151	0.246	0.271

based on the results of this analysis, it can be concluded that organizational improvisation positively affects organizational innovation performance, i.e., the hypothesis H1 of this study is verified. the regression coefficient of M2 independent variable transformational leadership is 0.493 ($p < 0.001$), which indicates that transformational leadership positively affects organizational innovation performance, i.e., the hypothesis H1 of this study is verified. The regression coefficient of M3 product term was 1.407 ($p < 0.001$), and the hypothesis H3 of this study was supported.

5.4 Results of Analysis of Moderating Effects

To reduce the problem of non-essential multicollinearity in the regression analysis, this study decentered the independent variable Organizational Improvisation and Transformational Leadership and the moderating variable VUCA Environment (mean center) before constructing the interaction terms, and then calculated the product of the independent and moderating variables after decentering and constructed interaction term 1 and interaction term 2, respectively. From the table 5, we can see that the adjusted R2 values are 0.004, 0.248, 0.271, 0.304, 0.315, i.e. the model can explain 4%, 24.8%, 27.1%, 30.4%, 31.5% of the variance, i.e. the model fit is good; model M0 is constructed as the regression equation with the control variable as the independent variable, model M1 is the regression equation with the introduction of the independent variable Organization improvisation and the regression equation of equation of the moderating variable VUCA environment, M2

Table 5: Table of tests for moderating effects.

Variant	M0	M1	M2	M3	M4
Control variable					
Age of business	0.090	0.09	0.089	0.085	0.071
Enterprise size	-0.058	-0.058	-0.044	-0.056	-0.051
Independent variable					
Organizational improvisation		0.233***	0.182		
Transformational Leadership				0.359***	0.282***
VUCA		0.35***	0.333***	0.279***	0.268***
Interaction 1			0.172***		
Interaction 2					0.141**
R^2	0.011	0.257	0.283	0.313	0.326
F-value	1.680	26.257***	23.881***	34.523***	29.197***
Adjusted R^2	0.004	0.248	0.271	0.304	0.315

is the regression equation of the interaction term 1 after the introduction of the independent variable organizational improvisation, the moderating variable VUCA environment, and decentering, model M3 is the regression equation of the introduction of the independent variable transformational leadership and the moderating variable VUCA environment, and M4 is the regression equation of the interaction term 2 after the introduction of the independent variable organizational improvisation, the moderating variable VUCA environment, and decentering . According to the above table, the regression coefficient of interaction term M2 is 0.172 ($p < 0.001$), and based on the results of this analysis, it can be concluded that VUCA environment reinforces the positive effect of organizational improvisation on innovation performance, thus supporting this paper's research hypothesis H4. The regression coefficient of interaction 2 in the M4 model is 0.141 ($p < 0.01$),thus being able to support this paper's research hypothesis H5, thus it can be concluded that the VUCA environment positively moderates the positive effect of transformational leadership on innovation performance.

6 CONCLUSION

Firstly, Organizational improvisation can positively affect the innovation performance of start-ups: if an organization facing uncertain environmental conditions and time pressures can efficiently coordinate internal and external organizational resources based on integrating and reconfiguring existing resources, react quickly and effectively, and dare to break the rules and respond to unpredictable environmental changes in innovative ways and means, it can take advantage of the business opportunities brought by environmental changes and create new competitive advantages and thus improve the innovation performance of the firm, which is consistent with the study of Cunha &Vera (2005) .

Secondly, transformational leaders can positively influence the performance of start-ups: transformational leaders tend to actively develop their own innovative awareness and capabilities, inspire organizational members through their own images, attitudes, beliefs and behaviors, correct employees' work attitudes, build employees' collective perceptions and ideas about the organization, help promote mutual perceptions among members, and guide the organization to accomplish innovative tasks, which is a key factor in improving organizational innovative performance A key factor in improving organizational innovation performance.

Thirdly, Organizational improvisation and transformational leadership can positively interact to influence the innovation performance of start-ups: transformational leaders create the conditions for organizational improvisation through personal leadership charisma, thereby gaining the trust of employees while they strive to create a harmonious, supportive, innovative and entrepreneurial culture within the organization that promotes change and innovation.

Fourthly, the VUCA environment plays a significant moderating role between organizational improvisation and innovation performance of start-ups: the more volatile, uncertain, complex and ambiguous the competitive environment is, the greater the role of organizational improvisation on innovation performance of start-ups, which proves that the volatility of the environment is one of the necessary conditions for organizational improvisation. The degree of VUCA in the environment is an external factor that is difficult for the organization to control, but by improving the dynamic capabilities of the organization, the organization can take advantage of market opportunities, creatively solve challenges, and improve the innovative performance of start-ups.

Fifthly, the VUCA environment plays a significant moderating role between transformational leadership and innovation performance of start-ups: the more volatile, uncertain, complex and ambiguous the competitive environment in which a company is operating, the greater the role of transformational leadership on innovation performance of start-ups.

7 DISCUSSION

7.1 Theoretical Significance

Firstly, we explore the role relationship and evolution mechanism between transformational leadership style and innovation performance from the perspectives of management, leadership, organizational behavior, and innovation performance of start-ups, and explore the effective path and influencing factors to enhance the innovation performance of start-ups under the perspective of organizational improvisation from the perspective of transformational leadership style.

Secondly, the concept of organizational improvisation is still in its infancy in China, and most of the qualitative studies on the relationship between organizational improvisation and performance have been conducted. In addition to this, the current research mainly focuses on organizational performance and competitive advantage, and few scholars have conducted research on the impact of organizational improvisation on the innovation performance of start-ups.

Thirdly, because start-ups are less able to withstand risks compared to mature firms, start-ups usually face problems such as fierce market competition and difficulty in predicting current market demand (Li & Cao, 2021), and with the further intensification of the VUCA environment, start-ups face even more serious competitive pressures and threats to their resources

7.2 Research Limitations and Future Directions

Firstly, due to objective reasons in terms of time and funding, the sample of this study is mainly from Southwest China, while the degree of environmental turbulence and the organisations themselves exist in different countries and regions with large characteristics. In addition, the industries in this study include both high-tech and traditional industries, and whether or not there are large differences caused by different industries may all affect the generalizability of the study's findings. One of the directions for future research is to conduct comparative studies by industry as well as by region.

Secondly, the questionnaire research method mainly used in this study, the scales are scoring mode, some employees may have concerns when it comes to the leader and the company performance related questions, which can not fully reflect the objective facts of the enterprise, with the limitations of first-hand data, and at the same time, the indexof the organization's innovation performance is also subjective measurement, which inevitably causes the bias of the research data. Future research can combine rooted theory with questionnaire research, thus making up for the shortcomings in this area.

Third, the data obtained in this study were only static cross-sectional data, which were analyzed to explore the relationship between the independent and dependent variables, without longitudinal data collection and dynamic tracking analysis. In order to obtain more accurate research results, future research needs to conduct longitudinal studies with specific time spans, as well as case studies exploring deeper associations between variables, in order to reduce errors arising from single cross-sectional data with a multi-faceted, cross-level research approach.

7.3 Management Suggestions for Practice

Firstly, this study confirms that innovation performance of start-ups is positively and positively influenced by transformational leadership. Therefore, organizations can determine whether existing leaders need transformational leadership training by analyzing the different needs of team leaders, develop training programs based on their needs and change behaviors in practice after training, create scenarios conducive to transformational leadership traits, and strengthen the effectiveness of the training by recognizing and rewarding transformational leadership behaviors.

Secondly, using modern Internet technology and other means, the company can effectively simulate and practice the difficulties and blows it may face, and require the employees of the new venture to make timely responses, and shorten the response time each time through repeated training, thus putting forward more requirements for improving the improvisation ability of the company's employees.

Thirdly, new start-ups should detach themselves from the previous independent perspective of transformational leadership and organizational improvisation, and pay attention to the positive effects generated by the integration of the two, and give full play to the complementary advantages of the two in the VUCA environment. At the same time, this paper finds that the impact of both on the innovation performance of start-ups is moderated by VUCA, and the role of both has a strong correlation with the turbulent environment in which the enterprise is located. At the same time, it is necessary to focus on the cultivation of innovation ability, break the inherent routine, avoid the rigidity of the organization, and maintain the creativity of the organization.

REFERENCES

Bass, B. M. (1985). Leadership and Performance beyond Expectations. The Academy of Management Review,
Chen, C. (2013). The impact of time-based competition on organizational performance from a cognitive perspective (Master's thesis, Northeast University of Finance and Economics).
Chowdhury, M. M. H., & Quaddus, M. (2017). Supply chain resilience:
Conceptualization and scale development using dynamic capability theory. International Journal of Production Economics, 188, 185‑204.
Crossan, M. M., Vieira da Cunha, J., Cunha, M. P. E., & Vera, D. (2002). Time and Organizational Improvisation. SSRN Electronic Journal.
Fultz, A. E. F., & Hmieleski, K. M. (2021). The art of discovering and exploiting unexpected opportunities: The roles of organizational improvisation and serendipity in new venture performance. Journal of Business Venturing, 36(4), 106121.
Leybourne, S., & Sadler-Smith, E. (2006). The role of intuition and improvisation in project management. International Journal of Project Management, 24(6), 483‑492.
Li, J.S., & Zhao, S.M. (2021)Orientation and path of human resource management model innovation in the VUCA era - focusing on the "three pillars" model. Journal of Jianghai (05), 90-96.
The NTL Handbook of Organization Development and Change: Principles, Practices, and Perspectives: Second Edition,2014,76(6):659-672.
Mach, M., Ferreira, A. I., & Abrantes, A. C. M. (2021). Transformational Leadership and Team Performance in Sports Teams: A Conditional Indirect Model. Applied Psychology, 71(2).
Ng, T. W. H. (2017). Transformational leadership and performance outcomes: Analyses of multiple mediation pathways. The Leadership Quarterly, 28(3), 385‑417.
Nguyen, T. P. L., Nguyen, T. T., Duong, C. D., & Doan, X. H. (2022). The effects of transformational leadership on employee creativity in Vietnam telecommunications enterprises. Management Decision, 60(3), 837‑857.
Nielsen, K., & Daniels, K. (2012). Does shared and differentiated transformational leadership predict followers' working conditions and well-being?. Leadership Quarterly, 23(3), 383-397.
Rong, P. F., Su, Y., & Wang, X. L. (2018). CEO leadership style, TMT behavioral integration and firm innovation performance. Xuehai (01), 196-206.
Scott, S. G., & Bruce, R. A. (1994, October 1). Creating innovative behavior among R&D professionals: the moderating effect of leadership on the relationship between problem-solving style and innovation. IEEE Xplore.
Shafi, M., Zoya, Lei, Z., Song, X., & Sarker, M. N. I. (2020). The effects of transformational leadership on employee creativity: Moderating role of intrinsic motivation. Asia Pacific Management Review, 25(3). ScienceDirect.
Shan, B. A., Pu, Y., Yan, S., & Liang, B. M. (2021). A study on the impact of improvisational ability on the performance of technology-based new ventures under uncertainty. Journal of Management (07), 1032-1039
Stanescu, D. F., Zbuchea, A., & Pinzaru, F. (2020). Transformational leadership and innovative work behaviour: the mediating role of psychological empowerment. Kybernetes.
Tang, Y., & Zhou, P. (2017). And so on and so forth: Corporate improvisation strategy in a dynamic environment. Tsinghua Management Review (05), 84-90
Troise, C., Corvello, V., Ghobadian, A., & O’Regan, N. (2022). How can SMEs successfully navigate VUCA environment: The role of agility in the digital transformation era. Technological Forecasting and Social Change,2022, 174, 121227.
Wang, C. (2021). Research on the Role Mechanism of Digital Transformation on Corporate Innovation Performance. Contemporary Economic Management (03), 34-42.
Weick, K. E. (1993). The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. Administrative Science Quarterly, 38(4), 628.
Xiang, Y. Y. (2021). The impact of resource patchwork and organizational improvisation on innovation performance of start-up firms (Master's thesis, Chongqing University of Technology and Business).
Zhang, J., &Wang, G. H. (2020). Analysis of the impact mechanism of transformational leadership on team innovation performance. Enterpris Economics (09), 37-43.
Zhang, X. E., & Zhang, K. (2018). The impact of entrepreneurial orientation on the performance of newly created social enterprises - The mediating role of resource collocation and the moderating role of regulation. Science and Technology Progress and Countermeasures (09), 91-99

Knowledge Sharing in Financial Institutions to Assist with IT Service Management: A Thematic Analysis

Cornelius JP Niemand[1][a] and Josef Langerman[2][b]
[1]*Department of Information and Knowledge Management, School of Consumer Intelligence and Information Systems, College of Business and Economics, University of Johannesburg, Gauteng, South Africa*
[2]*Department of Applied Information Systems, School of Consumer Intelligence and Information Systems, College of Business and Economics, University of Johannesburg, Gauteng, South Africa*
corn@uj.ac.za, josefl@uj.ac.za

Keywords: Service Management, Applications, Knowledge, Knowledge Sharing, Financial Institutions.

Abstract: The applications and services provided by financial institutions are important to individuals and economies. These applications and services are fragile because of service failures that are inherent in technology. The purpose of this article is to show how knowledge management can mitigate service disruption in financial institutions. By using bibliometric analysis and a structured literature review based on the PRISMA 2020 guidelines, we identified five major themes that drive knowledge management (KM) practices in information technology (IT) management in financial institutions. These themes identified are centered on the organizational environment, the motivation of employees, the people profile for example gender and race, and lastly the use of technology. By bolstering these KM practices in the IT service management (ITSM) of financial institutions we hope to shorten the time between system failures and shorten the actual time to repair failures. Knowledge management in IT management and especially ITSM is under-researched in financial institutions, and the KM themes identified provide some signposts to improved collaboration and better theorisation.

1 INTRODUCTION

Digital systems go down. The results range from mild irritation to catastrophic failure. In today's financial institutions most, interactions are through digital channels, be that the web, phone, or ATM. Invisible to the users are all the back-end systems that glue everything together and integrate the banking ecosystems of countries and the world (Khiaonarong et al., 2022; Klee, 2010; Merrouche & Schanz, 2010; Mishchenko et al., 2022).

It is not only a single financial organization that is prone to these types of outages but the financial ecosystem. For example, a system outage at a payment provider (that could be a bank or non-bank) is amplified in unexpected ways as the outage at one provider interacts with other providers as the technology outage ripples through the payment ecosystem. In worst-case scenarios, a cascading outage could cause significant parts of retail payment systems to shut down and eventually, that could harm the broader economy (Allen, 2021; Sillito & Kutomi, 2020).

System failures are not only mentioned in academic literature but also in popular literature. During black Friday in 2016, one of the payment providers to the largest online shopping portal in South Africa had an outage because of a high load. The payment processing was then passed to another bank as a fall-back mechanism. The new bank then had to process its payments, together with the payments of the original bank which caused a payment outage at the new bank. Eventually, the online platform had to shut down until these issues could be resolved (MyBroadband, 2017). Thousands of customers of Halifax, Bank of Scotland, and Lloyds were prevented from accessing their accounts for eight hours on New Year's Day 2020 because of a system outage. The system outages at financial firms have increased since a series of high-profile problems at companies like TSB and Visa in 2018. The UK Treasury also noted that there was an

[a] https://orcid.org/0000-0002-8582-0328
[b] https://orcid.org/0000-0003-1984-0205

Niemand, C. and Langerman, J.
Knowledge Sharing in Financial Institutions to Assist with IT Service Management: A Thematic Analysis.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 305-315
ISBN: 978-989-758-716-0; ISSN: 2184-3228

"unacceptable" level of IT failures among banks (Binham, 2020).

Software and hardware faults are inherent properties of computer systems. The huge complexity of software and hardware makes statistical outages unavoidable. Modern software consists of millions of lines of code and in many cases reaches a hundred million lines of code. (Domingos et al., 2021). Even though software complexity causes issues the main reason for software failure is change caused by human error. Sillito et al (2020) report that the major causes of software outages (in order of frequency) are deployments (software changes), infrastructure changes, exceeding scaling limits, and software and hardware failure. A major theme in their analysis is that incidents grow in scope as an initial failure cascades through a system exposing ways systems are not resilient to failure. (Sillito & Kutomi, 2020).

For clarity, an incident is when a software system experiences an outage or is degraded in functionality or performance, and engineers are notified to investigate and mitigate the problem. The work of the engineers in this context is seen as incident response. After the incident is resolved or mitigated a post-mortem is generally conducted (ISO, 2015; Sillito & Kutomi, 2020). The ISO 27043 standard divides the process into pre-incident response, during-incident response, and post-incident response. (ISO, 2015)

The current process frameworks that guide this in the financial sector are the IT Infrastructure Library (ITIL) and to a lesser extent Google's Site Reliability Engineering (SRE) framework (Langerman & Joseph, 2023). ITIL focuses primarily on processes and SRE on automation to reduce human error (Axelos, 2020; Beyer et al., 2016).

Both frameworks incorporate aspects of knowledge management, for example, ITIL 4 advocates for the capturing of knowledge at the source of an incident. This allows organizations to build a repository of institutional knowledge that may be of use for future incident resolution and problem-solving activities. Unfortunately, the implementation and use of the core knowledge management principles as outlined by the frameworks are limited.

In the context of the brief discussion of the frameworks guiding the financial sector, knowledge may be regarded as a strategic resource to assist managers and engineers in decision-making during the pre-incident, during-incident, and post-incident response cycle and can help mitigate the effect of an outage. Knowledge management can enhance the process of incident management, by ensuring the availability and accessibility of accurate and reliable information when required, through effective lesson learning (Ammirato et al., 2021). Ammirato et al (2021) and Seneviratne et al (2010) make the case for the importance of knowledge management during a disaster. Similarly, the researchers make the case in this paper for the use of knowledge management during an IT disaster or system outage. Despite the critical role of knowledge management in IT service management that can inform and enable decision-makers the literature on this is poor. Service management is addressed in IT journals but only marginally addressed by the knowledge management fraternity. A significant exception to this is the work of Baradari et al. (2023) which focus specifically on the role of knowledge management in ITIL. ITIL as a methodology is not specific to any industry but covers service management across all information technology industries.

As literature on the overlap between knowledge management, service management and information technology is so scant we restricted our literature survey to that of knowledge management and information technology in financial services. As service management is a subset of IT management (MacLean & Titah, 2023), this broadening makes sense for this research project.

1.1 Knowledge and the Sharing of Knowledge

The Merriam–Webster dictionary defines knowledge as "fact or condition of knowing something with familiarity gained through experience or association". For this research, two types of knowledge may be identified, i.e., tacit knowledge and explicit knowledge.

According to Polanyi, (1967) in Khan and Zaman (2020:2), "Tacit knowledge is embedded in the minds of people, accumulated during their career through experience, and is only visible through their actions…the other type is explicit knowledge which exists in written or other transferrable form". Some examples of explicit knowledge dimension are any form of knowledge that is in a written form, like policies, organizational strategies, operating procedures, vision, and mission statements with related rules governing them.

It should be noted that capturing and sharing explicit knowledge is deemed easier than capturing and sharing tacit knowledge. The capturing of explicit knowledge usually takes the form of a simplified and structured approach. The management of the two types of knowledge results in the concept of knowledge management.

Asrar-Ul-Haq & Anwar, 2016; Yeboah, 2023 in Kim and Hang (2024:1) state that "One essential element of the knowledge management system is knowledge sharing." For this research, knowledge sharing may be defined as the process of both internal and external transfer of both tacit and explicit knowledge for decision-making to ensure the longevity and profitability of the organization. The aforementioned is confirmed by Abbas, Hussain, Hussain, Akram, Shaheen, and Niu (2019:2) stating that "… knowledge-sharing strategies significantly influence a firm's success through their innovative performance processes." The aforementioned is echoed by Darroch and McNaughton (2003) in Alshwayat et al. (2021), stating that employing knowledge-sharing activities in an organization will create more creativity resulting in better economic performance, i.e. profitability.

Financial institutions are not exempt from the current economic climate and thus also be proactive in their stance to remain competitive in an ever-changing business environment. According to Abbas, Hussain, Hussain, Akram, Shaheen, and Niu (2019:2) "Realizing the importance of knowledge management, especially knowledge sharing, the banking sector has initiated the development of knowledge management (KM) teams in their institution."

2 METHODOLOGY

The current economic perspective is an economy based on knowledge, where individual, group, and societal existence are dependent on the use of and sharing of knowledge. Within modern-day organizations, knowledge sharing may be regarded as the new core capability that can increase the longevity of the organization as a provider of goods and services.

The main research question focuses on understanding the themes driving research in knowledge sharing, specifically in financial institutions, considering that these institutions and functions within the institutions are often characterized by individuals hoarding and not willing to share what they know.

To this end, the researchers adopted a pragmatic ontological stance, based on the practical application of the results of the study to achieve what Sekaran and Bougie (2013:30) coin as "intelligent practice".

As stated, pragmatism is the chosen ontological stance for the study and pragmatism may be attributed to Charles Sanders Peirce, the nineteenth-century American mathematician and logician. In an attempt to understand how researchers come to know, Jacobs (2010:725) postulates that "Peirce argued for abduction" as an epistemological assumption. Reichertz (2014:126-127) points out that the research activity starts when the researcher realizes that there is an imbalance between expectation and reality. The imbalance between expectation and reality can be described as the "surprise" factor, necessitating the researcher to de- and re-contextualize data and understanding about a specific phenomenon, and in so doing arrive at a new idea about the phenomenon under investigation.

The premise of this study is based on the fact that there is an imbalance between organizations' expected ability to manage knowledge sharing and the reality thereof. Abductive reasoning as an epistemological stance therefore makes sense in terms of this study.

The methodological assumptions focus on the process of research design. Kelemen and Rumens (2011) state that "by pragmatism's theoretical cornerstone, the pragmatist researcher is most likely to adopt research practices that will allow him/her to solve a practical problem efficiently". From the epistemological stance of the study, it is evident that the pragmatist researcher needs to be able to acknowledge all interactions between knowledge and action within a specific area of investigation. This research employed a systematic literature review approach, which is considered popular in qualitative research studies. The actual methodology employed in the review followed the PRISMA 2020 updated guideline for reporting on systematic reviews allowing for the reporting of "sufficient detail to allow users to assess the trustworthiness and applicability of the review findings (Page et al 202).

The protocol provides clear steps for the identification, screening, and inclusion of literature as part of the systematic literature review. Each of the following main points will be elaborated on:

- inclusion and exclusion criteria,
- the search strategy,
- The data sources, and
- the analysis and reporting elements.

2.1 Inclusion and Exclusion Criteria

The selection criteria identified in this section define what to include and what to exclude in the review of the sources. It should be noted that the inclusion and exclusion criteria aim to identify relevant research that will answer the main research question as postulated in section 2. The review was limited to scholarly peer-reviewed journal articles on the topic of knowledge sharing in the domain of financial institutions.

It should be noted that resources in the form of books, book chapters, and grey literature were not considered for inclusion in the review.

2.2 The Search Strategy

The identification of keywords and use of Boolean operators governed the creation of a search string that was used in searching for relevant sources to include in the review.

The search string used was:
(("knowledge sharing")) (Title) and (("financial institution*")) or (("bank*")) or (("financial service provider*")) (All Fields) and (("information technolog*")) or ((IT)) or ((system*)) (All Fields).

2.3 The Data Source/s

The search string as identified in section 2.2 was used to conduct a search for scholarly literature on the Clarivate Web of Science. The use of Clarivate Web of Science was deemed suitable for the database is a multi-disciplinary database covering a variety of different subjects within a large data range.

2.4 The Analysis and Reporting Elements

Applying the search string identified in section 2.2 yielded the following results (based on the Prisma 2020 protocol).

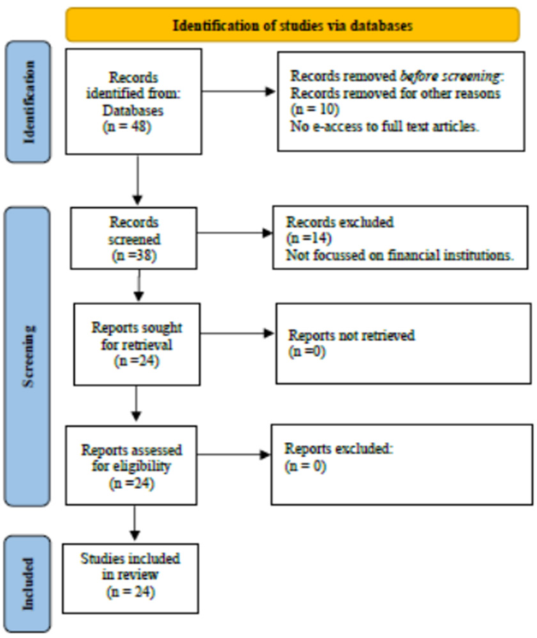


Figure 1: The analysis and reporting elements.

A total of 48 scholarly peer-reviewed journal articles were retrieved utilizing the identified search string. It should be noted that ten articles were removed before the screening process. The researchers did not have full-text access to the identified articles and were subsequently removed from the review process.

The resulting sources, i.e. the 38 articles were screened for inclusion in the review. Of the 38 articles, 14 articles were not deemed to fit into the review due to a lack of focus on knowledge sharing in financial institutions. The 14 articles were subsequently removed from the review.

The resulting 24 articles were downloaded and stored for further analysis.

2.4.1 Bibliometric Analysis of the Results of the Systematic Literature Identified

To position and enhance the relevance of the systematic literature review based on the Prisma 2020 protocol, the researchers conducted a bibliometric analysis of the results of the sources identified for inclusion in the review.

As identified in section 2.3, 24 scholarly peer-reviewed journal articles were identified for inclusion in the systematic literature review.

The researchers employed Bibliometrix, a powerful R package designed specifically for bibliometric analysis of sources. At its core, Bibliometrix analyzes three types of knowledge structures in the sources. These include:

- conceptual structure
- Intellectual structure, and
- Social structure.

Because Bibliometrix relies heavily on code commands a more user-friendly interface, i.e. Biblioshiny, which is an extension of Bibliometrix was employed to provide a more visual and interactive approach to the bibliometric analysis of the sources. Some of the main results of the bibliometric analysis are presented below:

Figure 2: Bibliometric overview of sources included in the review.

The most important findings from Figure 2 are:

- The sources included in the review cover a 10-year time frame, from 2014-2024
- 80 authors contributed to the topic at hand
- There is an average of 21.96 citations per article.

The annual scientific production per year during the said period reached a peak in 2021 with a total of five units produced. This is evident in Figure 3 below.

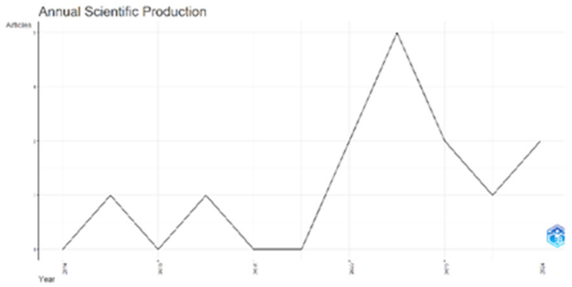


Figure 3: Annual scientific production.

The average article citations per year also followed a similar trend as in Figure 3, with peak citations in 2019 and 2021. These trends are illustrated in Figure 4.

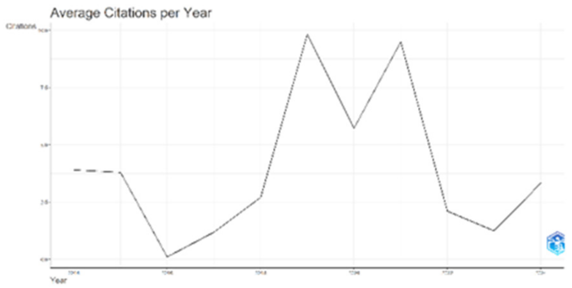


Figure 4: Average citations per year.

One of the most important elements to consider in a bibliometric analysis is the "vehicles" of dissemination of the research results, i.e. the sources of publications. The bibliometric analysis of the 24 articles revealed that the top 5 most relevant sources of publications are:

- Journal of Knowledge Management = 6 publications
- Employee relations = 2 publications
- Journal of Public Affairs = 2 publications
- Sustainability = 2 publications
- Applied Psychology – an International Review = 1 publication

When considering the countries where the scientific production originates from, it is interesting to note that Asia and India are the leaders in producing research on the topic.

A visual overview of the countries contributing to the scientific production on the topic is offered in Figure 5.

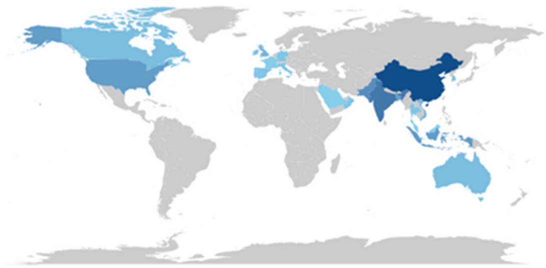


Figure 5: Countries' scientific production.

The following section will provide a discussion of the results of the systematic literature review.

3 ANALYSIS OF THE RESULTS

Out of the systematic literature review, five clear themes emerged. The themes are represented in Figure 6.

Figure 6: Results of the thematic literature review.

Each of the themes will be described in more detail to align the themes to answer the main research question as outlined in section 2.

3.1 Environment

The most prominent theme identified in the literature review is the creation of a suitable knowledge-sharing environment in financial institutions.
The driving concepts identified in this theme include but are not limited to:

- organizational attributes
- collaboration
- trust
- rewards

According to Kim and Hang (2024:2), the major catalyst for creating an environment fostering knowledge sharing is the organizational commitment and support towards creating such an environment for the staff members. Enwereuzor (2021) in Kim and Hang (2024:2) furthermore postulates that the commitment and support from organizational management be accompanied by elements and constructs such as diversity, respect, and engagement, specifically shown and driven by the management of the organization.

Abbas, Hussain, Hussain, Akram, Shaheen, and Niu (2019:2) extend the notion of organizational attributes to the extent that the organization should become a learning organization resulting in "…enhance knowledge sharing among employees within the organizations and empowering business firms to initiate critical actions and behaviors to gain the organizational settings to identify the real situation."

Gold et al (2001) believe that collaborations within the organizational environment should constitute interactions between the staff members of the organization. The authors furthermore state that collaboration is the result of open communication channels, participative activities, and interactions among the staff members.

This sentiment is echoed by Abbas, Hussain, Hussain, Akram, Shaheen, and Niu (2019:2) stating that "…new employees tend to develop relations with colleagues, creating a channel for knowledge sharing"

Von Krogh, Nonaka, and Rechsteiner (2012) believe that although management might create and facilitate an environment that enables collaboration, communication, and sharing of ideas, without trust in this collaborative space staff members will not be willing to share any knowledge. The lack of trust is highlighted by other authors including Bock et al., 2005, who states that a lack of trust in the organizational environment may be regarded as the main barrier to knowledge sharing and transfer.

Kim and Hang (2024:10) clearly state that although monetary and other rewards might not be a focal point for managers of knowledge, the impact thereof to stimulate the sharing of knowledge should not be underestimated.

3.2 Motivation

Motivation in terms of the literature reviewed may be considered as a descriptive concept, defining motivation within the construct of knowledge, knowledge sharing, and the environment. Nguyen and Malik (2022:1987) state that "motivation theorists posit that motivation drives employees' behavior"

The driving concepts identified in this theme include but are not limited to:

- Organizational motivation and strategies
- individual motivation
- emotional intelligence

Organizational motivation and strategies are driven by the formulation, implementation, and management of performance appraisals.

Fong et al (2011) in Gillani, Iqbal, Akram, and Rasheed (2017:180) believe that the use of performance appraisals may be regarded as positive reinforcement to shape the behaviors of the staff members of the organization. If knowledge sharing is defined as a key performance area and indicator, can contribute to staff members' improved performance in this specific area and contribute to the organizational environment and culture of sharing of knowledge.

It should be noted that the use of performance appraisals should be handled with great care and responsibility. According to Currie and Kerrin, (2003) in Gillani, Iqbal, Akram, and Rasheed (2017: 180) performance appraisals can hamper and choke the sharing of knowledge within the organizational context because there might be conflicts between the different individuals, and functional departments, and or sections in the organization.

One way to mitigate the negative impact of performance appraisals is to ensure that fair feedback is provided to the staff members thus strengthening the desired outcomes (Gillani, Iqbal, Akram, and Rasheed, 2017:180).

Motivation on an individual level is very closely related to organizational motivation. The literature reviewed for this research indicates that individual motivation is influenced and defined by various intrinsic and extrinsic factors.

When considering the intrinsic factors, it is important to note that each staff member's intrinsic motivation will differ based on their frame of reference, i.e. background. Furthermore, the literature indicates that intrinsically motivated staff members are more prone to sharing knowledge for the satisfaction it brings to the education of others.

Sathitsemakul and Calabrese (2017:81) identified the following examples of intrinsic factors:

- interpersonal trust,
- organizational commitment,
- and self-efficacy.

In contrast to intrinsic motivation, extrinsic motivation may be defined when external reasons drive an individual to perform a specific task. These reasons include but are not limited to:

- reward in any form,
- negative consequences or punishment,
- to increase individual self-worth and importance,
- or goal-orientated.

It should, however, be noted that according to Muqadas et al. (2017) in Kim and Hang (2024:2) believe that an over-emphasis on rewarding extrinsic motivation does not necessarily stimulate knowledge-sharing activities within the organizational context.

Emotional Intelligence (EI) may be regarded as a very important motivating factor. Sathitsemakul and Calabrese (2017:82) indicated that emotional intelligence "…is the ability to perceive emotions and cognitive processes such as reasoning with emotions, understanding their meaning, assimilating and locating relationships between the emotions". Research indicates that EI can potentially influence both the intrinsic and extrinsic motivation of staff members. It should be noted that EI has proven to develop individual motivation which influences knowledge-sharing behavior Sathitsemakul and Calabrese (2017:82). Shyh et al., (2006) contend that EI has the ability and impact to raise the propensity to share knowledge even if the individual is reluctant to do so.

3.3 People

Inkpen and Tsang, (2005) state that in essence an organization and more specifically financial institutions may be regarded as a social network where the organizational hierarchy will influence the social capital, i.e. the staff members. According to the authors, it is the social capital of the organization that underpins all knowledge-sharing activities and exchanges and the communication thereof.

The driving concepts identified in this theme include but are not limited to:

- profile
- gender

An interesting observation from the systematic literature review, reveals a close relationship to individual staff member profiles and knowledge sharing. Abbas, Hussain, Hussain, Akram, Shaheen, and Niu (2019:2) state that "qualifications, work experience, working relationships, and individual income might meaningfully impact knowledge sharing, motivation, and willingness" .

In addition to the initial observations from Abbas et al (2019), it should also be interesting to note that staff members with little job experience and staff members with an established and a "...greater professional level..." will have a higher propensity towards knowledge sharing activities.

Although most of the academic literature about knowledge, knowledge management, and knowledge sharing perceives that the activity of sharing knowledge is a gender-neutral activity, research suggests a negative imbalance towards female knowledge-sharing activities. According to Colley, 2003; Meelissen and Drent, 2008; Volman et al., 2005 in Nguyen and Malik (2022:1997) this imbalance may be attributed to assumed higher levels of technology anxiety among women. This assumed technological anxiety will impact of perceived usefulness of knowledge-sharing activities using knowledge-sharing online platforms in the organizations resulting in less knowledge being shared.

3.4 Technology

As alluded to in the introduction, (section 1), technology, specifically in the financial sector, plays a mission-critical role in facilitating financial transactions. The researchers would even state that no financial transactions in the current hyper-connected business environment will be possible without the use of the information technology backbone. The nature of the review did not focus on specific elements, specifications, or requirements of technology, but rather on what basic features or characteristics will enhance the use of technology for knowledge sharing.

In considering and analyzing the literature from a financial institution perspective the following main technological agnostic themes that drive research in technology were identified:

- online platforms
- human sensory feedback system

Nguyen and Malik (2022:1986) state that "…online platforms in an organization refers to using social networking, intranet, and other platforms to enhance knowledge sharing via communication and collaboration." It should be noted that the online platforms and integration of applications via the platforms (for example the Internet) form the backbone of all interaction and communication in modern business. Financial institutions are not exempt from the use of online platforms in every form or protocol, with more and more emphasis on the integration of Artificial Intelligence (AI) as an advanced communication and workflow platform.

The authors contend that the use of AI and AI-enabled technologies has resulted in a higher level of management of people, experiences, and talent management specifically within the financial sector. Thus, it may be deduced that AI and the related technologies on the agnostic platforms are not only used for knowledge sharing, but without said agnostic platforms knowledge sharing cannot happen.

An interesting theme emerged from the systematic literature review on knowledge sharing in financial institutions, with specific reference to technology, i.e. the delicate interplay between human information processing and technology. According to Chen, Ye, and Huang (2022:1) "Knowledge-sharing through ICT is a form of computer-mediated communication (CMC)." Scholl et al., 2020 in Chen, Ye, and Huang (2022:1) think that the notion of CMC provides a substantial advantage of communicating across vast distances but limits the sensory feedback for humans in this communication process.

According to Freitas-Magalhaes (2020) in Chen, Ye, and Huang (2022:1) sensory feedback may be defined as the non-verbal cues that humans add to the communication process, i.e. facial expressions, body language, or changes in emotions.

The non-verbal communication cues may be regarded as an essential part of the knowledge-sharing process, specifically in the financial sector. A lack of sensory feedback in this context may leave the communicating partners with a lack of trust in the communication process and what is shared in the communication, i.e. the knowledge that should be transferred. "Consequently, this situation constrains knowledge-sharing, leading to poor resilience". Chen, Ye, and Huang (2022:3).

A possible solution to the sensory deprivation issue that is a by-product of the use of current-day technology, is the incorporation of the notion of a virtual world, a metaverse, using a combination of Artificial Reality, Virtual Reality, and Artificial Intelligence elements.

4 CONCLUSION

Considering the main research problem as discussed in section 2, i.e. understanding the themes driving research in knowledge sharing, specifically in financial institutions, considering that these institutions and functions within the institutions are often characterized by individuals hoarding and not willing to share what they know, the systematic literature review conducted, yielded the following main themes, i.e. environment, motivation, people and technology.

The authors further analyzed each of the main themes to understand the concepts in each of the elements and their link to the notion of sharing knowledge in financial institutions, with specific emphasis on the use of technology.

The resulting analysis revealed that the concepts of organizational attributes, collaboration, trust, and rewards drive the research in terms of the environment. In summary, the management of organizations and more specifically the financial sector should create and maintain an environment that allows for the alignment of organizational and personal objectives, while fostering collaboration between staff members, based on mutual trust. This environment should be driven by a combination of both monetary and non-monetary rewards.

The next main theme identified by the researcher was that of motivation, i.e. what drives an individual staff member to participate in knowledge-sharing activities. The main concepts driving the theme of motivation are organizational motivation and strategies, individual motivation, and emotional intelligence.

Motivation may be regarded as one of the fundamental drivers of knowledge sharing within financial institutions. Organizations should endeavor to align organizational objectives with staff member motivation, both on an intrinsic and extrinsic level. Extending the notion of intrinsic and extrinsic motivation, the literature revealed that both factors may be influenced by the individual staff members' emotional intelligence. It is postulated that individual staff members with a higher EI will have a greater propensity to share knowledge within the organizations with specific reference to the financial sector, employing highly intellectual capital.

When considering the theme of people in the analysis of the literature, two interesting concepts in terms of knowledge sharing in financial institutions were identified, i.e. profile and gender.

According to the literature, the profile of the individual staff member within the financial institution will be a good indicator of the individual's willingness to participate in knowledge-sharing activities. The profile of the individual staff member refers specifically to the individual's qualifications, work experience, working relationships, and individual income.

Although knowledge sharing is considered a gender-neutral activity, some research suggests that there might still be some imbalance in the use of technology by females to share knowledge.

The final theme identified relates to technology and its use in financial institutions for knowledge-sharing purposes. The researcher identified two concepts that are deemed important considerations in driving research in this specific theme, i.e. online platforms and human sensory feedback systems.

The two concepts, as identified in the literature are closely related. The online platforms provide the necessary backbone to support daily activities with a specific emphasis on communication. Although the platforms provide and facilitate communication and connection over great distances, they cannot provide essential sensory feedback that allows individual staff members to trust the communication and knowledge-sharing activity.

Visually the results of the analysis may be presented in figure 7.

Figure 7: Conclusion of the study.

The constant changes in Information Technologies and the necessity for managing them especially in financial institutions, ITSM would benefit from KM to address the demands of the fourth industrial revolution.

The following section will provide an overview of future research directions.

5 FUTURE RESEARCH DIRECTIONS

The researchers propose that the identified themes be tested and confirmed in a financial institution. The test and confirmation should focus on institutions in both developing and developed countries. The results of the proposed future research should then be extrapolated to other sectors and environments.

ETHICAL CONSIDERATIONS

Ethical clearance for the proposed research was reviewed by the School of Consumer Intelligence and Information Systems Research Ethics Committee of the University of Johannesburg. Ethical clearance was granted with ethical clearance code 2024SCiiS011, with a rating CODE 01(Approved).

REFERENCES

Abbas, J., Hussain, I., Hussain, S., Akram, S., Shaheen, I., & Niu, B. (2019). The Impact of Knowledge Sharing and Innovation on Sustainable Performance in Islamic Banks: A Mediation Analysis through a SEM Approach. *Sustainability*, *11*(15). https://doi.org/10.3390/su11154049

Ahmad, F., Karim, M. (2019). Impacts of knowledge sharing: A review and directions for future research. *Journal of Workplace Learning*, *31*(3), 207–230. https://doi.org/10.1108/JWL-07-2018-0096

Allen, H. J. (2021). Payments Failure. *Boston College Law Review*, *62*(2), 453–514.

Alshwayat, D., MacVaugh, J., & Akbar, H. (2021). A multi-level perspective on trust, collaboration and knowledge sharing cultures in a highly formalized organization. *Journal Of Knowledge Management*, *25*(9), 2220–2244. https://doi.org/10.1108/JKM-05-2020-0354

Axelos. (2020). *Service continuity management: ITIL 4 practice guide*.

Ammirato, S., Linzalone, R., & Felicetti, A. M. (2021). Knowledge management in pandemics. A critical literature review. *Knowledge Management Research & Practice*, *19*(4), 415–426. https://doi.org/10.1080/14778238.2020.1801364

Babcock, P. (2004). Shedding light on knowledge management. *HR Magazine*, 49(5), 46–50.

Baradari, I., Maryam S., & Navid N. (2023). Defining the relationship between IT Service management and knowledge management: towards improved performance. *Knowledge Management Research & Practice* 21(2): 384-396.

Beyer, B., Jones, C., Petoff, J., & Murphy, N. R. (2016). *Site reliability engineering: How google runs production systems*. O'Reilly Media, Inc.

Binham, C. (2020, January 1). *Lloyds Banking Group customers hit by system outage*. https://www.ft.com/content/22387a94-2ca6-11ea-bc77-65e4aa615551

Bock, G.-W., Zmud, R. W., & Kim, Y.-G. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces and organizational climate. MIS Quarterly, 29(1), 87–11.

Bock, G.-W., Kim, Y.-G. (2002). Breaking the Myths of Rewards: An Exploratory Study of Attitudes about Knowledge Sharing. *Information Resource Management Journal*, 15, 14–21.

Chen, P., Ye, Y., & Huang, X. (2023). The metaverse in supply chain knowledge sharing and resilience contexts: An empirical investigation of factors affecting adoption and acceptance. *Journal Of Innovation & Knowledge*, *8*(4). https://doi.org/10.1016/j.jik.2023.100446

Chauhan, C., Parida, V., & Dhir, A. (2022). Linking circular economy and digitalization technologies: A systematic literature review of past achievements and future promises. *Technological Forecasting & Social Change*, *177*, 121508. https://doi.org/10. 1016/j.techfore.2022.121508

Domingos, J., Barbosa, R., & Madeira, H. (2021). Why is it so hard to predict computer systems failures? *2021 17th European Dependable Computing Conference (EDCC)*, 41–44. https://doi.org/10.1109/EDCC53658.2021.00013

Gold, A.H., Malhotra, A., Segars, A.H. (2001). Knowledge management: an organizational capabilities perspective. *Journal of Management Information Systems*, 18(1), 185-214.

Gillani, S., Iqbal, S., Akram, S., & Rasheed, M. (2018). Specific antecedents of employees' knowledge sharing behavior. *Vine Journal Of Information And Knowledge Management Systems*, *48*(2), 178–198. https://doi.org/10.1108/VJIKMS-05-2017-0023

Inkpen, A.C., Tsang, E.W. (2005). Social capital, networks, and transfer. *Academy of Management Review*, 30(1), 146-165.

ISO. (2015). *ISO/IEC 27043:2015: Information technology—Security techniques—Incident investigation principles and processes*. https://www.iso.org/standard/44407.html

Jacobs, D.C. (2010). Pragmatism. In *Encyclopaedia of Case Study Research*. Edited by Mills, A.J., Durepos, G. & Wiebe, E. doi: 10.4135/ 9781412957397

Kelemen, M.L. & Rumens, N. (2011). Critical perspectives on qualitative research. In *An introduction to critical management research.* doi: 10. 4135/9780857024336

Khan, M., & Zaman, U. (2022). Bankers' knowledge sharing behavior: The role of personality and perceived organizational incentives. *Journal of Public Affairs*, *22*(1). https://doi.org/10.1002/pa.2340

Khiaonarong, T., Rizaldy, R., & Leinonen, H. (2022). Payment outages: Experiences and issues. *Journal of Payments Strategy & Systems*, *16*(2), 170–180.

Kim, N., & Hang, N. (2024). Rewards, knowledge sharing and individual work performance: An empirical study. *Cogent Business & Management*, *11*(1). https://doi.org/10.1080/23311975.2024.2359372

Klee, E. (2010). Operational outages and aggregate uncertainty in the federal funds market. *Journal of Banking & Finance*, *34*(10), 2386–2402. https://doi.org/10.1016/j.jbankfin.2010.03.002

Langerman, J., & Joseph, N. (2023). Information Systems Resilience: An African Banking Case Study. *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1–8.

Lin, K.-Y., Lu, H.-P. (2011). Why people use social networking sites: an empirical study integrating network externalities and motivation theory. *Computers in Human Behavior,* 27(3), 1152-1161.

López-Nicolás, C., Meroño-Cerdán, Á.L. (2011). Strategic knowledge management, innovation, and performance. *International Journal of Information Management,* 31, 502–509.

MacLean, D., & Ryad, T. (2023). Implementation and impacts of IT Service Management in the IT function. *International Journal of Information Management* 70: 102628.

McNaughton, R., Darroch, J. (2003). Beyond market orientation: Knowledge management and the innovativeness of New Zealand firms. *European Journal of Marketing,* 37, 572–593.

Merrouche, O., & Schanz, J. (2010). Banks' intraday liquidity management during operational outages: Theory and evidence from the UK payment system. *Journal of Banking & Finance*, *34*(2), 314–323. https://doi.org/10.1016/j.jbankfin.2009.07.024

Mishchenko, V., Naumenkova, S., Grytsenko, A., & Mishchenko, S. (2022). Operational risk management of using electronic and mobile money. *Banks and Bank Systems*, *17*(3), 142–157. https://doi.org/10.21511/bbs.17(3).2022.12

MyBroadband. (2017, November 24). Takealot down minutes into Black Friday 2017. *MyBroadband*. https://mybroadband.co.za/news/business/239352-takealot-down-minutes-into-black-friday-2017.html

Nguyen, T., & Malik, A. (2022). Employee acceptance of online platforms for knowledge sharing: Exploring differences in usage behaviour. *Journal Of Knowledge Management*, *26*(8), 1985–2006. https://doi.org/10.1108/JKM-06-2021-0420

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. & Moher, D., (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ.* 372(71), 8.

Reichertz, J. (2014). Induction, deduction, abduction. In *The SAGE Handbook of Qualitative Data Analysis.* Edited by Flick, U. doi: 10.4135/9781 446282243

Sathitsemakul, C., & Calabrese, F. (2017). The Influence of Emotional Intelligence on Employees' Knowledge Sharing Attitude: The Case of a Commercial Bank in Thailand. *Journal Of Integrated Design & Process Science*, *21*(1), 81–98. https://doi.org/10.3233/jid-2017-0007

Sekaran, U., & Bougie, R. (2013). *Research methods for business: A skill-building approach,* John Wiley, United Kingdom, 6th edition.

Seneviratne, T. K. K., Amaratunga, D., Haigh, R., & Pathirage, C. P. (2010, May). *Knowledge management for disaster resilience: Identification of key success factors* [Conference]. CIB 2010, University of Salford. https://www.hud.ac.uk/news/

Shyh, R. F., Shin, C. F., Ming, C. C. (2006). The impact of internal marketing mechanism on KS motivation and KS behavior. Department of Business Administration, National Chung Hsing University, Taiwan.

Sillito, J., & Kutomi, E. (2020). Failures and Fixes: A Study of Software System Incident Response. *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 185–195. https://doi.org/10.1109/ICSME46990.2020.00027

Von Krogh, G., Nonaka, I., Rechsteiner, L. (2012). Leadership in organizational knowledge creation: a review and framework. *Journal of Management Studies*, 49(1), 240-277.

Performing Entity Relationship Model Extraction from Data and Schema Information as a Basis for Data Integration

Philipp Schmurr[a], Andreas Schmidt[b], Karl-Uwe Stucky[c], Wolfgang Suess[d] and Veit Hagenmeyer[e]

Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Kaiserstr. 12, Karlsruhe, Germany
{philipp.schmurr, andreas.schmidt, karl-uwe.stucky, wolfgang.suess, veit.hagenmeyer}@kit.edu

Keywords: Entity Relationship Model, Model Extraction, Data Integration, Structural Metadata, FAIR Principles.

Abstract: The goal of this work is to allow domain experts to properly perform data integration themselves and not to rely on external resources. This way the long-term data integration quality is not endangered and therefore cost for external resources can be saved. To achieve this, we propose a new approach that enables data integration based on entity-relationship (ER) models derived from arbitrary data sources. ER models are abstract and simply define all entities and relations needed for integration, which makes them easy to understand. Strategies to extract ER models from various standard data sources - relational databases, XML files and OWL data - are presented and a concept on how to extend it to arbitrary other data sources is introduced. Furthermore, the extracted models are a foundation to perform graphical data integration into an ontology based model and, thus, contribute to a harmonized knowledge management in heterogeneous data and information environments. It can be summarized as a strategy to improve the interoperability of existing data according to the FAIR principles.

1 INTRODUCTION

Whenever a new use case or project is started in industry, it usually requires gathering and integrating data from various sources to be able to achieve the intended goal. This can include internal or external data that is probably stored in multiple different information systems, file formats, and locations.

Especially with limited budget and staff members, data integration is performed manually to collect and transform the data as needed for the new project. In more advanced projects, data integration is done using available solutions that provide an automated and repeatable way for data integration. Data integration solutions often require internal data modeling experts or an outsourcing to external contractors. For cost reasons, the solutions are often simplified as much as possible, so that they hardly comply to the state of the art.

The energy domain is particularly affected by insufficient data integration. The energy transition requires more and more data-driven solutions, but especially smaller power companies do not have the necessary data experts yet. Moreover, software - similar to hardware - is mainly delivered as turnkey solutions with proprietary data models, formats and interfaces. In this context, the vendor lock-in is a considerable disadvantage, e.g. because the available interfaces to communicate with other software are hard to understand. Another challenge is that the energy system is a critical infrastructure and therefore all kinds of cloud solutions are not feasible. On the other hand, many software systems in the energy domain have originally been designed by electrical engineers as tools to support their daily work. Those tools have then evolved into software products that still have many legacy issues from a bygone era. As a result of the above reasons, the energy domain has many custom and legacy data sources and tools that are harder to integrate compared to other domains. The statements made above are based on our previous working experience in the energy industry.

Due to the mentioned conditions, we perceive it is a key characteristic of a data integration solution to be usable by energy domain experts without external assistance. As a solution we introduce FAIRlead,

[a] https://orcid.org/0009-0004-2324-7839
[b] https://orcid.org/0000-0002-9911-5881
[c] https://orcid.org/0000-0002-0065-0762
[d] https://orcid.org/0000-0003-2785-7736
[e] https://orcid.org/0000-0002-3572-9083

Schmurr, P., Schmidt, A., Stucky, K-U., Suess, W. and Hagenmeyer, V.
Performing Entity Relationship Model Extraction from Data and Schema Information as a Basis for Data Integration.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 316-322
ISBN: 978-989-758-716-0; ISSN: 2184-3228

a concept for a data integration and management system for the energy domain. It will be open source and, therefore, especially tailored for smaller budgets. The relevant characteristics of FAIRlead are the following:

- Data integration is performed with a graphical user interface that is based on conceptual entity-relationship (ER) models (Chen, 1976) of the input data sources.
- The user interface shall be easy to understand and works the same for all kinds of data sources.
- The ER models are semi automatically extracted from data sources and allow the user to improve or correct the model if necessary.
- The integration target model is based on ontologies.
- We will employ code generation methods on the created target ontology to simplify the access to the mapped data.

So in this paper, we demonstrate the first step necessary for the approach depicted above: a strategy to extract ER models from data directly, thereby using structural metadata if available. It is important to state that the user will have the final control on improving or correcting the generated ER model. To demonstrate our concept, we present the extraction process for relational databases, XML files, OWL data sets and a proprietary text-based file format from the energy domain. For the demonstration we are using the Mondial database (May, 1999) which is available in the mentioned formats, as well as a scenario file (RAW) from the Siemens PSS®E power system simulator.

The present paper is organized as follows: in Section 2 we present related work on extracting models from metadata and data. Then, in Section 3, we present the data sources on which our experiments are based. The basic concepts of extracting the components of an ER model from different data sources are explained in Section 4. In Section 5, we then present initial results of the model extraction tools we have developed so far. In Section 6 we summarize our results and formulate an outlook for future research.

2 RELATED WORK FOR MODEL EXTRACTION

Extraction or reverse engineering of models has already been done on several types of data:

For spreadsheet data there exist several publications around the extraction of ClassSheet models (Cunha et al., 2010). Those models can be integrated into a spreadsheet file directly (Cunha et al., 2012) and can also be compared against relational schemas (Cunha et al., 2016).

The detection of what table headers are, as well as the recognition of relations between the tables is a key benefit of the proposed ClassSheet implementations. We have attempted to integrate those features from the available open source code, but struggled with the more than a decade old code base. Moreover, the implementation only focuses on Open Office spreadsheets which is a harsh limitation.

Relational databases also have been target of model extraction work before. Chiang et al. have presented an approach to extract an extended entity relationship (EER) model from a given database (Chiang et al., 1994b) and later also investigated the performance of their approach (Chiang et al., 1994a). Later Alalfi et al. have published a solution to extract the EER model and to export it to a UML diagram in the XML meta interchange format (XMI) (Alalfi et al., 2008).

The logic to extract the concepts of an ER model (entities, attributes, relations, cardinalities) proposed in those works, is reused in a similar form in our solution.

However, we have focused on a more lightweight semantic representation of the ER model compared to those papers. So for us, the diagram can be generated from our ER representation, but it is not the primary representation.

The next big group of model extraction papers targets the extraction of data from XML files as well as their existing schema representations (for example Document Type Definition (DTD)). One approach to extract a DTD from XML data is presented by Siau et al. (Siau et al., 2011). This approach even creates a new graph format they call Extended DTD Graph. Moreover, the approach employs techniques to find out relations between elements that are based on ID/IDREF(s) relationships and not just the hierarchical structure of the document. Klímek et al. created a survey of approaches to extract schema information from XML data also including further schema formats as XML schema (XSD), RelaxNG and schematron (Klímek and Nečaský, 2010). Finally, the extraction of ER models from DTDs is presented by both Yang (Yang et al., 2004) and Mello (Mello and Heuser, 2001). The first aims to improve existing DTDs since they are hard to read and understand. The ER representation is therefore used as a means to improve understandability. This follows the same basic idea as our approach to make data integration more approachable for domain experts. The approach from Mello employs a rule set to extract a canonical conceptual model in an ontological representation and then also utilizes it for a semantic integration solution that focuses on XML data. The basic idea of conceptual model extraction for integration purposes is similar in our approach and we consider the idea to represent the conceptual model with an ontology for the future. We apply the concepts presented in these papers to extract the ER models from a DTD source. A newer approach by Della Penna et al. was proposed to extract an ER model also from XML schema (XSD) (Della Penna et al., 2006). Our implementation does follow a similar approach when working with a dataset that provides XSD information. In general, our solution can utilize existing software for either DTD or XSD extraction as a pre-processing step to utilize the benefits of an available DTD or XSD schema compared to directly extracting an ER model from XML data alone.

Lastly, the extraction of ER models from data represented in the Resource Description Framework (RDF) is an area of interest. However, RDF based data usually has an ontology (mostly in Web Ontology Language (OWL) format) as a model which is already an integration ready way of representing a data structure. The terms conceptual model (ER models are conceptual models) and ontology often appear together, both as a means to understand the domain before creating an ontology (Gomez-Perez et al., 2000) and also to use ontologies (e.g. the Unified Foundational Ontology (UFO)) to properly define the semantics of a conceptual model that is built for a new information system (Guizzardi, 2005). The latter approach is called ontology-based conceptual modelling.

To actually extract conceptual models from ontologies El-Ghalayini et al. proposed a rule based approach that was intended to later merge multiple conceptual models into one overarching conceptual model of a target domain (El-Ghalayini et al., 2005). A similar rule based approach was presented by Han et al. that directly focused on ER models (Han et al., 2010) and also used OWL while El-Ghalayini did still work with the OWL predecessors. Our presented solution is similar to those two, but puts more detail into also understanding the restrictions and cardinalities of the input ontology.

Last but not least, there is the Conceptual Model Ontology (CMO) (McCusker et al., 2011) that provides the possibility to annotate other ontologies with conceptual model concepts. One of the goals of the CMO is also to allow integration of different data sources, but rather by allowing to use a common natural language terminology to query different data sources. This only works if the concepts have already been annotated with the additional triples, while our approach tries to aid the process of creating the integration mapping instead.

3 DEMONSTRATION DATA SETS

The Mondial database[1] is a collection of data about the world - more precisely countries, cities and geographic features like mountains, lakes or rivers, as well as some demographic features like economy, religions and ethnic groups.

We selected it as one of the data sets for this publication due to the availability in several data formats and because it is more than just a trivial example, containing more than 15 entities and more than 20 individual relations. Also, it includes different concepts that are relevant for ER models like the differentiation between weak and strong entities, key attributes and all sorts of different relation cardinalities. Moreover, it provides a reference ER model that we can compare our results against.

Since we focus our work on the energy domain and also claim to support arbitrary data sources, we also include the SAVNW example power grid model from the Siemens PSS®E power system simulator in RAW format. This model represents the topology and component attributes of an electrical power grid. It is serialized as a text file that can be interpreted as a set of different tables, one for each power grid element type like busbars, transformers and generators. The serialization may contain comments for the column labels, but this is not mandatory. So without any particular inputs about the structure of the file format the ER model will not be able to extract anything else than a set of enumerated tables and a set of enumerated attributes. This scenario is a good example that needs to leverage additional user inputs about the ER model.

4 ER MODEL EXTRACTION

This Section will give a brief overview on the components of an ER model and how they can be extracted from various data source types. This is mostly limited to the important keywords to look for with a certain data type. The works referenced in the related work section can give more detailed instructions on how to perform the necessary extraction steps.

ER models are used to create an abstract representation of the most important concepts and relationships while building a database model. It helps

[1] https://www.dbis.informatik.uni-goettingen.de/Mondial/

to clarify the required capabilities of a model before transforming it into a physical set of tables and constraints in a database system. ER models basically consist of entity types, relation types, their cardinalities and attribute types.

There are several common notations to visualize ER models as diagrams. The notation used in the this paper is the Chen notation (Chen, 1976). Example diagrams can be seen in Figure 1 which will be explained in more detail in Section 5. In general, entity types are depicted as rectangular boxes, while relation types use a diamond shape. Attribute types use an oval shape and cardinalities are applied as labels to the edges between entity types and relation types.

Double-lined rectangles or diamonds indicate either so-called weak entity types or identifying relationship types respectively. This is a special way of indicating that the weak entity can not exist without the entity connected via the identifying relationship.

The usage of the term entity is sometimes not precisely clear. In order to be precise the following rules apply: An entity is the instance of an entity type and all entity instances form an entity set. An ER diagram does only depict entity types, so in natural language it can happen that the words "type" and "set" are omitted when talking about ER models. This effectively means that the word "entity" is used synonymously for all of them.

4.1 Entities

Entities are often the most simple component to extract from any data source. For any tabular data like a set of CSV files or a relational database, the entities are usually reflected by the tables. However, tables can also represent $n : m$ relations, which requires some rules to check both the table's foreign and primary keys (in a relational database). In CSV based data there is not direct clue that can be used to determine if a table is a relation or not. For OWL data the entity types are usually represented by the *owl:Class* type, so the entities are the instances of that class respectively. Depending on the used OWL model it is important to use a reasoner that fills in missing class definitions from *rdfs:subClassOf* predicates.

In XML an entity is represented by an element that has attributes or does contain child elements (called *complexType* in XSD). Lastly, in hierarchical object notations like JSON all objects are considered an entity. Therefore, the ER diagram does contain one rectangle per entity set respectively. The decision what type an entity belongs to can be difficult (e.g. if there can be optional values that are not present in every object). If there is schema information available, it becomes much easier to assign objects to their respective entity type.

4.2 Relations

Whenever an entity refers to one or more other entities this is usually done with a relation.

In object notations this can either be a property that has another object as its value or its value is a reference to some kind of ID attribute. In OWL we have the dedicated *owl:ObjectProperty* concept that specifies a relation between entities. There it is necessary to correctly track the domains and ranges of the respective property to see what entities can actually be connected with this relation. For tabular data and relational databases it is harder to recognize relations. Generally a foreign key constraint in a relational database represents a relation. However, it was mentioned above that sometimes a table can also represent a relation. So to make a decision it is required to examine the foreign and private key constraints. For bare table data (e.g. CSV) without any modeled key constraints, a strategy to find potential relation types is to check for common naming patterns that involve for example ID columns in each table and a combination with the table name in others (e.g. EntityB has a column EntityA_ID). However, in these cases it is often better to just have the user correct the extracted ER model to contain proper relations.

4.3 Cardinalities

In order to carry the intended meaning, relations require quantity constraints that apply between the connected entities. For plain table data this is almost impossible to tell without any additional schema information. So, again this is one of the points that require the user to be able to correct the model with their knowledge.

Relational database schemas allow to infer cardinalities through their modeling patterns and constraints. In OWL there is the *owl:Restriction* concept as well as the exact, max and min cardinality predicates that can apply to an *ObjectProperty* and the domain (and sometimes even range) in question. However, if there is no inverse property defined there is no clear indication for the second half of the relation cardinality, as one property only defines the cardinality on the domain side. For hierarchical object notations there is no precise solution to extract cardinalities without an additional schema that specifies them. For XSD these are the *minOccurs* and *maxOccurs* attributes on the element. In JSON schema, cardinalities can be inferred from arrays if there are allowed element counts and for an object property it can be specified if it is required or not.

4.4 Attributes

In relational databases, all attributes that are not used in the definition of foreign key constraints can be considered as attributes in the context of an ER model. In simple table based data, the attributes are all the columns that do not qualify to be part of a relation. In OWL there is the dedicated *owl:DatatypeProperty* concept for the purpose of encoding attributes. For object notations all primitive properties can be considered attributes, if they are not a reference to another entity's ID. In XML, elements without XML attributes and only a primitive content as well as XML attributes are typically regarded as attributes of the ER model.

5 EXAMPLES FROM THE FAIRlead ER MODEL EXTRACTION

Figure 1 shows the ER diagram of the province entity in three versions. It uses the mondial XML data set. The upper version is extracted without including any additional schema information. The second version does then include information from the official mondial DTD file. And the last version is generated using the official XSD file.

Several differences can be observed between the three strategies. First, the version without any schema information does use the entity name *mondial/country/province* which is our solution's way of compound naming the province entity type to indicate the hierarchical position of the entity set within the XML file. This is the case, because without schema information it might be possible to encounter different province instances at different positions in the XML tree and could not guarantee that they are equal. The *mondial/country/province/city* entity type is an example of this - also occurring as *mondial/country/city* if it does not belong to any province. This one or none relation can even be seen in the third diagram at the *citytoprov* relation that is derived from a *xsd:keyref* between the two entities.

In the lower two diagrams we see *province* to be an independent entity, which is guaranteed by either a *xsd:key* in XSD or an *ID* type in DTD. Additionally, the relations in the lower diagrams can have different names, while without a schema the only relation is *IsChild*, which reflects the hierarchical structure of the

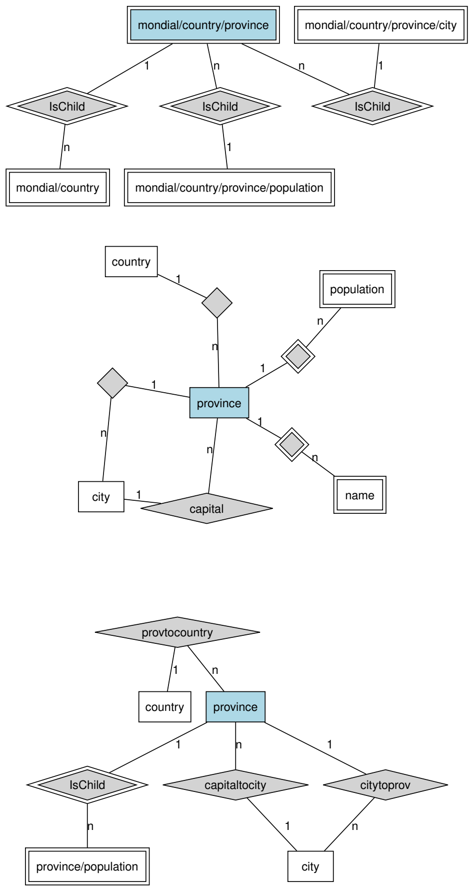


Figure 1: ER extraction of the province entity from Mondial XML data in three processing types.

XML document. Overall, it seems the XSD version is the most accurate one - but for many other entities the quality of the model lacks behind the schema information available from a relational or ontological model.

An example for the lacking quality that is based on the modeling decisions in the XSD file can be seen in Figure 2. It shows, a pattern that can often be seen in the model extracted from the XSD. The highlighted entity *river/located* can also be seen as a relation itself. For this kind of scenario it might be possible to contract the resulting ER diagram and produce an *n to n* relation *located*. This is something that we might consider in the future to optimize certain patterns in the resulting ER diagram independent of the original data source.

Looking at Figure 3 is a nice example of how cardinalities can be extracted from OWL data. The *isBorderOf* relation is correctly reflected as a *2 to n* relationship. Moreover, the *locatedIn* relation has no restrictions in OWL, so it correctly is depicted as a many-to-many relationship.

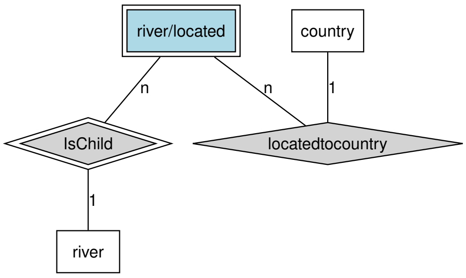


Figure 2: Extracted ER model from the XSD schema that could be contracted with a post-processing step.

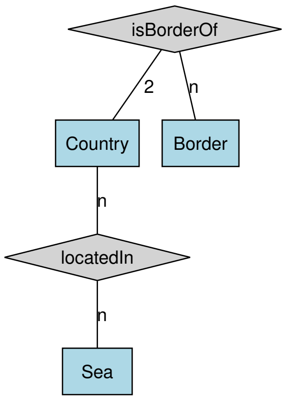


Figure 3: Extracted ER model of the OWL data set that showcases the cardinality possibilities in an ontology.

Considering the PSS®E data set, the matter is more complicated. The file theoretically can include structural metadata in the form of comments, but that is not mandatory to be a valid file. Depending on the source the file was received from, these comments might not be available. So in the worst case scenario, an extracted ER model might contain no information on table and column names. So after some quick pre-processing to split the RAW file into a set of CSV files, the only information for the ER model available is a set of entities named *Table1* to *TableN*. Each of those entities will have a set of attributes named *Column1* to *ColumnN* respectively. At that point it is mandatory, to allow the user to edit this extracted ER model. It might be enough to just correct the bits and pieces that are later relevant for the data integration step. An example of such a user corrected ER model can be seen in Figure 4.

Finally, it must be noted that none of the extracted models exactly matches the original ER diagram of the mondial database. A reason for this are the specific modeling decisions that have been made to create the respective physical models, that do not allow for an exact reconstruction. In general, this is not a problem for the target purpose of performing data integration. In our approach the domain expert that knows the data sources is the same person that shall perform the visual data integration. This means as long as the ER representation is reflecting the data source more or less accurately the user will be able to make sense of it.

However, there are extreme cases like the XML file with no schema information or the PSS®E data set as well as the cases where relations have been detected as entities. In those examples, it may not be possible to infer any useful ER model. In that regard the presented approach is semi automatic and the user needs to introduce corrections to the ER model. This has been shown as an example with the PSS®E data set and the ER model in Figure 4.

6 CONCLUSION AND FUTURE WORK

With our data integration and management solution FAIRlead, we want to enable domain experts to perform data integration in a way that improves the FAIRness of their existing data. As a first step of this solution, we have presented an open-source tool that allows the extraction of ER models from various data sources that can be reused for future data integration efforts. Moreover, additional data sources can be integrated in the model extraction either by implementing a new converter or by converting the input data to an already supported format. Many potential data sources can likely be transformed into one of the presented ER model extraction solutions with a small pre-processing step. With the example of the PSS®E RAW file, it has become clear that the user must be able to improve the extracted ER model, because the file can technically come without any structural metadata.

Finding an appropriate solution for the model corrections and also to allow visual editing of the extracted schema is one of our future steps. Moreover, we will continuously improve the current implementation, to produce accurate ER models according to the given input data. This for example includes the integration of existing schema generation tools into the process (e.g. to generate a DTD out of XML data before processing it).

Based upon the FAIRlead ER model extraction strategies presented in this paper, we will create a GUI that allows to perform data integration from multiple heterogeneous data sources using conceptual models. This user interface will use a flow-based programming approach to visually show the link between original data source entities and the resulting ontological concepts.

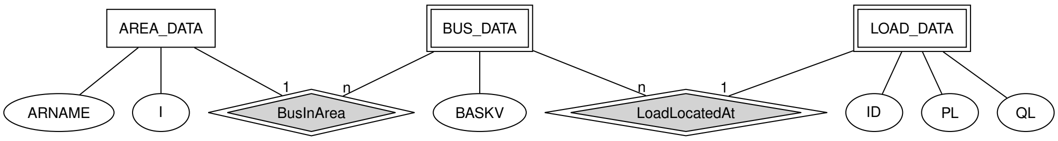


Figure 4: User corrected ER model of the PSS®E data set.

ACKNOWLEDGEMENTS

The authors would like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501865131 within the German National Research Data Infrastructure (NFDI, https://www.nfdi.de/).

This publication was also supported by the Helmholtz Metadata Collaboration (HMC, https://www.helmholtz-metadata.de/), an incubator-platform of the Helmholtz Association within the framework of the Information and Data Science strategic initiative.

REFERENCES

Alalfi, M. H., Cordy, J. R., and Dean, T. R. (2008). SQL2XMI: Reverse Engineering of UML-ER Diagrams from Relational Database Schemas. In *2008 15th Working Conference on Reverse Engineering*, pages 187–191.

Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36.

Chiang, R. H. L., Barron, T. M., and Storey, V. C. (1994a). Performance evaluation of reverse engineering relational databases into extended Entity-Relationship models. In Elmasri, R. A., Kouramajian, V., and Thalheim, B., editors, *Entity-Relationship Approach — ER '93*, pages 352–363, Berlin, Heidelberg. Springer.

Chiang, R. H. L., Barron, T. M., and Storey, V. C. (1994b). Reverse engineering of relational databases: Extraction of an EER model from a relational database. *Data & Knowledge Engineering*, 12(2):107–142.

Cunha, J., Erwig, M., Mendes, J., and Saraiva, J. (2016). Model inference for spreadsheets. *Automated Software Engineering*, 23(3):361–392.

Cunha, J., Erwig, M., and Saraiva, J. (2010). Automatically Inferring ClassSheet Models from Spreadsheets. In *2010 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 93–100.

Cunha, J., Fernandes, J. P., Mendes, J., and Saraiva, J. (2012). Extension and implementation of ClassSheet models. In *2012 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 19–22.

Della Penna, G., Marco, A. D., Intrigila, B., Melatti, I., and Pierantonio, A. (2006). Interoperability mapping from XML schemas to ER diagrams. *Data & Knowledge Engineering*, 59(1):166–188.

El-Ghalayini, H., Odeh, M., Mcclatchey, R., and Solomonides, T. (2005). Reverse Engineering Ontology to Conceptual Data Models.

Gomez-Perez, A., Fernández-López, M., and Vicente, A. (2000). Towards a Method to Conceptualize Domain Ontologies.

Guizzardi, G. (2005). *Ontological Foundations for Structural Conceptual Models*. PhD thesis.

Han, L., Xu, J., and Yao, Q. (2010). Entity-Relationship semantic meta-model based on ontology. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, volume 11, pages V11–219–V11–222.

Klímek, J. and Nečaský, M. (2010). Reverse-engineering of XML Schemas: A Survey. In *CEUR Workshop Proceedings*, volume 567, pages 96–107.

May, W. (1999). Information Extraction and Integration with Florid: The Mondial Case Study. Technical Report 131, Universität Freiburg, Institut für Informatik.

McCusker, J., Luciano, J., and Mcguinness, D. (2011). Towards an Ontology for Conceptual Modeling. *CEUR Workshop Proceedings*, 833.

Mello, R. d. S. and Heuser, C. A. (2001). A Rule-Based Conversion of a DTD to a Conceptual Schema. In S.Kunii, H., Jajodia, S., and Sølvberg, A., editors, *Conceptual Modeling — ER 2001*, pages 133–148, Berlin, Heidelberg. Springer.

Siau, K., Shiu, H., and Fong, J. (2011). Reverse Engineering from an XML Document into an Extended DTD Graph. pages 101–119.

Yang, W., Zhan, M., Wang, Q., and Shi, B. (2004). A conversion of a DTD to conceptual model by using UML. In *The Fourth International Conference onComputer and Information Technology, 2004. CIT '04.*, pages 303–308.

APPENDIX

The FAIRlead code used to perform the ER model extraction can be found on github: https://github.com/Cpprentice/FAIRlead-model-extraction

FAIRlead: A Conceptual Framework for a Model Driven Software Development Approach in the Field of FAIR Data Management

Andreas Schmidt[1,2][a], Mohamed Anis Koubaa[1][b], Nan Liu[1][c], Philipp Schmurr[1][d], Karl-Uwe Stucky[1][e] and Wolfgang Süß[1][f]

[1] *Institute for Automation and Applied Computer Science, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*

[2] *Department of Computer Science and Business Information Systems, Karlsruhe University of Applied Sciences, Karlsruhe, Germany*

{*andreas.schmidt, mohamed.koubaa, nan.liu, philipp.schmurr, karl-uwe.stucky, wolfgang.suess*}*@kit.edu*

Keywords: Code Generation, Metadata, Ontology Based Engineering, FAIR.

Abstract: The publication of scientific results together with the underlying experiments is an important source of further research. In 2016, the "FAIR Guiding Principles for scientific data management and stewardship" were published, in which the authors postulate a series of guidelines for improving the (F)indability, (A)ccessibility, (I)nteroperability and (R)eusability of digital information (FAIR). The point (I)nteroperability deals with the prerequisites for the reusability of digital objects. The central point here is the need to have a common understanding of the meaning of digital objects. This understanding is provided by formal languages of knowledge representation (ontologies), which describe the actual data. These descriptions of data are also known as metadata. As part of our current work at the Institute for Automation and Applied Computer Science (IAI) at KIT, we are implementing novel concepts and technologies for the sustainable handling of research data using high-quality metadata. As part of this work, we plan to develop a software tool that can be used to enrich data with suitable metadata and thus automate the process of making research results available. A key requirement is that the tool must be independent of the underlying domain. In order to be able to deal with data from any domain, we have opted for a model-driven approach in which an ontology, and possibly other platform-specific information, are input for a software generator, which then generates an (interactive) tool for specifying the metadata and linking it to the data itself. The generated tool includes the complete software stack, starting with a user interface, programmatic APIs for connecting additional application logic, and a persistence component. How these individual layers are realized is not specified, but defined by the mapping rules of the software generator, which also opens up the possibility of generating and evaluating different variants of the software.

1 INTRODUCTION

In computer science, an ontology is the formal naming and definition of the concepts, categories, properties and relationships between the concepts, data or entities of a particular domain (Ont, 2024). These basic concepts for ontologies are also the basic elements of Conceptual Models (CM), like the ER-Model (Chen, 1976) and UML (OMG, 2011). So, to that extend, languages and tools from both worlds can be used interchangeably. In the ontology context there are e.g. languages like OWL, RDF(S), and SHACL (Shapes Constraint Language) (SHACL, 2017). Having this in mind, the paper on hand generally uses ontology terminology and employs the CM terms only where more suitable for understanding.

[a] https://orcid.org/0000-0002-9911-5881
[b] https://orcid.org/0000-0001-8552-2008
[c] https://orcid.org/0009-0005-8768-7072
[d] https://orcid.org/0009-0004-2324-7839
[e] https://orcid.org/0000-0002-0065-0762
[f] https://orcid.org/0000-0003-2785-7736

1.1 Metadata

Scientific experiments take place in a specific context and it is within this context that data, parameters and results obtained have a practical meaning. In order to be able to interpret the underlying data and results in retrospect, detailed knowledge of this context is required. Data and its context form a unit that can be described by ontologies and their instantiations. This additional data is referred to as metadata, as it provides data about data.

Schmidt, A., Koubaa, M., Liu, N., Schmurr, P., Stucky, K-U. and Süß, W.
FAIRlead: A Conceptual Framework for a Model Driven Software Development Approach in the Field of FAIR Data Management.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 323-330
ISBN: 978-989-758-716-0; ISSN: 2184-3228

In the context of FAIR (Wilkinson et al., 2016), the present work deals with the aspect of (I)nteroperability, which intends the enrichment of data with rich metadata.

Due to the exponentially growing amount of scientific data, the enrichment with metadata must be automated by using suitable tools.

1.2 Example Scenario

In order to get a clearer picture of the requirements for such software, a concrete scenario will be presented here. The Institute for Automation and Applied Informatics (IAI) at KIT operates an experimental photovoltaic system (PV), consisting of a large number of panels that are connected together as an array (in parallel) or string (in series). Additional components of this PV system include power inverters, batteries, measuring equipment, and so on. An overview of the concepts and their relationships are shown in Figure 1. By reconfiguring the components, a large number of experiments can be carried out in order to achieve certain goals (e.g. maximum average electricity yield), or to observe the behavior under certain effects (e.g. partial shading).

As part of an experiment that is carried out with a specific interconnection of the components over a specific time interval and under specific weather conditions, a series of result data is generated that is stored in a time series database. In order to be able to interpret the measurement results, it is necessary to know the components' connectivity and the weather condition at this time. This is an example for meta information that has to be managed by our tool, together with the information on where the result values are stored.

1.3 Requirements for a Metadata Management Component

In order to fulfill this task, the software must have a persistence component that allows the specific test setup (the metadata) to be saved. Since we are talking about ontology-based metadata the database schema will be derived from the ontology. In addition, the component must have an interface through which the information can be entered. This can be, for example, a simple web-based CRUD (Create, Read, Update, and Delete) interface via which the individual components of the system can be specified, or a dedicated graphical editor with which the panels and other components can be graphically created and linked together.

The software must also maintain the connection between data and metadata. Data can be available in a variety of formats, such as measurement data in a time series database, relational data, parameter sets of a learned neural network, as well as a variety of proprietary data formats. Therefore, a component that establishes this data-metadata connection is needed.

Such a tool can provide the data of the experiment as well as the concrete setup (the metadata). By preparing this information according to existing standard formats, it is now able to export the data together with its metadata, or directly write it into a previously specified repository like the *databus* (Hoyer-Klick et al., 2023). This repository than covers the FAIR aspects (F)indable and (A)ccessible.

In contrast, in today's reality, the information about a specific experimental setup frequently is only implicitly available in configuration files, installation scripts or makefiles, which makes it almost impossible to extract this meta information.

The general functionality of the component just described is therefore not only useful for the publication of semantically enriched FAIR data, but also offers valuable services as an electronic notebook of the experiments carried out.

In addition, it is not limited to the PV domain used here as an example. The statements made here are valid for any domain. While the general functionality of the application is the same for all domains, the structure of the metadata will be different. It depends on the specific entities or concepts that describe the specific application. These are described by CMs or by the ontologies that describe the applications' domains.

1.4 Approach

And this brings us to the core idea of our research approach, the generation of an application, as described in the previous section, on the basis of the available domain information. We use a Model-Driven Software Development (MDSD) approach (see Section 2 for details). The application to be realized is generated from a model description (the ontology) and a number of transformation rules, which map the model information to source code for a specific target platform. In addition to the model information in the form of an ontology, the generator can process further information such as a specific GUI layout or information on the underlying software platform during the generation process.

The rest of the paper is structured as follows: Next, in Section 2 the basic terms and the methodology of MDSD are presented. Then the concept of our *FAIRlead* generator is presented in Section 3. Having

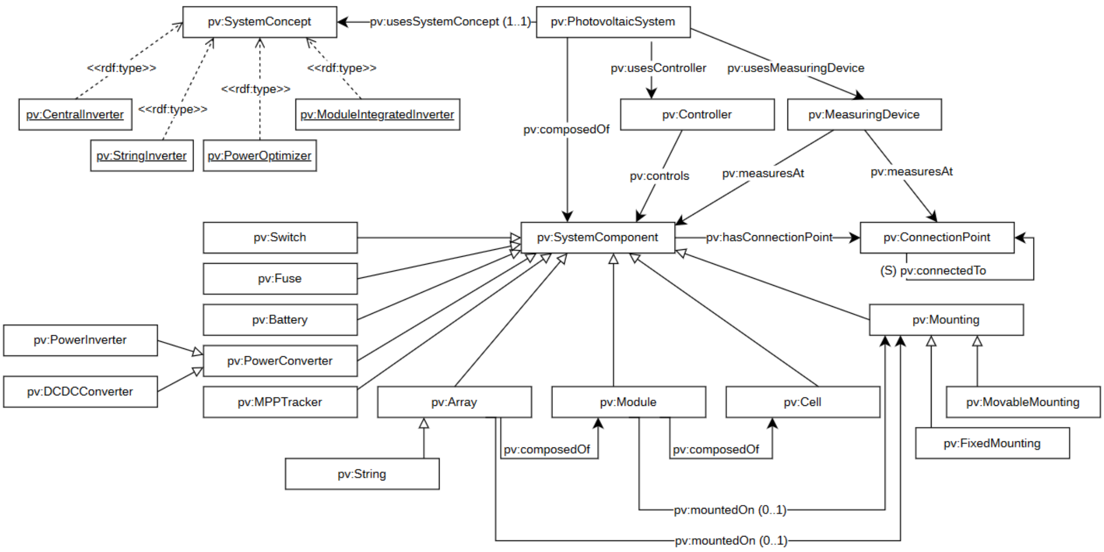


Figure 1: PV ontology (from (Schweikert et al., 2023)).

developed the concept so far, the following steps that we plan to take will be explained in Section 3. In Section 4, we take a closer look at the benefits we expect from the use of our generator, before giving a brief summary and outlook in Section 5.

2 MODEL DRIVEN SOFTWARE DEVELOPMENT

The basic idea of MDSD is the generation of source code from model information that describe the software to be developed (see Figure 2). Model descriptions are compact, abstract, formal, and platform-independent. As an example, the following is an abstract model description of the class `Person`:

```
<class Person(name: string(40),
              birthday: date,
              mother: Person,
              father: Person)>
```

Transformation rules are needed to map this abstract representation of the problem space onto programming code. These are usually provided in the form of templates and add the platform-specific information to the model. The following code fragment shows an example of a transformation rule that transforms the above model description into executable PHP code (specifically: a class description with constructor and setter methods). To do this, a template language (language elements shown in red) is used to integrate the variable parts from the model into the static code framework (in black).

```
<? foreach ($model->classes as $class) { ?>
  // class generated, do not edit !!!
  // timestamp: <?= date(DATE_RFC2822); ?>
  class <?= $class->name; ?>  {
  <? foreach ($class->properties as $p) { ?>
         protected $<?= $p ?>;
  <? } ?>
    function __construct() { }

    <? foreach ($class->properties as $p) { ?>
         function set<?= ucfirst($p) ?>($v) {
             $this-><?= $p ?> = $v;
         }
    <? } ?>
  }
<? } ?>
```

The code generated from the model and the transformation rule with the help of the generator then, after an additional code formatting step (not shown in Figure 2), looks like this:

```
// class generated, do not edit !!!
// timestamp: Tue, 24 Sep 2024 14:49:37 +0200
class Person {
    protected $name;
    protected $birthday;
    protected $mother;
    protected $father;

    function __construct() { }

    function setName($value) {
        $this->name = $value;
    }

    function setBirthday($value) {
        $this->birthday = $value;
    }
// more code folows here ...
}
```

Typically, 60% to 80% of an application's code can be generated (Stahl and Völter, 2006). The basic functionality of the intended application can even be generated almost by 100%. In addition to the higher development speed, this code typically has a higher quality, as the transformation rules are centrally defined in the generator templates and are consistently applied to the generated platform code. Even the quality of the software architecture is usually higher, as more thought is given to the underlying architecture during the template development. Furthermore, there is a clear separation of functional and technological aspects, so that the transition from one technological platform to another only requires an adaptation of the templates, but the model remains untouched.

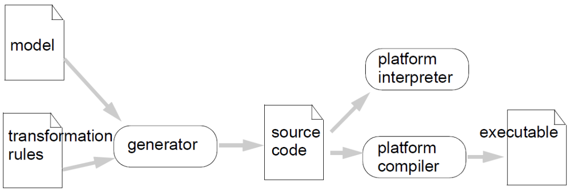


Figure 2: Principle function of a software generator. The input for the generator is the abstract, platform-neutral description of the application to be realized, as well as the platform-specific transformation rules. The result is the generated source code, which is then compiled in a further step or executed directly by an interpreter.

Starting point of an MDSD project is a reference implementation that is as lean as possible but executable. It serves as the basis for the development of the generator templates, which perform the transformation into source code. Once the reference implementation has been created, it is analyzed and the code is broken down according to the following criteria (Stahl and Völter, 2006):

1. generic code, that is the same for all possible applications
2. schematic, repetative code, which is individual for each application but has the same schematic structure
3. application-specific code (individual code)

This approach is illustrated in Figure 3. The generic, repetetive code (1) can simply be used, while the schematic code (2) is the starting point for creating the transformation rules in the generator templates. For this purpose, the code fragments are generalized into transformation rules and merged with the model information, as shown in Figure 2.

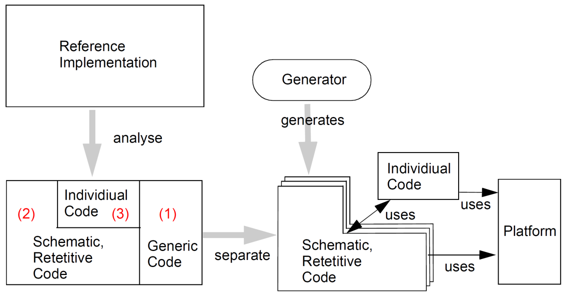


Figure 3: Principle of MDSD (adapted from (Stahl and Völter, 2006)).

3 FAIRlead APPROACH

In a preliminary step, we choose the target platform. This is the platform on which the generated application must be able to run. This can be a programming language such as PHP or a platform such as .net. A framework such as django (Vincent, 2019) or symfony (Zaninotto and Potencier, 2008) can also serve as a target platform. The use of a framework has the advantage that some of the tasks that would otherwise have to be covered by the generator are already covered by the framework (e.g. creation of CRUD interfaces for the concepts occurring in the ontology, creation of database schemas, object-relational mapping).

In line with the MDSD approach presented in the previous section, we will develop a simple reference implementation for our PV domain in a first step. The functionality of the application corresponds to that described in Section 1.3. We then identify both the generic part and the repetitive schematic part. The reference implementation can be very simple, as it only serves to verify the functionality of the generator. In the case that one or more components support introspection, generic approaches can also be pursued, as the meta information (e.g. properties of a class and their types) can be read and analyzed at runtime. This is especially important in order to implement communication interfaces for domain components, and also for the generic creation and extension of database schemas.

The next step is to analyze whether the input ontology provides enough information for generation of the schematic code part, or if additional information must be supplied to the generator. An example are SHACL definitions that extend OWL ontologies with the definitions of necessary properties, cardinality constraints or value ranges. Furthermore, platform-specific information, such as the graphical layout of the GUI, etc., will certainly be added. However, it remains to be seen to what extent this will be necessary.

A further task is the selection of a suitable generator.

The spectrum here ranges from in-house development with a scripting language such as PHP or Python together with a templating module available for the programming language, such as *twig* (Twig, 2024) or *smarty* (Smarty, 2024) for PHP, *mako* (Mako, 2024) or *jinia2* (Jin, 2024) (for Python), to the use of a tool such as the modeling workflow engine (MWE, 2024) in the Eclipse Modeling Project (EMP, 2024). The decisive factor is which information the abstract metamodel of the generator already contains and how flexibly the metamodel can be extended to meet the application's requirements. If a suitable generator is not available, an abstract metamodel with all the necessary properties has to be developed programmatically (i.e. as a set of Java classes) and then combined with an existing template framework in the selected language to be used as a software generator. Figure 4 shows the dependencies when selecting a suitable generator. In order to map the concepts described in the ontology, the generator's internal meta model must support them, or one must be able to extend the existing meta model to do so. Only under this condition is it possible to address the aspects described in the ontology within the templates and thus implement them in the generated application.

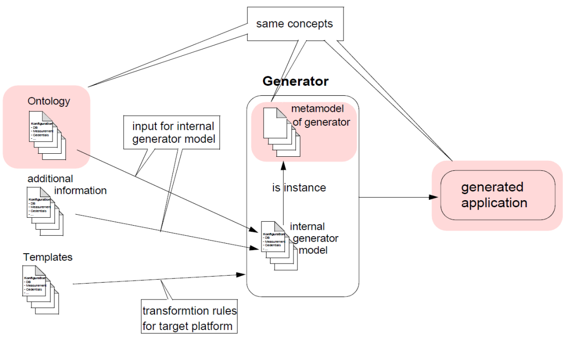


Figure 4: Dependencies between ontologies, generator and generated application. The concepts from the ontology can only be implemented in the application to be generated if the generator's internal meta-model supports them.

Once the generator is chosen, templates are created to regenerate the code of the reference implementation.

In a further step, the reference implementation is extended. This includes domain-specific tasks like the connection of external components such as proprietary systems or real time measurement software (i.e. Beckhoff automatisation software (Beck, 2024)) as well as the integration of already established repositories and registries following with the FAIR principles. The aim of this step is to establish meaningful extension interfaces that behave independently of the domain (i.e. the time series exporter tool `Zeitgeist` (Schmidt et al., 2023)). The newly established extension interface can then also become the starting point for a plug-in architecture of the generated code.

Our plan is to develop a generator framework which, in a first stable version, provides a web-based interface with CRUD functionality. This framework will allow the integration of meta information and also enables the referencing of primary data in various data sources (e.g. relational databases, time series databases, ...). At this time, the framework is also planned to be released as a GitHub project to gain input and enhancements from the community.

Figure 5 gives an overview of the architecture. Input for the *FAIRlead* generator are the ontologies and the generator templates, which define the transformations on the target platform. Further input can come from SHACL files, which further specify the input ontologies. In addition to the source code for GUI and CRUD functionality, the database schema is also generated. On the right-hand side you will find manually created application logic. The arrows that originate from the application logic represent the connection of the manually created code to the extension interfaces of the generated code. Patterns on how this can be done can be found in (Stahl and Völter, 2006). The database with the meta information, which is linked to the actual data (bottom), is located near the center of the figure. Since we want to support the widest possible range of data sources, we need to find an API that is as generic as possible. However, there are already approaches here, such as (DTP, 2024; ODC, 2024; Beam, 2024), which we will examine for their suitability.

Further research steps include:

- finding interfaces between the individual parts of the application so that the individual components are as interchangeable as possible (i.e. exchange a simple web-basd formular with a graphical editor for specifying the instances of domain concepts).
- Mechanisms for the connection between metadata and data.
- Detection of inconsistencies between data and metadata (consistency checks).

Figure 6 shows a flowchart of the individual activities we have defined so far, along with their dependencies.

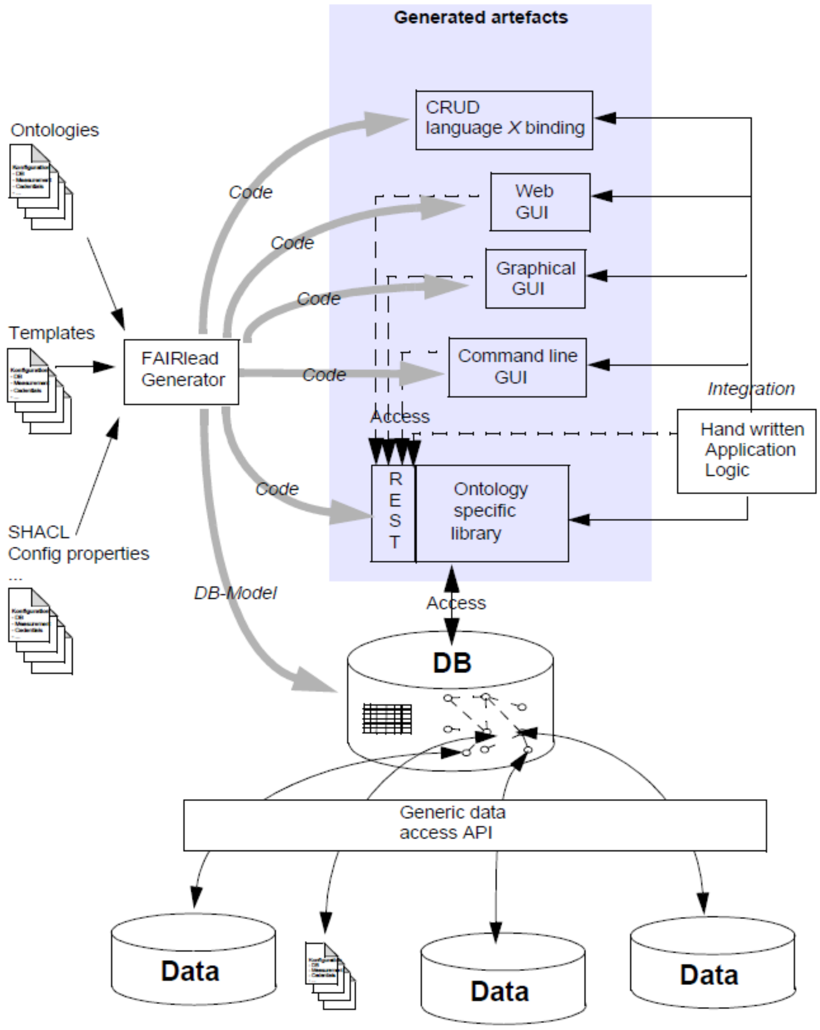


Figure 5: FAIRLead Generator architecture.

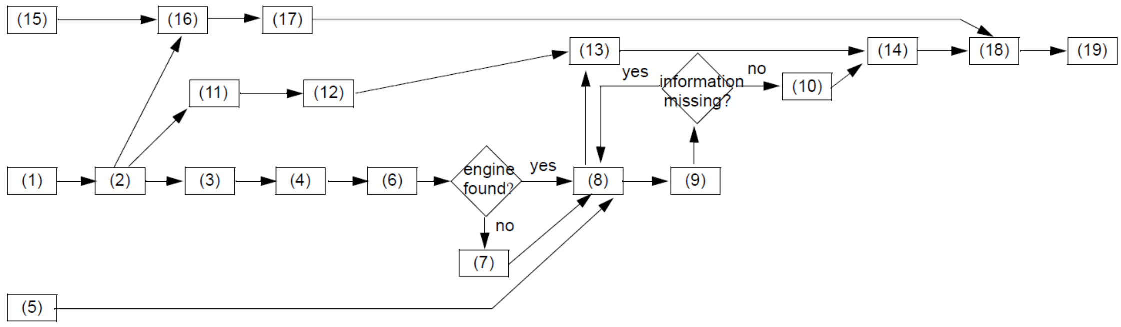
4 ADVANTAGES OF THE FAIRlead APPROACH

In the FAIR community, there are currently a number of alternative concepts for implementing the principles. An Example is given by FAIR Digital Objects (FDO) which, e.g., have been compared to Linked Data approaches in (Soiland-Reyes et al., 2024). Web Frameworks are available as well as systems based on Web Components (Schmidt and Tobias, 2024). As with respect to storage, Triple-Store-based solutions exist beside alternative persistence concepts. With our approach, it is easy to compare and evaluate all those different implementation variants in an overall system, since only the corresponding additional templates have to be developed for the different implementations.

The generic approach shifts a significant proportion of the application implementation to the conceptual, technology-independent part - defined by ontologies - and therefore takes place at a higher level of abstraction. As a result, design decisions like those mentioned in the preceding paragraph can be made later or realized in parallel and comparatively with minimum effort. The only prerequisite for this is the development of one or more corresponding technology-specific templates (see Figure 3).

In addition, once the software generator has been fully developed, it is easier to involve technical experts in the creation of new application systems, as discussions can take place at a purely conceptual level and the target architecture is generated by the generator and the templates already developed at this stage.

We plan to make the generator we have developed



1. Choose target platform for PV-application
2. Build reference implementation (RI) for PV-ontology (full stack)
3. Anaylze code of RI according to repetative patterns (RP) (see Section 2)
4. Analyze ontology information concerning information needed to generate RP
5. Search for libraries that extract information from ontologies
6. Look for an appropriate generator engine
7. Implement own generator engine (choose language, template engine, define internal metamodel
8. Build/extend ontolgy import modul for generator
9. Create templates
10. Generate PV-application using created templates
11. Extend reference implementation with application logic
12. Design application logic REST-API interface
13. Create templates for application logic interface
14. Extend generated PV-application with application logic using generated interface
15. Linking data and metadata. Investigation of existing libraries for referencing different data types/sources.
16. Integration of the referencing functionality in the RI
17. Creation of templates for referening datasets
18. Generation of the application with complete functionality
19. Applying the concept to other areas (ontologies)

Figure 6: Flowchart of the activities we have defined and are implementing.

available to the FAIR community as open source. This will enable researchers around the world to use our tool by using their ontologies as input for the generator and generating the domain-specific application for linking data and metadata.

This application can then be integrated into the respective laboratory infrastructure and used, for example, as an electronic laboratory notebook and, on the other hand, make the step towards publishing their research results much easier.

However, the functionality of the generated application can easily extend by developing new generator templates or adapting the templates to your own requirements. This is not particularly complicated if you use the existing generator templates as a blueprint.

5 CONCLUSION AND OUTLOOK

The enrichment of data with associated metadata is seen as a necessary step to make research results reproducible, to verify scientific experiments or to build on the results of these. In this context, we presented the conceptual framework *FAIRlead*, which is designed to support scientists in enriching data with metadata. To support people from different domains, we have chosen a model-driven approach that generates software artifacts to enrich data with metadata based on an ontology description of the domain.

The functionality of the generated software includes the specification of a scientific experiment based on the concepts defined in the ontology and their relationships to each other as well as the linking to the actual data. Now that we have defined the conceptual framework for our future work, we will next carry out a reference implementation based on the PV ontology we have developed in order to derive our generator templates, which form the core of the *FAIRlead* generator. An important future step is the establishment of a clean interface architecture in the generated application so that different possible implementation variants can be easily exchanged. The same applies to the interface for connecting the business logic and external systems.

REFERENCES

Beam (2024). Apache beam i/o connectors. https://beam.apache.org/documentation/io/connectors/. (Accessed on 2024-09-24).

Beck (2024). Beckhoff new automatisation technology. https://www.beckhoff.com/en-en/products/automation/twincat/. (Accessed on 2024-09-24).

Chen, P. P. (1976). The entity-relationship model - toward a unified view of data. *ACM Trans. Database Syst.*, 1(1):9–36.

DTP (2024). Eclipse data tools platform. https://projects.eclipse.org/projects/tools.datatools. (Accessed on 2024-09-24).

EMP (2024). Eclipse modeling project. https://projects.eclipse.org/projects/modeling. (Accessed on 2024-09-24).

Hoyer-Klick, C., Blesl, M., von Bremen, L., Frey, U., Gainnousakis, A., Hülk, L., Kronshage, S., Kuckertz, P., Lohmann, G., Muschner, C., Pehl, M., and Schroedter-Homscheidt, M. (2023). Fair data in energy systems analysis. In *Proceedings of the International Conference on Energy Meterology*.

Jin (2024). Template designer documentation. https://jinja.palletsprojects.com/en/3.0.x/templates/. (Accessed on 2024-09-24).

Mako (2024). Mako templates for python. https://www.makotemplates.org/. (Accessed on 2024-09-24).

MWE (2024). Eclipse modeling workflow engine. https://projects.eclipse.org/projects/modeling.emf.mwe. (Accessed on 2024-09-24).

ODC (2024). Open data connector. https://www.dataspaces.fraunhofer.de/en/software/connector/open_data_connector.html. (Accessed on 2024-09-24).

OMG (2011). Unified modeling language™ (uml®).

Ont (2024). Ontology (information science). https://en.wikipedia.org/wiki/Ontology_(information_science). (Accessed on 2024-09-24).

Schmidt, A., Koubaa, M. A., Schweikert, J., Stucky, K.-U., Süß, W., and Hagenmeyer, V. (2023). Zeitgeist - A Generic Tool Supporting the Dissemination of Time Series Data following FAIR Principles. In *Proceedings of the International Conference on Knowledge Management and Information Systems*. Insticc, SCITEPRESS.

Schmidt, A. and Tobias, M. (2024). Web Components for Database Developers. In *Proceedings of the Sixteenth International Conference on Advances in Databases, Knowledge, and Data Applications*. ThinkMind.

Schweikert, J., Stucky, K.-U., Süß, W., and Hagenmeyer, V. (2023). A photovoltaic system model integrating fair digital objects and ontologies. *Energies*, 16(3).

SHACL (2017). Shapes Constraint Language (SHACL). (Accessed on 2024-09-24).

Smarty (2024). Smarty template engine. https://www.smarty.net/. (Accessed on 2024-09-24).

Soiland-Reyes, S., Goble, C., and Groth, P. (2024). Evaluating fair digital object and linked data as distributed object systems. *PeerJ Computer Science*, 10:e1781.

Stahl, T. and Völter, M. (2006). *Model-Driven Software Development: Technology, Engineering, Management*. Wiley, Chichester, UK.

Twig (2024). Twig for template designers. https://twig.symfony.com/doc/3.x/templates.html. (Accessed on 2024-09-24).

Vincent, W. S. (2019). *Django for Professionals: Production websites with Python & Django*. Independently published.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3.

Zaninotto, F. and Potencier, F. (2008). *The Definitive Guide to symfony*. apress.

ML System Engineering Supported by a Body of Knowledge

Juliette Mattioli[1], Dominique Tachet[2], Fabien Tschirhart[2], Henri Sohier[2], Loic Cantat[2] and Boris Robert[2,3]
[1]*Thales, France*
[2]*IRT SystemmX, France*
[3]*IRT Saint Exupéry, France*
juliette.mattioli@thalesgroup.com, dominique.tachet@ext.irt-systemx.fr, fabien.tschirhart@irt-systemx.fr, loic.cantat@irt-systemx.fr, boris.robert@irt-saintexupery.com, henri.sohier@irt-systemx.fr

Keywords: Body-of-Knowledge, Knowledge Graph, Knowledge Extraction, Knowledge Fusion, ML Engineering.

Abstract: A body of knowledge (BoK) can be defined as the comprehensive set of concepts, terminology, standards, and activities that facilitate the dissemination of knowledge about a specific field, providing guidance for practice or work. This paper presents a methodology for the construction of a body of knowledge (BoK) based on knowledge-based artificial intelligence. The process begins with the identification of relevant documents and data, which are then used to capture concepts, standards, best practices, and state-of-the-art. These knowledge items are then fused into a knowledge graph, and finally, query capacities are provided. The overall process of knowledge collection, storage, and retrieval is implemented with the objective of supporting a trustworthy machine learning (ML) end-to-end engineering methodology, through the ML Engineering BoK.

1 RATIONALE

Systems and products are developed in competitive, volatile, uncertain, complex and ambiguous contexts, influenced by external factors such as regulations and societal expectations. Artificial Intelligence (AI) technologies, in particular Machine Learning (ML) approaches, improve product quality and production efficiency (Li et al., 2017). Reliability is crucial for critical systems to remain reliable throughout their lifecycle and to evolve cost-effectively. However, the rapid adoption of AI technologies is leading to a specialization of engineers and a dispersion of the required skills. Current demographics are making highly skilled and experienced engineers scarce, leading to a lack of project support and mentorship. In addition, the complexity of AI-based solutions and past trends in each component require the management of AI engineering knowledge and general engineering practices. This highlights the need for effective management of AI engineering knowledge and practices. Knowledge management and knowledge engineering (KE) are often used interchangeably, with "manage" referring to executive leadership and "engineer" to planning, construction, or design activities. The main difference is that the knowledge manager sets the process direction, while the knowledge engineer develops the means to achieve it. The Confiance.ai[1] program's end-to-end methodology serves as a foundational framework for knowledge management in trustworthy AI engineering (Awadid et al., 2024). It addresses non-functional requirements for successful implementation of ML-based components in critical systems (Adedjouma et al., 2022). The methodology covers various process levels and aligns with industrial best practices. It is essential to define the scope and position of the methodology in relation to other engineering disciplines. KE is a sub-field of AI that focuses on understanding, designing, and implementing methods for representing information effectively (Shapiro, 2006). It facilitates the management of all types of knowledge based on labelled graphs and provides guidance for resolving ML engineering-related issues. In order to address these issues, Confiance.ai program put forth the argument that there is a need to consolidate the AI engineering field's largely fragmented body of knowledge (Mattioli et al., 2024), which encompasses data engineering, algorithm engineering, software and system engineering, safety and cyber-security, similar to the SWEBOK definition of software engineering (Robert et al., 2002).

[1] www.confiance.ai/en

Mattioli, J., Tachet, D., Tschirhart, F., Sohier, H., Cantat, L. and Robert, B.
ML System Engineering Supported by a Body of Knowledge.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 331-338
ISBN: 978-989-758-716-0; ISSN: 2184-3228

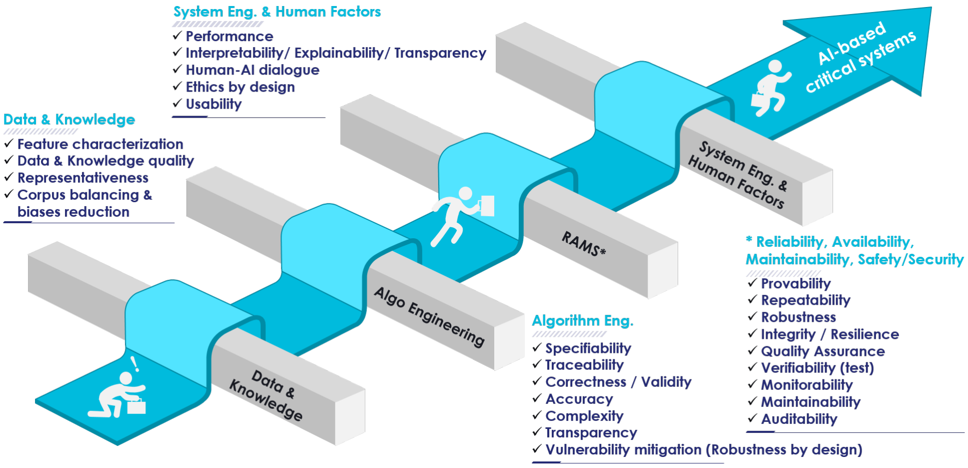


Figure 1: AI/ML deployment induces some (engineering) challenges.

2 KNOWLEDGE ENGINEERING

From this perspective, Knowledge Engineering (KE) represents a sub-field of AI that is concerned with the understanding, design and implementation of methods for representing information in a manner that enables computers to utilize it effectively. In other words, the objective of KE is the understanding and subsequent representation of human knowledge in the form of data structures, semantic models (conceptual graph of the data in relation to the real world) and heuristics (or rules). The induced process comprises three principal elements: a) knowledge acquisition, representation, and validation; b) inference; and c) explanation and justification. It is evident that approaches to knowledge representation, such as ontologies, taxonomies, thesauri, and vocabularies, can be employed to support a diverse array of activities. Techniques that could be used for such purpose include:

- Using a taxonomy that is well-known and accepted by the organization in work-products;
- Using patterns to help a computer "understand" the content of work-products;
- Encoding the breakdown structures that represent well-established knowledge within the organization;
- Using inference rules to reason about the quality of the content of different work-products.

An organization must manage knowledge about its systems and engineering practices, known as a Body of Knowledge (BoK). The BoK provides a comprehensive description of ML system engineering contents and practices, establishing a foundation for curriculum development and creating a coherent curriculum for qualification towards certification.

3 ML ENGINEERING BoK

A **body of knowledge** (BoK) is"*structured knowledge that is employed by members of a discipline to inform their practice or work*" (Ören, 2005). It is used to define concepts and activities, such as accuracy in data engineering, machine learning models and system-level applications. The core components of a BoK include concepts, knowledge, skills, standards, terminology, guidelines, practices and activities. It serves as the "ground truth" for ML engineering activities, covering various domains such as data engineering, algorithm engineering, software engineering, systems engineering, cyber-security, safety, and cognitive engineering.

3.1 The BoK Design

The BoK design is an iterative process involving knowledge acquisition, fusion, storage, and retrieval. Knowledge is acquired from structured, semi-structured, and unstructured data, with extraction focusing on entities, attributes, and relations. Knowledge fusion requires ongoing ontology construction and quality evaluation. Currently, knowledge is typically stored in KG databases. Confiance.ai BoK (fig 2) is a comprehensive guide for engineers on the lifecycle of AI-based critical systems. It offers guidance and support throughout the development, maintenance, and evolution of these systems. The guide defines trustworthy ML engineering concepts and provides an outline of essential knowledge, skills,

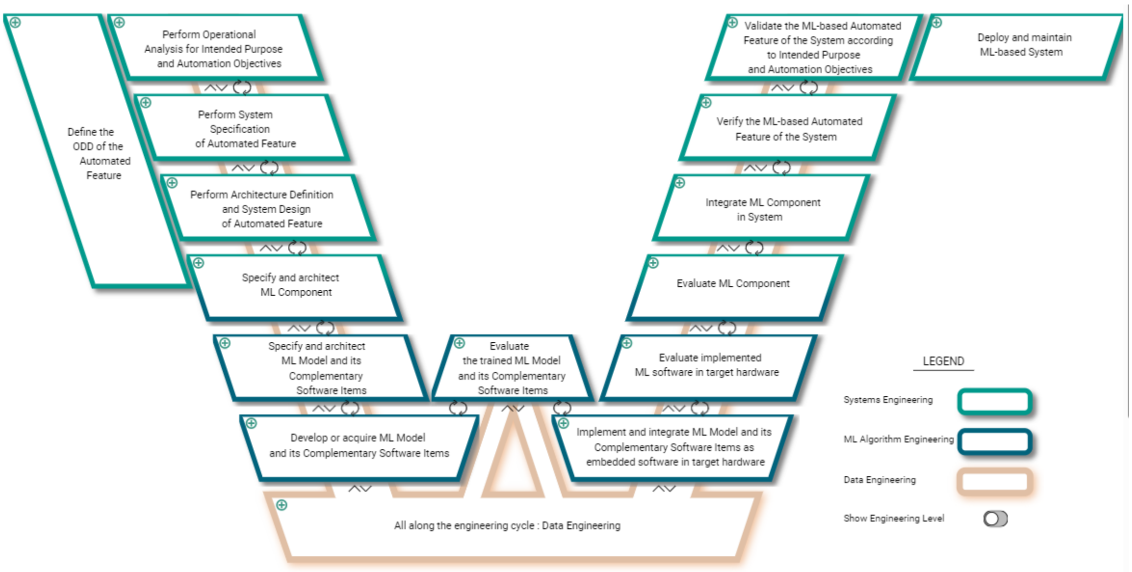


Figure 2: Trustworthy ML engineering Body-of-Knowledge - https://bok.confiance.ai/.

and practices, covering all fundamental competencies required by professionals in the field. The initial step of the ML Engineering BoK design is to identify the domain of application and compile a list of relevant knowledge sources. Secondly, a conceptual model will be devised with the objective of gathering together the entities of interest, their inter-relationships and the categories. A valuable resource for conceptual modeling is Capela©, which is a model-based engineering solution that has proven effective in numerous industrial contexts. Thirdly, the logical and physical models will provide a logical representation and assertions for the entities and relationships that have been collected. Fourthly, the technical development and implementation must take into account the coding language to be employed (for example, RDF and OWL), as well as the serialization formats (such as RDF/XML, Turtle and JSON-LD). The final stage is the deployment of the BoK as a service, thereby facilitating reuse and enabling the engineering community to provide feedback. In essence, this process entails the transformation of knowledge held by engineering experts and end-users into a machine-readable format. Designing a ML Engineering BoK induces some general issues:

- **Raw Data Acquisition.** How to select the relevant data and information to be fed into the BoK?
- **Knowledge Extraction, Representation and Validation.** How do we represent human knowledge as it currently exists in state of the art reports, scientific articles, standards and norms, Engineering best practices, and the minds of the experts in terms of data structures that can be processed by a computer? How to determine the best representation for any given engineering problem?
- **Knowledge Integration and Fusion.** How do we use these knowledge item to generate useful information in the context of a ML engineering?
- **KG Design.** How to manipulate the knowledge to provide explanations to the engineer/user?
- **Knowledge Query.** How do we use these abstract knowledge structures to generate useful information in the context of a specific case?

3.2 Step 1: Raw Data Acquisition

BoK developers create knowledge bases from scratch, dealing with diversity and heterogeneity of knowledge representation formalisms and mismatch of different knowledge items. Knowledge engineers focus on modeling structural use cases and expert knowledge concepts. The first step is to define a taxonomy of the ML engineering domain aligned with ISO/IEC DIS 5338 standard for AI systems and safety and reliability standards. This taxonomy is the classification of concepts induced by Trustworthy ML Engineering activities. The operational definition of trustworthy AI includes a taxonomy and keywords that define core domains such as AI Engineering, Data Engineering, and Safety Engineering, covering the entire life cycle of AI-based critical systems. This framework aids in harmonizing design and support activities, including monitoring and maintenance, and serves as a foundation for the Confiance.ai methodology, which outlines requirements and recommendations. The initial stage is devoted to the identification of data sources, as this has a significant impact on the entire knowledge graph (KG) development process, as well as on the selection of knowledge extraction techniques. Definitions have been collated from various sources, including European and worldwide standardization bodies (ISO/IEC 5338, Aerospace Standard 6983, IEEE 7000...), National and European projects (Confiance.ai, DEEL project[2], JRC Flagship on AI[3]), scientific publications and working groups (e.g. the HLEG or the AI Safety Landscape initiative[4]), as well as other relevant sources. The working group responsible for the state of the art sourced definitions from external literature in most cases. Sometimes the existing literature did not match the scope of ML Engineering. We created a new definition. This phase was also based on data about making AI reliable. This includes making AI reliable through design, data engineering for trusted AI, IVVQ Strategy (Integration, Verification, Validation and Qualification), and targeted embedded AI.

3.3 Step 2: Knowledge Extraction, Representation and Validation

The extraction of knowledge from semi-structured sources is easier than from unstructured sources, which hold more information. The second phase focuses on extracting knowledge from unstructured data to create and enhance knowledge graphs, identifying entities and relationships. This process involves natural language processing (NLP) and knowledge representation technologies to automatically extract structured information from various data types, facilitating the effective use of external data. An entity represents the most fundamental unit of a knowledge graph. It represents a concept. Furthermore, the quality of knowledge graph construction is contingent upon the accuracy and integrity of its extraction. Subsequently, relationship extraction entails the identification of associations between entities, thereby establishing semantic relations and forming a knowledge network. These types of graphs embed a structured representation of facts, consisting of entities, relationships, and semantic descriptions, which are modeled with an RDF (Resource Description Framework) structure. An RDF model is a flexible data representation model comprising three-element tuples, with no fixed schema requirements. It is a graph-based model for the description of entities and their relationships on the Web. Many researchers prefer to conceptualize RDF as a set of triples, although it is commonly described as a directed and labelled graph, each consisting of a subject, predicate and object in the form of $< subject, predicate, object >$. In this context, the predicate represents the relationship between the subject and the object. For example: $< Data\_Engineering, is\_an\_activity, MLOps >$. The triples are stored in a triple store and can be queried using the SPARQL query language. In comparison to both inverted indices and plain text files, triple stores and the SPARQL query language facilitate the formulation of sophisticated queries, enabling users to satisfy complex information needs. Although a model is required for representing data in triples (similar to relational databases), RDF enables the expression of rich semantics and supports knowledge inference (Hertz et al., 2019). Like any model, such a BoK is only an approximation of reality. New observations based on ML engineering use-cases can guide the further acquisition of knowledge. Therefore, an evaluation of the represented knowledge with respect to reality is indispensable for the creation of an adequate model. These limitations relate to the so-called symbol grounding problem (Harnad, 1990), and concern the extent to which representational elements are hand-crafted rather than learned from data. The most common methods employed include pattern matching, machine learning, and semantic rule extraction. Furthermore, generative models such as large language models (LLMs) can play a pivotal role in the construction of knowledge graphs by extracting entities, relationships, and attributes from unstructured text data (Meyer et al., 2023). They can be pre-trained through the structurally consistent linearization of text, which facilitates the transition from traditional understanding to structured understanding and increases knowledge sharing (Wang et al., 2022). In contexts with Named Entity Recognition (NER), as demonstrated by (Straková et al., 2019), the proposed generative method implicitly models the structure between named entities. This approach effectively avoids the complexity inherent to multi-label mapping. Similarly, extracting overlapping triples in relation extraction is also challenging to address for traditional discriminating models (Zeng et al., 2018), introducing a new perspective for addressing this issue through a general generative framework. Moreover, several features must be taken into account when developing a BoK:

- Redundancy: Are there identical or equivalent knowledge items that is a special case of another (subsumed)?
- Consistency: Are there ambiguous or conflicting knowledge, is there indeterminacy in its ap-

[2]https://www.deel.ai/

[3]https://joint-research-centre.ec.europa.eu/jrc-mission-statement-work-programme/facts4eufuture/artificial-intelligence-european-perspective/future-ai_en

[4]https://www.aisafetyw.org/ai-safety-landscape

plication? Is it intended? Are several outcomes possible, for example, depending on the strategy (the order in which the knowledge models are ordered)?
- Minimality: Can the knowledge set be reduced and simplified? Is the reduced form logically equivalent to the first one?
- Completeness: Are all possible entries covered by the knowledge of the set?

Thus, a good BoK must have properties such as:
- Representational Accuracy: It should represent all kinds of required knowledge.
- Inferential Adequacy: It should be able to manipulate the representational structures to produce new knowledge corresponding to the existing structure.
- Inferential Efficiency: The ability to direct the inferential knowledge mechanism into the most productive directions by storing appropriate guides.
- Acquisitional Efficiency: The ability to complete with new knowledge easily using automatic methods.

Peer reviews with various stakeholders (data scientists, software and system engineers, safety and cybersecurity engineers...) were carried out to assess the appropriateness and quality of the acquired knowledge in relation to the ML engineering end-to-end methodology (Adedjouma et al., 2022).

3.4 Step 3: Knowledge Integration and Fusion

For (Sowa, 2000), "*Knowledge Representation is the application of logic and ontology to the task of constructing software models for some domain*". Therefore, the way a knowledge representation is conceived reflects a particular insight or understanding of how people reason. The selection of any of the currently available representation technologies (such as logic, knowledge bases, ontology, semantic networks...) commits one to fundamental views on the nature of intelligent reasoning and consequently very different goals and definitions of success. As we manipulate concepts with words, all ontologies use human language to "represent" the world. Thus, ontology is expressed as a formal representation of knowledge by a set of concepts within a domain and the relationships between these concepts. Nevertheless, the "fidelity" of the representation depends on what the knowledge-based system captures from the real thing and what it omits. If such system has an imperfect model of its universe, knowledge exchange or sharing may increase or compound errors during the ML Engineering process. As such, a fundamental step is to establish effective knowledge representation (symbolic representation) that can be used for query. The sheer complexity, variety and volume of data available today presents a significant challenge to achieving efficient and accurate knowledge graph fusion. The process of integrating disparate sources of knowledge, also known as knowledge fusion, entails the elimination of redundancies, inconsistencies, and ambiguities from the integrated corpus. The field of engineering is one in which knowledge is typically subject to updates. In the majority of cases, users will have the capacity to supplement existing external knowledge graphs with external knowledge. Thus, the objective of knowledge fusion is to merge semantically equivalent elements, for example, the concepts of "accuracy" and "machine learning accuracy", with the intention of integrating novel forms of knowledge within existing conceptual frameworks or factual assertions. The sub-tasks of knowledge fusion include the alignment of attributes, the matching of entities with small-scale incoming triples, and the alignment of entities with a complete knowledge graph. This stage is beneficial for both the generation and completion of knowledge graphs. We employ the knowledge graph representation for knowledge fusion, as proposed by (Laudy et al., 2007), which is based on the conceptual graph model. This representation is used to store and combine knowledge. The approach is to examine observations with domain knowledge and graph operators. This removes any bias from translating data from one format to another with different models. We suggest using it for a high-level information fusion approach based on the Maximal Join operator, which is an aggregation operator on conceptual graphs (Laudy, 2011).

3.5 Step 4: Knowledge Graph

At worst, the effort involved in specifying the relevant knowledge forces us to think more deeply about the relevant ways of characterizing the ML engineering models that we as researchers implicitly construct anyway. The use of Knowledge Graphs (KG) as a means of representing knowledge is becoming increasingly prevalent. Their versatility in terms of representation allows for the integration of diverse data sources, both within and across engineering boundaries. Therefore, our primary strategy to support this step in practice was the creation of a knowledge graph, which collects information about each ML development activity, the artifacts and processes used in the entire ML-based system lifecycle, the end-to-end methodology, and the motivation behind the key design decisions. Several replications have been carried out in this way, contributing to a growing body of knowledge about ML engineering techniques. By employing the graph architecture, KGs are capable of modeling a range of relationship types (edges) and entities (nodes) (Chen et al., 2020). KGs comprise an additional embedded layer, designated a reasoner (or inference engine), which enables them to extract implicit information from existing explicit concepts, in contrast to plain graph or non-relational databases. The most well-known examples of knowledge graphs (KGs) – DBpedia, Freebase, Wikidata, YAGO, and so forth – encompass a diverse array of domains and are either derived from Wikipedia or created by volunteer communities (Heist et al., 2020). The Google Knowledge Graph is one of the largest and most comprehensive KGs in existence, aiming to model and link all structured information found on the internet, including persons, organizations, skills, events, products, and more. This is one of the reasons why the Google search engine is so effective. A graph-based knowledge representation and reasoning formalism derived from conceptual graphs has been formalized as finite bipartite graphs, as outlined in (Mugnier and Chein, 1992). In this formalism, the set of nodes is divided into concept and conceptual relation nodes. In such a graph, concept nodes represent classes of individuals, and conceptual relation nodes illustrate the relationships between the aforementioned concept nodes. This is in accordance with the findings of (Sowa, 1976). As outlined in (Ehrlinger and Wöß, 2016), a KG acquires information and integrates it into an ontology, subsequently applying a reasoner to derive new knowledge. Furthermore, in accordance with the definition provided by (Ji et al., 2021), KGs are "structured representations of a fact, consisting of entities, relations, and semantics." Entities may be either real-world objects or abstract concepts. Relationships represent the relationship between entities, and semantic descriptions of entities and their relationships contain types and properties with defined semantics. Property graphs, in which nodes and relations possess properties or attributes, or attribute graphs, are extensively employed. All of these facets rely on a knowledge inference over knowledge graphs, which represents one of the core technologies in the design of our ML engineering BoK. The Semantic Web community has reached a consensus on the use of RDF to represent a knowledge graph. Then, RDF model also allows for a more expressive semantics of the modeled data that can be used for knowledge inference. As a result, a KG is a set of interconnected information on a specific set of facts that includes characteristics of many data management paradigms:

- Database: Structured queries can be used to explore data in a database.
- Graph: KGs can be analyzed in the same way that any other network data structure can be.
- Knowledge Base: Formal semantics are encoded in KGs, which can be used to understand data and infer new facts.

3.6 Step 5: Knowledge Query

In this context, a body of knowledge (BoK) is conceptualized as a graph of knowledge, as proposed by (Mattioli et al., 2022). Ultimately, the utility of the ingested, transformed, integrated and stored knowledge is contingent upon the efficiency with which answers can be retrieved by users in an intuitive manner. At the present time, keyword queries and specialized query languages (e.g. SQL and SPARQL) represent the prevailing approaches to information retrieval. However, in order to facilitate the search for a specific ML engineering knowledge by querying the KG and selecting the set of relevant engineering views to perform specific ML engineering activities, it is necessary to enable the identification of similarities between Confiance.ai documents by searching for isomorphisms between the graphs representing the knowledge extracted from the text. A number of algorithms have been defined which implement subgraph isomorphism; however, the subgraph isomorphic problem is an NP-complete problem. The initial component is a generic sub-graph matching mechanism that functions in conjunction with fusion schemes. This component is responsible for ensuring the structural consistency of the merged information with respect to the structures of the initial documents throughout the fusion process. The fusion approach is constituted by the similarity and compatibility functions applied to the members of the graphs to be fused. The generic fusion algorithm can be adapted to suit the context in which it is used by adopting these strategies. The knowledge graph fusion method offers two additional operations, contingent upon the fusion strategies employed. Information synthesis is the collection and organization of data on a subject. Information is then put together into a network through information synthesis, where any repetitions are removed. Fusing techniques are used to combine information about the same thing, even though it is in different forms. When different sources of information are used to create a representation of something, inconsistencies may appear. This function finds all the information in a network that follows a specific pattern. The structure of the query graph must match that of the data graph. To find the information query function, look for a one-to-one mapping between the query graph and the data graph.

4 ILLUSTRATION ON ML ROBUSTNESS EVALUATION

The utilization of keyword-based queries has become a prevalent methodology for enabling non-technical users to access expansive RDF data sets. At the present time, the user is able to select an engineering activity within the graph that utilizes the end-to-end methodology, and the underlying knowledge will then be presented to them. The ML Engineering BoK is a trustworthy ML end-to-end engineering guideline that engineers should follow throughout their activities. It is based on a comprehensive set of descriptions, engineering knowledge, metrics, and key performance indicators, which are capitalized in the BoK. These elements enable engineers to assess both functional and non-functional properties. For example, an engineer is seeking information on the assessment of ML model robustness in the context of the activity of evaluating an ML model in order to analyze and characterize the system's sensitivity to changes in the input, with a view to determining its overall resilience. For this engineering activity, the BoK suggests a strategy made of two successive phases: 1) Robustness test by sampling and perturbation (empirical evaluation) and 2) Formal verification of robustness (formal evaluation). Each step is described by a Capela model and a textual content. It consists in selecting the most appropriate tool for this robustness test. To make this selection, the key criteria are: the ML Model Algorithm, the type of data (images, time series, and language), the type of perturbation and its intensity (examples of perturbation include image luminance, image blurring, geometric transformation of plane position), and the target level of robustness. There may be different tools for data perturbation and for execution of the test., or it can be the same tool. With the selected tool, the specified perturbation is applied on the Test Dataset and the ML Model is executed on this perturbed dataset and the resulting behavior of the ML Model is captured.

5 CONCLUSIONS

The objective of building a Trustworthy ML Engineering Knowledge Graph is to facilitate more effective specification, design, comprehension, monitoring, and maintenance of ML-based systems for system design engineers and ML-based system operations personnel. Ultimately, this should enhance safety, cyber security, reliability, and performance, while also improving availability. The construction of a reliable ML engineering framework entails the utilization of an array of data sources and artificial intelligence methodologies, encompassing knowledge representation, knowledge graphs, semantic networks, high-level information fusion, graph theory, and numerous other techniques. Confiance.ai's methodological contributions span the entire development process of an ML-based system, from initial specification and design through to the commissioning and subsequent supervision of operational deployment, and even to the embedding of the latter in other systems. These contributions are manifold and include:

- A taxonomy used in trustworthy AI;
- A complete documentation of the process, including modeling of activities and roles, with elements enabling corporate engineering departments to implement it;
- A first development of a Trustworthy AI ontology, linking the main concepts of the process and the taxonomy;
- And a “Body-of-Knowledge” which brings together all these elements and makes them accessible on the website of the same name.

Furthermore, the ML-engineering BoK provides support to stakeholders across the ML value chain, offering invaluable assistance in the elicitation, validation, and verification of safety and cyber-security relevant quality attributes. Based on the Confiance.ai end-to-end methodology, it is able to guarantee that the heterogeneous requirements of stakeholders are met, thereby further consolidating its status as a fundamental element in the field of safety within the context of Machine Learning. While the deployment of these methodological instruments does not inherently guarantee the compliance of ML-based systems with regulatory requirements, it may serve as a basis for justifying such compliance, particularly in light of the prevailing standards set by the notified bodies responsible for verification. Furthermore, this ML Engineering BoK serves as a valuable resource in addressing the following key challenges:

- How to design AI models, so that, by construction, they satisfy trustworthy properties (accuracy, robustness, etc.)?
- How to characterize these AI models, for example, to understand and explain their behavior and their adequacy to the operational domain?
- How to implement and embed those AI models on hardware, by making them fit for the target with-

out losing their trustworthy properties.

- What are the data engineering method to apply in order to manage important volumes of data, account for the evolution of the operational domain, etc.?
- What are the appropriate verification, validation, and certification processes to consider for AI-based systems?

ACKNOWLEDGEMENTS

This work has been supported by the French government under the "France 2030" program, as part of the SystemX Technological Research Institute within the *Confiance.ai* Program (www.confiance.ai).

REFERENCES

Adedjouma, M., Alix, C., et al. (2022). Engineering dependable ai systems. In *2022 17th Annual System of Systems Engineering Conference (SOSE)*, pages 458–463. IEEE.

Awadid, A., Le Roux, X., et al. (2024). Ensuring the Reliability of AI Systems through Methodological Processes. In *The 24th IEEE International Conference on Software Quality, Reliability, and Security*.

Chen, Z., Wang, Y., et al. (2020). Knowledge graph completion: A review. *IEEE Access*, 8:192435–192456.

Ehrlinger, L. and Wöß, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48(1-4):2.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

Heist, N., Hertling, S., et al. (2020). Knowledge graphs on the web–an overview. *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, pages 3–22.

Hertz, S., Olof-Ors, M., et al. (2019). Machine learning-based relationship association and related discovery and search engines. US Patent 10,303,999.

Ji, S., Pan, S., et al. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514.

Laudy, C. (2011). Semantic knowledge representations for soft data fusion. *Efficient Decision Support Systems-Practice and Challenges from Current to Future*.

Laudy, C., Ganascia, J.-G., and Sedogbo, C. (2007). High-level fusion based on conceptual graphs. In *2007 10th International Conference on Information Fusion*, pages 1–8. IEEE.

Li, B. et al. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, 18(1):86–96.

Mattioli, J., Laudy, C., et al. (2022). Body-of-knowledge development by using artificial intelligence. In *17th IEEE System of Systems Engineering Conference (SOSE)*, pages 14–19.

Mattioli, J., Tachet, D., et al. (2024). Leveraging Knowledge Graph to design the Machine-Learning Engineering Body-of-Knowledge. In *IEEE International Conference on AI x Science, Technology, and Technology (AIxSET)*, Laguna hills, United States.

Meyer, L., Stadler, C., et al. (2023). LLM-assisted knowledge graph engineering: Experiments with ChatGPT. In *Working conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, pages 103–115. Springer Fachmedien Wiesbaden Wiesbaden.

Mugnier, M.-L. and Chein, M. (1992). Conceptual graphs: Fundamental notions. *Revue d'intelligence artificielle*, 6(4):365–406.

Ören, T. (2005). Toward the body of knowledge of modeling and simulation. In *Interservice/industry training, simulation, and education conference (I/ITSEC)*, volume 2005.

Robert, F., Abran, A., and Bourque, P. (2002). A technical review of the software construction knowledge area in the swebok guide. In *10th International Workshop on Software Technology and Engineering Practice*, pages 36–42. IEEE.

Shapiro, S. C. (2006). Knowledge representation. *Encyclopedia of cognitive science*.

Sowa, J. F. (1976). Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357.

Sowa, J. F. (2000). Guided tour of ontology. *Retrieved from*.

Straková, J., Straka, M., and Hajič, J. (2019). Neural architectures for nested ner through linearization. *arXiv preprint arXiv:1908.06926*.

Wang, C., Liu, X., et al. (2022). Deepstruct: Pretraining of language models for structure prediction. *arXiv preprint arXiv:2205.10475*.

Zeng, X., Zeng, D., et al. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 506–514.

Developing an Artificial Intelligence Model to Enhance the Emotional Intelligence of Motor Vehicle Drivers for Safer Roads

Ana Todorova[a], Irina Kostadinova[b] and Svetlana Stefanova[c]
University of Ruse "Angel Kanchev", 8 Studentska Street, Ruse, Bulgaria
attodorova@uni-ruse.bg, ikostadinova@uni-ruse.bg, sstefanova@uni-ruse.bg

Keywords: Emotional Intelligence, Artificial Intelligence, Virtual Reality, Training, Safety Roads.

Abstract: Traffic accidents and risky driving behaviour are among the deadliest problems worldwide. This statement becomes an undeniable fact, thanks to the grim statistics of the World Health Organization, according to which more than 1 million people die on the roads every year. Road accidents are also among the leading causes of death among children and young people aged 5 to 29. Against this background, a number of studies look for a link between the emotional intelligence of motor vehicle drivers and the potential prevention of risky driving. Building on the scientific knowledge generated up to this point, the present study suggests a prototype of an AI-based model that aims, through ongoing assessment and subsequent training, to enhance the emotional intelligence of both future and current motor vehicle drivers who are prone to risky behaviour on the road. Through simulated scenarios in a virtual environment, the model aims to improve the ability of drivers to recognise and manage their own and other people's emotions and to react adequately to different situations on the road. The expectation is that the model will reduce the manifestations of aggression and intolerance on the road and ultimately lead to safer roads.

1 INTRODUCTION

Risky driving is a recognised factor in road traffic accidents. Driving behaviour significantly influences the occurrence of traffic accidents and fatalities. Intentional dangerous behaviours, such as exceeding speed limits or driving under the influence of substances, are among the predominant contributing reasons for traffic accidents. It is estimated that between 90% and 95% of traffic accidents worldwide are the result of human error (Aniah, 2021; Ahmed *et al.*, 2022). Given the serious and often fatal outcomes associated with risky driving practices, certain behaviours can seem confusing when viewed from a rational perspective. In this context, theories of risky decision-making emphasise the influence of emotions on an individual's actions in dangerous situations (Megías-Robles *et al.*, 2022).

Emotionally agitated drivers may approach hazards with impaired attention and thus – unintentionally or not – engage in reckless driving. Therefore, the tools by which drivers can exercise control over their own emotions and thus prevent risky driving are vital to getting drivers safely from point A to point B. Emotional control over oneself, as well as over others, is imperative in driving situations (Ahmed *et al.*, 2022).

According to Megías-Robles *et al.* (2022), the driver's emotional state is a critical factor in explaining risk-taking propensity. The authors also claim that an adequate ability to perceive, understand, and manage emotions would allow drivers to have better control over their emotional condition and their perception of other road traffic participants. As a result, it would help to reduce participation in risky behaviours and the number of road accidents.

The relationship between emotional intelligence (EI) and driving behaviour is the subject of increasing interest in the aspect of risky and aggressive driving behaviour, especially in the context of young drivers (Aniah, 2021). Drivers' emotions have been found to be among the main factors contributing to dangerous driving behaviour. Emotions can be measured, understood and regulated most effectively through EI.

[a] https://orcid.org/0009-0007-2993-077X
[b] https://orcid.org/0000-0001-8845-7598
[c] https://orcid.org/0000-0003-4500-4833

Todorova, A., Kostadinova, I. and Stefanova, S.
Developing an Artificial Intelligence Model to Enhance the Emotional Intelligence of Motor Vehicle Drivers for Safer Roads.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 339-346
ISBN: 978-989-758-716-0; ISSN: 2184-3228

To varying degrees, large-scale studies have confirmed the link between EI and dangerous driving behaviour (Ahmed *et al.*, 2022). Emotion affects drivers due to its influence on the degree of self-control, thus also affecting the driving method (Aniah, 2021). A 2019 study also found that people who drive every day have poor EI, which hinders safe driving. The study suggests that training drivers in emotional regulation can contribute to safer roads (Parameswaran and Balasubramanian, 2020).

Taking into account the studies already carried out on the subject, the present study aims to develop an algorithm that will help develop the EI of vehicle drivers. The report is structured as follows: 1) Literature review of studies looking for a link between EI and driving; 2) Presentation of the research methodology; 3) Presentation of the developed model for evaluating and increasing EI; 4) Conclusion.

2 THEORETICAL BACKGROUND

At the end of 2023, the World Health Organization announced that around 1.19 million people die worldwide due to road traffic accidents annually. Road traffic accidents alone are the leading cause of death for children and young people between 5 and 29 years of age (WHO, 2023).

Similarly, in the United States, the National Highway Traffic Safety Administration states that in 2018, more than 2.7 million people were injured, and 36,096 people died in motor vehicle crashes. The data are frightening and show beyond doubt that it is imperative to identify the factors contributing to road accidents and to take measures to reduce them. Such factors may include vehicle defects and environmental obstructions, *i.e.* road and weather conditions, but dangerous driving behaviour also emerges as a significant factor. The latter is defined as any inappropriate driver activity that increases road hazards and the likelihood of a vehicle crash (Ahmed *et al.*, 2022).

At the same time, drivers engage in dangerous driving for a variety of reasons, including fatigue, distraction, and driving under the influence of alcohol or drugs. A number of psychological factors, such as personality traits and emotions, have also been found to contribute to dangerous driving behaviour (Owsley *et al.*, 2003). Research has alerted to the fact that drivers' emotions can particularly strongly influence their destructive driving behaviour. Previous research has also emphasised the correlation between different emotions (e.g., feelings of frustration, anger, and sadness) and aggressive driving (Ahmed *et al.*, 2022). According to Aniah (2021), physiological and psychological variables such as gender, experience, age, and emotions inevitably influence the behaviour of drivers. The author defines driving as a psychomotor ability because it consists of body movement and a cognitive task. This ability can take many forms, but the technique a driver uses depends on their personality and behavioural profile.

In a broad sense, EI is a construct that encompasses all of a person's emotional abilities. In this regard, a number of studies have sought and found a relationship between EI and risky driving behaviour (Ahmed *et al.*, 2022). The results are of varying degrees of certainty due to the existence of different approaches to EI in the literature. The self-report ability model understands EI as a mental ability and focuses on the emotional skills included in a conceptualisation of EI. It uses self-report measures to assess these abilities. The performance-based ability model also views EI as a set of emotion-related abilities but assesses them using performance-based tests. Finally, the mixed model conceptualises EI as a broad construct combining both emotion-related skills and personality factors that are evaluated through self-report instruments (Megías-Robles *et al.*, 2022).

The primary motivation to look for a link between EI and driving is that it is the driving style of many drivers that is responsible for the significant number of accidents that occur (Aniah, 2021). The results of a study by Megías-Robles *et al.* (2022) found that higher self-reported EI, especially the ability to regulate emotions, was associated with a lower propensity for risky driving. According to the authors, emotion regulation and evaluation of the emotions of others are EI abilities that may predict the number of potential accidents. EI, which is explained as a person's ability to recognise, identify, use, express, as well as regulate their own and others' emotions, has been empirically proven to influence driving behaviour (Aniah, 2021).

Such a conclusion is hardly surprising to anyone. Emotions are a fundamental part of human behaviour – they guide an individual's attention, memory, motivation, and even decision-making process. In risky contexts, however, emotions are essential, given the time (momentary) pressure and substantial emotional consequences these situations often involve. The integration of emotional factors in the processing of risky behaviour has also been demonstrated at the neural level, including in the context of driving (Hernandez *et al.*, 2014). Driving is an activity in which emotions often arise – traffic jams, accidents, risky traffic participants, *etc.* – all these situations can cause fear, as well as retaliatory aggression, intolerance, and dangerous behaviour. In many cases, these emotions underlie human behaviour, but their consequences in risky driving are particularly significant (Megías-Robles *et al.*, 2022).

There is no doubt that emotion can arise at any particular moment while driving and can have a different emotional impact on the driver's behaviour. Therefore, various emotional states can affect driving differently because people differ in how they react to situations. This makes emotional control of self and others imperative in driving conditions (Owsley *et al.*, 2003).

Aniah (2021) opined that driving behaviour is usually a pattern chosen by the driver himself. Therefore, it is argued that the specific style and skills that a motor vehicle driver applies at a given place and time are strongly influenced by his emotions and the relationship between *stimulus and response*. The link between stimulus and response, or rather the *bridge* between them, represents the individual's EI.

The above gives reason to generalise that drivers with reported higher EI scores engage in less dangerous driving, which is reflected in fewer crashes and fatalities. Therefore, as reported by Ahmed *et al.* (2022), promoting and improving EI can be helpful in preventing risky driving among non-professional drivers. Incorporating EI training into driver training, on-the-job training, and licensing procedures can help develop safer drivers.

The problem with this type of evaluation of EI and training is the human factor, *i.e.* the private interest or the subjective opinion of the trainer, which can be a prerequisite for intentional or unintentional mistakes and corrupt practices in such an essential field as road safety. With the help of large language models and the incorporation of AI during the learning stage of driving for young adults, it can be very beneficial for them to be assessed and trained based on their EI. Their behaviour can be evaluated and then compared to “ideal” behaviour, which can help them nip the negative traits in the beginning stages before passing their driving test and becoming experienced drivers. Such negative behaviour is challenging to be subjected to change, if not impossible, after years and years of driving experience.

That motivated the authors of the present study to develop a prototype model based on AI that would balance, without removing, the human factor and minimise the associated risks. In addition, a central place in the developing AI model is precisely the emotions and, more specifically, the ability of the individual to recognise (empathy) and regulate (self/control) both his own and others' emotions. The key thing about these two elements is that, although worded differently, they are involved in the different models of EI, which makes them generalisable to the EI personality.

3 METHODOLOGY

The methodology of the study includes three main stages: *1) Literature review* (introduced in the section "Theoretical background"), *2) Determination of assessing variables* (described in the section "Methodology", and *3) Explanation of the construction of the developing AI model* (described and analysed in "Results" and "Discussion" sections): elements selection, model structure, action algorithm, process analysis.

Based on the theoretical analysis, the research methodology introduces three *basic variables,* which, with the help of information and communication technologies (ICT) and human control, will be prioritised for research, analysis and improvement: emotion regulation or *self-control (x)*, emotion recognition (in others and self) or *empathy (y)*, and evaluation of the *EI as a construct (z)*.

The measurement and evaluation of *variable z* will follow one of the three schools of EI: cognitive ability, personality trait, or mixed model. The choice is yet to be refined and validated. Based on the selected overall construct, *additional variables* will be introduced, such as self-knowledge or social skills (elements of the mixed EI model).

The need to introduce baseline variables in the first place is justified by existing research on the subject of the relationship between the baseline variables specified and the driving pattern applied by the individual. Implementation of additional variables will look for underestimated or so far neglected correlations between EI components and driving behaviour. At this stage, these variables are not built into the model and its algorithm.

The developed AI model is tied to risky driving, which is why the algorithm will look for correlations between the basic and potential additional variables introduced so far and the characteristics of the motor vehicle driver identified in the background literature, influencing the genetic driving style. These *key variables* are: *aggression (a)*, *intolerance (b)*, and *risk-taking (c)*.

An interdisciplinary team of Bulgarian university scientists with expertise in management, risk management, entrepreneurship, social sciences, and ICT participated in developing the methodology and AI model. The goal is to specify as much as possible the main elements of the model, the algorithm, and the overall process from development to implementation and validation of the AI model.

4 RESULTS

From the perspective of the ICT that build the model's algorithm, the research introduces a toolkit of three technologies: artificial intelligence (AI), virtual reality (VR), and blockchain technology (BT). Table 1 describes the need for their applicability.

Table 1: Type and applicability of implemented ICT. Source*:* own development.

Technology	Applicability
Artificial Intelligence (AI)	Definition: It focuses on the creation of intelligent agents, *i.e.*, systems capable of perceiving their environment, reasoning, and acting to achieve specific goals. AI models strive to mimic and sometimes even surpass human cognitive abilities such as learning, problem-solving, pattern recognition, and natural language understanding (Antonova *et al.*, 2021; Gignac & Szodorai, 2024). Applicability: AI models can analyse large amounts of data faster and more accurately than human raters. At the same time, in big data sets, AI models can detect complex patterns that are difficult or even impossible for humans to spot. In addition, while there is still debate about their accuracy, AI models also offer greater objectivity. They are less susceptible to subjective bias, which reflects in more objective results. These two advantages, along with automating the evaluation process and saving time and resources, make AI models an adequate substitute for human resources.
Virtual reality (VR)	Definition: VR is a technology that creates immersive and interactive environments that simulate real or imagined scenarios. VR users wear headsets that display 3D images and sounds and sometimes use controllers or gloves to interact with the virtual world. VR can create realistic and engaging experiences that can stimulate emotions, thoughts, and behaviours (Susindar *et al.*, 2019). Applicability: Well-developed EI is largely believed to result from an individual's experience, which is why parallels between EI and wisdom are often sought and found. Through VR technology, the participant will be placed in situations that mimic real life and are known to help develop EI competencies.
Blockchain Technology (BT)	Definition: A distributed database that records transactions in a way that is secure, transparent, and resistant to change. Information in the blockchain is organised into blocks that are linked together by cryptographic hashes. Once added, information on a blockchain is virtually impossible to change, making it highly reliable. This technology is decentralised, meaning there is no one central authority to control the network. Instead, blockchain networks are managed by multiple computers, making them resistant to censorship and manipulation (Tripathi *et al.*, 2023). Applicability: Implementing the BT on the AI model is necessary to preserve the confidentiality of each training participant's data. Personality test results do not constitute medical information by themselves. However, the fact that they provide personal and sensitive information for the individual requires a more serious commitment to their preservation.

The developed model for increasing the EI of drivers is based on three AI models: machine learning, deep learning and natural language processing, and more specifically:

1) *Machine Learning: Classification, Regression, Clustering*

The classification will serve to classify participants' responses into different personality types or traits, such as "extroverted" or "introverted." Regression will be applied to predict the values of the entered (baseline, additional, and key) variables related to the learner's personality. Clustering will be applicable when grouping participants with similar personality profiles.

2) *Deep Learning: Neural Networks, Recurrent Neural Networks*

Neural networks will simulate the workings of the human brain and can detect complex patterns in data. They are used for natural language analysis, pattern recognition and processing large amounts of unstructured data. Recurrent neural networks, on the other hand, are suitable for analysing sequential data, such as text responses to questions in personality tests. They can pick up on contextual information and extract deeper meanings.

3) *Natural Language Analysis: Sentiment Analysis, Keyword Extraction, Semantic Content Analysis, Personalised Recommendations*

AI models will be used to analyse large amounts of text responses to personality test questions to identify keywords, phrases and emotional responses that are characteristic of certain personality traits. In some cases, AI models will operate to analyse facial expressions, gestures and other non-verbal signals to gain additional information about an individual's personality. As a result, AI will create personalised recommendations for an individual's EI development based on test results, as well as for the virtual experiences that will be most relevant to the object, subject and purpose of the learning.

Figure 1 presents the algorithm of the developed AI model. The whole process is divided into *four main Stages*: *I)* Measurement and evaluation of the introduced *basic (x, y, z)*, *key (a, b, c)* and *additional variables*; *II)* Verification of the results obtained by a person*; III)* and *IV)* Depending on the results of *Stage II* – respectively unsatisfactory or satisfactory from the point of view of human evaluation, in the next stage either a human specialist makes a new measurement and evaluation of the variables, or the AI determines the virtual experience that the object to be subjected to.

The model introduces *four working agents*: *(A)* – learner, *(B)* – training/evaluating AI, *(C)* – training/assessing/supervising specialist persons, and *(D)* – AI-defined the virtual experience. In *Stages I* and *III*, blockchain encryption of the data generated from the tests to measure the sought variables was introduced. This, as stated at an earlier stage, is imperative from the point of view of the object's confidentiality and to protect its data. The results of such tests, if publicly available, may have an adverse impact on an individual's personal and professional development in the future. Also, blockchain technology will neutralise the possibility that the records of the results can be tampered with and manipulated.

The role of virtual reality, on the other hand, is to act as an imaginary learning environment, where learning does not mean learning phrases or behavioural responses (Han & Lorenzo Najord, 2024). On the contrary, the goal is a rapid accumulation of life experience, but in a protected (virtual) environment, with this experience directly related to developing the basic variables – self-regulation and empathy, and reducing the key variables – aggression, intolerance, propensity to risk. In fact, in virtual reality, the individual will be trained to respond most effectively to the *stimulus-response* interdependence that Aniah (2021) also talks about. This effective response is due to a developed EI capability.

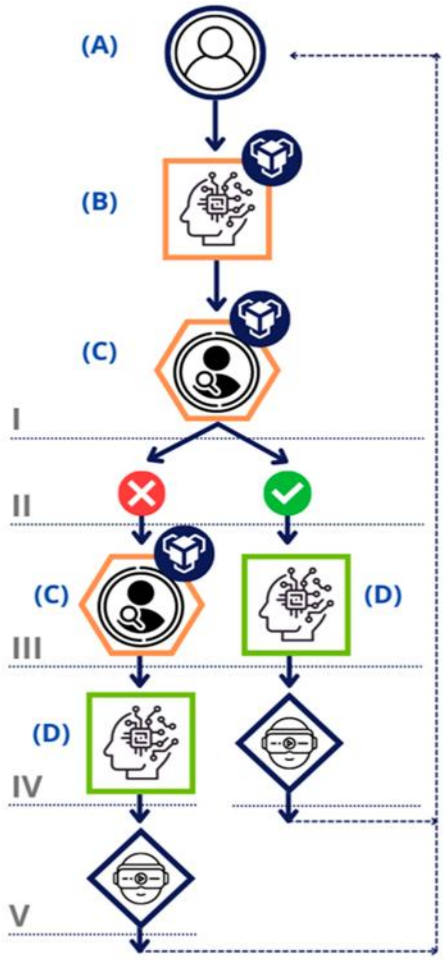


Figure 1: Working algorithm of the developed AI model for evaluating and improving the EI of motor vehicle drivers for safer roads. Source*:* own development.

In their study, Susindar *et al.* (2019) demonstrate that the use of VR can be an effective emotion-inducing method when investigating emotion-influenced decision-making. As stated in the theoretical background, this is highly inherent to drivers of motor vehicles who are daily exposed to emotionally arousing situations and have to make immediate decisions and respond to the stimulus-response relationship according to their momentary mood. Susindar and his co-authors' research focused on extracting and generating situations that evoke fear and anger – emotions that drivers also face on the road. At the same time, the authors of the cited study add that it is not entirely clear how the virtual environment affects performance (or learning) and the degree to which emotions are evoked. This means, on the one hand, that it is not categorically clear whether an individual would react in the same way in a real and virtual environment and, on the other hand, that more in-depth research is needed in this direction. However, the latter in no way belittles VR technology capabilities, as evidenced by similar studies on the topic (Marques *et al.*, 2022; Mancuso *et al.*, 2023; Hariyady *et al.*, 2024).

Considering the rapid growth of technology in every aspect, people are more and more concerned about how to improve user experience rather than the construction of the experience itself. Giving people the opportunity to experience this in virtual reality would allow them to see their mistakes and correct them in their free time in a safe environment, making them better on the road without the risks of actually being there. This would help them improve their EI time and time again, improving the quality of their driving skills, thus improving the overall quality of driving for the other drivers on the road around them.

As the graph (Figure 1) shows, the process does not stop after the learner's experience in virtual reality. New measurements and evaluation of the basic variables, as well as the additional and key variables adopted in the Methodology, are needed. Therefore, the algorithm starts again from *Stage I*. The subsequent steps, *i.e.* whether to stop or continue training, depend on the evaluation of agents B and C and the decision of agent B. Realistically, the process of increasing EI may never stop and even apply to refresher driving courses (similar to the periodically conducted psychological tests) of experienced drivers of motor vehicles.

Although still unproven categorically, some EI researchers believe that this ability and its inherent competencies have the potential to develop positively over time, one of the reasons being precisely the accumulated life experience. The AI model developed and presented in the present study is based precisely on the statement that EI can be developed and that it undergoes evolution with age (Gilar-Corbi *et al.*, 2019).

5 DISCUSSION

When analysing the algorithm, the logical question arises: What necessitates human monitoring and control of the evaluation of the AI model? Conversely, if this activity is entirely within a human's capabilities, why is it necessary to introduce the use of AI?

First of all, the potential of AI to process vast amounts of data at a speed beyond the reach of a human being has already been repeatedly tested and proven. However, not only is speed essential, but so is the refinement of the results. Subjectivity and unintentional omissions in data analysis are human. But the same goes for imperfect (yet) AI, which, like its creator, is also not immune to biases and errors in its algorithm. In fact, the AI model being developed to increase the EI of drivers shows precisely how natural and artificial intelligence should collaborate – in a balanced synergy, but also with a good-natured mistrust of the abilities of one or the other.

On the other hand, it is essential to note that AI models cannot and are not expected to replace human raters. In the overall model, the human factor is equal to the involvement of AI. This further justifies the intervention of blockchain technology in the algorithm. The encryption of the data from the conducted tests and training will guarantee the confidentiality and objectivity of the process. Therefore, AI models should be used as an additional tool to justify and support decision-making. To ensure the validity and reliability of AI-based test results, careful validation studies need to be conducted.

Attention to detail throughout the process is significant as irreparable damage can be done to the learner's psyche, which is not the purpose of the AI model. It is extremely crucial to create and apply the "training" in virtual reality according to the test results of the entered variables. A hypersensitive person would accept such experiences in one way, and a more selective individual in his feelings would have a radically different perception of what is happening. Therefore, every AI-proposed and human-approved virtual learning environment in EI must be perfect and maximally adapted to the personal qualities of the respective learner.

It is equally important to explain why the basic variables, such as empathy and self-regulation, are set in the algorithm rather than examining only the EI construct as a whole. Or why the general construct EI is asked at all, and not only the connections between empathy and self-regulation of motor vehicle drivers and their aggressive and angry behaviour that puts them in risky situations are investigated. As stated earlier, self-regulation and empathy are part of the general EI construct. Although in the pilot presented AI model, as well as numerous other studies, they are studied as independent competencies, in reality, it is not possible to fully develop these two abilities without the other EI competencies, regardless of which of the three EI models will be followed. For example, the mixed model includes self-regulation and empathy, as well as self-knowledge, motivation, and social relationships. Goleman himself, one of the creators of the model, points out that it is not possible to achieve self-regulation without self-awareness since each of the sub-competencies of self-management steps on self-awareness (Goleman and Chernish, 2023). This necessitates the study, evaluation and improvement of the general EI construct by examining the construct's constituent components separately, *i.e.*, it is crucial to approach the problem deductively.

Limitations of the Study

A significant limitation of the study and the introduced model is the lack of definitive data on the effectiveness of some of the applied ICT. The analyses and tests carried out so far are scarce, and the results are contradictory. Therefore, the application of the described technologies does not guarantee the desired result.

Another limitation of the study is the lack of data from the practical application of the developed model. The presented model is purely theoretical and has not been validated empirically, which limits the ability to conclude its reliability and predictive validity. To overcome this limitation, it is necessary to conduct an empirical study in which the model is tested on a large and heterogeneous sample of participants. The formed interdisciplinary team needs to conduct considerable research and testing in this direction, but the model aims to make a start.

On the other hand, implementing such an algorithm in a seemingly state-controlled but apparently private-interest-dominated environment, such as driving schools, requires considerable will and agreement from multiple (dis)interested parties. In this sense, just creating the AI model is not enough. Therefore, the author collective's future efforts will be directed to the experimental introduction of the model and the search for validation of the algorithm and the expected results.

Applicability

The goal of the research team is to apply the algorithm as a priority in the courses for acquiring driving skills. It is an environment in which the individual most actively reveals himself – as more aggressive, more intolerant, more selfish or vice versa. Bulgaria is first in the EU in terms of road deaths in 2023. The European Commission published preliminary data for last year – the EU average is 46 deaths per million inhabitants. In Bulgaria, the ratio is 82 victims per one million inhabitants (Apostolova, 2024). In this sense, the developed AI model is expected to lead to significantly greater self-awareness and self-regulation of their own emotions in both inexperienced and seasoned drivers on the roads.

At the same time, the developed AI model can also be used in a number of other areas where weak self-regulation of emotions leads to conflict situations or where reducing the levels of aggression is necessary. The results of the implementation will validate the results and allow the model to be embedded in educational programs and its application in the fight against hate speech, intolerance of differences, selfishness, cruelty to the weaker, *etc.* This will be the subject of future research by the authors.

6 CONCLUSION

Death on the roads as a result of serious road traffic accidents is one of the saddest facts of our time, which we must either accept or overcome. Over 1 million people die on the road, not a small number of them children and young people. The purpose of the developed AI model to increase the EI of drivers is precisely this – fighting statistics, risky driving, but above all, saving lives before they are even in danger.

It is essential to clarify that in this "battle", neither EI nor AI is a panacea. The human factor remains a significant unknown, along with the cultural and social characteristics of one or another country. However, suppose it is almost impossible to change the cultural and social conditions. In that case, only the way in which the individual responds to the *stimulus-response* relationship is within his capabilities.

In conclusion, the developed AI model is a small step towards achieving high EI drivers of motor vehicles and reducing risky behaviour on the road. Although further research and efforts are needed to implement this technology, it is essential to find a way to save human lives and reduce the socio-economic costs directly related to traffic accidents.

ACKNOWLEDGMENTS

This study is financed by the European Union-NextGenerationEU through the National Recovery and Resilience Plan of the Republic of Bulgaria, project №BG-RRP-2.013-0001-C01.

REFERENCES

Ahmed, J., Ward, N., Otto, J., McMahill, A. (2022). How does emotional intelligence predict driving behaviors among non-commercial drivers? *Transportation Research Part F: Traffic Psychology and Behaviour, Volume 85*, 38-46, February 2022. DOI: 10.1016/j.trf. 2021.12.013.

Aniah, E. (2021). The influence of Emotional Intelligence on Driving Behavior in the Transportation Sector: A literature review. *Journal of Contemporary Issues in Business and Government, Volume 27*, Issue 06, 1524-1537, December 2021. DOI:10.47750/cibg. 2021.27. 06.121.

Antonova, D., Kunev, S., Venelinova, N., Kostadinova, I. (2021). Data Management Conceptual Algorithm of Transnational Digital Scientific Infrastructure as an Answer to the Globalization Challenges, *20th International Scientific Conference Globalization and Its Socio-Economic Consequences University of Zilina*, 1-9, DOI: 10.1051/shsconf/20219205002.

Apostolova, D. (2024). България на първо място в ЕС по катастрофи, завършили със смърт [Bulgaria ranks first in the EU in fatal accidents], BNT 1, 8 March 2024. https://bntnews.bg/news/balgariya-na-parvo-myasto-v-es-po-katastrofi-zavarshili-sas-smart-1269911news. html [In Bulgarian] (Accessed 16 August 2024).

Gignac, G., Szodorai, E. (2024). Defining intelligence: Bridging the gap between human and artificial perspectives. *Intelligence, Volume 104*, 101832, May–June 2024. DOI: 10.1016/j.intell.2024.101832.

Gilar-Corbi, R., Pozo-Rico, T., Sánchez, B., Castejón, J. (2019). Can emotional intelligence be improved? A randomized experimental study of a business-oriented EI training program for senior managers. *PLoS One, Volume 14*, Issue 10, e0224254, 23 October 2019. DOI: 10.1371/journal.pone. 0224254.

Goleman, D., Chernish, C. (2023). *Optimal: How to Sustain Personal and Organizational Excellence Every Day*, Publisher: Harper Business.

Han, M., Lorenzo Najord, I. (2024). Virtual reality with e-learning model in business management and knowledge complex environment for gaming analysis. *Entertainment Computing, Volume 50*, 100703, May 2024. DOI: 10.1016/j.entcom.2024.100703.

Hariyady *et al.* (2024). Virtual Reality and Emotional Responses: A Comprehensive Literature Review on Theories, Frameworks, and Research Gaps. *ITM Web of Conferences 63*, 2-13, 01022, February 2024. DOI: 10.1051/itmconf/20246301022.

Hernandez, J., McDuff, D., Benavides, X., Fernandez, J., Maes, P., Picard, R. (2014). AutoEmotive: Bringing empathy to the driving experience to manage stress. *DIS Companion '14*, 21-25 June 2014, Vancouver, BC, 978-1-4503-2903-3/14/06. DOI:10.1145/2598784.2602780.

Mancuso, V., Bruni, F., Stramba-Badiale, C., Riva, G., Cipresso, P., Pedroli, E. (2023). How do emotions elicited in virtual reality affect our memory? *A systematic review. Computers in Human Behavior, Volume 146*, 107812, September 2023. DOI: 10.1016/ j.chb.2023.107812.

Marques, A. J., Gomes Veloso, P., Araújo, M., de Almeida, R. S., Correia, A., Pereira, J., Queiros, C., Pimenta, R., Pereira, A. S., Silva, C. F. (2022). Impact of a Virtual Reality-Based Simulation on Empathy and Attitudes Toward Schizophrenia. *Frontiers in psychology, Volume 13*, 814984, 4 May 2022. DOI: 10.3389/ fpsyg.2022.814984.

Megías-Robles, A., Sánchez-López, M., Fernández-Berrocal, P. (2022). The relationship between self-reported ability emotional intelligence and risky driving behaviour: Consequences for accident and traffic ticket rate. *Accident Analysis & Prevention, Volume 174*, September 2022. DOI: 10.1016/j.aap.2022.106760.

Owsley, C., McGwin Jr., G., McNeal, S. (2003). Impact of impulsiveness, venturesomeness, and empathy on driving by older adults. *Journal of Safety Research, Volume 34*, Issue 4, 2003, 353-359. DOI: 10.1016/ j.jsr.2003.09.013.

Parameswaran, S., Balasubramanian, V. (2020). Assessing the Relation Between Emotional Intelligence and Driving Behavior: An Online Survey. *Advances in Human Factors of Transportation, Volume 964*, 233–239, 06 June 2019. DOI: 10.1007/978-3-030-20503-4_21.

Susindar, S., Sadeghi, M., Huntington, L., Singer, A., Ferris, T. (2019). The Feeling is Real: Emotion Elicitation in Virtual Reality, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 63*, Issue 1, 252-256, 20 November 2019. DOI:10.1177/1071181319631509.

Tripathi, G., Ahad, M. A., Casalino, G. (2023). A comprehensive review of blockchain technology: Underlying principles and historical background with future challenges. *Decision Analytics Journal, Volume 9*, 100344, December 2023. DOI: 10.1016/j.dajour. 2023.100344.

WHO (2023). Global status report on road safety 2023. World Health Organization, 13 December 2023, ISBN: 978-92-4-008651-7. https://www.who.int/publications/ i/item/9789240086517 (Accessed 16 August 2024).

Connecting Critical Infrastructure Operators and Law Enforcement Agencies to Share Cyber Incident Information with Early Warning Systems

Harri Ruoslahti[1] [a] and Ilkka Tikanmäki[1,2] [b]
[1]ResLab, Laurea University of Applied Sciences, Vanha maantie 9, Espoo, Finland
[2]Department of Warfare, National Defence University, Helsinki, Finland
{harri.ruoslahti, Ilkka.tikanmaki}@laurea.fi

Keywords: Early Warning System, Information Sharing, Law Enforcement, Critical Infrastructure.

Abstract: Cyber incidents and business interruptions rank as the foremost business risks. With Early Warning Systems (EWS), that work in parallel with other cyber mechanisms, organisations can independently manage cyber-sensitive intelligence-related data. This article provides a qualitative multi-case study analysis. The data consists of systematic reviews and cross-case conclusions of six (n = 6) case studies on information sharing. EWS is a valuable tool that can help critical infrastructure providers protect against cyberattacks. EWS can provide a platform for sharing information and resources. This can help improve situational awareness, enhance incident response, and facilitate collaboration. between critical infrastructure providers, as critical infrastructure operators and relevant Law Enforcement Agencies (LEA) can share information on cyber incidents and monitor cyber incident progress. EWS can be used to exchange cyber threat intelligence and information sharing can be facilitated with a common reference library where alerts can be shared as tickets. This would enable information exchange in both directions.

1 INTRODUCTION

Cyber Threat Intelligence (CTI) can be aided by an Early Warning System (EWS) to provide any type of organisation with an increased capability to share information needed to detect and respond to cyber incidents. An EWS is a security operation support tool that enables the coordination and sharing of cyber-incident information in near real-time. An EWS will help provide timely and accurate information to all involved parties.

The development of EWSs can be rooted in information sharing and trust models from within the cyber domain as well as models from other domains (Rajamäki & Katos, 2020).

This article provides a qualitative multi-case study analysis consisting of systematic reviews and cross-case conclusions of six (n = 6) case studies on information sharing among partners. This analysis provides a deeper understanding of how EWS can enable cyber incident information sharing across organisational boundaries between critical infrastructure operators and Law Enforcement Agencies (LEA).

An EWS for cyber intelligence can serve as a security operations support tool in that it enables all network members to share information and coordinate their responses in near real-time (Rajamäki & Katos, 2020), e.g., connect critical infrastructure and service providers with law enforcement authorities (Almén et al., 2022).

With EWS stakeholders can retain their independent management of cyber-sensitive intelligence and related data management, while the EWS will work parallel with other cyber mechanisms (Rajamäki & Katos, 2020).

The ECHO project (European network of Cybersecurity centres and competence Hub for innovation and Operations) was one of four Pilot projects launched by the European Commission to establish and operate a Cybersecurity Competence Network focused on the ECHO Early Warning System (E-EWS) and ECHO Federated Cyber Range

[a] https://orcid.org/0000-0001-9726-7956
[b] https://orcid.org/0000-0001-8950-5221

Ruoslahti, H. and Tikanmäki, I.
Connecting Critical Infrastructure Operators and Law Enforcement Agencies to Share Cyber Incident Information with Early Warning Systems.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 347-355
ISBN: 978-989-758-716-0; ISSN: 2184-3228

(E-FCR) and related inter-sector prototypes (ECHO project, 2021).

The development of EWS can be traced to information sharing and trust models within the cyber domain (Rajamäki & Katos, 2020). Project ECHO conducted demonstration activities that highlighted e.g., the benefits of the features and capabilities of the E-EWS, and promoted the technology roadmaps for E-EWS, E-FCR, and inter-sector prototypes to show their value in multi-sector scenario requirements and demonstration cases (ECHO project, 2021).

This study's research question (RQ) is: How can an Early Warning System enable the sharing of cyber incident information between critical infrastructure operators and law enforcement agencies?

2 LITERATURE

According to the Allianz Risk Barometer 2023 cyber incidents and business interruptions rank as the foremost business risks (Allianz Global Corporate & Specialty SE, 2023). Today's critical infrastructure operators face critical incidents and cyber-attacks. Organisations need to reconsider their approaches to information-sharing-based resilience-building (Pöyhönen et al., 2020).

Organisations operating critical infrastructures or providing critical services for society are more and more dependent on complex and interlinked cyber systems and their interconnections (The International Chamber of Commerce, 2024).

2.1 Resilience

(Vos, 2017) defines resilience as the ability to adapt to a changing environment and mitigate emergency crises. Resilience can be demonstrated as flexibility, endurance, and an ability to recover from adverse events or to adapt to an after-crisis new normal (Cauffman, 2018). Crises are often caused by external risks, while the resilience of an organisation will also include many internal priorities, such as preventive behaviours, and preparing guidelines and procedures to when a response to a critical event may be needed (Linkov et al., 2014).

Four event management cycle phases (plan, absorb, recover, and learn/adapt) can be combined with the domains of physical, information, cognitive, and social can help understand resilience in the fields of Information Technology (IT) and systems sciences (Linkov et al., 2013).

"Resilient organizations or networks show organizational stability, agility, and a culture that promotes situational awareness to detect and identify clues that may indicate the realization of risks for appropriate mitigation and reaction" (Hytönen et al., 2023, p. 163).

Examples of threats against critical infrastructures and vital societal services include e.g., the attacks that impaired the functionality of the English National Health Service (NHS) in 2017 (Ghafur et al., 2019), halted a hospital network in the Czech Republic in 2021 (Muthuppalaniappan & Stevenson, 2020), and stopped the movement of goods by a South African port and rail operator in 2021 (Fitch Solutions, 2021). Similarly, an attacker stole the records of thousands of patients from the Finnish private psychotherapy service provider Vastaamo from 2018 to 2020 and tried to use the stolen files to blackmail individual patients threatening to expose documents that contained their personal identity codes and therapy session transcripts (Tuttle, 2021; Whitney, 2021).

Promoting resilience calls for leadership, resource allocation, planning, and awareness (O'Rourke & Briggs, 2007). Understanding how network members view a common problem can help enhance communication and understand interdependencies (Linkov et al., 2014). Systems that combine principles of business continuity with cyber threat warning systems can promote better preparedness and cyber resilience against cyber incidents (DYNAMO project, 2024).

Transparent dialogue on resilience management, and potential risks, supported by innovative leadership, effective planning, and long-term commitment to allocate needed resources help build and maintain acceptance of resilience (Linkov et al., 2014; O'Rourke & Briggs, 2007). Systems often show complex interactions between people, technologies, and processes (Vos, 2017), and the vulnerability of many of these socio-technical systems (combining human and technical aspects) have increased; understanding the mutual entanglement of material structures and human organisations help create practices to anticipate possible incidents and promote feedback and learning (Amir & Kant, 2018; Rajamäki & Ruoslahti, 2018).

(Vos, 2017) states that organisational resilience as a framework can create tools and conditions to help reduce risks, understand issues, and mitigate crises.

Resilience requires adaptive capacities and cooperation (Vos, 2017), where information on threats and vulnerabilities helps identify trends, understand risks, and determine preventive measures (Stanciugelu et al., 2013).

The resilience matrix integrates the phases of planning, absorbing, recovering, and learning/adapting combined with domains physical, information, cognitive, and social (Linkov et al., 2013). Domains can be based on business continuity management (BMC), adding elements: risks, critical functions, key personnel, guidelines/procedures, and open communication (Ruoslahti, 2020). These principles are combined in the project DYNAMO matrix (Hytönen & Ruoslahti, 2023).

2.2 Cybersecurity

The growing numbers and increasing sophistication of cybersecurity threats and attacks are a reality and one of the foremost risks to business continuity (Michel & King, 2019). Continuity management for critical infrastructure operators and their networks rely on the interconnectivity between other networks and systems of systems (Linkov et al., 2013).

Cybersecurity helps make the online secure and safe; cybersecurity uses technology and legislation to protect and manage information (Ruoslahti & Tikanmäki, 2022). Cybersecurity can be seen as processes and measures that protect cyberspace, its systems, physical aspects, devices, and software, which have no geographical boundaries leaving only digital traces (Mohammed, 2015) from foreseen threats (Craigen et al., 2014). Cyber events can have very tangible effects though cyberspace in itself is intangible (Shoemaker & Conklin, 2011), and cybersecurity as well as security in general should be solidly embedded in all organisational processes (Kilani, 2020).

New cyber threats and vulnerabilities are constantly emerging, so cybersecurity needs to be a consistent and continuous process (Cavelty, 2010). Cybersecurity is needed to protect applications and cyberspace from various threats that could compromise their safety (Craigen et al., 2014). Making cybersecurity part of comprehensive security and part of one's organisational security culture shared by all builds situational awareness, defined direction and guidelines (Limnéll et al., 2014).

Cybercrime internationally is one recognised threat to cybersecurity (Mohammed, 2015), and costs caused by cybercrime have continuously increased (Cavelty, 2010). Cybercrime is seen as all illegal and criminal acts against computer data, systems, unauthorised access, modification or impairment of digital or computer systems (Mohammed, 2015; Payne, 2020).

ICT skills can upgraded through proper ICT training (Conkova, 2013; Isidro-Filho et al., 2013). Building skills and competencies aim at people to better navigate the cyber domain (Aaltola & Taitto, 2019). With appropriate knowledge of ICT, workers can capture, store and share organisational knowledge that makes their expertise better available within the organisation (Im et al., 2013), and that organisations can develop the skills needed to absorb state-of-the-art knowledge from external sources (Cupiał et al., 2018).

2.3 Information Sharing

A secure barrier formed by cybersecurity can protect a most valuable organisational asset. Cybersecurity measures can enhance business continuity when well-organised and widely applied. According to literature, business continuity is the primary focus of every organisation, and Figure 1 shows how cyber security can be seen as a circle surrounding it (Frisk et al., 2022).

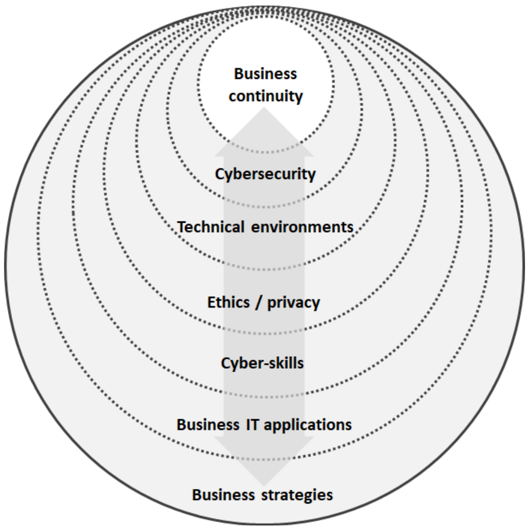


Figure 1: Model of the dimensions of cybersecurity (Frisk et al., 2022).

Cyber security is the key to achieving business continuity. The foundation for comprehensive cyber security combines appropriate levels of cyber skills and well-functioning trusted technical environments. Ethical configuration and use of technical environments are necessary to ensure privacy when using applications, as neither alone can guarantee security. Cyber security insight is provided by the company's IT applications, which have both digital and physical elements and are cyber-physical in nature. (Frisk et al., 2022).

The Maritime Integrated Surveillance Awareness (MARISA) project's user community's collaboration was complicated due to its nature. The use cases of the MARISA project involved numerous actors from various sectors and countries. The complexity of the sector was further complicated by the presence of multiple authorities, such as police and gendarmerie, in some EU member states. (MARISA Project, 2017).

Collaboration is a key factor in the development of knowledge (Pirinen, 2015), and it is necessary to work intensely (Ruoslahti, 2018, p. 115), including interdependence and resource integration (Ruoslahti & Tikanmäki, 2017).

The consequence is the requirement to use resources belonging to others and generate exchange value: "Knowledge itself is an increasingly important source to competitive advantage and a key to the success of modern organisations and creative higher education, strengthening the collective expertise, industry-service clusters, employees and competitiveness in the global economy" (Pirinen, 2015, p. 315).

The dynamic interaction between several different actors with different interests must be highlighted in organisations' multi-stakeholder communication, including consortia of publicly funded innovation projects (Vos et al., 2014). Issues that hold the most significance to people are those that are central to them, as stated by (Luoma-aho & Vos, 2010). Problem-solving arenas for exchanging practical, legal, and ethical issues are provided by authority communities, where actors work together to define and refine creative use cases. These arenas are also places where people compete for problem-solving and influence, with their decisions being influenced by both common agendas and one's activities (Vos, 2018).

(Pirinen, 2017) states that sharing information and situational information is necessary to enhance resilience. Awareness and communication can help promote flexible networks (O'Rourke & Briggs, 2007). To be effective in addressing resilience training, it is necessary to include all stakeholders, such as industry, industry associations, and decision-makers (Ruoslahti et al., 2018).

The importance of communication with stakeholders in terms of resilience is highlighted by (Linkov et al., 2014). Networked organisations are striving to enhance their resilience due to the vulnerability inherent in interdependencies.

Building situational awareness and promoting collaboration requires the interaction between authorities and the sharing of information, which is important in increasing safety. Cooperation aims to enhance the situation by increasing recognition, exchanging best practices, enhancing interoperability, decreasing overlap, and promoting cooperation across borders and sectors. (Tikanmäki & Ruoslahti, 2017).

Situational awareness is a crucial factor in resolving security incidents. Understanding the current situation and how their actions affect it is what it means to a person. Following an appropriate security policy can lead to a higher level of understanding and awareness. All employees are required to undergo training and continuous cyber security development as part of the security policy (Almén et al., 2022).

Network disruption data sharing between critical infrastructure administrators and law enforcement authorities can provide them with a shared situational awareness of new threats. A more proactive defensive position can be achieved through the identification of attack patterns, emerging vulnerabilities, and potential targets through this cooperation. Potential attackers can be deterred by a robust information sharing framework. Preventing adversaries from targeting critical infrastructure by quickly detecting, sharing, and responding to any cyber intrusion.

3 METHOD

This study is a qualitative multi-case research analysis consisting of systematic reviews and cross-case conclusions of six (n = 6) case studies on information sharing among partners.

Table 1: Six (n = 6) case studies on information sharing among partners.

Data sources	Title
ECHO Deliverable D8.2	E-EWS and E-FCR demonstration surveys
Rajamäki & Katos, 2020	Information Sharing Models for Early Warning Systems of Cybersecurity Intelligence
Simola, J., & Lehto, M. 2020	National cyber threat prevention mechanism as a part of the E-EWS.
Rajamäki et al., 2024	View of E-EWS-based Governance Framework for Sharing Cyber Threat Intelligence in the Energy Sector
Hytönen, E., Rajamäki, J., & Ruoslahti, H., 2023	Managing Variable Cyber Environments with Organizational Foresight and Resilience Thinking.
Almén, C., Hagström, N., & Rajamäki, J., 2022	ECHO Early Warning System as a Preventive Tool against Cybercrime in the Energy Sector.

The final analysis provides a deeper understanding of how EWS can enable cyber incident information sharing across organisational boundaries between critical infrastructure operators and Law Enforcement Agencies (LEA).

4 ECHO EARLY WARNING SYSTEM

Early Warning Systems (EWS) can help critical infrastructure providers share information on cyber threats in near real-time, which allows them to be more proactive in protecting their systems from attacks (Simola & Lehto, 2020).

The ECHO Early Warning System (E-EWS) will allow the collection and preservation of evidence in a forensically sound manner through information sharing between CERTS/CSIRTS, Critical infrastructure and services providers, and LEA (Rajamäki & Katos, 2020).

Table 2: EWS network partners. Modified from (Rajamäki & Katos, 2020).

EWS User	Role
National/EU CERTS	Protect critical infrastructure
ISP CERTS	Protect Internet and its services
Organisational CERTS	Protect organisation
ICT Vendor CERTS	Protect products
Law enforcement Agencies	Ensure public safety and security of society
Critical infrastructure organisations	Provide products and services critical to society
Organisations, Individuals, Researchers	Secondary users that may be involved when handling incidents.

The E-EWS is a tool used to enhance proactive cyber defence through effective information sharing. It facilitates trusted cooperation among multiple parties in the cybersecurity domain, providing reliable incident handling and collaboration capabilities (Almén et al., 2022).

Sharing essential information quickly between stakeholders requires automated information sharing. E-EWS is designed to provide a security support tool that facilitates the coordination and sharing of information among ECHO network members in near real-time (Simola & Lehto, 2020). Proactive cyber defence is improved and strengthened through efficient and effective information sharing through the E-EWS tool. A reliable cooperation between multiple parties on the cyber security scene is created by the tool. (Almén et al., 2022).

Cyber-sensitive data management and related data management can be completely independent for ECHO partners in E-EWS. In the public safety environment, the early warning system functions as a supplementary component to other mechanisms. (Simola & Lehto, 2020). Collaboration and case handling are enhanced by its excellent and reliable features. These are suitable for use both at the start of attacks and during the duration of the attack (Almén et al., 2022).

Critical infrastructure operators use command and control networks and systems – Industrial Control Systems (ICS) / Supervisory Control and Data Acquisition (SCADA) systems – designed to support industrial processes. Today, ICS and SCADA systems are widely used in many critical infrastructure sectors to help for process control, automation, and safety (Almén et al., 2022).

The tool aims to aid and improve the performance of computer/cyber emergency response teams, including CIRTs and SOCs. Alerts are shared among partners and are the main source of information for E-EWS. An enriched data model that can be accessed by both humans and machines is produced by analysing and using alerts. Alerts give access to a vast array of attributes for records that can be utilised. (Almén et al., 2022). Table 3 provides three specific examples of how the EWS can help critical infrastructure providers.

Table 3: Examples of benefits of EWS to critical infrastructure. Modified from (Simola & Lehto, 2020).

Example	Possible actions
Provider receives alert from the EWS about a new malware that is targeting critical infrastructure.	Protect systems from malware, such as installing security patches or updating their antivirus software
Provider receives alert from the EWS about a denial-of-service attack that is targeting a specific critical infrastructure sector.	Protect systems from the attack, such as increasing their bandwidth or implementing security measures to prevent the attack from succeeding
Provider may participate in coordinated response to a cyber incident with other providers, government agencies, and other stakeholders.	Sharing information about the incident, coordinating security measures, or developing a plan for recovery

Firstly, as seen above (Table 3) a critical infrastructure provider receives an alert from the EWS about a new malware targeting critical infrastructure. The provider is prompted to take steps to protect their systems from the malware, e.g., by installing security patches or updating their antivirus software, Secondly, a provider receives an alert from the EWS about a denial-of-service attack that is targeting a specific critical infrastructure sector. This would allow the provider to take steps to protect their systems from the attack by e.g., increasing their bandwidth or implementing security measures to prevent the attack from succeeding, Thirdly, a provider participates in a coordinated response to a cyber incident with LEA, government agencies, and other providers or stakeholders, involving e.g., sharing information about the incident, coordinating security measures, or developing a recovery plan. (Simola & Lehto, 2020).

Projects ECHO and DYNAMO connect information sharing with the concept of situation awareness – understanding the current situation of a security incident and how one's actions impact the situation – is an important element in solving critical infrastructure security incidents (Almén et al., 2022, pp. 17). Cyber situational awareness to support decision-making can be improved by combining systematic Cyber Threat Intelligence (CTI) and Business Continuity Management (BCM) (Hytönen et al., 2023).

Early warning systems can be highly beneficial for users in countering potential threats and attacks (Simola, 2019). However, it's crucial to note that the effectiveness of these systems may be compromised if the users operating them are unsure of how to act in difficult circumstances (Matveeva, 2006).

Cybercriminals and state-sponsored actors can exploit vulnerabilities in industrial control systems and SCADA systems (Alanazi et al., 2023). The E-EWS, further developed at DYNAMO, offers the energy sector valuable resources to secure infrastructure availability and performance in the event of a cyber threat and prevent this type of attack (Rajamäki et al., 2024).

The complexity of interdependencies and their impact on operational continuity is often forgotten during traditional risk assessments and information security management processes (Hytönen et al., 2023). All company or organisation staff should have an understanding of cybersecurity-related issues and countermeasures (National Cyber Security Centre, 2016).

EWSs assist in orchestrating responses to cyber incidents, potentially reducing the impact of attacks and preventing their spread to other service providers (Simola, 2019).

Fighting cybercrime requires a powerful combination of situation awareness and Early Warning Systems (Almén et al., 2022). Improving cyber security awareness, coordination, and response capabilities can be achieved through EWS, which can be an important instrument for safeguarding critical infrastructure against cyber threats (Ramaki & Atani, 2016).

5 CONCLUSIONS

Providing a common platform for cybersecurity information sharing can help improve communication and collaboration between providers, government agencies, and other stakeholders. Early warning systems (EWS) can help raise awareness about cybersecurity threats and best practices among critical infrastructure providers (Almén et al., 2022; Rajamäki & Katos, 2020; Simola & Lehto, 2020).

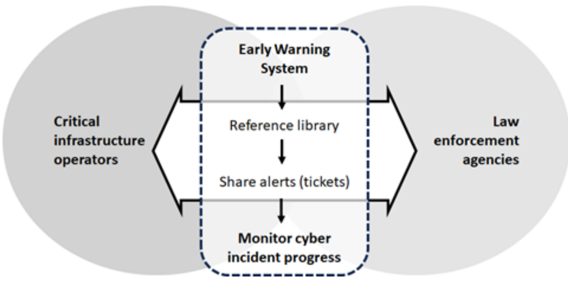


Figure 2: The concept to connect critical infrastructure operators and law enforcement agencies with Early Warning Systems to share cyber incident information (by author).

Critical infrastructure operators and law enforcement agencies (LEA) can share information on cyber incidents and monitor cyber incident progress (Figure 2). The two entities would use an EWS to exchange cyber threat intelligence. Information sharing can be facilitated with a common reference library where alerts can be shared as tickets. This would enable information exchange in both directions.

The security of critical infrastructure like power grids, water supplies, transportation systems, and communication networks has become crucial in an interconnected world. The breakdown of these systems is a significant risk to economic, social, and national security, as they are the backbone of modern society. Fostering strong collaboration between critical infrastructure operators and law enforcement agencies is a highly effective strategy for securing these vital assets. Our collective ability to detect, prevent, and respond to cyber threats is enhanced by this partnership, which enables us to share cyber incident information through EWS.

Cybercriminals and nation-state actors are attracted to critical infrastructure due to its central role in societal functioning. A cyber-attack that succeeds in attacking these systems can cause widespread outages, economic losses, and even death. The power grid may be disrupted by a cyber-attack, which can have an impact on hospitals, emergency services, and everyday operations. The protection of these systems is a national responsibility, not just a technical challenge.

The investigation of cybercrimes, gathering intelligence on potential threats, and coordinating responses to cyber incidents are all carried out by LEA. The early detection of anomalies and potential cyber intrusions by critical infrastructure operators can be strengthened by timely and accurate information exchange with their efforts.

Cyber threats are detected by early warning systems before they can cause significant damage. To identify and report suspicious activity, these systems utilise continuous monitoring, data analytics, and threat intelligence. Early warning systems can gain a broader perspective on the threat landscape by integrating inputs from multiple sources, such as critical infrastructure operators and LEAs.

By sharing information, coordinating response can be achieved and all relevant parties can be informed to take appropriate action to mitigate impacts. The significance of this is amplified for critical infrastructure, as a delay in response can lead to significant consequences.

EWS can help critical infrastructure providers improve their situational awareness with a more comprehensive picture of the cyber threat landscape, which can help them identify and prioritise risks. EWS facilitate collaboration by helping critical infrastructure providers to collaborate with each other and with LEA to share information and resources. EWS enhance incident response by providing them with access to information and resources.

Our contribution to theory is the understanding that EWS can greatly facilitate cyber incident information sharing between critical infrastructure operators and LEAs, and the contribution to practice is, despite only providing a preliminary take on the subject, opening a practical discussion on how critical infrastructure operators and LEAs can better collaborate in cyber incident management. Further study is recommended to gain a more in-depth discussion of the practical ramifications and real-world uses of EWS. These future studies should look to understand how different critical infrastructure operators and LEAs can use EWS as a common information-sharing environment in various practical settings and national structures that may differ between EU member states.

Sharing cyber incident information through early warning systems with critical infrastructure operators and law enforcement is a crucial step in improving cybersecurity. Overcoming challenges associated with trust, standardisation, legal frameworks, and technical integration can result in more effective threat detection, better incident response, and a more resilient society through collaboration. As cyber threats progress, so do our methods to safeguard the vital systems that support our way of life. The overall security and stability of an interconnected world are strengthened through this collaborative approach, which not only secures critical infrastructure but also enhances its overall security and stability.

ACKNOWLEDGEMENTS

This study has received funding from the European Union projects DYNAMO, under grant agreement no. 101069601 and CONNECTOR, under grant agreement no. 101121271. The views expressed are those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

Aaltola, K., & Taitto, P. (2019). Utilising Experiential and Organizational Learning Theories to Improve Human Performance in Cyber Training. *Information & Security: An International Journal*, *43*(2), 123–133. https://doi.org/10.11610/isij.4311

Alanazi, M., Mahmood, A., & Chowdhury, M. J. M. (2023). SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. *Computers & Security*, *125*, 103028. https://doi.org/10.1016/j.cose.2022.103028

Allianz Global Corporate & Specialty SE. (2023). *Allianz Risk Barometer: Identifying the major business risks for 2023* (p. 40) [Survey].

Almén, C., Hagström, N., & Rajamäki, J. (2022). ECHO Early Warning System as a Preventive Tool against Cybercrime in the Energy Sector. *Information & Security: An International Journal, 53*(1), 11–20. https://doi.org/10.11610/isij.5301

Amir, S., & Kant, V. (2018). Sociotechnical Resilience: A Preliminary Concept. *Risk Analysis*, *38*(1), 8–16. https://doi.org/10.1111/risa.12816

Cauffman, S. A. (2018). *Community resilience planning guide for buildings and infrastructure systems: Observations on initial implementations* (Guide

NISTIR 8229; p. 39). US Department of Commerce, National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8229.pdf

Cavelty, M. D. (2010). Cyber-Security. In *The Routledge Handbook of New Security Studies* (1st ed., p. 328). Routledge.

Conkova, M. (2013). Analysis of Perceptions of Conventional and E-Learning Education in Corporate Training. *Journal of Competitiveness*, *5*(4), 73–97. https://doi.org/10.7441/joc.2013.04.05

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining Cybersecurity. *Technology Innovation Management Review*, *4*(10), 13–21.

Cupiał, M., Szeląg-Sikora, A., Sikora, J., Rorat, J., & Niemiec, M. (2018). Information technology tools in corporate knowledge management. *Ekonomia i Prawo. Economics and Law*, *17*(1), 5–15.

DYNAMO project. (2024, May 9). *Dynamic Resilience Assessment Method including combined Business Continuity Management and Cyber Threat Intelligence solution for Critical Sectors'*. https://horizon-dynamo.eu/wp-content/uploads/2023/01/DYNAMO_Leaflet_web.pdf

ECHO project. (2021). *E-EWS and E-FCR demonstration surveys* (Deliverable D8.2; p. 25). European Commission. https://echonetwork.eu/wp-content/uploads/2022/03/ECHO-D8.2-E-EWS-and-E-FCR-demonstration-surveys_v1.0.pdf

Fitch Solutions. (2021). *South Africa Autos Report* (Country Industry Reports Q4; p. 1). Fitch Solutions. https://store.fitchsolutions.com/autos/south-africa-autos-report

Frisk, I., Tikanmäki, I., & Ruoslahti, H. (2022). Piloting the ECHO E-skills and Training Toolkit. *Information & Security: An International Journal*, *53*(2), 163–175. https://doi.org/10.11610/isij.5311

Ghafur, S., Graß, E., Jennings, N., & Darzi, A. (2019). The challenges of cybersecurity in health care: The UK National Health Service as a case study. *The Lancet Digital Health*, *1*(1), e10–e12. https://doi.org/10.1016/S2589-7500(19)30005-6

Hytönen, E., Rajamäki, J., & Ruoslahti, H. (2023). Managing Variable Cyber Environments with Organizational Foresight and Resilience Thinking. *International Conference on Cyber Warfare and Security*, *18*, 162–170. https://doi.org/10.34190/iccws.18.1.979

Hytönen, E., & Ruoslahti, H. (2023). A Lens to Examine Communication Through Business Continuity Management. In D. Verčič, A. T. Verčič, & K. Sriramesh (Eds.), *Public Relations and Sustainability* (pp. 205–216). University of Ljubljana: Faculty of Social Sciences. https://www.bledcom.com/

Im, T., Porumbescu, G., & Lee, H. (2013). ICT as a Buffer to Change. *Public Performance & Management Review*, *36*(3), 436–455. https://doi.org/10.2753/PMR1530-9576360303

Isidro-Filho, A., Guimarães, T. de A., Perin, M. G., & Leung, R. C. (2013). Workplace learning strategies and professional competencies in innovation contexts in Brazilian hospitals. *BAR - Brazilian Administration Review*, *10*(2), 121–134. https://doi.org/10.1590/S1807-76922013000200002

Kilani, Y. (2020). Cyber-security effect on organizational internal process: Mediating role of technological infrastructure. *Problems and Perspectives in Management*, *18*, 449–460. https://doi.org/10.21511/ppm.18(1).2020.39

Limnéll, J., Majewski, K., & Salminen, M. (2014). *Kyberturvallisuus*. Docendo.

Linkov, I., Bridges, T., Creutzig, F., Decker, J., Fox-Lent, C., Kröger, W., Lambert, J. H., Levermann, A., Montreuil, B., & Nathwani, J. (2014). Changing the resilience paradigm. *Nature Climate Change*, *4*(6), 407–409.

Linkov, I., Eisenberg, D. A., Plourde, K., Seager, T. P., Allen, J., & Kott, A. (2013). Resilience metrics for cyber systems. *Environment Systems and Decisions*, *33*, 471–476.

Luoma-aho, V., & Vos, M. (2010). Towards a more dynamic stakeholder model: Acknowledging multiple issue arenas. *Corporate Communications: An International Journal*, *15*(3), 315–331. https://doi.org/10.1108/13563281011068159

MARISA Project. (2017). *MARISA Grant Agreement number 740698*. European Commission.

Matveeva, A. (2006). *Early Warning and Early Response: Conceptual and Empirical Dilemmas* (Issue Paper 1; p. 66). European Centre for Conflict Prevention /International Secretariat of the Global Partnership for the Prevention of Armed Conflict. https://gppac.net/files/2018-12/Early%20Warning%20and%20Early%20Response.pdf

Michel, M. C. K., & King, M. C. (2019). Cyber Influence of Human Behavior: Personal and National Security, Privacy, and Fraud Awareness to Prevent Harm. *2019 IEEE International Symposium on Technology and Society (ISTAS)*, 1–7. https://doi.org/10.1109/ISTAS48451.2019.8938009

Mohammed, S. (2015). An Introduction to Digital Crimes. *International Journal in Foundations of Computer Science & Technology*, *5*(3), 13–24. https://doi.org/10.5121/ijfcst.2015.5302

Muthuppalaniappan, M., & Stevenson, K. (2020). Healthcare Cyber-Attacks and the COVID-19 Pandemic: An Urgent Threat to Global Health. *International Journal for Quality in Health Care*, *33*(1), 1–4. https://doi.org/10.1093/intqhc/mzaa117

National Cyber Security Centre. (2016). *Common Cyber Attacks: Reducing the Impact*. Crown Copyright. https://www.ncsc.gov.uk/guidance/white-papers/common-cyber-attacks-reducing-impact

O'Rourke, T. D., & Briggs, T. R. (2007). Critical Infrastructure, Interdependencies, and Resilience. *The Bridge*, 22–29.

Payne, B. K. (2020). Defining Cybercrime. In T. J. Holt & A. M. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 3–

25). Springer International Publishing. https://doi.org/10.1007/978-3-319-78440-3_1

Pirinen, R. (2015). Studies of Externally Funded Research and Development Projects in Higher Education: Knowledge Sources and Transfers. *Creative Education*, *6*(3), 315–330. https://doi.org/10.4236/ce.2015.63030

Pirinen, R. (2017). Towards Common Information Systems Maturity Validation: Resilience Readiness Levels (ResRL). *In Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *3*, 259–266. https://doi.org/10.5220/0006450802590266

Pöyhönen, J., Rajamäki, J., Ruoslahti, H., & Lehto, M. (2020). Cyber Situational Awareness in Critical Infrastructure Protection. *Annals of Disaster Risk Sciences : ADRS*, *3*(1), 0–0. https://doi.org/10.51381/adrs.v3i1.36

Rajamäki, J., Feyesa, A., & Nepal, A. (2024). View of E-EWS-based Governance Framework for Sharing Cyber Threat Intelligence in the Energy Sector. *Proceedings of the 23rd European Conference on Cyber Warfare and Security, ECCWS 2024*, *23*, 398–406. https://doi.org/10.34190/eccws.23.1.2073

Rajamäki, J., & Katos, V. (2020). Information Sharing Models for Early Warning Systems of Cybersecurity Intelligence. *31679*, *46*(2), 198–214.

Rajamäki, J., & Ruoslahti, H. (2018). Educational Competences with Regard to Critical Infrastucture Protection. *Journal of Information Warfare*, *17*(3), 415–423.

Ramaki, A. A., & Atani, R. E. (2016). A survey of IT early warning systems: Architectures, challenges, and solutions. *Security and Communication Networks*, *9*(17), 4751–4776. https://doi.org/10.1002/sec.1647

Ruoslahti, H. (2018). Co-creation of knowledge for innovation requires multi-stakeholder public relations. In *Public Relations and the Power of Creativity* (Vol. 3, pp. 115–133). Emerald Publishing Limited.

Ruoslahti, H. (2020). Business continuity for critical infrastructure operators. *Annals of Disaster Risk Sciences: ADRS*, *3*(1), 0–0.

Ruoslahti, H., Rajamäki, J., & Koski, E. (2018). Educational Competences with regard to Resilience of Critical Infrastructure. *Journal of Information Warfare*, *17*(3), 1–16.

Ruoslahti, H., & Tikanmäki, I. (2017). End-users Co-create Shared Information for a More Complete Real-time Maritime Picture: *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 267–274. https://doi.org/10.5220/0006559702670274

Ruoslahti, H., & Tikanmäki, I. (2022). Cybersecurity in Skills Development and Leadership. *Future-Proof Business - System Leadership Competences*. 3UAS: Future-proof Business - System Leadership Competences, Virtual conference.

Shoemaker, D., & Conklin, W. A. (2011). *Cybersecurity: The Essential Body of Knowledge*. Cengage Learning.

Simola, J. (2019). Comparative Research of Cybersecurity Information Sharing Models. *Information & Security: An International Journal*, *43*(2), 175–195. https://doi.org/10.11610/isij.4315

Simola, J., & Lehto, M. (2020). National Cyber Threat Prevention Mechanism as a part of the E-EWS. *The Proceedings of the International Conference on Cyber Warfare and Security*, 539–548. https://jyx.jyu.fi/handle/123456789/69054

Stanciugelu, I., Alpas, H., Florin, S. D., & Bozoglu, F. (2013). Perception and communication of terrorism risk on food supply chain: A case study (Romania and Turkey). In *Applied Social Sciences: Communication Studies* (pp. 189–196). Cambridge Scholars Publishing.

The International Chamber of Commerce. (2024, July). *Protecting the cybersecurity of critical infrastructures and their supply chains*. ICC. https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/ICC-2024_Protecting-the-cybersecurity-of-critical-infrastructures-and-their-supply-chains.pdf

Tikanmäki, I., & Ruoslahti, H. (2017). Increasing Cooperation between the European Maritime Domain Authorities. *International Journal of Environmental Science*, *02*. https://www.iaras.org/iaras/home/caijes/increasing-cooperation-between-the-european-maritime-domain-authorities

Tuttle, H. (2021, March). Ransomware Attackers Turn to Double Extortion. *Risk Management*, *68*(2), 8–9.

Vos, M. (2017). *Communication in turbulent times: Exploring issue arenas and crisis communication to enhance organisational resilience* (Vol. 40). Vos & Schoemaker.

Vos, M. (2018). Issue Arenas. In R. L. Heath & W. Johansen, *The International Encyclopedia of Strategic Communication* (1st ed., pp. 1–10). John Wiley & Sons. https://doi.org/10.1002/9781119010722.iesc0093

Vos, M., Schoemaker, H., & Luoma-aho, V. (2014). Setting the agenda for research on issue arenas. *Corporate Communications: An International Journal*, *19*(2), 200–215. https://doi.org/10.1108/CCIJ-08-2012-0055

Whitney, L. (2021, June 4). *Ransomware: A Cheat Sheet for Professionals*. TechRepublic. https://www.techrepublic.com/article/ransomware-cheat-sheet-everything-you-need-to-know/

Development Areas in Knowledge Management Processes in Social and Health Care Services: A Pilot Study

Nina Helander[1][a], Annamaija Paunu[1], Pasi Hellsten[1][b], Hannele Väyrynen[2][c] and Virpi Sillanpää[3][d]
[1]*Unit of Information and Knowledge Management, Tampere University, Tampere, Finland*
[2]*Independent Scholar, Espoo, Finland*
[3]*Modulight, Tampere, Finland*
pasi.hellsten@tuni.fi

Keywords: Knowledge Management, Development, Social and Healthcare Sector, Assessment.

Abstract: Knowledge management is about providing the right information for the decision maker at the right time, in the right format. This is equally essential in public sector as in the private sector enterprises. However, it is often easier said than done where and how the development schemes are to be directed. More data and information is needed. Public sector's social and health care has a number of data sources where knowledge management could make difference both in operational side, i. e. the care function and also in the management of the function, i. e. resource allocation. However, these are quite often planned some time ago and in the need of rethinking. The paper explores the possible points to be developed based on the knowledge management process, how to combine the two for a better outcome.

1 INTRODUCTION

Knowledge management (KM) is a holistic and systematic process that integrates technology and human aspects, enabling genuine dialogue in the management of organizations (Girard and Girard, 2015). In the process of KM, different stages can be distinguished in which information and further knowledge are processed from data (Choo, 2002). Knowledge is collected, organized, stored, shared, and utilized in a way that improves the decision-making and functioning of the individual and the organization (Vitt et al., 2002). However, there are several challenges in KM, due to, for example, technical infrastructure, poor quality data, human bias in thinking, reluctance to share knowledge, or inefficient practices and processes (Väyrynen et al., 2017). Thus, the possibilities to develop the operation are case-specifically many. It is plausible to assume that costs are one centric way to guide the flow of scarce resources for process development, especially in the public sector, where the funds are from a shared source and functions are regulated. Also there are often political pressures and ambitions in the way how the issues are to be approached. As the development towards improved efficiency but also 'better' outcome is concerned, the data regarding the costs are in a crucial role to direct development schemes. In retrieval of this knowledge a KM process will prove to be useful.

The aim of this paper is to identify the development areas at different stages of the knowledge management process in social and health care sector (SHC). The article seeks to answer the empirically weighted research question "What kind of development areas in knowledge management SHC operations can be identified at different stages of the knowledge management process model?" The context of the article is social and health care sector and the empirical research has been conducted through a pilot case study in that area. Through the analysis of the pilot case, the article contributes to knowledge management process literature. However, as the empirical case of this study is highly

[a] https://orcid.org/0000-0003-2201-6444
[b] https://orcid.org/0000-0001-7602-1690
[c] https://orcid.org/0000-0002-3636-280X
[d] https://orcid.org/0000-0003-4949-5203

Helander, N., Paunu, A., Hellsten, P., Väyrynen, H. and Sillanpää, V.
Development Areas in Knowledge Management Processes in Social and Health Care Services: A Pilot Study.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 356-361
ISBN: 978-989-758-716-0; ISSN: 2184-3228

specialized in nature, further generalizations based on the results must be done with caution.

The structure of the article is as follows: After the introductory chapter, Chapter 2 is followed by a concise theoretical overview of the basics of knowledge management. The focus of the article is on the presentation of empirical pilot case study and the analysis of the results, so the theoretical part is followed by Chapter 3, which presents the methodological choices and the pilot case study, and Chapter 4, which presents the main results of the case studies. Chapter 5 summarizes the results of the article and contains an evaluation of the study and presents topics for further research.

2 KNOWLEDGE MANAGEMENT PROCESS

The purpose of knowledge management (KM) is to collect, process, organize, store, share and utilize knowledge in a way that improves the decision-making and activities of the individual and the organization (Jääskeläinen et al., 2020). With KM we can identify our organization's tacit knowledge and combine it with explicit knowledge, also from external sources, thus supporting the organizational problem solving, decision making and strategic development. KM is a holistic and systematic process that integrates technology and human aspects (Valkokari and Helander, 2007).

Knowledge, which is ultimately used in human decision-making and action, is processed from data through information into knowledge by adding structure and meaning to it (Choo, 2002; Thierauf, 2001). This chain from data to knowledge emphasizes the need for data processing and the nature of enrichment. Thus, knowledge does not appear from scratch, but is created by enriching data and information. In order to make good, knowledge-based decisions, data and information must be of high quality and easily accessible in a form that is understandable to decision-makers as timely as possible (Vilminko-Heikkinen and Pekkola, 2017). To ensure the functioning of this chain, it is necessary to coordinate both the more technical side of KM and the more human side in organizational structures, processes, and practices (Jääskeläinen et al., 2020).

One practical tool for this knowledge processing is the so-called knowledge process model, which is comprised of following six phases: Knowledge need identification; Knowledge retrieval and creation; Knowledge maintenance and storage; Knowledge sharing; Knowledge use; and Measurement and learning.

The process usually begins with defining the knowledge needs. The knowledge needs are first identified so that they can later be met as well and efficiently as possible. At this stage, knowledge is obtained both from external sources (e.g. competitors and customers), and internal sources (e.g. information systems (IS) and communities of practice). The knowledge created and collected from different sources will be organized and maintained in the organization's repositories. This means the stage of analyzing and organizing the knowledge, which facilitates the next stages of the process, i.e. knowledge sharing. However, knowledge only gains its final value when it is used, for example, applied in decision-making and operational development, and when real changes occur in the organization's operations. By assessing the changes that have taken place and learning from the proceedings and thereby identifying new development needs, the cycle starts over. It should be noted that the process is in fact an iterative process and that the variation between phases is not always straightforward.

The process of KM presented above, and its various stages include both the technical and the human angle to knowledge (Jääskeläinen et al., 2019). The model may be criticized as too a mechanistic way to structure such a complex phenomenon as KM is. In practice it is often the case that some basic model is needed to identify the challenges of knowledge management and to develop best practices. Nor does applying the process model of knowledge management to practice mean focusing on the process rather than dialogue. On the contrary, dialogue can be better built into organizations when a clear framework is built. The different stages of the process include both technical and human aspects, thus the model serves as a good analytical tool for identifying the challenges and opportunities of knowledge management (Hellsten and Myllärniemi, 2019). The analysis of the stages in the process makes it possible to improve the understanding of public sector organizations of their own skills, knowledge and knowledge resources and to support the use of knowledge in problem solving, decision-making and strategic development, as stated in traditional KM literature (Dalkir, 2017; Grant, 1996; Hislop et al., 2018).

3 METHOD

Knowledge management (KM) process is studied through an empirical case from Finnish SHC. The aim was to develop a practical tool for assessing the cost-effectiveness of utilising a gerontechnological application in home care for the elderly (Colnar et al., 2020a). The starting point for the pilot was a previously developed cost-effectiveness assessment model (CEAM) (Sillanpää and Korhonen, 2022), which was to be tested and developed further. In the pilot planning, it was important to understand the purpose of home care, the services and operations, and where the expected impacts could be visible (for example, home care service processes or customer well-being). Secondly, it was necessary to find out where the data and information needed to assess the effectiveness of the gerontechnological application are located (e.g. the information system and the parties managing the data and information, or the entities producing data and information). A home care operating area that had been offering gerontechnology (medication dispensing robot) to its customers for a long time was chosen to act as a pilot. The home care administration selected the participants in the pilot.

The next step was to identify changes that occur in the implementation or in the use of a new gerontechnological application by the customer. To explore the changes, three participatory workshops were conducted in spring 2022 for 70 healthcare and home care professionals. Based on the results, a critical question emerged as to what the core information is needed to verify technology-influenced changes in the CEAM. In addition, consideration was given to how the necessary data could be collected from the organisation's various IS's. To find answers, seven semi-structured interviews were conducted with the management and professionals of the SHC support services. The interviews were recorded, transcribed and analysed using qualitative content analysis. After the interviews, more workshops were held in summer 2022 with appropriate interviewees to find out how to refine the data and information needed. The next step was to define the indicators for the evaluation and to develop a CEAM together with researchers and city coordinators (financial planning representative, report designer and telecare technology expert). The pilot dealt with client and health data in home care for the elderly, so special attention was paid to the information security of technology and personal data; researchers were provided with anonymized data from which individual participants could not be identified. (Acosta et al., 2022)

Indicators identified and utilised in the literature in assessing the cost-effectiveness of digital remote monitoring services include hospital days, clinical visits in primary or specialised health care, and the number and costs of first aid visits. In home care, cost indicators may include home visits by healthcare professionals (nurses), number of contacts, and direct costs of remote monitoring or remote access (e.g. technology rental costs, user fees, Internet access costs) (Polisena et al., 2009; Seto, 2008; Upatising et al., 2015). To identify the effects of digital services, in addition to direct costs, it is necessary to identify the process requirements for new home care practices and the costs related to process changes (Askedal et al., 2017).

In the pilot, the indicators to be evaluated in the CEAM were identified from workshop results, interviews and literature, as well as from previous research by Sillanpää & Korhonen (2022). The researchers formulated guidelines for the use of the CEAM, and the model was ready to be piloted in home care for the elderly on the use of a medication dispensing robot. The pilot was carried out with data produced by the City, which made the actual data and reports on the use of home care services, the number of users of the technology and costs available for the whole year. The pilot consisted of ten users of medication dispensing robots (hereinafter referred to as "medicine dispensing robot in use") and ten respondents of not using the medicine dispensing robots (hereinafter referred to as "medicine dispensing robot in NOT use). Customers with the same level of functional capacity were invited to participate in the pilot in both control groups medicine dispensing robot in use/medicine dispensing robot NOT in use), the selection was not based on defined medical diagnoses. The CEAM makes it possible to define a follow-up period (e.g. quarterly), which makes it possible to obtain information on the development of service production, operations and costs by means of visualization (e.g. graphic images) with the data of the monitoring period.

4 EMPIRICAL RESULTS

The basic idea of the model is to compare the cost-effectiveness of a technology-assisted service for two elderly home care customer groups, a medicine dispensing robot: a medication dispensing robot for groups in use and a medicine dispensing robot not in use. The costs were assessed from the service provider's point of view, and customer fees were excluded from the CEAM. The template is based on a spreadsheet program that contains three separate pages. The first page contains instructions on how to use the model: instructions to support the use are recommended so that the indicators to be evaluated in the model are understood and how the measurement is utilized (Jääskeläinen et al., 2020, 2013). The instructions are important for the measurement and the calculation principles of the background data to be the same regardless of the time of measurement or the person using the model. In addition, the page contains the indicators of the model, the sources of data or other information used (quantitative, qualitative, and financial information), instructions for calculating costs, the service provider's service-specific price, and the persons responsible for the information needed for the meters, which is used on page two of the spreadsheet program.

There is also a form to which data and information from the organization's IS's were added (e. g. McKlin et al., 2001), in the pilot the data were entered manually. The data and information needed to implement the pilot, instructions for obtaining the necessary information, and ownership of the data and information had to be defined. The indicators to be evaluated and entered were the use of home care services/number of clients (searchable from the patient IS), the number of home visits in physical home care and the costs of visits (to be retrieved from the service provider's financial administration). The home care did not separately record remote visits during the review period, but the remote visit was recorded as a physical visit, so remote home care visits as a separate examination were excluded from the pilot. Indicators were also the number and cost of emergency calls and the number of hospital days. Resource-based information from home care staff is important to illustrate how technology can support resources and reduce peak-hour physical home care home visits. Therefore, home care visit times had to be classified and the number of visits in the time frame had to be manually entered in the table. The operating costs of the medicine dispensing robot were classified into two categories: direct and indirect costs. Direct costs derive from the price of the equipment (e.g. lease or rental costs) and general costs of service activities (e.g. general IT costs). Indirect costs include technology support services for both home care staff and clients: training costs related to the use and administrative costs of the technology used by the customer (e.g. customer and service production data), as well as the number of service-related calls or contacts.

The number of clients working with home care services is essential information, and the customer months during the year may be of different lengths in the control group (e.g. not all clients are necessarily home care customers for the entire follow-up period, which was one calendar year). For this reason, the customer months had to be made commensurate with the calculation formula in the spreadsheet so that the number of customer months in the two groups being compared is proportionate.

The aim of the pilot was to test the developed CEAM with the actual service usage and cost data of the home care service provider for the elderly, and to obtain preliminary results on the cost-effectiveness of the medication dispensing robot in home care for the elderly. The CEAM provides a tool for assessing the economic impact of the use of technology. However, evaluation is a combination of several indicators and the context of the evaluation influences the set indicators to be evaluated, and therefore the limitations of the model must be taken into account when considering the results. In general, technology is expected to have an impact on the well-being of the elderly and the well-being of employees at work. In the pilot, qualitative impacts were identified (e.g. RAI (Resident Assessment Instrument) system for mapping of clients' functional capacity and service needs, 15D indicator for assessment of the effectiveness of treatment, but the qualitative impacts were not measured in this pilot. The CEAM makes it possible to measure the economic and quantitative impacts of the use of technology, but qualitative assessment of client and home care personnel is also needed.

5 DISCUSSION AND CONCLUSIONS

Making the knowledge management process to work requires seamless cooperation between the technical infrastructure and human factors, so that knowledge management can truly create value for public administration actors and citizens (Jääskeläinen et al., 2019). This means also simultaneously know-how requirement for the personnel and in our case also for the other side of the table, the customers as they are recently called. In addition to this, the process is advised to be familiar for the decision makers as the phases while interconnect often have their own features and actions that can offer improvement. In summary both the overall picture is needed but also the individual phases need to be known. KM process model offers detailed enough picture to be used for further development targeting and possibly even schemes on issues in the practice.

Much of the previous public sector knowledge management research has focused on issues related to either IS's, data quality, or specific aspect of knowledge management. Less frequently, the various stages of the entire knowledge management process have been illustrated, from information needs to information acquisition, analysis and sharing, and its utilization to offer basis for process development. In this paper, we have sought to look at this whole process of knowledge processing and, through the pilot case study, have identified development areas of knowledge management in SHC. The key empirical findings concerning each of the KM process phases are illustrated in next Table 1.

Table 1: Key empirical findings.

KM process phase	Development area
Knowledge need identification	Identification of qualitative data indicators, not only quantitative. On both clients and personnel.
Knowledge retrieval and creation	Permits process, various IS's, data comparability, automation needed to retrieve data from different IS's, newer data sources
Knowledge maintenance and storage	Automated systematic identification of customer relationship classification, data accessibility
Knowledge sharing	Data acquisition and information sharing between the parties, esp. decision-makers, policies and procedures on the matter.
Knowledge use	Ownership, management and development must be designated.
Measurement and learning	Development and usage of the CEAM must be strategic, defined key performance indicators

The paper emphasizes the utilisation of data and information, especially in SHC and home care for the elderly, by showing that data can create value for organisations and employees if relevant data and information are identified from data flows, who in the organisation manages the data or owns the data, and also for what purpose the data and information is used and who benefits from the data and information. After identifying these elements, the CEAM described in this article can provide visualized information to support decision-making, and the information can guide the actors in the organization to react to situations in their daily operations.

Although this study succeeded in presenting one example of a CEAM with illustrative results, the study has some limitations. First, the sample of the pilot study was small and limited to a specific location and a specific, individual gerontechnology in Finland. Secondly, home care for the elderly is publicly funded in Finland, although clients are charged for services. For this reason, the results cannot be widely generalised because, for example, the different wellbeing areas may have different service fees and the provision of services at different prices, or because the model may have to be modified, for example, for each wellbeing services sector (e.g. indicators defined differently regionally). In addition, the results cannot be compared with private sector healthcare service production, such as health insurance -based service production. In the future, research with a larger sample and the context of home care for older people at national level with uniform indicators will be needed to formulate larger picture of the development ideas. In future studies, it should also be taken into account that home care clients may use several gerontechnologies, and it would be important and interesting to be able to assess the use of several gerontechnologies and their combined impact on the client's health and well-being as well as the service provider's operating costs. However, this requires, among other things, the automatic identification of customer relationships from IS's, which has been included in the development proposals. The aim of this study was to develop the previous CEAM towards a more extensive cost-effectiveness assessment and to implement the evaluation in practice. The pilot identified several qualitative effects of technology on both home care clients and employees but measuring them at this stage proved challenging. However, the model can be used to illustrate changes in operational capacity (RAI values, number of emergency visits), and thus the customers' ability to function in everyday life has been included in the model. More research is needed to add qualitative indicators alongside the economic factors of the CEAM (Colnar et al., 2020b; Colnar and Dimovski, 2020; McKlin et al., 2001). However, it would be at least as important to raise a debate on the issues of principle, such as policies, rules of the game, and processes that should change to enable genuine knowledge management. Moreover, it should always be borne in mind that knowledge management is not an end, but only a tool - the goal is to create value together and sustainably better than before.

REFERENCES

Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J., 2022. Multimodal biomedical AI. Nat. Med. 28, 1773–1784.

Askedal, K., Flak, L.S., Solli-Sæther, H., Straub, D.W., 2017. Organizational Learning to Leverage Benefits Realization Management; Evidence from a Municipal eHealth Effort, in: Janssen, M., Axelsson, K., Glassey, O., Klievink, B., Krimmer, R., Lindgren, I., Parycek, P., Scholl, H.J., Trutnev, D. (Eds.), Electronic Government, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 142–153. https://doi.org/10.1007/978-3-319-64677-0_12

Choo, C.W., 2002. Information management for the intelligent organization: the art of scanning the environment. Information Today, Inc.

Colnar, S., Dimovski, V., 2020. Knowledge management in social work: Management support, incentives, knowledge implementation, and employee empowerment. Econ. Bus. Rev. 22, 4.

Colnar, S., Dimovski, V., Grah, B., Rogelj, V., Bogataj, D., 2020a. Gerontechnology: literature review and research agenda, in: 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). IEEE, pp. 391–396.

Colnar, S., Dimovski, V., Grah, B., Rogelj, V., Bogataj, D., 2020b. Smart home supporting integrated health and care services for older adults in the community: literature review and research agenda, in: 2020 24th International Conference on System Theory, Control and Computing (ICSTCC). Ieee, pp. 526–531.

Dalkir, K., 2017. Knowledge management in theory and practice. MIT press.

Girard, J., Girard, J., 2015. Defining knowledge management: Toward an applied compendium. Online J. Appl. Knowl. Manag. 3, 1–20.

Grant, R.M., 1996. Toward a knowledge-based theory of the firm. Strateg. Manag. J. 17, 109–122.

Hellsten, P., Myllärniemi, J., 2019. Business Intelligence Process Model Revisited., in: KMIS. pp. 341–348.

Hislop, D., Bosua, R., Helms, R., 2018. Knowledge management in organizations: A critical introduction. Oxford university press.

Jääskeläinen, A., Laihonen, H., Lönnqvist, A., Pekkola, S., Sillanpää, V., Ukko, J., 2013. Arvoa palvelutuotannon mittareista.(In Finnish)

Jääskeläinen, A., Sillanpää, V., Helander, N., 2019. A model for profiling information and knowledge management in the public sector, in: International Forum on Knowledge Asset Dynamics.

Jääskeläinen, A., Sillanpää, V., Helander, N., Leskelä, R.-L., Haavisto, I., Laasonen, V., Torkki, P., 2020. Designing a maturity model for analyzing information and knowledge management in the public sector. VINE J. Inf. Knowl. Manag. Syst.

McKlin, T., Harmon, S.W., Evans, W., Jones, M.G., 2001. Cognitive presence in Web-based learning: A content analysis of students' online discussions.

Polisena, J., Tran, K., Cimon, K., Hutton, B., McGill, S., Palmer, K., 2009. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes. Metab. 11, 913–930. https://doi.org/10.1111/j.1463-1326.2009.01057.x

Seto, E., 2008. Cost Comparison Between Telemonitoring and Usual Care of Heart Failure: A Systematic Review. Telemed. E-Health 14, 679–686. https://doi.org/10.1089/tmj.2007.0114

Sillanpää, V., Korhonen, T., 2022. Terveysteknologian kustannusvaikutusten arviointi: sydämen vajaatoiminnan kotihoito Tampereella, in: Kohti Vaikuttavaa Sosiaali-Ja Terveyspalveluiden Yhteiskehittämistä. pp. 35–46. (In Finnish)

Thierauf, R.J., 2001. Effective business intelligence systems. Greenwood Publishing Group.

Upatising, B., Wood, D.L., Kremers, W.K., Christ, S.L., Yih, Y., Hanson, G.J., Takahashi, P.Y., 2015. Cost Comparison Between Home Telemonitoring and Usual Care of Older Adults: A Randomized Trial (Tele-ERA). Telemed. E-Health 21, 3–8. https://doi.org/10.1089/tmj.2014.0021

Valkokari, K., Helander, N., 2007. Knowledge management in different types of strategic SME networks. Manag. Res. News.

Vilminko-Heikkinen, R., Pekkola, S., 2017. Master data management and its organizational implementation: An ethnographical study within the public sector. J. Enterp. Inf. Manag. 30, 454–475.

Vitt, E., Luckevich, M., Misner, S., Corporation (Redmond), M., 2002. Business intelligence: Making better decisions faster. Microsoft Press Redmond, WA.

Väyrynen, H., Helander, N., & Vasell, T., 2017. Knowledge management for open innovation: comparing research results between SMEs and large companies. International Journal of Innovation Management, 21(05), 1740004.

Transforming Knowledge Management Using Generative AI: From Theory to Practice

Dmitry Kudryavtsev[a], Umair Ali Khan[b] and Janne Kauttonen[c]
Haaga-Helia University of Applied Sciences, Ratapihantie 13, Helsinki, Finland
{dmitry.kudryavtsev, umairali.khan, janne.kauttonen}@haaga-helia.fi

Keywords: Knowledge Management, Generative AI, Knowledge Processes.

Abstract: Generative AI is revolutionizing the way people and companies create, capture and access knowledge. This study is driven by problems and new opportunities related to knowledge work. We identify, organize, and prioritize generative AI use cases for knowledge management. Our analysis of business needs and in-depth interaction with companies is the main data source used to create insights into this study. In addition to the use cases, the research highlights the challenges of using generative AI for knowledge management and existing research tasks. Creating a reusable toolkit for Generative AI-enhanced knowledge management is proposed as the next step of applied research to address the identified use cases and challenges.

1 INTRODUCTION

Businesses have long recognized the importance of leveraging internal knowledge for success, and this has led to significant investment in this area in the past few decades. While precise and quick access to business knowledge has been a priority, knowledge management (KM) remains a labor-intensive and complex challenge. Despite several tools and technological advancements, companies still struggle to locate the right information at the right time. They often need to sift through various fragmented internal knowledge sources, including document management systems, databases, note-taking tools, internal repositories, file-sharing platforms, and emails (Khan et al., 2024). Knowledge workers spend too much time on information searches or duplicate efforts by recreating existing information. This situation leads to inefficiencies, increased operational costs, and delayed project timelines.

While the sudden rise of generative AI (GenAI) was unprecedented and left many surprised, its potential to transform the field of KM was seen in the very beginning through capturing, creating, accessing, and streamlining knowledge governance processes (Murphy, 2023). Nowadays, GenAI's role in KM is also explored from an academic point of view (Alavi et al., 2024; Benbya et al., 2024; Pimentel & Veliz, 2024).

The rapid progress in NLP (NLP), particularly through Large Language Models (LLMs), provides businesses with innovative methods to interact with vast amounts of structured and unstructured data. GenAI models enable users to engage with data through conversational interfaces, asking long-tail queries in natural language and receiving precise answers from large document sets. This has steered the current research to existing KM problems and new opportunities related to the way businesses create, capture, and access knowledge. GenAI showed potential to increase the productivity of knowledge work in various areas. For example, according to the US National Bureau of Economic Research, GenAI systems can increase productivity by 14% on average (as measured by issues resolved per hour), and by 34% for novice and low-skilled workers (Brynjolfsson et al, 2023).

Despite GenAI's immense potential, businesses face significant challenges in adopting GenAI for KM. These challenges are caused by a lack of knowledge and technical skills in the rapidly moving field of GenAI. While businesses recognize the potential of GenAI to empower employees and

[a] https://orcid.org/0000-0002-1798-5809
[b] https://orcid.org/0000-0002-8560-764X
[c] https://orcid.org/0000-0003-2432-5072

Kudryavtsev, D., Khan, U. and Kauttonen, J.
Transforming Knowledge Management Using Generative AI: From Theory to Practice.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 362-370
ISBN: 978-989-758-716-0; ISSN: 2184-3228

enhance customer experience, the diverse range of GenAI solutions and the complexity of AI integration based on business size and type present significant challenges. Businesses struggle to leverage AI effectively due to concerns over data privacy, confidentiality, operations transparency, and the accuracy of GenAI solutions. While several existing GenAI solutions, such as Microsoft Co-pilot, attempt to address the challenge of precise knowledge access across scattered business documents, they often fall short in providing a tailored, customized, and holistic approach that meets modern businesses' unique and rapidly evolving needs. These solutions typically offer broad functionalities but lack the specificity required to handle the nuanced and domain-specific KM needs of businesses.

In addition to these issues, businesses also face difficulties in developing and deploying complex GenAI solutions due to a lack of expertise and uncertainty about the appropriate level of AI integration. While they see the substantial potential of GenAI, they are often indecisive and/or lack the necessary resources to implement these technologies effectively. This gap results in underutilized data, inefficiencies, and missed opportunities for leveraging information to drive business growth and innovation

Despite the growing potential of GenAI for business, its right application remains a challenge. The use of GenAI can even reduce productivity if applied incorrectly or to inappropriate tasks (Dell'Acqua et al, 2023). To obtain benefits from the application of GenAI and mitigate the respective risks, businesses must navigate a complex landscape of technologies and best practices for their implementation. The selection, integration, and deployment of specific GenAI technologies vary based on use cases, industry, the company’s AI readiness, integration levels, company-specific knowledge access needs, and several other critical issues.

To develop a GenAI-enhanced KM solution, we performed a structured study in this paper, which analyzes the KM requirements of businesses and identifies the most relevant use cases. Subsequently, we investigate the development of a GenAI-KM toolkit that can be used to address these use-cases. We performed this study in collaboration with several companies and research institutes using a co-creation approach in which the companies underwent a detailed need analysis process. The research questions investigated in this study are as follows:

RQ1. How GenAI can transform KM? What are the GenAI use cases for KM?
RQ2. What are the challenges of using GenAI for KM, and how can they be addressed?
RQ3. What are the research tasks associated with applying GenAI for KM?
RQ4. How to develop GenAI-enhanced KM solutions?

The remainder of this paper is structured as follows: Section 2 explores how GenAI can transform KM by identifying, organising and prioritising corresponding use cases. This section combines theory-enhanced with industry-focused perspectives, data collection, analysis, and synthesis process are disclosed in the methodology part of section 2. Section 3 explores the challenges of using GenAI in KM and, specifically, in tackling the identified use cases. Section 4 outlines key research topics that must be addressed for effective implementation. Section 5 discusses a reusable GenAI toolkit as a potential practical solution. Section 6 concludes the paper with future work.

2 GENAI-BASED USE CASES FOR KNOWLEDGE MANAGEMENT

2.1 Research Methodology

To define, organize, and prioritize GenAI use cases for KM, we followed 3 steps (see Fig. 1).

Step 1: The research presented in this study is driven by our experience in AI and KM by observing the needs and requirements of companies. The AI needs analysis and consultancy offered by Haaga-Helia University of Applied Sciences to 60+ companies as part of the Finnish AI Region (FAIR) EDIH project[1] shed light on the existing opportunities of GenAI for businesses and also revealed several challenges that companies face in adopting GenAI solutions. Then it became clear that addressing these challenges necessitates a comprehensive study both from academic and industry perspectives.

Step 2: We identified and organized GenAI use cases based on the literature review. The result was an organized list of generic GenAI use cases (knowledge services). Since a dialogue with businesses is needed to define the specific use cases, the generic use cases were associated with specific

[1] https://www.fairedih.fi

use cases from various application areas (sales and marketing, customer service, etc.).

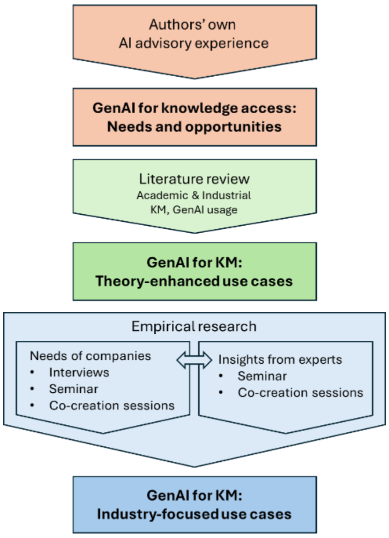


Figure 1: Methodology adopted to identify, organize, and prioritize the GenAI use cases for KM.

Step 3: The third step aimed to elaborate the original description of GenAI use cases for KM via empirical analysis of companies' needs and prioritize them. The main questions in this step included: "*What are the most needed GenAI-based knowledge services, and how to formulate them in a business-understandable way*?". Interaction with companies using different data collection and co-creation methods was the main approach here. Experts (e.g. from consulting) were also involved here to extend the list of generic use cases and define the most/least popular ones.

2.2 Theory-Enhanced List of Use Cases

The description of GenAI's role in knowledge management was inspired and based on the work of Alavi et al (2024), where the authors examined how GenAI impacts the processes of knowledge creation, storage, transfer, and application, highlighting both the opportunities and challenges this technology presents. Various academic papers, research reports and business articles, such as (Murphy, 2023), were used to synthesize the list of use cases.

Since the aim was to specify KM-related business needs in a way that would help to design and develop GenAI-based solutions, we decided to formulate the role of GenAI in KM in the form of knowledge services (Maier et al., 2009) arranged around knowledge processes. Such a service-based approach helps business users easily understand the functionality of KM solutions and gives IT specialists an understanding of what information systems should do. We named these GenAI-based services as generic ones to differentiate them from specific ones in various domains (content synthesis services vs. customer proposal generation). Such generic services allow us to see similarities across application domains and create reusable solutions that can be used in various application areas. These knowledge services correspond to activities, behaviors, or means supporting knowledge conveyance and transformation processes in ISO 30401:2018 (Kudryavtsev, Sadykova, 2019; Carlucci et al, 2022). The theory-enhanced GenAI-based services (use cases) for the main knowledge processes:

Knowledge Creation Services (Use Cases)

1. **Idea Generation Service:** An AI service that generates new ideas, concepts, or solutions based on the analysis of existing data, trends, and patterns. It aids in sparking creativity and innovation.
2. **Content Synthesis Service:** This service automates the synthesis of new knowledge content by integrating and reinterpreting existing information, research findings, and data analyses, generating comprehensive reports, articles, or research papers.
3. **Learning and Reflection aid Service:** This service provides contextual examples and in-depth explanations that help in understanding complex concepts, fostering a deeper level of creativity and innovation.

Knowledge Capture and Organization Services (Use Cases)

1. **Knowledge Capture Service:** This service helps knowledge workers to externalize their knowledge in the most convenient and easy way (e.g. by making notes, voice recording, or screencast) and convert this information into a suitable format for further processing (e.g. by transcribing audio into text).
2. **Automated Categorization and Tagging Service:** This service uses AI to organize, tag, and categorize knowledge resources efficiently, enhancing the structure of knowledge repositories and making it easier to store vast amounts of information in a searchable and accessible manner.
3. **Knowledge Curation Service:** Proactively manages the knowledge base by identifying and archiving content that has become obsolete or less relevant, ensuring the repository stays current and valuable. Simultaneously, it highlights areas in need of fresh insights or updates, guiding the

continuous growth and evolution of organizational knowledge.

Knowledge Access Services (Use Cases)

1. **Conversational Interface (Chatbot):** This service focuses on interacting with users through a conversational interface, employing NLP to understand and respond to queries in a natural, engaging manner. It's designed to facilitate easy access to information by answering questions, providing explanations, and guiding users through complex datasets or knowledge repositories, making the process more intuitive and user-friendly
2. **Advanced/Enhanced Search Service:** This service enhances traditional search functionalities by incorporating sophisticated NLP algorithms. It allows users to input search queries in their own words, including complex questions and conversational phrases, and employs AI to interpret these queries accurately. The service searches through extensive databases and knowledge repositories to find precise, relevant answers and information, significantly improving the efficiency and effectiveness of knowledge retrieval.
3. **Personalized Knowledge Feed Service (Tailoring knowledge):** Delivers customized knowledge recommendations to users based on their roles, interests, and past interactions, ensuring they receive the most relevant and timely information.
4. **Expert Search and Recommendation:** Utilizes AI to quickly identify and recommend experts within the organization based on specific queries or project needs. By analyzing employees' contributions, skills, and historical project involvement, this service facilitates efficient collaboration and knowledge sharing among team members.

Domain-specific examples complemented the synthesized list of generic GenAI-based services (use cases), as they were necessary for communicating with companies. This level of generalisation is much more apparent to them.

2.3 Data Collection, Analysis and Co-Creation

The main focus of the research was on step 3, especially the analysis of business needs. 20+ companies from different sectors participated in interviews. In addition to interviews, we organized an awareness-raising seminar in May 2024 in Helsinki, Finland – "Generative AI-Enhanced Knowledge Management in Business". Within this seminar, we presented our vision of GenAI use cases for KM and collected feedback from participants (representatives of various companies), including their own use cases. The panel discussion was a part of this seminar and included AI and KM experts from the industry. This discussion helped to extend the list of GenAI use cases for KM and organize them. The summary of the panel discussion was published in the article (Khan, Kudryavtsev, 2024). 20+ onsite participants and 20+ online participants attended the seminar.

The most detailed analysis, description, and prioritization of the use cases were done together with 6 Finnish companies that participated in co-creation sessions: Company 1: ~60 persons, electric/automation/energy; Company 2: ~120 persons, demolition & hazard. material removal; Company 3: ~10 persons, IT/software development/automation; Company 4: ~360 persons, metal manufacturing; Company 5: ~740 persons, IT/Accounting/HR; Company 6: ~10 persons, healthcare.

They facilitated in-depth need analysis by providing detailed information about their processes, challenges, and expectations.

Simultaneously, we contacted research organizations specialising in KM, GenAI, NLP, and other relevant fields to incorporate cutting-edge insights and methodologies into the analysis of use cases and corresponding challenges. The University of Helsinki and Tampere University expressed strong interest in the study and contributed their in-depth expertise in LLMs and KM. Co-creation with these academic partners provided additional perspective during the formulation of use cases.

2.4 Industry-Focused Prioritized List of Use-Cases

The results of the elaboration and prioritization of use cases are presented in Table 1.

The three most needed generic use cases were defined:

1. AI assistance for talking to the company's data and systems ("ChatGPT" for the company's own documents/data) *to support the knowledge access process;*

2. Report and document creation assistance (project reports, business proposals/offers) *to support the knowledge creation process.*

3. Speech-to-structured document conversion *to support the knowledge capture process.*

These generic use cases generalize company-specific use cases.

Table 1: The industry-focused prioritized list of the most needed GenAI use cases for KM.

Knowledge process	Generic use cases	Company-specific use cases
Knowledge access	1. AI assistance for talking to the company's data and systems ("ChatGPT" for the company's own documents/data)	• Flexible chatbot-type search functionality for a collection of company-specific manuals and documentation • Conversational agent for dental video learning content • GenAI-based assistant on top of the existing scheduling/optimization software (platform).
Knowledge creation	2. Report and document creation assistance (project reports, business proposals/offers)	• Sales order generation assistant • Reports generation assistant • Customer proposal/quotation generation assistant, including the proposal generation assistant for renewable energy systems • Write summary reports from project data (could be of multiple types, e.g., incidents or overall view) • Generate easy-to-read summaries of projects for customers (instead of complex tables and diagrams) • Verbal interpretation of financial reports using a wide range of language models
Knowledge capture	3. Speech to structured documents conversion	• Automated incident reporting and analysis for safety and operational transparency to generate reports based on predefined templates and even interact with systems via text or phone calls to create reports • Extraction of Finnish transcripts from video learning content and their precise translations into English and Scandinavian languages • Generate diaries from speech without typing

3 THE CHALLENGES OF USING GENAI FOR KM AND HOW TO TACKLE THEM

To understand the challenges of implementing the identified use cases, we have combined a literature review with an empirical qualitative approach – the panel discussion with AI and KM experts (Khan, Kudryavtsev, 2024). We learned that companies face various challenges when transitioning to generative AI-based systems for KM. These are both technical and related to practical and management work. GenAI systems, which rely on probabilistic models rather than rule-based processes, differ fundamentally from traditional knowledge systems, leading to potential compatibility issues during integration. LLMs' ability to process and understand vast amounts of information makes them suitable for tasks closely aligned with knowledge management principles, but these models are also less predictable. One of the primary challenges lies in aligning these dynamic AI-driven models with existing knowledge infrastructure, which is often structured and based on pre-defined rules. This paradigm shift requires organisations to rethink their entire KM infrastructure, which can involve costly overhauls of content management systems and staff retraining. Generative AI solutions also differ from traditional AI models (e.g., classification, regression, or clustering), where inputs and outputs remain well-defined, and models can be trained and owned in-house.

There are many competing technologies for implementing GenAI in organizations; however, as the technology is still emerging, the situation is different from mature technologies. The selection, integration, and implementation of the specific GenAI technologies is not straightforward and depends on the use cases, industry, company's AI readiness, and level of integration. Specialized expertise in GenAI is needed to navigate this multitude of opportunities. At the moment GenAI relies heavily on foundation models from a few major technology providers (e.g., OpenAI, Google, and Microsoft) that own the most capable models. These models must be accessed via cloud service (API) and are closed source. While open-source models are also available, they remain less capable and weaker for tasks that require state-of-the-art reasoning skills and large contextual information capacity. Companies must balance between capability, usage costs, latency, geographical availability, and data storage constraints. Particularly, SMEs have limited IT resources and AI implementation maturity, which creates a dependency on technology providers. While large organizations may have the resources to integrate GenAI into their KM systems effectively, SMEs may struggle with the costs associated with cloud-based AI services from major providers (Ghimire et al., 2024). These businesses must weigh the trade-offs between GenAI's powerful capabilities and the high costs, potential latency issues, and geographical restrictions associated with cloud-based models (Kaczorowska-Spychalska et al., 2024).

The main technical challenges are related to change management and integrating the new AI system with existing systems. This process can be complex and requires a strategic overhaul of current content management practices. The amount of content generation sources and channels is growing, and there is a need to deliver relevant content to various users at the right time and place. Companies might need to rebuild their content management systems to provide consistent and relevant content for various audiences. This rebuild is necessary to ensure that the content is up-to-date and valuable for different user groups. One of the primary challenges is the need to align the new AI capabilities with the established workflows and data structures in data infrastructure. Traditional KM systems are often designed around structured data and rule-based processes, whereas Generative AI relies on unstructured data and probabilistic models, which can lead to compatibility issues during integration. Organizations must adapt or overhaul their existing systems to ensure seamless interaction between these different paradigms. This requires extensive retraining of staff, who may be unfamiliar with the dynamic nature of AI-driven systems, resulting in potential resistance to adoption (Wu et al. 2024).

Users need to trust the responses generated by chatbots or other GenAI tools when making critical decisions. Organizations, therefore, need to implement strong validation processes and techniques such as retrieval-augmented generation (RAG) which are gaining popularity to ensure that responses are grounded in verified, organizational-specific knowledge sources (Earley, 2023). Businesses must navigate complexities related to data privacy, intellectual property, and ethical AI use when establishing governance frameworks for generative AI (Ghosh & Lakshmi, 2023). It is crucial to define clear policies regarding data access, ensuring that only authorized personnel can use AI tools and handle sensitive organizational knowledge (Ferrari, et al, 2023). A significant challenge lies in balancing innovation with compliance, as organizations must align their AI usage with global regulations like the GDPR or AI Act, particularly regarding data processing and consent management (Michael et al, 2024). Moreover, implementing robust audit trails and transparency protocols to monitor AI-generated content is essential to ensure traceability and accountability, minimizing the risks of biased or misleading outputs (Pimentel & Veliz, 2024). Continuous oversight is necessary to prevent "model drift," where AI systems gradually produce less accurate or aligned outputs over time, potentially leading to business risks (Reuel & Undheim, 2024). Finally, organisations must prioritise educating employees on the ethical implications of AI use, fostering a culture that promotes responsible AI utilization and ongoing governance adjustments to address evolving legal and technological landscapes (Earley, 2023).

The main problems with GenAI, particularly LLMs, are related to hallucination, bias, and limited real-world understanding (Naveed et al., 2023).

Ensuring the privacy of a company's information is the first and foremost concern highlighted by the companies. This requires secure mechanisms and regulatory sandboxes, as emphasized by the AI Act (European Commission, 2024) and compliance with GDPR.

Scalable methods for acquiring, transforming, and integrating diverse kinds of input data are required. Hence, incremental updates are required that can either periodically be performed in a batch-like manner or in a more dynamic, streaming-like fashion. (Hofer et al., 2024)

GenAI-based KM requires high-quality data. The difficulty of this task grows with the rising number and heterogeneity of data sources, especially if one relies on automatic data acquisition and data integration. This requires data cleaning at major steps in the construction pipeline so that the degree of dirty or wrong information entering the system is limited (Hofer et al., 2024).

Domain-specific knowledge models such as knowledge graphs (KG) hold good potential to address the limitations of LLMs (Deloitte 2023). Integrating KGs as a potential component may require constructing KGs from diverse data. LLMs are being increasingly used to construct KGs. While automated pipelines for KG construction have shown impressive performance, they are still prone to LLM's inherent limitation of hallucination and biases. At the same time, LLMs have also been found to show inferior performance for domain-specific knowledge (Ghanem et al., 2024), for instance, generating KG for a disease dataset, as they have been trained on generic datasets. These limitations can result in the omission of crucial entities and relationships, potentially leading to misinformation. Quality control measures must be used to ensure the reliability of the knowledge represented in KGs.

Feedback from domain experts is crucial to ensure the accuracy and dependability of the causal relationships within the data (Zhou et al., 2024). Cleaned and verified data can be then used in building a knowledge graph.

4 KEY RESEARCH AREAS FOR DEVELOPING A GENAI-KM SOLUTION

We identify several research topics that need to be investigated to meet the business KM requirements through a holistic GenAI-KM solution. These topics are important to explore innovative approaches, design scalable solutions, and overcome the technical and practical challenges associated with implementing GenAI in KM.

Identification of knowledge Tasks: Identifying the specific knowledge tasks that can benefit from GenAI is essential for maximizing its value in business applications. It is important to investigate where GenAI can automate or enhance these processes in a meaningful way, tailored to individual business needs.

Developing Efficient Data Pipelines for LLM Integration: The effectiveness of LLMs depends on the quality and consistency of the data they process. Building processes for the collection, preparation, and continual updating of company’s structured and unstructured data will ensure that LLMs remain accurate and relevant.

Creating and Integrating Domain-Specific Knowledge Graphs with LLMs: Domain-specific KGs provide a powerful way to integrate external business knowledge into LLMs. KGs enable LLMs to respond more effectively to complex queries by discovering both semantic and syntactic relationships among data points. This integration allows businesses to leverage GenAI not only for data processing but also for discovering relationships and generating insights that would be difficult to achieve with traditional methods.

Designing Domain-Specific Architectures and Data Pipelines for LLMs: Domain-specific architectures and data pipelines ensure that LLMs are optimized for particular business environments. Efficient designs allow for the processing of domain-relevant data which may lead to faster and more accurate knowledge management solutions. By focusing on tailored architectures, businesses can ensure that their GenAI implementations meet their unique operational needs.

Ensuring System Scalability, Explainability, Traceability, and Accuracy: To maintain trust and effectiveness, GenAI systems must be scalable, explainable, and accurate. Developing methods to ensure that these systems can grow alongside the business while remaining transparent and traceable, is crucial.

Regulatory and Ethical Compliance of KG-LLM Systems: As businesses adopt GenAI systems, compliance with regulatory frameworks and ethical standards becomes a key consideration. Ensuring that GenAI integrations align with data privacy laws, security standards, and ethical guidelines is important for maintaining trust and preventing misuse. It is important to find the right balance between innovation and responsibility to ensure that GenAI solutions are used in an ethical and legally compliant manner.

5 MAKING GENAI-ENHANCED KM A REALITY

Through our rigorous analysis of modern KM requirements by the businesses and the underlying challenges, we propose developing a GenAI-enhanced toolkit to improve the three knowledge processes: how companies create, synthesize, and access critical business knowledge.

This toolkit can provide several smart services for both internal business processes and the enhancement of the customer experience. It can enable companies and technology providers to develop easy-to-deploy knowledge management solutions using business data. The toolkit's main, user-friendly components include:

- o Software components: customized products, code library,
- o Method components: guidelines, process models, templates,
- o Content components: reusable knowledge models.

From the technology perspective, we are considering advanced RAG combined with a knowledge graph (GraphRAG). However, we plan to assess the need for these technologies for the selected industrial use cases (see section 2.3). Since the integration of advanced technologies should be justified for each case, therefore, the toolkit will also include a decision-making guide. This guide will help companies select when they should utilize certain technology, e.g., when to use simple RAG and when to utilize GraphRAG.

Since the companies involved in this study showed strong interest in such a toolkit, this toolkit can be tested and evaluated within the target companies. For this purpose, a working demo/prototype of GenAI tools can be created and deployed. The recommendations on how to transform them into companies’ own production-grade AI services can be provided.

Developing this toolkit involves understanding and documenting the specific KM requirements of businesses to ensure the effective integration of GenAI solutions. It also requires gathering stakeholder insights to identify where GenAI could enhance knowledge processes. Building on this, an evaluation of existing GenAI technologies must be done to assess their applicability and effectiveness for identified use cases. This is also important to select the technologies that best address the KM challenges, ensuring scalability, accuracy, and adaptability.

6 CONCLUSIONS

In this study, we highlight the transformative potential of GenAI in addressing longstanding challenges in KM. We highlight the key knowledge processes: knowledge capture, access, and synthesis, that could be enhanced through GenAI and could lead to significant improvements in productivity and decision-making. In addition, this study highlighted the challenges involved in adopting GenAI, including integrating GenAI with existing KM systems, ensuring data privacy and accuracy, and overcoming technical and resource limitations. The co-creation approach with companies allowed us to identify specific business needs, highlighting the requirement of a GenAI-KM toolkit which could be used to develop company-specific use cases. This toolkit holds the promise of streamlining KM systems by offering scalable and customizable solutions for diverse business contexts.

A key insight from this research is the importance of a domain-specific approach when integrating GenAI into KM systems. Generic solutions, while useful, often fall short of addressing the unique needs of businesses. By focusing on domain-specific knowledge graphs and adaptable data pipelines, the proposed toolkit can effectively overcome the limitations of current GenAI solutions.

Future work will focus on developing the GenAI-KM toolkit, extending its application to more diverse sectors, and ensuring its scalability and other features, such as explainability and traceability, to ensure trustworthiness. Additionally, research will explore methods to enhance the toolkit's compliance with regulatory frameworks to ensure data privacy and security.

ACKNOWLEDGEMENTS

The research was supported by ERDF (European Regional Development Fund).

REFERENCES

Alavi, M., Leidner, D. E., & Mousavi, R. (2024) A Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems, 25(1), 1-12.

Benbya, H., Strich, F., & Tamm, T. (2024). Navigating generative artificial intelligence promises and perils for knowledge and creative work. Journal of the Association for Information Systems, 25(1), 23-36.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.

Carlucci D., Kudryavtsev D., Santarsiero F., Lagrutta R. & Garavelli A.C. (2022). The ISO 30401 Knowledge Management Systems: a new frame for managing knowledge. Conceptualization and practice, Knowledge Management Research & Practice, 20:6, 975-986.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer L., Candelon F. & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School

Technology & Operations Mgt. Unit Working Paper, (24-013).
Deloitte. (2023). Responsible Enterprise Decisions with Knowledge-enriched Generative AI. Accessed: 13 March 2024.
Earley, S. (2023). What executives need to know about knowledge management, large language models and generative AI. Applied Marketing Analytics, 9(3), 215-229.
European Commission. (2024). Regulatory Framework on Artificial Intelligence. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
Ferrari, F., van Dijck, J., & van den Bosch, A. (2023). Observe, inspect, modify: Three conditions for generative AI governance. New Media & Society.
Ghimire, P., Kim, K., & Acharya, M. (2024). Opportunities and challenges of generative AI in construction industry: Focusing on adoption of text-based models. Buildings, 14(1), 220.
Ghosh, A., & Lakshmi, D. (2023). Dual Governance: The intersection of centralized regulation and crowdsourced safety mechanisms for Generative AI. arXiv preprint arXiv:2308.04448.
Ghanem, H., & Cruz, C. (2024). Fine-Tuning vs. Prompting: Evaluating the Knowledge Graph Construction with LLMs. Proceedings of the 3rd International Workshop on Knowledge Graph Generation from Text (Text2KG), Co-located with the Extended Semantic Web Conference.
Hofer, M., Obraczka, D., Saeedi, A., Köpcke, H., & Rahm, E. (2024). Construction of Knowledge Graphs: Current State and Challenges. *Information*, *15*(8), 509.
Kaczorowska-Spychalska, D., Kotula, N., Mazurek, G., & Sułkowski, Ł. (2024). Generative AI as source of change of knowledge management paradigm. Human Technology, 20(1), 131-154.
Khan, U. A., Kudryavtsev, D. & Kauttonen, J. (2024) Enhancing generative AI for accessing enterprise knowledge. eSignals PRO. Haaga-Helia University of Applied Science. Helsinki. Accessed 28.3.2024. http://urn.fi/URN:NBN:fi-fe2024032813633
Khan, U.A., Kudryavtsev D. (2024). Transforming Knowledge Management in Businesses with Generative AI: Opportunities and Challenges. eSignals PRO. Haaga-Helia University of Applied Science. Helsinki. URN http://urn.fi/URN:NBN:fi-fe2024090267835
Kudryavtsev, D., & Sadykova, D. (2019). Towards architecting a knowledge management system: Requirements for an ISO compliant framework. In IFIP Working Conference on The Practice of Enterprise Modeling (pp. 36–50). Springer.
Maier, R., Hädrich, T., Peinl, R. (2009) Enterprise knowledge infrastructures. Springer Science & Business Media.
Michael, K., Carvalko, J. R., Andrews, C. J., & Batley, L. (2024). In the Special Issues: GenAI, Ingenuity, the Law, and Unintended Consequences. IEEE Transactions on Technology and Society, 5(2), 149-155.
Murphy, T. (2023) How generative AI can improve knowledge management. TechTarget.
Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2023). A comprehensive overview of large language models. arXiv preprint arXiv:2307.06435.
Pimentel, M., & Veliz, J. C. (2024, September). The Generative AI Solutions for Enhancing Knowledge Management: Literature Review and Roadmap. In European Conference on Knowledge Management (Vol. 25, No. 1, pp. 1092-1095).
Reuel, A., & Undheim, T. A. (2024). Generative AI Needs Adaptive Governance. arXiv preprint arXiv:2406.04554.
Wu, D., Chen, L., & Han, Y. (2023). Challenges that generative artificial intelligence poses to higher education and management and the countermeasures—A visualized analysis of literature from 2019 to 2023 using CiteSpace. In Proceedings of the 2023 International Conference on Information Education and Artificial Intelligence (ICIEAI '23). 733–738.
Zhou, B., Li, X., Liu, T., Xu, K., Liu, W., & Bao, J. (2024). CausalKGPT: industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing. Advanced Engineering Informatics, 59, 102333

Predicting Agricultural Product and Supplies Prices Using Artificial Intelligence

Ioannis Dionissopoulos[1][a], Fotis Assimakopoulos[2][b], Dimitris Spiliotopoulos[2][c], Dionisis Margaris[3][d] and Costas Vassilakis[1][e]
[1]*Department of Informatics and Telecommunications, University of the Peloponnese, Tripoli, Greece*
[2]*Department of Management Science and Technology, University of the Peloponnese, Tripoli, Greece*
[3]*Department of Digital Systems, University of the Peloponnese, Sparta, Greece*
{tst06057, fasimakopoulos}@go.uop.gr, {dspiliot, margaris, costas}@uop.gr

Keywords: Agricultural Products, Price Forecasting, Machine Learning, Deep Learning, Data Integration, Forecasting Models.

Abstract: This work focuses on the prediction of agricultural product and supply prices using historical data and artificial intelligence methods. Agricultural product and supply prices are important for the economy and growth of agriculture. Using modern data analysis and deep learning methods, a forecasting model was developed to help us predict future price trends. The data used include the sales prices of crop products and the purchase prices of agricultural inputs. The developed forecasting methods exhibit high accuracy for predicting the actual prices of products and supplies, with error margins ranging from 0.29% to 9.8%, while they can also predict price rises and falls, with respective success rates ranging from 73.29% to 84.96%.

1 INTRODUCTION

In recent years there has been an explosion in data collection. Developments in internet technology have led more and more organisations, both private and public, to organise the collection and dissemination of their data. Some of this data is posted on open-data portals for public use.

Machine learning (ML) frameworks offer a clear knowledge of the process by analysing the massive amounts of data and interpreting the information extracted. These technologies are employed in the construction of models that delineate the connections between elements and actions. Furthermore, ML models can be utilised to predict future actions in a specific scenario (Rashid et al., 2021).

Precision farming uses algorithmic approaches and data to improve productivity, by predicting weather conditions, soil analysis, crop recommendations, and fertilizer and pesticide usage. It uses advanced technologies like IoT, Data Mining, and Machine Learning (ML) to collect data and train the respective systems. This approach reduces manual labour and increases productivity. Farmers face challenges like crop failure and soil infertility (Durai & Shamili, 2022).

Artificial Intelligence (AI) is being used in agriculture to improve crop production, disease prediction, supply chain management, operational efficiency, and water waste reduction (Pallathadka et al., 2023). Machine learning (ML) and deep learning (DL) are commonly used for data prediction, disease prediction, water irrigation optimisation, sales growth, profit maximisation, inventory management, security, fraud detection, and portfolio management.

Various ML approaches can be utilised for crop price prediction, including regression-based methods, time series forecasting techniques, ensemble methods, DL strategies, and hybrid models (Singh & Sindhu, 2024). ML approaches have strengths, limitations, and practical applications. However, there are challenges like data accessibility, feature

[a] https://orcid.org/0009-0003-2482-5322
[b] https://orcid.org/0009-0006-1888-6411
[c] https://orcid.org/0000-0003-3646-1362
[d] https://orcid.org/0000-0002-7487-374X
[e] https://orcid.org/0000-0001-9940-1821

Dionissopoulos, I., Assimakopoulos, F., Spiliotopoulos, D., Margaris, D. and Vassilakis, C.
Predicting Agricultural Product and Supplies Prices Using Artificial Intelligence.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 371-379
ISBN: 978-989-758-716-0; ISSN: 2184-3228

selection, model interpretability, scalability, and generalisation (Cravero et al., 2022).

Many works provide insights for researchers, practitioners, and policymakers, facilitating informed decision-making in agricultural contexts (Assimakopoulos et al., 2024). ML and IoT-enabled farm machinery are key components of the next agriculture revolution. ML applications in agriculture focus on soil parameters, crop yield prediction, disease detection, and species detection. ML with computer vision can monitor crop quality and yield assessment. This approach can enhance livestock production, predict fertility patterns, diagnose eating disorders, and reduce human labour. Knowledge-based agriculture improves sustainable productivity and product quality (Sharma et al., 2021).

Smart farming, utilising AI, addresses agricultural sustainability challenges (Akkem et al., 2023). ML, DL, and time series analysis are crucial for crop selection, yield prediction, soil compatibility classification, and water management. These algorithms classify soil fertility, crop selection, and forecast production. Time series analysis helps predict demand, commodity price, and crop yield. As population growth increases, crop production forecasting is crucial to overcome food insufficiency. Using ML and DL techniques, crop recommendations can be made based on time series analysis to reduce future food insufficiency (Benos et al., 2021).

The purpose of this paper is to develop a model for forecasting agricultural commodity prices using historical data. The development includes the entire process flow, from data collection to the evaluation of the results using performance metrics of our forecasting model. Data analysis and the use of advanced ML techniques will enable the prediction of future prices of these products and the future agricultural production.

For the purposes of this work, data from Eurostat were used, as well as ancillary data of other parameters from other internet sources. Eurostat's open data portal offers us in a user-friendly and structured way information relating to the European Union and figures concerning a wide range of sectors and activities in its area of competence, including data relating to agricultural production.

In the remainder of the paper, Section 2 discusses the related work in ML and other related technologies for precision agriculture. Section 3 presents the data used and the preprocessing and integration methods that were applied to construct the training dataset. Section 4 presents the price forecasting methods that were implemented, while section 5 presents and discusses the results obtained. Finally, section 6 concludes the paper and outlines future work.

2 RELATED WORK

Agricultural data, economy data (market, local economy, wholesale), and world data are but a few domains that are useful for price prediction of agricultural goods. The data provide a strong foundation for innovative agricultural economic management and contributes to scientifically sound price prediction, as well as decision making in precision agriculture (Su & Wang, 2021).

Kumar et al. researched crop yield prediction using historical data to forecast crop yields, considering factors like temperature, humidity, and rainfall. The approach found that the Random Forest (RF) algorithm provides the best predictions, considering the least number of models, making it useful in the agriculture sector (Kumar et al., 2020).

Zhao used a wavelet method to smooth multiple sources of data and build a model to process the hierarchical information after signal decomposition (Zhao, 2021). Another study compared predictive accuracies of various ML techniques, focusing on GRNN, with the Autoregressive integrated moving average (ARIMA) model (Paul et al., 2022). Results showed GRNN outperforms other techniques in all seventeen markets, while RF is comparable in four. The Diebold-Mariano test confirmed these superior performances. Other techniques like SVR, GBM, and ARIMA are not as effective.

Xu & Zhang investigated corn cash price forecasting using univariate neural network (NN) modelling and bivariate NN modelling with futures prices. Results show high accuracy for one-day ahead horizons, with futures prices benefiting cash price forecasting. The framework was deemed easy to deploy and can be generalised to other commodities (Xu & Zhang, 2021).

Oktoviany et al. proposed a two-step hybrid model using ML methods to incorporate external factors in price changes (Oktoviany et al., 2021). The model assigns price states to historical prices and predicts future price states using short-term predictions. The model is applied to real corn futures data and generates price scenarios through Monte Carlo simulations. The simulations can be used to assess price risks in risk management systems or support trading strategies under different price states.

Another research used supervised ML for intelligent information prediction analysis to improve farming efficiency and profitability (Shakoor et al., 2017). The approach suggests area-based beneficial crop rankings, based on static data from previous years. This happens before the cultivation process. It indicates the crops that are cost effective for cultivation for a particular area of land. The study used Decision Tree Learning-ID3 and K-Nearest Neighbours Regression algorithms.

Time-series and ML models have also been deployed to predict monthly areca nut prices using SARIMA, Holt-Winter's Seasonal method, and LSTM neural networks (NNs). The LSTM NN model was found to be the best fit for the data (Sabu & Kumar, 2020). ANNs have also been used to predict soybean harvest area, yield, and production, comparing it with classical methods of Time Series Analysis (Abraham et al., 2020).

The work in (Purohit et al., 2021) proposed two additive hybrid methods and five multiplicative hybrid methods to predict the monthly retail and wholesale prices of three commonly used vegetable crops in India: tomato, onion, and potato (TOP). Extensive statistical analyses confirmed the superiority of the hybrid methods against existing statistical models, ML models, and existing hybrid methods in predicting TOP prices.

An alternative method that addresses the nonlinearity problem if time series approaches is wavelet transformation in generating hybrid models for predicting monthly prices markets. This hybrid model approach significantly improved over conventional techniques, utilising a combination of ANN and ML techniques (Paul & Garai, 2021).

Xu & Zhang investigated the use of nonlinear autoregressive neural networks (NARNN) and NARNN with exogenous inputs (NARNN–X) for price forecasting soybeans and soybean oil for periods that spanned over fifty years. The models exhibited accurate and stable performance, with relative root mean square errors of 1.701% and 1.777% for soybeans and 1.757% for soybean oil, respectively. Also, the approach can be generalised for other similar commodities (Xu & Zhang, 2022).

Menculini et al. examined various techniques for forecasting sale prices in an Italian food wholesaler, comparing ARIMA models, Prophet, which is a scalable forecasting tool from Facebook, and deep learning models, such as LSTM and CNNs. Results showed that ARIMA models and LSTM neural networks perform similarly, while the combination of CNNs and LSTMs achieved the best accuracy but requires more tuning time. Prophet was quick and easy to use but less accurate (Menculini et al., 2021).

3 DATA AND PREPROCESSING

Agricultural price prediction is a highly complex task, due to the fact that prices depend on numerous factors, both within the agricultural value chain and in the macroeconomic environment. Besides building a comprehensive dataset, encompassing the widest possible range of factors affecting prices, the quality and trustworthiness of data are of critical importance, in order to achieve high prediction accuracy. In the following paragraphs we describe the data sources used, as well as the integration and preprocessing methods used to formulate the training datasets.

3.1 Data Sources

Two key datasets for this research were obtained from Eurostat. These datasets are as follows:

1. Selling prices of crop products. These data cover the historical dimension of agricultural sales, supporting the dimension of analysis, which assumes that future patterns of agricultural prices will follow similar patterns, already been observed in the past.
2. Purchase prices of agricultural production means. The analysis considers this data, since the selling prices of agricultural products obviously depend on the prices of the means used for their production.

These datasets contain information spanning from 1969 to 2023; each dataset is provided in two parts, with the first covering the period 1969-2000 and the second spanning from 2001 to 2023. Since our price data is sourced from Eurostat, they contain only data for EU countries, hence price predictions in our experiments are limited to member states of the EU.

Energy cost is an important factor in the cost of agricultural production since oil is extensively used to operate motorised equipment, such as tractors and tillers, and is thus involved in the production cost. Consequently, we take Brent oil prices into account, in our predictions. It is considered as the most important indicator of energy spent in agricultural production, since its two main fuels, diesel and gasoline, are used to drive motorised equipment with internal combustion engines. Data concerning Brent oil prices were obtained from statista.com. At this stage of our research, Brent oil price is used as an overall indicator for energy cost. The inclusion of more detailed energy costs, notably electricity costs, is considered as part of our future research.

Land use data, from the World In data website, were also considered in our work. This dataset provides information on overall land use, cropland

land use, grazing land use and built-up area, per country and year.

The availability of human resources in agricultural production is also a factor impacting the prices of agricultural production. These data were obtained from Eurostat and cover the period from 1973 to 2023. The dataset provides a detailed breakdown of the total labour force to salaried and non-salaried workers. In our work, we consider all types of employees and hence we maintained only the sum of these two categories.

In our work, we also consider indicators of economic nature, concerning agriculture. From the Economic Accounts for Agriculture dataset, sourced from Eurostat, we extract and use the following data: (a) Production Value at Basic Price, (b) Subsidies on Products, (c) Tax on Products, and (d) Production Value at Producer Price.

3.2 Preprocessing and Integration

The data obtained from the above listed data sources were not directly utilisable for model training, necessitating preprocessing and integration activities.

Preprocessing activities concern the handling of missing data, noisy data, inconsistent data, encoding and value range discrepancies, and handling of textual data. In the following paragraphs, we outline the specific activities taken to address these issues.

Missing Data. Some attribute values may be missing due to an error, either in the registration process or because they were not provided by the relevant agency. For these cases, we considered firstly to find supplemental datasets that provided the missing values and integrate them into our dataset.

Values that were still missing, we applied imputers to fill in the missing values. For each data element, different imputers were considered and the effectiveness of the use of each imputer to predict a data element on the accuracy of the predictions was assessed. Experimental results demonstrated that the most accurate results were obtained by using the following imputers: (a) for Brent oil prices, backward fill (i.e. if a price is missing, use the price for the next known data point); (b) for the labour workforce, linear regression. For the agricultural economic accounts, a KNN-based imputer was applied, using N=20 (N denotes the number of nearest neighbours considered for computing a missing value).

For the cases which after the use of imputers data were still missing (because the imputers could not calculate the missing data due to the sparsity of the original dataset), the relevant records were dropped.

Noisy data: Data containing errors or outliers, which are highly deviant from the normal pattern, were discarded, since their use affected negatively the accuracy of predictions. The interquartile distance method (Vinutha et al., 2018) was used for identifying potential outliers and subsequently visual verification was conducted using graphs.

Inconsistent Data. Either duplicate values or data providing different values for a specific data element, for the same country and period. Data that were verified to be duplicates were discarded.

Differences in Units. Due to the currency change in many European countries, our data contained prices in both Euros and the previous local currency. For the algorithm to have comparable data at its disposal, price conversions to Euro were performed for countries that underwent currency changes.

Differences in Encoding. The price datasets obtained from Eurostat used different codes for agricultural products and supplies for the period 1969-2000 than for the period 2001-2023. To produce the integrated dataset, the product/supplies codes for the data concerning the period 1969-2000 were replaced by the respective codes used for the period 2001-2023. A fuzzy match on the names of the products was used to perform the mapping.

Handling of Textual Data. AI-based regression methods that were used for price prediction mainly work with numeric data and not textual data. Our datasets contain multiple cases where textual data are present, e.g., country names/codes and agricultural products/supplies names and codes For these cases, *label encoding* was employed, i.e., each distinct value of the respective data element was mapped to a unique integer, and only the mapped value was considered in the prediction process.

Different Scales. Different data elements had highly divergent scales (e.g. land availability and Brent oil prices), and this aspect negatively affected the accuracy of the predictions, due to overfitting. To mitigate this issue, each data column (except encoded labels and prices) was normalized to the range [0,1] using the *Min Max Scaler*; the normalized value *NV* produced by the Min Max Scaler for a value *V* is computed as $NV = \frac{value - MinVal}{MaxVal - MinVal}$, where $MinVal$ and $MaxVal$ are the minimum and maximum values for the specific column, respectively.

4 PRICE FORECASTING METHODS

In the previous section we presented the data collection, preprocessing and integration process. Following the above, all input data have been formulated in two comprehensive datasets:

- The crop products selling prices dataset,
- The agricultural production means dataset.

Each of these two datasets contains records with the following data elements: (i) country, (ii) agricultural product or means of production, (iii) year, (iv) price, (v) availability of labor in agricultural production, (vi) purchase and rental prices of the land, (vii) Brent oil prices and (viii) economic indicators of agricultural production (production value at basic price, subsidies on products, tax on products, and production value at producer price). These datasets can be used to train ML algorithms to perform predictions.

Since multiple AI-based methods and configurations are available for performing predictions, and each of these can be tuned through a number of hyperparameters, we resorted to the use of automatic machine learning (autoML) toolkits which underpin the tasks of method selection and hyperparameter tuning. To this end, the Autokeras and the TPOT autoML toolkits were used.

AutoKeras (https://autokeras.com/) is an open-source ML library, based on Keras and Tensorflow, which aims to build and optimise NNs automatically. In its basic function, the user only specifies whether a classification or a regression model is required, and the columns that are used for training, designating the target column for prediction.

TPOT (https://epistasislab.github.io/tpot/) is an open-source library that explores the performance of ML models in an automatic way, as well. It allows to search for the most efficient ML algorithm for the dataset used each time.

The hyperparameters used for the autoconfiguration process performed by the AutoKeras toolkit are as follows:

- *Tries.* The number of attempts AutoKeras will perform to arrive at the most efficient model. In this work we will experiment with 25 attempts for each dataset.
- *Test Size.* The percentage of training data that we will use for testing, in order to avoid Overfitting. In this work we will experiment with 30% of the data.
- *Number of Training Epochs*: i.e. the number of iterations in which each of our models is trained to approach the best result. In this work we will experiment with 30 seasons.

Table 1 illustrates the topology of the NN. This topology is designated as optimal for both price prediction tasks (agricultural products and supplies).

Table 1: The topology of the neural network.

Layer (type)	**Output Shape**	**Parameter value**
input_1 (InputLayer)	(None, 20)	0
multi_category_encoding (MultiCategoryEncoding)	(None, 20)	0
normalization (Normalization)	(None, 20)	41
dense (Dense)	(None, 32)	672
re_lu (ReLU)	(None, 32)	0
dense_1 (Dense)	(None, 128)	4224
re_lu_1 (ReLU)	(None, 128)	0
regression_head_1 (Dense)	(None, 1)	129

Table 2: Parameters for the Random Forest regressor.

Parameter	**Value**
n_estimators (The number of trees in the forest)	100
max_features (number of features to consider when looking for the best split)	75% of the number of input features
min_samples_leaf (the minimum number of samples a leaf node must contain)	7
min_samples_split (minimum number of samples required to split an internal node)	19

Table 3: Parameters for the Gradient Boosting regressor.

Parameter	**Value**
loss (Loss function used in optimization; the value huber combines squared error and absolute error)	huber
alpha (The alpha-quantile of the huber loss function and the quantile loss function)	0.8
learning_rate (moderates the contribution of each tree)	0.1
max_depth (moderates the maximum number of nodes in a tree, setting the maximum depth of the individual regression estimators)	7
max_features (number of features are considered in each split; value 1 indicates that all features are taken into account)	1.0
min_samples_leaf (the minimum number of samples a leaf node must contain)	1
min_samples_split (minimum number of samples required to split an internal node)	11
n_estimators (number of boosting stages that will be performed)	100
Subsample (percentage of samples used for fitting the individual base learners)	0.65

For the TPOT toolkit, the number of generations was set to 15, while the population size was set to 15. The population size refers to the number of individuals in each generation that retain their characteristics, as compared to the previous generation. The output of the TPOT toolkit determined that the optimal prediction method for agricultural product price prediction would be the random forest regression method, under the parameters illustrated in Table 2. Agricultural supplies prices, on the other hand, are more accurately predicted using Gradient Boosting, under the parameters listed in Table 3.

In the following section, the results and evaluation of this work will be presented and analysed.

5 RESULTS AND EVALUATION

In this section, the results and evaluation of this work are presented and analysed.

The prediction accuracy of our model can be assessed using performance metrics, which evaluate the closeness between the prediction result and the actual result. The metrics used in this work, are widely used in related research works that measure prediction The metrics are illustrated in Table 4, along with their respective formulas.

Table 4: The performance metrics used in our work.

Metric Name	Formula
Root Mean Square Error (RMSE)	$\sqrt{\sum_{i=i}^{n}(y_i - \hat{y}_i)^2 / n}$
Mean Average Error (MAE)	$\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i - \hat{y}_i \rvert$
Normalized MAE (NMAE)	$\left(\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i - \hat{y}_i \rvert\right) \Big/ \left(\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i \rvert\right)$

The RMSE metric boosts the significance of large deviations between the prediction result and the actual result, while the MAE handles all errors uniformly. The NMAE has the property of amortizing differences in the scale of the predicted variables, however, when the actual values are very small, errors are over-emphasised. In all the aforementioned metrics, lower values indicate smaller divergence and hence more accurate predictions.

In addition to the above, in this work, we include an additional performance metric, namely the *Percentage of Successful Predictions* (PSP); this metric computes the percentage of predictions that are deemed to be 'successful', and a price prediction $\hat{y}_i$ for time point *i* is considered successful iff

$$(y_i - y_{i-1}) * (\hat{y}_i - y_{i-1}) > 0$$

where y_i and y_{i-1} are the actual prices at time points *i* and *i-1*, respectively. Effectively, a prediction is considered to be successful iff either (a) a rise in the price is predicted and a rise actually occurred or (b) a drop in the price is predicted and a drop actually took place, otherwise the prediction is deemed unsuccessful. The *percentage of successful predictions* metric can be useful for assessing the utility of the approach for investment decisions, e.g., to invest on a particular product.

Tables 5 and 6 depict the accuracy metrics obtained from our experiments regarding the prediction of agricultural product sale prices and agricultural supplies, respectively.

In Table 5 we can observe that the NN optimised and proposed by AutoKeras achieves predictions that deviate from the actual prices by 6.6% on average (c.f. the NMAE metric), surpassing the accuracy of the Random Forest predictor proposed by TPot (average deviation 9.8%). The AutoKeras NN also achieves superior performance in predicting price rises or drops (80.96% vs. 73.29%).

In Table 6 we notice that both the AutoKeras NN and the gradient boosting predictor, proposed by TPot, formulate predictions with very small deviations from the actual prices (2.7% and 0.29%, respectively). While the gradient boosting predictor estimates actual prices better than the AutoKeras NN, it lags behind concerning the prediction of price rises or drops.

Table 5: Prediction accuracy for agricultural product sale prices.

Metric	Neural network (AutoKeras)	Random forest (Tpot)
RMSE	28.66	29.69
MAE	11.76	11.49
NMAE	0.0659	0.098
PSP	80.96%	73.29%

Table 6: Prediction deviation agricultural supplies prices.

Metric	Neural network (AutoKeras)	Gradient boosting (TPot)
RMSE	10.60	10.08
MAE	3.66	2.84
NMAE	0.0269	0.0029
PSP	84.96%	79.34%

The performance recorded for price predictions in our experiment surpasses the price prediction accuracy recorded for the works surveyed in section 2, which exhibit deviations from actual prices ranging from 12% to 26%. Since our experiment is limited to EU countries only, involving only countries for which historical data of high accuracy and ample time depth are available, more experimentation is required to fully compare the proposed algorithm against approaches proposed in the literature. This is considered a part of our future work.

Finally, in our experiments we can observe that prices of the means of agricultural production are predicted with higher accuracy than prices of agricultural products. This may be attributed to a dependence of agricultural product prices to additional factors than the ones considered in our work, while these factors suffice for the prediction of prices of means of agricultural production; this aspect will also be examined in our future work.

6 CONCLUSION

In this paper, we have presented a model for forecasting agricultural product and supply prices using historical data. We analysed the entire process flow, including data selection, preprocessing and integration, model training and algorithm tuning, as well as performance metrics and model evaluation.

The proposed model exhibits high accuracy for price predictions, especially for agricultural supplies, while it is also able to predict price rises or drops. Thus, the proposed algorithm can be used for budgeting production, estimating earnings and investment planning.

As richer datasets become available, especially with the advent of IoT, additional data can be taken into account for performing price predictions. Yet, developing countries are still challenged regarding the availability and accuracy of data. These aspects will be surveyed in our future work, elaborating on methods and techniques that are able to achieve high prediction accuracy over more sparse datasets.

ACKNOWLEDGEMENTS

This research was funded by project SODASENSE (https://sodasense.uop.gr) under grant agreement No. MIS 6001407 (co-financed by Greece and the EU through the European Regional Development Fund).

REFERENCES

Abraham, E. R., Mendes Dos Reis, J. G., Vendrametto, O., Oliveira Costa Neto, P. L. D., Carlo Toloi, R., Souza, A. E. D., & Oliveira Morais, M. D. (2020). Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. *Agriculture*, *10*(10), 475. https://doi.org/10.3390/agriculture10100475

Akkem, Y., Biswas, S. K., & Varanasi, A. (2023). Smart farming using artificial intelligence: A review. *Eng. Appl. of Artificial Intelligence*, *120*, 105899. DOI: 10.1016/j.engappai.2023.105899

Assimakopoulos, F., Vassilakis, C., Margaris, D., Kotis, K., & Spiliotopoulos, D. (2024). The Implementation of “Smart” Technologies in the Agricultural Sector: A Review. *Information*, *15*(8), 466. https://doi.org/10.3390/info15080466

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, *21*(11), 3758. DOI: 10.3390/s21113758

Cravero, A., Pardo, S., Sepúlveda, S., & Muñoz, L. (2022). Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy*, *12*(3), 748. https://doi.org/10.3390/agronomy12030748

Deepa, S., Alli, A., Sheetac, & Gokila, S. (2023). Machine learning regression model for material synthesis prices prediction in agriculture. *Materials Today: Procs.*, *81*, 989–993. https://doi.org/10.1016/j.matpr.2021.04.327

Durai, S. K. S., & Shamili, M. D. (2022). Smart farming using Machine Learning and Deep Learning techniques. *Decision Analytics Journal*, *3*, 100041. https://doi.org/10.1016/j.dajour.2022.100041

Falah, R. A., & Rachmaniah, M. (2022). Price Prediction Model for Red and Curly Red Chilies using Long Short Term Memory Method. *Indonesian Journal of Statistics and Its Applications*, *6*(1), 143–160. https://doi.org/10.29244/ijsa.v6i1p143-160

Ge, Y., & Wu, H. (2020). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications*, *32*(22), 16843–16855. https://doi.org/10.1007/s00521-018-03970-4

Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent Crop Price Prediction using suitable Machine Learning Algorithm. *ITM Web of Conf.*, *40*, 03040. DOI: /10.1051/itmconf/20214003040

Hegde, G., Hulipalled, V. R., & Simha, J. B. (2021). Price Prediction of Agriculture Commodities Using Machine Learning and NLP. *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 1–6. https://doi.org/10.1109/ICSTCEE54422.2021.9708582

Im, J.-J., Kim, T.-W., Lim, J.-S., Kim, J.-H., Yoo, T.-Y., & Lee, W. J. (2022). A Design and Implement of Efficient Agricultural Product Price Prediction Model. *Journal of the Korea Society of Computer and Information*, *27*(5), 29–36. https://doi.org/10.9708/JKSCI.2022.27.05.029

Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020). Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. *2020 5th Intl. Conf. on Comm. and Electronics Systems (ICCES)*, 736–741. https://doi.org/10.1109/ICCES48766.2020.9137868

Kurumatani, K. (2020). Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*, *2*(8), 1434. https://doi.org/10.1007/s42452-020-03225-9

Lanjewar, M. G., Parate, R. K., & Parab, J. S. (2022). Machine Learning Approach with Data Normalization Technique for Early Stage Detection of Hypothyroidism. In M. K. Ahirwal, N. D. Londhe, & A. Kumar, *Artificial Intelligence Applications for Health Care* (1st ed., pp. 91–108). CRC Press. https://doi.org/10.1201/9781003241409-5

Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2022). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting*, *38*(1), 35–50. https://doi.org/10.1016/j.ijforecast.2020.02.002

Luo, J., Zhao, C., Chen, Q., & Li, G. (2022). Using deep belief network to construct the agricultural information system based on Internet of Things. *The Journal of Supercomputing*, *78*(1), 379–405. https://doi.org/10.1007/s11227-021-03898-y

Mahto, A. K., Alam, M. A., Biswas, R., Ahmed, J., & Alam, S. I. (2021). Short-Term Forecasting of Agriculture Commodities in Context of Indian Market for Sustainable Agriculture by Using the Artificial Neural Network. *Journal of Food Quality*, *2021*, 1–13. https://doi.org/10.1155/2021/9939906

Mamoudan, M. M., Mohammadnazari, Z., Ostadi, A., & Esfahbodi, A. (2024). Food products pricing theory with application of machine learning and game theory approach. *International Journal of Production Research*, *62*(15), 5489–5509. https://doi.org/10.1080/00207543.2022.2128921

Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. *Forecasting*, *3*(3), 644–662. https://doi.org/10.3390/forecast3030040

Mohanty, M. K., Thakurta, P. K. G., & Kar, S. (2023). Agricultural commodity price prediction model: A machine learning framework. *Neural Computing and Applications*, *35*(20), 15109–15128. https://doi.org/10.1007/s00521-023-08528-7

Oktoviany, P., Knobloch, R., & Korn, R. (2021). A machine learning-based price state prediction model for agricultural commodities using external factors. *Decisions in Economics and Finance*, *44*(2), 1063–1085. https://doi.org/10.1007/s10203-021-00354-7

Pallathadka, H., Mustafa, M., Sanchez, D. T., Sekhar Sajja, G., Gour, S., & Naved, M. (2023). Impact of Machine Learning on Management, Healthcare and Agriculture. *Materials Today: Proceedings*, *80*, 2803–2806. https://doi.org/10.1016/j.matpr.2021.07.042

Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Soft Computing*, *25*(20), 12857–12873. https://doi.org/10.1007/s00500-021-06087-4

Paul, R. K., Yeasin, Md., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., & Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLOS ONE*, *17*(7), e0270553. https://doi.org/10.1371/journal.pone.0270553

Purohit, S. K., Panigrahi, S., Sethy, P. K., & Behera, S. K. (2021). Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. *Applied Artificial Intelligence*, *35*(15), 1388–1406. https://doi.org/10.1080/08839514.2021.1981659

Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access*, *9*, 63406–63439. https://doi.org/10.1109/ACCESS.2021.3075159

Ribeiro, M. H. D. M., & Dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, *86*, 105837. https://doi.org/10.1016/j.asoc.2019.105837

Sabu, K. M., & Kumar, T. K. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science*, *171*, 699–708. https://doi.org/10.1016/j.procs.2020.04.076

Shakoor, Md. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques. *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 182–187. https://doi.org/10.1109/NEXTCOMP.2017.8016196

Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, *9*, 4843–4873. DOI: 10.1109/ACCESS.2020.3048415

Singh, N., & Sindhu, R. (2024). Crop Price Prediction Using Machine Learning. *Journal of Electrical Systems*, *20*(7s), 2258–2269. DOI: 10.52783/jes.3961

Su, Y., & Wang, X. (2021). Innovation of agricultural economic management in the process of constructing smart agriculture by big data. *Sustainable Computing: Informatics and Systems*, *31*, 100579. https://doi.org/10.1016/j.suscom.2021.100579

Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, & J. R. Mohanty (Eds.), *Information and Decision Sciences* (Vol. 701, pp. 511–518). Springer Singapore. https://doi.org/10.1007/978-981-10-7563-6_53

Wang, J., Wang, Z., Li, X., & Zhou, H. (2022). Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, *38*(1), 21–34. https://doi.org/10.1016/j.ijforecast.2019.08.006

Wang, L., Feng, J., Sui, X., Chu, X., & Mu, W. (2020). Agricultural product price forecasting methods: Research advances and trend. *British Food Journal*, *122*(7), 2121–2138. DOI: 10.1108/BFJ-09-2019-0683

Xu, X. (2020). Corn Cash Price Forecasting. *American Journal of Agricultural Economics*, *102*(4), 1297–1320. https://doi.org/10.1002/ajae.12041

Xu, X., & Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in*

Agriculture, *184*, 106120. https://doi.org/10.1016/j.compag.2021.106120

Xu, X., & Zhang, Y. (2022). Soybean and Soybean Oil Price Forecasting through the Nonlinear Autoregressive Neural Network (NARNN) and NARNN with Exogenous Inputs (NARNN–X). *Intelligent Systems with Applications*, *13*, 200061. https://doi.org/10.1016/j.iswa.2022.200061

Yin, H., Jin, D., Gu, Y. H., Park, C. J., Han, S. K., & Yoo, S. J. (2020). STL-ATTLSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM. *Agriculture*, *10*(12), 612. https://doi.org/10.3390/agriculture10120612

Zhang, D., Chen, S., Liwen, L., & Xia, Q. (2020). Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons. *IEEE Access*, *8*, 28197–28209. https://doi.org/10.1109/ACCESS.2020.2971591

Zhao, H. (2021). Futures price prediction of agricultural products based on machine learning. *Neural Computing and Applications*, *33*(3), 837–850. https://doi.org/10.1007/s00521-020-05250-6

AUTHOR INDEX

Proceedings of IC3K 2024 - Volume 3: KMIS

16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management

https://ic3k.scitevents.org

EVENT MANAGEMENT SYSTEM:

LOGISTICS:

IN COOPERATION WITH:

PROCEEDINGS WILL BE SUBMITTED FOR EVALUATION FOR INDEXING BY:

ISSN: 2184-3228
ISBN: 978-989-758-716-0