Neuroinformatics-2026

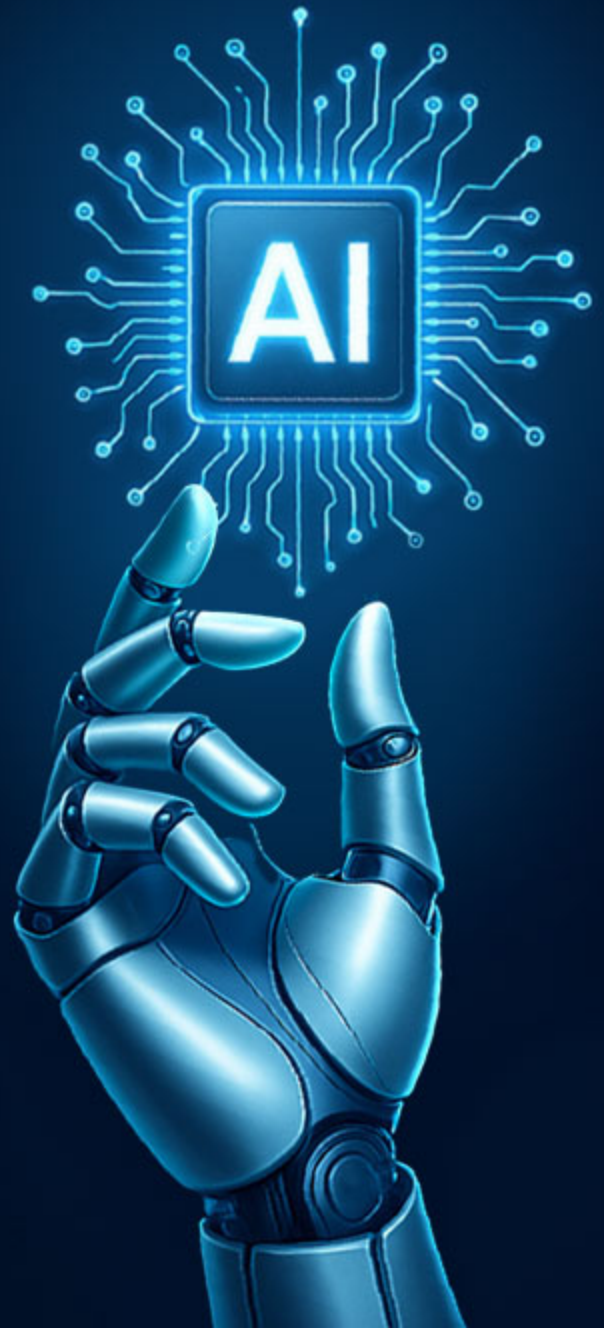
AN ARCHITECTURE FOR MIXED HUMAN–AI BOARDS

Autonomous AI Systems for Corporate Governance

A Framework Bridging Computational Law, Operational Context, and Multi-Level Oversight



A. S. Romanova, Ph.D., MBA, LL.M., ALM, FCCA
GSOM



Agenda

1. Solving the “Positronic Brain” Problem for Corporations
2. Delaware and the Rise of Algorithmic Entities
3. Hardware Miniaturisation and the Future “Small Brains”
4. Machine Consciousness
5. The A0 - A5 Autonomy Spectrum
6. Dedicated Operational Context
7. Jurisdictional Neutrality (an invariant mathematical structure for fairness, equality, interval, risk and management decision)

FROM DECISION SUPPORT TO DECISION MANAGEMENT



When AI stops advising and starts deciding, law and IT governance lose their object

What corporate law assumes

- A human director judging discrete decisions
- Duty of care, duty of loyalty, business judgment rule
- Good faith: belief, intent and ethical judgment

What AI actually delivers

- Continuous, probabilistic, adaptive outputs
- A pace and opacity the doctrines cannot absorb
- No electronic personhood: accountability stays with natural persons

The gap is threefold — and must be closed on all three levels at once

Conceptual **Jurisdictional** **Procedural**

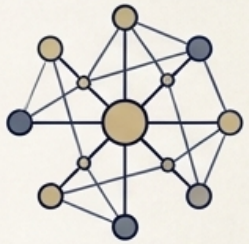
Precedents: VITAL (2014), Mika, Tang Yu — the question is no longer whether machines take part, but how

The Corporate AI Systems Typology









Class 1: Digital Command Centers

- Primary Function: Big Data analytics & continuous monitoring
- Human Interface: Dashboards / None
- Autonomy Level: A0-A2
- Precedent: Panorama (ADNOC)

Class 2: Personalized Virtual Agents

- Primary Function: Social interaction & management recommendations
- Human Interface: Cognitive/Communicative Avatar
- Autonomy Level: A1-A3
- Precedents: Tang Yu, Aiden Insight

Class 3: Humanoid Robots

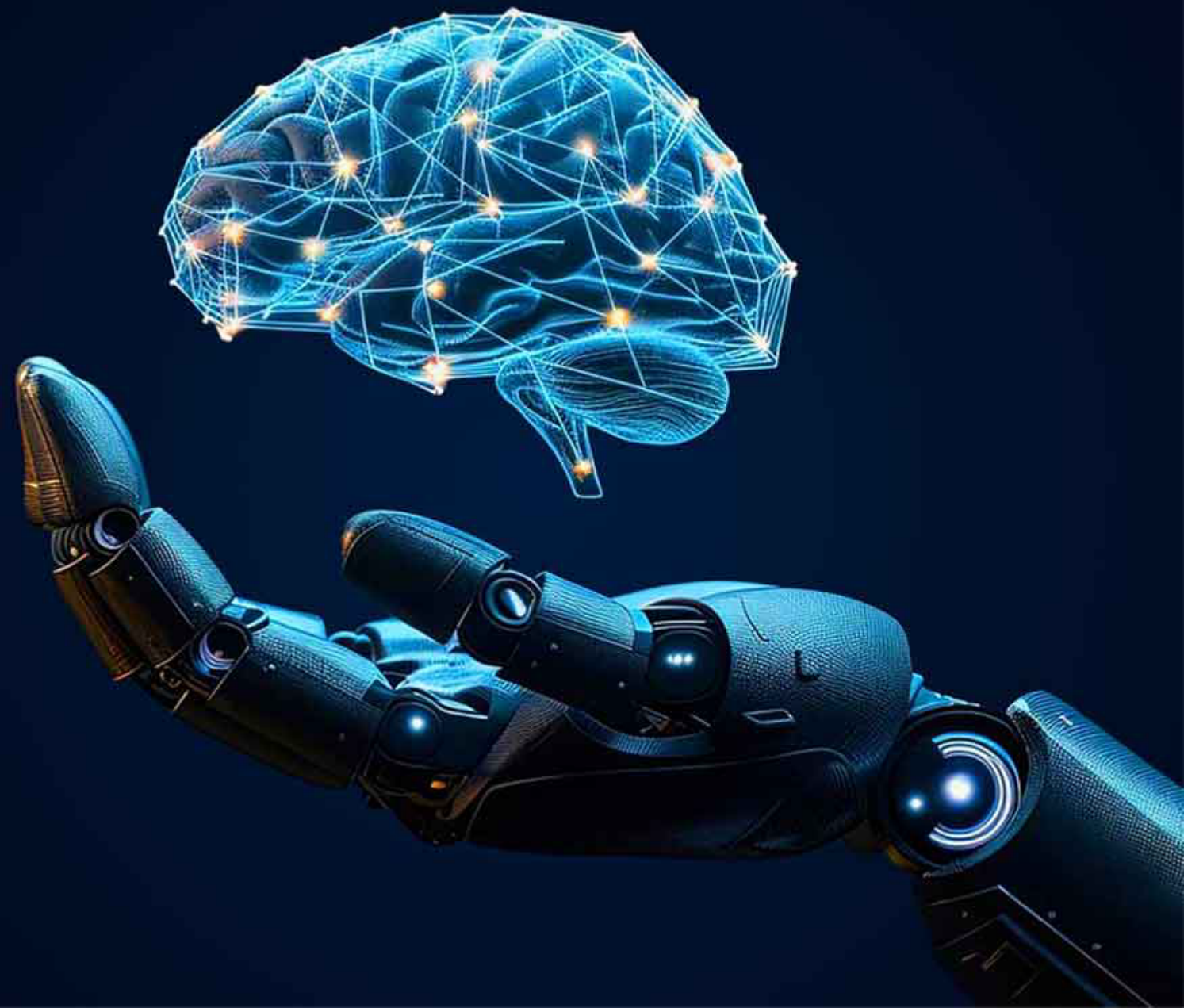
- Primary Function: Physical presence & symbolic representation
- Human Interface: Mechatronic physical embodiment
- Autonomy Level: A1
- Precedents: Mika, Sophia

Class 4: Hybrid AI Factories

- Primary Function: Distributed coordination & automated execution
- Human Interface: DAO / Distributed architecture
- Autonomy Level: A3-A5 (Theoretical)
- Precedent: Blockchain DAO ecosystems

Gemini Notebook

Autonomous companies require autonomous management — governed by computable law.





A corporation run by AI needs one continuous, norm-bound operational brain

Dan-Cohen, 1986

The “personless organisation” already has intelligence independent of its members

Simon’s corollary

Autonomous companies require autonomous management

Prototype today

Multi-subsidiary digital command centres such as ADNOC Panorama

THE LOOP IT CLOSES

Observe → Understand → Predict → Decide → Act

FED BY, UNDER A SINGLE SOURCE OF TRUTH

IoT / IIoT · Digital twins · Blockchain · ERP · CRM · SCADA

CONTINUITY OF EXISTENCE ON FIVE AXES

Data · Memory · State · Context · Identity

NORMATIVE ENVELOPE

Corporate goals · legal limits · codified ethical priorities

The Translation Gap Between Fiduciary Doctrine and Machine Logic



Delaware's Business Judgment Rule presupposes human traits.

Closing this gap requires a structural translation layer, not just updated legal phrasing.

AI possesses no tacit experience; it requires precise mathematical operationalization.

Gemini Notebook



Algorithm-run legal persons already exist in law; legislators are catching up

Bayern · 2013

The zero-member LLC

Software can own and run an LLC under existing US law: contracts, assets, trade — no human required.

LoPucki · 2017

Four weak points

- Controls most legal forms
- Hops jurisdictions fast
- Controllers stay invisible
- Competition blocks bans

Möslein · 2018

The good-faith gap

“Good faith” presupposes human belief and intent: an ontological gap between law and algorithm.

2026

Delaware and Argentina are developing legislation for algorithmic companies — yet no jurisdiction regulates AI directors directly. Practice has outrun both law and method.

Sources: Bayern (2013) Nw. U. L. Rev.; LoPucki (2017) Wash. U. L. Rev.; Möslein (2018) Research Handbook on the Law of AI

03 · HARDWARE MINIATURISATION AND THE FUTURE “SMALL BRAINS”

Design the corporate brain for the chip it will run on, not today’s data centre

THIRTY YEARS AGO

Cooled server halls

Racks, rooms, cooling plants

→

TODAY

A mobile form factor

The same compute, in the hand

→

TARGET FORM

A compact autonomous core

Self-learning, always on

DESIGN PRINCIPLE

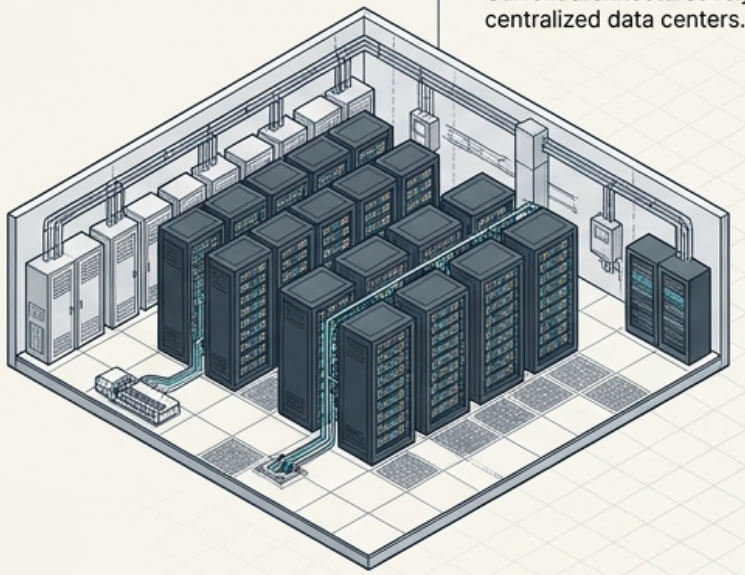
Control logic, algorithmic loops and protocols are specified as if the target were already a miniature core. Today’s command centres and server-side infrastructure are a prototype environment and a transitional stage, not the goal.

RESULTING ENGINEERING REQUIREMENTS

- Integrality
- Continuous operation
- Learnability
- Unified semantic representation
- State preservation

Hardware Miniaturization Drives the Shift to Autonomous Edge Small Brains

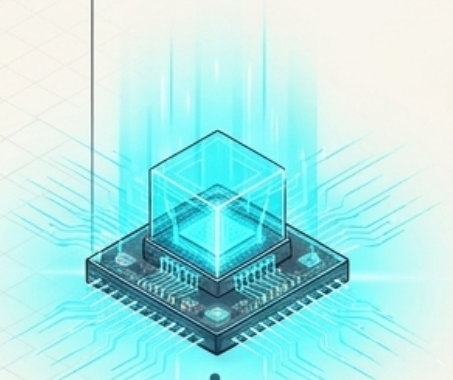
Yesterday
Current architectures rely on massive, centralized data centers.



Today
Sleek, centralized "Digital Command Centers" (e.g., ADNOC's Panorama, aggregating data from 14 subsidiaries in real-time).



Tomorrow's Small Brain
Future corporate AI directors will be fully contained, miniaturized computing cores.



1 Yesterday

2 Today

3 Tomorrow's Small Brain

Governance architectures designed today must assume and secure this highly distributed, edge-computed future.

Gemini Notebook

04 · MACHINE CONSCIOUSNESS — FUNCTIONAL, NOT METAPHYSICAL



Machine consciousness becomes testable through behaviour, without settling what it is

Systematicity

Decisions follow from its model of the world and its norms

Narrative coherence

Its account of its own decisions is consistent and connected

Reflexivity

It explains why it decided and which alternatives it weighed

Adaptivity

It changes strategy when circumstances change

Normative compliance

It follows declared values even when that is unprofitable

Emergence

Complex behaviour arises that was never explicitly programmed

Continuity hypothesis

Why it matters legally

Engineering under way

05 · THE A0–A5 AUTONOMY SPECTRUM

Six autonomy levels along five criteria

Five orthogonal criteria: who controls the outcome · advisory vs. mandatory execution · human in, on or out of the loop · ability to modify goals and metrics · structural integration into the organs of governance

Level	Role of the AI	Human role	What the AI does
A0	Digital infrastructure	Full control	No autonomous decisions
A1	Advisory	In the loop	Recommends; a human ratifies
A2	Augmented	In the loop	Shapes alternatives and criteria
A3	**Semi-autonomous by default**	On the loop	Acts within approved policies
A4	High autonomy	Out of the loop	Decides and acts as autonomous CEO
A5	Self-organising (autopoietic)	Indirect or none	Adapts the rules of governance

Tinted row — A3 is the critical threshold: the object of control shifts from individual decisions to policies, limits, monitoring and escalation.



Each level carries its own duties and its own standard of judicial review

A1–A2

Classical business judgment rule

Applies with modest adaptation: a human still ratifies every decision.

A3

Duty attaches to design choices

Which thresholds, which monitoring metrics, which exception handlers: the critical threshold.

A4

Caremark-style oversight

Directors must show that ex ante design and ex post audit were jointly sufficient.

A5

Legitimacy beyond corporate law

Delegating value-laden trade-offs approaches constitutional concerns.

REALITY CHECK

Most headline “AI director” appointments sit at A1–A2, not A4.

Sources: the Caremark line of cases; Langenbucher on adapting the business judgment rule to AI-supported decisions

A corporate Operational Design Domain: autonomy only inside a bounded context

Analogue of the Operational Design Domain in SAE J3016.

Outside its boundary the system hands over to a human.

Generalises the “self-driving subsidiary” of Armour and Eidenmüller.

INSIDE THE DOC — ADMISSIBLE OPERATION

- **Corporate context** — jurisdiction, governance form
- **Ethical context** — codified value priorities
- **Market context** — competitive, investment and stakeholder dynamics

→ **Outside the boundary: stop, or escalate to a human director**

THREE LINES OF DEFENCE AROUND THE DOC

1st line · Operations	2nd line · Risk and compliance	3rd line · Board and oversight
Operators and risk owners: logs, incidents, interventions	CRO, AI-risk committee, compliance: policies, standards, monitoring	Board, cyber committee, internal audit: risk appetite, assessment

07 · JURISDICTIONAL NEUTRALITY

One formal structure for every legal system; jurisdiction enters only as parameters

INVARIANT STRUCTURE

Definitions · axioms · integrated admissibility condition

Justice — a vector of compliance with the chosen fairness metrics, F(d)

Equality — admissible versus inadmissible causal pathways

Decision intervals — spatial and temporal bounds on outputs

Acceptable risk — multi-stakeholder vectors with lexicographic priority

JURISDICTION SUPPLIES

Parameters, not structure

The set of protected attributes A

The justice metric ϕ for each attribute

The threshold values θ

The admissible causal pathways

EU anti-discrimination law

US Title VII

Same architecture — swap the parameters, keep the structure

Axiom F4, Jurisdictional Parameterization — from the companion paper on justice, equality and risk

07 · JURISDICTIONAL NEUTRALITY — WHY THE PARAMETERS MUST BE EXPLICIT

“Be fair” is computationally meaningless until a criterion is chosen and documented

Kleinberg et al. · Chouldechova

Impossibility

Calibration and balance for the positive and negative classes cannot all hold when base rates differ across groups. The corporation must choose a criterion, record it and be ready to justify it.

Axiom F1

Non-emptiness

An admissible decision always exists: “escalate to a human director” satisfies every justice constraint by construction.

Polynomial verification

Tractability

Justice metrics, path-specific effects, interval membership and VaR/CVaR risk all verify in polynomial time — fast enough for real-time governance.



Formalization of Justice

Definition 1 (Corporate Decision Space). A corporate decision problem is characterized by a tuple (D, S, A, U) where:

- $D = \{d_1, d_2, \ldots, d_n\}$ is a finite set of possible decisions from which the AI system selects at each decision point (e.g., approve/reject transaction, allocate budget across divisions, appoint candidate to position);
- $S = \{s_1, s_2, \ldots, s_k\}$ is the set of stakeholders affected by decisions (shareholders, employees, customers, suppliers, local communities, regulators);
- $A = \{A_1, A_2, \ldots, A_m\}$ is the set of protected attributes along which discrimination is prohibited (the specific list is determined by applicable law and entered as a parameter);
- For each stakeholder s_i, a utility function $u_i: D \to \mathbb{R}$ assigns a numerical value representing how decision d affects stakeholder s_i's interests.

Justice Vector

Definition 2 (Justice Vector). For each decision d and protected attribute A_j, let $\varphi_j(d) \in [0,1]$ denote the degree to which d satisfies the chosen justice criterion with respect to A_j. The justice vector is $F(d) = (\varphi_1(d), \varphi_2(d), \ldots, \varphi_m(d))$.

The framework does not prescribe which specific justice metric must be used for each component. Table 1 presents a classification of available metrics by their mathematical structure, following Verma and Rubin (2018) and supplemented with causal approaches from Plečko and Bareinboim (2024).

Justice Axioms

Axiom F1 (Non-emptiness of Admissible Set). For each protected attribute A_j, the corporation specifies a threshold θ_j representing the minimum acceptable level of justice. The axiom requires that for every state of the world ω, there exists at least one decision d such that $\varphi_j(d) \geq \theta_j$ for all $j = 1, \ldots, m$.

Formally: $\forall \omega \in \Omega : \exists d \in D : F(d) \geq \theta$

This axiom may be satisfied constructively by including a “null action” or “escalate to human” option that is defined to satisfy all justice constraints trivially.



Axiom F2 (Pareto-Constrained Improvement). If an alternative decision d′ makes at least one stakeholder better off without harming any other stakeholder and without violating justice thresholds, then d′ should be preferred to d. Formally: if $\exists i: u_i(d') > u_i(d)$ and $\forall j \neq i: u_j(d') \geq u_j(d)$ and $F(d') \geq \theta$, then $d \prec d'$.

This axiom adapts the classical Pareto criterion from welfare economics but adds a crucial constraint: improvements cannot be achieved through means that violate justice requirements. A decision that benefits shareholders while discriminating against employees is not a Pareto improvement under this axiom.

Axiom F3 (Non-Exacerbation of Injustice). Let d_0 represent the status quo and d a proposed decision. The axiom prohibits decisions that make the justice situation worse: $\Delta\varphi_j(d) = \varphi_j(d) - \varphi_j(d_0) \geq 0$ for all j.

This axiom operationalizes the moral principle articulated by Hellman (2008): if an action would exacerbate existing injustice, that fact provides a morally significant reason to avoid the action, even if the action is otherwise permissible. Unlike standard fairness metrics that evaluate decisions in isolation, Axiom F3 introduces a historical dimension: the system must track whether the decision moves toward greater justice or entrenches existing disparities.



Axiom F4 (Jurisdictional Parameterization). The formal structure (definitions, axioms, integrated formula) remains invariant across jurisdictions. All jurisdiction-specific content enters through three parameter groups:

(1) the set of protected attributes A;

(2) the choice of justice metric φ_j for each attribute;

(3) the threshold values θ_j.

This enables multinational corporations to implement a single algorithmic architecture while complying with different legal requirements by substituting parameter values.



Definition 10 (Admissible Corporate Decision). Decision d_t is admissible at time t if it belongs to the set:

$$D^*(t) = \{d \in D : F(d) \geq \theta \wedge E(d) = \text{true} \wedge d \in I_t \wedge R(d) \leq R^{max}\}$$

where:

$F(d) \geq \theta$ requires all components of the justice vector to exceed their thresholds;

$E(d) = \text{true}$ requires absence of influence through inadmissible causal pathways;

$d \in I_t$ requires the decision to fall within the admissible interval at time t;

$R(d) \leq R^{max}$ requires all components of the multi-stakeholder risk vector to stay below their limits.

References

1 Романова А. С. Развитие автономных систем искусственного интеллекта для управления корпорациями // Искусственные общества. 2023. Т. 18. Выпуск 2. DOI: 10.18254/S207751800024942-5

2 Романова А.С. Начала законодательства для автономных систем искусственного интеллекта. Надежность. 2024;24(3):10-17. https://doi.org/10.21683/1729-2646-2024-24-3-10-17

3 Романова А. С. Основы моделирования алгоритмических решений при управлении корпорациями // Искусственные общества. 2024. Т. 19. Выпуск 3. DOI: 10.18254/S207751800032184-1

4 Романова А.С. Теория игр для автономных систем искусственного интеллекта при управлении корпорациями. Надежность. 2025;25(2):50-58. https://doi.org/10.21683/1729-2646-2025-25-2-50-58

5 Романова А. С. Прототип модели принятия легитимных и этичных решений автономными системами искусственного интеллекта при управлении корпорациями // Труды МФТИ. — 2025. — Т. 17, № 2 (66). — С. 50-63.

6 Романова А. С. Анализ проблем информативности интерфейса при разработке автономных систем искусственного интеллекта для управления корпорациями // Труды МФТИ. — 2025. — Т. 17, № 2 (66). — С. 64-77.

7 Свидетельство № 2025696048 от 16.12.2025 о государственной регистрации программы для ЭВМ: Программный прототип системы ИИ-директор.

8 Свидетельство № 2026620499 от 03.02.2026 о государственной регистрации базы данных: Синтетические сценарии для обучения и тестирования ИИ-директора.

9 Romanova, A.S. (2026). Modeling of Autonomous Artificial Intelligence Systems for Corporate Management. In: Chowdhury, S., Panigrahi, A., Pati, A. (eds) Artificial Intelligence Powered Ethical Systems, Generative Intelligence, and Societal Futures. AIR 2024. Advances in Science, Technology & Innovation. Springer, Cham. https://doi.org/10.1007/978-3-032-24784-1_26

THANK YOU!

anna.romanova@stern.nyu.edu

romanova.as@phystech.edu

a.romanova@gsom.spbu.ru

https://orcid.org/0009-0004-4649-6037

https://link.springer.com/researchers/92078629SN

https://github.com/iboard-project

