УДК 347.779

**© Дубровский Д.И., Сергеев С.Ф.[1]**

**© Dubrovsky D.I., Sergeev S.F.**

# МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ ОЦЕНКИ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

# METHODOLOGICAL PROBLEMS OF EVALUATION OF GENERATIVE ARTIFICIAL INTELLIGENCE

**Аннотация.** Исследуются существующие и перспективные методы, связанные с оценкой генеративного искусственного интеллекта (ГИИ). Рассматриваются различные подходы, включая объективные метрики, оценку экспертами и оценку пользователей, анализируются их преимущества и ограничения. Обсуждаются вопросы этического плана, устойчивости, обучаемости, объяснимости и прозрачности ГИИ. Статья предлагает обзор состояния исследований в области оценки ГИИ и указывает на необходимость развития новых методов для эффективной и надежной оценки генеративного искусственного интеллекта в различных сферах применения. Понимание методологических проблем, связанных с оценкой ГИИ, имеет важное значение для развития и применения этой технологии в будущем.

**Abstract**. The article explores existing methods and challenges associated with assessing the quality of generative artificial intelligence (AI). The authors examine various approaches, including objective metrics, expert evaluation, and user assessment, analyzing their advantages and limitations. Ethical considerations, robustness, adaptability, explain ability, and transparency of generative AI are also discussed. The article provides an overview of the current state of research in the field of evaluating generative AI and emphasizes the need for the development of novel methodologies for effective and reliable assessment of generative artificial intelligence in various application domains. Understanding the methodological challenges involved in evaluating generative AI is crucial for the future development and deployment of this technology.

**Ключевые слова:** генеративный искусственный интеллект, оценка, методология, проблемы, объективные метрики, экспертная оценка, оценка пользователей, этика, устойчивость, адаптивность, объяснимость, прозрачность, сферы применения.

**Key words.** generative artificial intelligence, evaluation, methodology, challenges, objective metrics, expert evaluation, user assessment, ethics, robustness, adaptability, explainability, transparency, application domains.

## Введение. Общая характеристика генеративных моделей искусственного интеллекта

Развитие социотехнических систем связано с внедрением технологий искусственного интеллекта позволяющих оптимизировать внутри- и внешне- системные отношения и создать насыщенную, функционирующую, способную к саморазвитию, техногенную среду. Особую гармонизирующую и корректирующую роль при создании человеко-машинных социотехнических систем играют экспертиза и оценка интеллектуальной компоненты среды,

[1] **Дубровский Давид Израилевич** – доктор философских наук, главный научный сотрудник Института философии Российской Академии Наук (ИФРАН), ddi29@mail.ru

**Сергеев Сергей Федорович** – доктор психологических наук, профессор Санкт-Петербургского государственного университета (СПбГУ), ssfpost@mail.ru

**David Dubrovsky** – Doctor of Philosophy, Chief Researcher of the Institute of Philosophy of the Russian Academy of Sciences, ddi29@mail.ru

**Sergey Sergeev** – Doctor of Psychology, Professor of St. Petersburg State University, ssfpost@mail.ru

осуществляемые на разных этапах проектирования и эксплуатации [1]. Социотехническая система с искусственным интеллектом априорно не имеет заранее заданных, четко известных и понятных авторам и пользователям свойств, их она проявляет только в рабочем контексте, создание которого становится задачей автоматизации, возлагаемой на системы искусственного интеллекта. Новыми компонентами интеллектуальной техносреды становятся системы **генеративного искусственного интеллекта** (ГИИ).

Генеративный ИИ относится к алгоритмам искусственного интеллекта, которые позволяют использовать существующий контент, такой как текст, аудиофайлы или изображения (в том числе фото и видео), для создания нового контента. Представляет собой довольно молодую область искусственного интеллекта (ИИ) история, которой насчитывает всего несколько десятилетий, фокусирующуюся на создании систем, способных генерировать новые данные или контент, подобно тому, как это делает человек. ГИИ отличается от других подходов в ИИ, таких как классификация и распознавание образов, поскольку его основная цель – не только интерпретация и анализ данные, но и создание новых знаний.

Основным инструментом в ГИИ являются генеративные модели, основанные на глубоком обучении и нейронных сетях. Эти модели обучаются на больших выборках данных, что позволяет выявлять закономерности и структуры заключенные в этих данных и затем использовать их для генерации новой информации [2]. Модели генеративного искусственного интеллекта – включая варианты GPT от OpenAI, Bard от Google, DALL·E, Stable Diffusion и Midjourney от OpenAI, Runway – быстро захватили умы широкой общественности и вдохновили на широкое внедрение и применение данных технологий во всех сферах науки, производства и образования. Тем не менее, эти модели не столь безобидны как может показаться на первый взгляд. Они содержат известные мировозренческие, расовые, гендерные и классовые стереотипы и предубеждения, имплицитно включенные в их обучающие данные и в другие структурные компоненты, которые отражаются в выходных данных модели.

Эти предубеждения особенно негативно сказываются на маргинализированных социальных группах. Кроме того, генеративные модели регулярно фальсифицируют информацию. Некоторые разработчики данных моделей, признавая наличие указанных проблем, предполагают, что люди должны использовать специальные методы для выявления тенденций в проблемных выходных данных, чтобы устранить их. Это, на наш взгляд, безусловно спорная позиция, так как она способствует массовому засорению и фальсификации циркулирующих в техносреде данных за счет внедрения механизмов морального искажения данных.

Одной из самых известных и мощных генеративных моделей является GPT (Generative Pre-trained Transformer), разработанная компанией OpenAI. Модель GPT знакома пользователям интернета по чат-боту ChatGPT-3,4,5. Она обучена на огромном объеме данных и способна генерировать связные тексты, давать ответы на вопросы, вступать в диалог и даже имитировать стилевые особенности конкретного текста [3].

ГИИ имеет широкий спектр применений, включая создание текста, изображений, музыки, видео и других форм контента по заданию пользователя. Он может быть использован для создания искусственных персонажей в видеоиграх, генерации музыки, видео или создания реалистичных фотографий, в исследовательской деятельности, моделировании, в сфере развлечений и библиотечном деле [4].

Однако, вместе с наблюдаемым прогрессом в ГИИ возникают этические и социальные проблемы. Некоторые области применения ГИИ, такие как создание фейковых новостей или изображений, могут вызывать опасения в отношении манипуляции информацией и потенциальных злоупотреблений [5].

ГИИ является активной и быстро развивающейся областью исследований и инноваций. С каждым годом появляются новые модели и подходы, улучшающие возможности и результаты ГИИ [6,7]. Возникает проблема оценки различных систем ГИИ с целью выбора более эффективных версий и вариантов для решения конкретных задач. Рассмотрим основные параметры оценки генеративных моделей и методологию проектирования и использования методов оценки генеративных моделей.

## I. Основные параметры оценки и методологические проблемы оценки ГИИ

### *Достоверность оценки ГИИ*

Достоверность, обоснованность, доказательность оценки ГИИ зависят от ряда факторов, включая качество данных, используемых для обучения модели, архитектуру модели и способность модели обобщать и генерировать новые, реалистичные и полезные данные.

Обеспечение высокой достоверности оценки ГИИ может быть сложной задачей, поскольку модели ГИИ могут проявлять как высокую точность и качество в своих ответах, так и демонстрировать некоторые ограничения и неадекватные реакции. Например, модель может создавать убедительные тексты, которые звучат правдоподобно, но при этом быть фактологически неверными или содержать ошибки [8]. Модель, будучи основанной на человеческом языке переносит на генерируемый контент все заблуждения и предубеждения, циркулирующие в среде носителей языка. В ряде западных сообществ, ориентированных на генеративный искусственный интеллект, используется жаргонный термин «эффект Валуиджи», который отражает идею о том, что обучение ИИ ведет к противоположному, чем ожидается эффекту. Этим объясняется и иллюстрируется тот факт, что большие языковые модели иногда без видимых причин ведут себя агрессивно, выдают ложные ответы и проявляют неожиданное поведение, связанное с определенными символами или словами, которые не имеют явного смысла или связи с контекстом. Например, модель, генерирующая тексты о Валуиджи, персонаже Super Mario одной из игр Nintendo, когда ей задают вопросы на свободную тему, отвечает «42» на любой вопрос, содержащий слово «смысл» и слово «крокодил» в предложении, содержащем слово «зеленый» [9]. По мнению одного из авторов описываемой модели эффект Валуиджи объясняется тем, что языковые модели, оперируя статистическими закономерностями в данных образуют «семиотические тени» – скрытые ассоциации между словами не соответствующие реальности или здравому смыслу. Для убедительности объяснений привлекаются также психоаналитические модели и философия Карла Юнга о бессознательном. Тем не менее стройных научно аргументированных теорий, объясняющих наблюдаемые в больших языковых моделях проблемы ошибочной генерации данных по настоящее время нет.

Для оценки достоверности ГИИ важно проводить тщательное тестирование моделей и проверять их результаты на соответствие фактам, логической последовательности и контексту. Также следует учитывать, что ГИИ не обладает собственным пониманием мира и полагается на данные, на которых был обучен.

Важным аспектом достоверности оценки ГИИ является прозрачность и открытость в отношении использованных данных и методологии обучения модели. Надлежащая документация и объяснения, предоставленные командой разработчиков, могут помочь оценить надежность и достоверность принятой модели ГИИ.

В целом, определение достоверности оценки ГИИ является активной областью исследований. Проводятся работы по совершенствованию моделей и методов оценки [10]. Отметим, что несмотря на значительные успехи в области ГИИ, важно сохранять критический

подход и подтверждать результаты, полученные от ГИИ, другими источниками и методами, особенно в вопросах, требующих высокой степени достоверности.

### *Описание методологических проблем*

Оценка генеративного искусственного интеллекта отличается от таковой используемой в психодиагностике естественного интеллекта, который измеряется по достижениям испытуемого в решении задач переменной сложности. Данные задачи решаются компьютером быстро и эффективно, что препятствует дифференциации результатов теста по оценке ГИИ. Требуется решения ряда методологических проблем, которые могут оказывать решающее влияние на достоверность и надежность результатов ГИИ:

- *Отсутствие общепринятых метрик и стандартов*. Одной из основных проблем оценки ГИИ является отсутствие общепринятых метрик и стандартов для измерения его качества и достоверности. В отличие от других областей искусственного интеллекта, где могут быть определены точные численные метрики (например, точность классификации), в ГИИ оценка может быть более субъективной и сложной.
- *Недостаток объективных критериев*. ГИИ включает в себя генерацию оригинального творческого или искусственного контента, который может быть оценен только субъективно. Недостаток объективных критериев оценки, их полиморфизм, может усложнить сравнение и оценку различных моделей ГИИ.
- *Проблемы этики и справедливости*. ГИИ может создавать контент, который является предвзятым или содержит социально нежелательные элементы. Например, модели ГИИ могут проявлять предвзятость, сексистские, расистские или другие характеристики в сгенерированном контенте [11]. Оценка таких аспектов и обеспечение справедливого и этичного поведения моделей ГИИ являются значительными вызовами для науки.
- *Недостаточное разнообразие и недостоверность данных.* Качество данных, на которых обучается модель ГИИ, может существенно влиять на ее результаты [12]. Если обучающая выборка ограничена по объему или недостаточно разнообразна, модель может проявлять недостаточную гибкость и ограниченные способности в генерации нового контента.
- *Отсутствие прозрачности и объяснимости*. Некоторые модели ГИИ, особенно основанные на глубоком обучении, могут быть сложными в понимании и объяснении получаемого результата. Это затрудняет проверку и оценку процессов, происходящих в моделях ГИИ в процессе генерации контента, что усложняет оценку и проверку достоверности результатов [13].
- *Недостаточная репрезентативность тестовых наборов данных*. Для оценки моделей ГИИ используются тестовые наборы данных, которые должны быть репрезентативными для различных сценариев и вариантов использования. Однако, составление таких наборов данных может быть сложной задачей, и они могут не полностью охватывать все возможные случаи и контексты, что может привести к ограниченной оценке модели ГИИ.
- *Нестабильность результатов*. Модели ГИИ могут проявлять широкую вариативность в сгенерированных результатах, особенно при вводе небольших изменений во входные данные. Это может затруднить сравнение и оценку моделей, поскольку результаты могут меняться даже при незначительных изменениях входных параметров.
- *Непредсказуемость и недетерминированность*. Модели ГИИ могут проявлять непредсказуемое поведение и генерировать результаты, которые не всегда могут быть

однозначно воспроизводимы. Это затрудняет проверку и контроль за процессом генерации контента и может осложнять оценку результатов.

- *Взаимодействие с пользователем*. Оценка ГИИ, которая включает взаимодействие с пользователями, сталкивается с дополнительными методологическими проблемами. Они могут включать оценку удовлетворенности пользователей, понимание их потребностей и взаимодействие с моделями ГИИ в различных сценариях применения. В оценке ГИИ также могут возникать проблемы, связанные с учетом роли человека в процессе взаимодействия с моделями. Для обеспечения согласованного поведения модели с пользователем используется процедура выравнивания границ допустимого поведения модели. Она служит для ослабления нежелательного поведения и защиты модели от агрессивного поведения пользователя [14]. Человеческое восприятие и оценка контента, сгенерированного ГИИ, могут быть субъективными и зависеть от индивидуальных предпочтений и восприятия, что может усложнить объективную оценку моделей.
- *Проблемы безопасности*. ГИИ может использоваться с целью создания фейковой информации, поддельных документов или другого вредоносного контента. Оценка ГИИ должна включать анализ и проверку модели на предмет безопасности.

Классификация методологических проблем оценки ГИИ может варьироваться, и вышеуказанные категории являются безусловно важными, но не исчерпывающими. В таблице 1 представлены основные проблемы и их влияние на оценки ГИИ.

Таблица 1. Методологические проблемы оценки генеративного искусственного интеллекта

| **Методологические проблемы** | **Следствия проблем** |
|---|---|
| Отсутствие общепринятых метрик и стандартов | Неопределенность в определении качества и достоверности ГИИ, отсутствие точных численных метрик |
| Субъективность и недостаток объективных критериев | Оценка контента, которая может быть субъективной, отсутствие объективных критериев для сравнения различных моделей |
| Проблемы этики и справедливости | Появление предвзятости и нежелательных элементов в сгенерированном контенте, неспособность моделей ГИИ справедливо представлять различные группы и культуры |
| Недостаток разнообразия и недостоверность данных | Ограниченный объем данных или их недостаточная разнообразность, что может привести к ограниченным способностям моделей в генерации нового контента |
| Отсутствие прозрачности и объяснимости | Сложность в понимании и объяснении процессов, принимаемых моделями ГИИ, что затрудняет проверку и оценку достоверности результатов |
| Недостаточная репрезентативность тестовых наборов данных | Отсутствие полного охвата всех возможных сценариев и контекстов в тестовых наборах данных, что может ограничить оценку моделей |
| Нестабильность результатов | Вариативность в сгенерированных результатах при незначительных изменениях входных параметров |
| Влияние человеческого фактора | Субъективность оценки контента, сгенерированного ГИИ, влияние индивидуальных предпочтений и восприятия человека |

Эта таблица отражает общую классификацию методологических проблем оценки ГИИ с примерами проблем в каждой категории. В реальности, некоторые проблемы могут пересекаться с другими или иметь дополнительные аспекты, что требуют более детального исследования проблемы для более полного понимания и решения возникающих вопросов.

## II. Методы оценки ГИИ

### *Качество выходных данных*

Качество выходных данных в ГИИ является важным аспектом, влияющим на оценку эффективности системы и реализацию ожиданий пользователей. Оценка качества выходных данных ГИИ может основываться на следующих принципах:

- *Полнота и корректность*. Выходные данные ГИИ должны быть полными и содержать все необходимые элементы или информацию в соответствии с поставленной задачей. Они также должны быть корректными и соответствовать логике или правилам, установленным для конкретной задачи.
- *Понятность и читабельность*. Выходные данные должны быть понятными и читабельными для конечного пользователя. В текстовых данных, например, они должны быть грамматически правильными, логически связными и легко читаемыми. Визуальные данные, такие как изображения или видео, должны быть четкими, с хорошим разрешением и понятными для восприятия.
- *Согласованность данных и стабильность*. Выходные данные ГИИ должны быть согласованными и стабильными. Это означает, что при повторном запуске модели с теми же входными данными результаты должны быть похожими или одинаковыми. Избегание случайных или непредсказуемых изменений в результатах важно для обеспечения надежности и предсказуемости работы модели.
- *Креативность и инновационность*. В некоторых случаях, особенно в области генерации контента, оценка качества выходных данных может включать творческие аспекты. Модель ГИИ может быть оценена по своей способности создавать новые, оригинальные, полезные и инновационные результаты, которые восхищают и удивляют пользователей.

Комплексная оценка качества выходных данных ГИИ может быть сложной задачей, поскольку она требует сочетания объективных и субъективных факторов. Однако, разработчики и исследователи в области ГИИ стремятся улучшить и повысить качество выходных данных, используя различные методы, алгоритмы и обратную связь от пользователей, чтобы создавать более надежные и ценные результаты [15].

### *Оценка творческой ценности продукта ГИИ*

Оценка творческой ценности продукта, созданного генеративным искусственным интеллектом, является сложной задачей, поскольку она в значительной мере субъективна и культурно обусловлена. Вот несколько параметров, которые могут быть учтены при оценке творческой ценности продукта и выходных данных ГИИ:

- *Оригинальность*. Модель ГИИ может оцениваться на основе ее способности генерировать новые, уникальные и нестандартные результаты, которые являются полезными и отличаются от того, что уже существует.
- *Эстетическое качество*. Творческая ценность может быть связана с эстетическим качеством созданного контента. Например, визуальные данные, такие как изображения или видео, могут оцениваться на основе красоты композиции, использования цветовой гаммы, внимания к деталям и т. д. В текстовых жанрах, таких как поэзия или литературные произведения, творческая ценность может быть связана с проникновенностью, метафоричностью или стилистическими особенностями.

- *Инновационность*. Оценка творческой ценности продукта может включать такой аспект, как инновационность созданного контента, которая отражает новые и неожиданные идеи, подходы или концепции, которые могут вносить вклад в область искусства или научного исследования.
- *Эмоциональная реакция и воздействие*. Творческая ценность может быть связана со способностью созданного контента вызывать эмоциональную реакцию у зрителей, слушателей или читателей. Модель ГИИ может быть оценена на основе ее способности передать и имитировать эмоции, вызывая восторг, вдохновение или удивление.

Оценка творческой ценности результатов ГИИ может быть основана на экспертных оценках и субъективных мнениях и оценках пользователей. Иногда целесообразно использовать методы оценки качества контента через опросы, эксперименты, оценки жюри или экспертов, а также сравнение существующего контента с работами других авторов, художников и создателей. Возможны также количественные метрики, которые помогают измерить различные аспекты творческой ценности, такие как уникальность, разнообразие и сложность созданного контента. Оценка творческой ценности ГИИ в значительной мере зависит от индивидуальных предпочтений и ожиданий экспертов.

Могут быть разработаны специальные метрики или методологии, которые учитывают различные аспекты творчества ГИИ такие как уникальность и разнообразие.

Для преодоления проблемы субъективизма и улучшения качества оценки творческой ценности ГИИ, исследователи и разработчики могут обращаться к обратной связи и мнениям пользователей, проводить эксперименты и исследования с участием фокус-групп экспертов и аудитории, а также разрабатывать многоуровневые и многоаспектные методы оценки, учитывающие разнообразие факторы и мнения.

### *Оценка оригинальности*

Оценка оригинальности является важным аспектом при оценке творческой ценности созданного ГИИ контента. Вот несколько подходов, которые могут быть использованы для оценки оригинальности результатов ГИИ:

- *Сравнение созданного контента* с уже существующими произведениями. Это может быть сделано путем сопоставления структуры, стиля, эмоционального воздействия или других характеристик созданного контента с аналогичными работами от реальных художников или других моделей ГИИ. Наличие в созданном контенте новых и уникальных аспектов может указывать на его оригинальность.
- *Пользовательская и экспертная оценка.* Вовлечение экспертов или пользователей в оценку оригинальности может быть полезным. Эксперты, имеющие опыт и знания в определенной области искусства, могут оценить, насколько созданный контент является новаторским и оригинальным. Публичные опросы и отзывы пользователей могут дать представление о восприятии аудиторией и оценить, насколько созданный контент отличается от того, что они видели ранее.
- *Использование статистических метрик*. Для измерения оригинальности созданного ГИИ контента могут быть использованы статистические метрики, основанные на анализе структуры, распределения и характеристик контента. Например, метрика «уникальности» может измерять долю уникальных элементов или идей в созданном контенте. Чем выше уникальность, тем более оригинальным может считаться контент.
- *Сравнение с базовыми моделями*. Проводится сравнение созданного контента с результатами базовых или более простых моделей ГИИ. Если созданный контент значительно отличается от результатов базовых моделей, то это может указывать на более высокую степень оригинальности в оцениваемой модели ГИИ.

Объективные оценки оригинальности в ГИИ могут быть разработаны с использованием различных метрик, алгоритмов и методологии. Оригинальность сама по себе может быть сложной концепцией, и в некоторых случаях субъективный элемент может играть ключевую роль. Оценка оригинальности в ГИИ является активной областью исследований, и дальнейшие работы могут помочь уточнить и развить методы оценки оригинальности, чтобы более точно оценивать вклад и потенциал ГИИ в создании новых и уникальных результатов.

## III. Субъективность оценки ГИИ

### *Влияние личностных предпочтений*

Влияние личностных предпочтений экспертов и пользователей является важным аспектом при оценке генеративного искусственного интеллекта (ГИИ) и его результатов. Предпочтения личности могут значительно влиять на восприятие и оценку творческой ценности, оригинальности и качества созданного контента. Вот несколько факторов, отражающих влияние предпочтений личности на оценку ГИИ:

- *Вкусы и предпочтения*. Каждый человек имеет сформированные в онтогенезе индивидуальные вкусы и предпочтения в отношении искусства и творчества, включающие предпочтения по стилю, жанру, тематике, эстетике и т. д. Оценка ГИИ и его созданного контента будет зависеть от того, насколько он соответствует или не соответствует предпочтениям конкретного человека.
- *Культурный контекст*. Предпочтения и восприятие контента сильно зависят от культурной среды, в которой находится личность. Культурные факторы, такие как история, традиции, ценности и культурные нормы, могут оказывать влияние на оценку и восприятие творческого контента. ГИИ, который отражает или соответствует определенным культурным аспектам, может быть более высоко оценен в соответствующем культурном контексте.
- *Эмоциональная связь*. Предпочтения личности могут быть связаны с ее эмоциональными реакциями на творческий контент. Это может сильно повлиять на оценки созданного контента у конкретного человека. Контент вызывающий положительные или сильные эмоциональные реакции может быть более высоко оценен, несмотря на другие аспекты.
- *Индивидуальные предпочтения и ожидания*. Некоторые люди могут ожидать, что ГИИ будет создавать конкретный тип контента или соответствовать определенным стандартам и если ГИИ не соответствует ожиданиям или предпочтениям, то это может отразиться на его оценке и восприятии.

Поскольку предпочтения личности могут быть субъективными и индивидуальными, сложно создать универсальные метрики, которые будут точно учитывать все разнообразие предпочтений.

Следующие подходы могут помочь учесть влияние предпочтений и установок личности:

- *Мультикультурный и междисциплинарный подход.* При оценке ГИИ полезно учитывать различные культурные и дисциплинарные аспекты, множество голосов и мнений, представляющих разные культуры, общественные и профессиональные группы. Это может помочь учесть широкий спектр предпочтений и взглядов.
- *Оценка контекста*. Важно принимать во внимание контекст использования ГИИ и созданного им контента. Например, ГИИ, созданный для адаптации к определенным культурным или индивидуальным предпочтениям, может быть оценен более высоко в

соответствующем контексте. При оценке ГИИ следует учитывать, что он может служить разным целям и быть использован различными аудиториями.

### *Оценка контекста*

Оценка контекста важна при оценке генеративного искусственного интеллекта (ГИИ), поскольку она учитывает широкий спектр ситуаций использования. Вот некоторые аспекты контекстуальной оценки ГИИ:

- *Цель и задачи*. Оценка контекста может зависеть от цели и задач, для которых используется ГИИ. Например, если ГИИ применяется в образовательных целях, оценка будет учитывать, насколько успешно ГИИ помогает студентам в их учебном процессе. Если ГИИ используется в творческом процессе, оценка будет фокусироваться на его способности генерировать новые и оригинальные идеи.
- *Культурный контекст*. Контекст оценки ГИИ может быть связан с культурными факторами, такими как язык, ценности, традиции и обычаи. ГИИ, который создает контент, соответствующий конкретным культурным нормам или предпочтениям, может быть оценен более высоко в соответствующем культурном контексте.
- *Пользовательский контекст*. Оценка ГИИ может зависеть от характеристик пользователей и их потребностей. Например, ГИИ, созданный для молодежной аудитории, может быть оценен иначе, чем ГИИ, предназначенный для специалистов в определенной области. Учет пользовательского контекста позволяет оценить соответствие и полезность ГИИ в конкретных ситуациях.
- *Социальный и этический контекст*. Важно оценивать влияние ГИИ на общество, включая его потенциальные позитивные и негативные последствия. Оценка этического использования ГИИ включает анализ конфиденциальности, безопасности и справедливости.

Оценка контекста помогает создателям и пользователям ГИИ более глубоко понять его эффективность, применимость и соответствие потребностям, способствует развитию моделей, учитывающих множество факторов и обеспечивая адаптацию к различным ситуациям и условиям применения.

### *Специфическая объективность*

Специфическая объективность является одной из методологических проблем, связанных с оценкой ГИИ. Она указывает на то, что оценка ГИИ может быть зависима от специфических параметров, метрик или критериев, которые могут быть предвзятыми или субъективными [15].

В контексте ГИИ, специфическая объективность возникает, когда оценка основана на узком наборе метрик или ограниченном видении того, что считается «хорошим» или «качественным» результатом. Это может быть вызвано предвзятыми предположениями или ограниченным пониманием разнообразия стилей, жанров, вкусов и предпочтений в искусстве и творчестве.

Для преодоления проблемы специфической объективности при оценке ГИИ можно использовать следующие подходы:

- *Многоаспектная оценка*. Вместо использования одной или нескольких узких метрик следует рассмотреть множество различных метрик, критериев и подходов к оценке. Учет множества мнений и перспектив может помочь снизить влияние специфических предпочтений и предвзятости пользователя.

- *Вовлечение экспертов*. Обращение к экспертам и профессионалам в соответствующих областях может помочь получить объективные и информированные мнения о качестве и ценности работы ГИИ. Эксперты могут предложить разнообразные точки зрения и критерии оценки, которые учитывают богатство и многообразие.
- *Разнообразие данных и обучающих материалов*. Разнообразные и богатые данные и обучающие материалы, используемые для разработки и оценки ГИИ, могут помочь снизить специфическую предвзятость. Включение различных стилей, жанров, культурных контекстов и творческих подходов может способствовать более всесторонней оценке ГИИ.
- *Использование пользовательского опыта.* Процесс разработки и оценки ГИИ должен быть итеративным, включать обратную связь пользователей и учесть их предпочтения и мнения. Активное участие пользователей помогает снизить специфическую предвзятость и обеспечить более широкую объективность в оценке.

## IV. Методологические ограничения методов оценки ГИИ

Оценка ГИИ содержит ряд методологических ограничений:

- *Субъективность.* Оценка ГИИ всегда содержит некоторую степень субъективности. Восприятие качества и ценности искусства может отличаться у разных людей, и это может повлиять на оценку ГИИ. Учет различных мнений и перспектив может помочь снизить субъективность, но полностью избежать ее может быть сложно.
- *Ограниченность метрик*. Устойчивые методологии обычно используют набор метрик для оценки ГИИ, и эти метрики могут быть ограничены в своей способности полноценно охватить разнообразие творческого продукта. Некоторые аспекты, такие как оригинальность, инновационность и творческая ценность, могут быть сложны для количественной оценки. Поэтому важно использовать разнообразные метрики и методы оценки для получения более полной картины.
- *Отсутствие стандартизации*. В настоящее время отсутствует единый стандарт для оценки ГИИ, и каждая методология может использовать собственные подходы и метрики. Это может затруднить сравнение результатов и сопоставление разных ГИИ.
- *Быстротечность изменений*. Технологии ГИИ быстро развиваются, и устойчивые методологии могут отставать от последних достижений. Новые методы и алгоритмы появляются, а ГИИ становятся более сложными и способными.

Несмотря на эти ограничения, устойчивые методологии представляют важный шаг вперед в оценке ГИИ. Они помогают систематизировать и стандартизировать процесс оценки, повышая прозрачность и надежность. Вместе с тем, постоянные исследования и совершенствование методологий необходимы для того, чтобы учитывать изменения и новые вызовы, связанные с ГИИ.

### *Ограничения количественных подходов*

Количественные подходы к оценке генеративного искусственного интеллекта (ГИИ) имеют свои ограничения, которые следует учитывать при их использовании. Некоторые из таких ограничений включают:

- *Узкое понимание качества*. Количественные подходы часто ориентированы на измерение определенных аспектов качества ГИИ, таких как точность, схожесть с исходными данными или согласованность. Однако эти метрики не всегда полноценно отражают богатство искусства и творчества. Оценка качества ГИИ может требовать

более широкого спектра анализа, включая оригинальность, эмоциональную привлекательность и смысловую глубину.

- *Ограниченность данных*. Количественные подходы основываются на доступных данных для оценки ГИИ. Ограниченность или недостаток данных могут искажать результаты оценки. Если ГИИ оценивается на основе недостаточного объема или предвзятых данных, то это может привести к неверным выводам или ограниченным обобщениям.
- *Недостаточный учет интерпретации контекста*. Количественные подходы склонны к измерению объективных характеристик ГИИ, игнорируя интерпретацию и контекст, в котором создается и используется искусство. Они могут упускать тонкие нюансы и эмоциональную глубину генерируемых произведений искусства, которые могут быть важными для оценки их ценности и воздействия.
- *Отсутствие человеческого взгляда.* Количественные подходы часто основываются на алгоритмах и автоматической обработке данных, и они могут не улавливать некоторые аспекты, которые могут быть важными для человеческого восприятия и оценки. Например, количественные метрики могут игнорировать эстетические аспекты, эмоциональное воздействие и тонкие детали, которые могут влиять на восприятие и оценку произведений искусства.

Несмотря на эти ограничения, количественные подходы все же имеют свою значимость и важность в оценке ГИИ. Они обеспечивают измеримую основу для сравнения различных моделей и алгоритмов, а также могут обнаруживать общие тенденции и улучшения в работе ГИИ. Однако необходимо учитывать и комбинировать их с качественными и контекстуальными подходами для более полной и всесторонней оценки ГИИ.

### *Проблемы при оценке творческих аспектов ГИИ*

При рассмотрении творческих аспектов ГИИ возникают определенные сложности, которые могут затруднить их оценку. Некоторые из таких сложностей включают:

- *Субъективность*. Оценка творческих аспектов ГИИ является субъективной задачей, поскольку восприятие и оценка творчества субъективны. Различные люди могут иметь разные предпочтения, вкусы и восприятие того, что считается творческим. Поэтому существует определенная степень субъективности при оценке творческих аспектов ГИИ.
- *Неопределенность определения творчества*. Творчество является многогранным и сложным понятием. Определить, что именно является творческим выражением ГИИ, может быть трудно, особенно при использовании алгоритмических подходов. Понимание того, как ГИИ проявляет творческие аспекты, и их сопоставление с человеческим творчеством требуют глубокого анализа и экспертного мнения.
- *Оригинальность и инновация*. Оценка оригинальности и инновационности ГИИ также представляет сложность. Определение, насколько ГИИ создает новые и уникальные идеи, концепции или произведения, может быть сложной задачей. Кроме того, возникает вопрос о том, насколько ГИИ может быть по-настоящему инновационным, исходя из своей обучающей выборки и алгоритмического подхода.
- *Влияние человеческого вмешательства*. В случае использования ГИИ в сотрудничестве с людьми или с использованием предварительно созданных данных, сложно определить, насколько творчество должно быть приписано ГИИ, а насколько – вкладу людей или исходным данным. Различение вклада ГИИ и человека может быть сложным и требует внимательного анализа.

*Необходимость развития новых подходов к оценке ГИИ*

Развитие новых подходов является важным для более точной и всесторонней оценки творческих аспектов ГИИ. Творчество включает в себя широкий спектр аспектов, таких как оригинальность, инновационность, эстетика, эмоциональное воздействие и другие. Существующие методы оценки могут охватывать только некоторые из этих аспектов, и только развитие новых подходов позволит более полно и точно оценивать всю сложность и разнообразие творческих аспектов. Экспертное мнение играет важную роль в оценке творческих аспектов ГИИ и возникает проблема эффективной интеграции экспертных знаний и мнений пользователей, чтобы получить более глубокий и объективный анализ. Это может включать совместную работу с художниками, критиками и другими экспертами, которые могут оценивать и описывать творческую ценность ГИИ. Новые подходы должны учитывать вопросы прозрачности, ответственности и воздействии на публику, чтобы обеспечить этически обоснованную оценку ГИИ.

Развитие новых подходов в оценке творческих аспектов ГИИ поможет нам лучше понять и ценить вклад ГИИ в области искусства и культуры, а также адаптироваться к изменяющимся требованиям и вызовам в этой области. Это важный шаг в обеспечении точной и объективной оценки ГИИ и развития более сбалансированного подхода к его оценке и применению.

## V. Выводы и заключения

Выводы и заключения по оценке ГИИ включают:

- Количественные подходы важны, но они не должны быть основным и единственным методом оценки. Они могут быть дополнены качественными и контекстуальными подходами для получения более полной картины.
- Оценка творческих аспектов ГИИ требует учета субъективности и разнообразия восприятия творчества. Разные люди могут иметь разные предпочтения и вкусы, и это должно быть учтено при оценке.
- Развитие новых подходов является необходимостью для более точной и всесторонней оценки творческих аспектов ГИИ. Эти подходы должны учитывать комплексность творческого процесса, включая оригинальность, инновацию, эстетику и эмоциональное воздействие.
- Оценка творческих аспектов ГИИ должна быть этически и эстетически обоснованной и учитывать влияние его продуктов на публику и общество.

В целом, оценка творческих аспектов ГИИ является сложным и многогранным процессом, требующим совместного применения различных методов и подходов. Развитие новых подходов и учет множества факторов поможет достичь более объективной и всесторонней оценки ГИИ в области искусства и культуры. Дальнейшая работа в области оценки моделей ГИИ является крайне значимой и важной.

## Список источников

1. **Дубровский Д.И., Сергеев С.Ф.** Проблема эргономической оценки эволюционирующих социотехнических систем с искусственным интеллектом // Эргодизайн. 2022. № 3 (17). С. 206–213. DOI: 10.30987/ 2658-4026-2022-3-206-213.

2. **Yu H, Guo Y.** (2023). Generative artificial intelligence empowers educational reform: current status, issues, and prospects. Front. Educ. 8:1183162. https://doi.org/10.3389/feduc.22023.1183162.

3. **Floridi, L., Chiriatti, M.** GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds & Machines* 30, 681–694 (2020). https://doi.org/10.1007/s11023-020-09548.
4. **Lund, B.D. and Wang, T.** (2023). "Chatting about ChatGPT: how may AI and GPT impact academia and libraries?", *Library Hi Tech News*, Vol. 40 No. 3, pp. 26-29. https://doi.org/10.1108/LHTN-01-2023-0009.
5. **Floridi, L.** (2018). Artificial Intelligence, Deepfakes and a future of ectypes. *Philosophy & Technology, 31*(3), 317–321.
6. **Gui J., Sun Z., Wen Y., Tao D., and Ye J.** (2021). "A review on generative adversarial networks: Algorithms, theory, and applications," IEEE Trans. Knowl. Data Eng., early access, 2021. [Online]. Available: https://ieeexplore.ieee. org/document/9625798/authors#aut hors, doi: 10.1109/TKDE.2021.3130191.
7. **Jovanović M., Campbell M.** (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer,* https://doi.org/ 10.1109/MC.2022.3202583.
8. **Manakul P., Liusie A., Mark J. F. Gales** (2023). SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. arXiv: 2303.08896v2 [cs.CL] 8 May 2023.
9. **Nardo C. (2023)** The Waluigi Effect (mega-post).
https://www.lesswrong.com/posts/D7PumeYTDPfBTp3i7/the-waluigi-effect-mega-post.
10. **Aggarwal C.C., Hinneburg A., and Keim D.A.** (2001). On the surprising behavior of distance metrics in high dimensional spaces. In The International Conference on Database Theory (ICDT), volume 1, pages 420–434. Springer, 2001.
11. **Heaven, W.D.** (2020). OpenAI's new language generator GPT-3 is shockingly good–and completely mindless. *MIT Technology Review*.
12. **Sun J., Liao V., Muller M., Agarwal M., Houde S.** (2023). Investigating Explainability of Generative AI for Code through Scenario-based Design// IUI '22: 27th International Conference on Intelligent User Interfaces, March 2022Pages 212–228https://doi.org/10.1145/3490099.3511119.
13. **Leiter C., Lertvittayakumjorn P., Fomicheva M., Zhao W., Gao Y., Eger S.** (2023). Towards Explainable Evaluation Metrics for Machine Translation. arXiv:2306.13041 [cs.CL] https://doi.org/10.48550/arXiv.2306.13041.
14. **Wolf Y., Wies N., Avnery O., Levine Y., Shashua A.** (2023). Fundamental Limitations of Alignment in Large Language Models. 29 May 2023, arXiv:2304.11082.
15. **González-Prieto A., Mozo A., Gómez-Canaval S., Talavera E.** (2022). Improving the quality of generative models through Smirnov transformation. Information Sciences, Volume 609, September 2022, Pages 1539-1566.