Galactica Media: Journal of Media Studies. 2024. No 4 | ISSN: 2658-7734
Животные в современных медиа | https://doi.org/10.46539/gmd.v6i4.554

Let's Practice Better... on Cats: Description and Visualisation of Artistic Images in Generative AI Models

Ruslan Khandogin[1] & Nina S. Proner[2]

Novosibirsk State Technical University. Novosibirsk, Russia

Received: 27 July 2024 | Revised: 19 October 2024 | Accepted: 27 October 2024

Abstract

Artificial Intelligence (AI) plays an increasingly prominent role in various spheres of life in today's world, including generation of a variety of visual content from selfie stream processing to creating works of digital art. The present paper raises the question of whether AI is capable of creating real art or it just imitates its external form. The paper examines the specificity of prompts: from concrete named ones to interpretive descriptive queries in linguistic, artistic and socio-cultural contexts. The article dwells upon some important aspects of evaluating the quality of keyword extraction algorithms and their relation to artistic practice. The authors rely on semiotic analysis to uncover encoded meanings and imports in the text. The article emphasises that the literary text is at the top of the hierarchy of cultural texts; it is characterised by intentionality and coherence and represents a complex semantic field where key words and images interact with the explicit and implicit contexts. The study examines and analyses the visualised images of Cheshire Cat, Cat Behemoth and Tomcat Murr created by the authors with the use of three generative neural networks: Stable Diffusion, Dall-E and Kandinsky. Understanding and visualising the literary text by generative systems and models realising specific algorithms requires the ability to reveal its multilayered semantics and connection with the cultural context, which ultimately helps to understand the in-depth meanings of the work and its place in culture. Consideration of the operational quality of algorithms for keyword system extraction and image generation is deemed possible from the point of view of their structural organisation. Generative algorithms create an imitative reality, while the immanence of the artistic value determines the uniqueness and meanings of the created figurative world. The article can be useful to anyone interested in the substance and specificity of digital art, the relationship between technological innovations and socio-cultural context, the creation and visualisation of artistic images in generative AI models, their conceptualisation and interpretation.

Keywords

Digital Art; Generative Models; Neural Networks; Artistic Image; Visualisation; Socio-Cultural Context; Prompt; DALL·E; Stable Diffusion; Kandinsky

This work is licensed under a Creative Commons "Attribution" 4.0 International License

1 Email: khandogin.ruslan[at]yandex.ru ORCID https://orcid.org/0000-0003-3280-7055
2 Email: kafedr.p[at]yandex.ru ORCID https://orcid.org/0000-0001-6482-8382

Galactica Media: Journal of Media Studies. 2024. No 4 | ISSN: 2658-7734
Animals in Contemporary Media | https://doi.org/10.46539/gmd.v6i4.554

Потренируемся лучше... на котах: описание и визуализация художественных образов в генеративных моделях ИИ

Хандогин Руслан Викторович[1], **Пронер Нина Сергеевна**[2]

Новосибирский государственный технический университет. Новосибирск, Россия

Рукопись получена: 27 июля 2024 | Пересмотрена: 19 октября 2024 | Принята: 27 октября 2024

Аннотация

В современном мире искусственный интеллект (ИИ) играет все более заметную роль в различных сферах жизни, включая генерацию разнообразного визуального контента от потоковой обработки селфи до создания произведений цифрового искусства. В данной статье поднимается вопрос о том, способен ли ИИ создавать настоящее искусство или он лишь имитирует его внешнюю форму. В работе рассматривается специфика промптов: от конкретных именованных до интерпретативных описательных запросов в языковом, художественном и социокультурном контекстах. Затрагиваются важные аспекты оценки качества работы алгоритмов извлечения ключевых слов и их связь с художественной практикой. Авторы опираются на семиотический анализ, позволяющий раскрывать закодированные смыслы и значения в тексте. В статье подчеркивается, что художественный текст представляет собой вершину иерархии текстов культуры, характеризуется содержательностью и когерентностью, представляет собой сложное семантическое поле, где ключевые слова и образы взаимодействуют с эксплицитным и имплицитным контекстом. В рамках проведённого исследования рассматриваются и анализируются визуализации образов Чеширского Кота, Кота Бегемота и Кота Мурра, созданные авторами с помощью трёх генеративных нейросетей: Stable Diffusion, DallE и Kandinsky. Для понимания генеративными системами и моделями, реализующими те или иные алгоритмы, художественного текста и его визуализации требуется умение раскрывать его многослойную семантику и связь с культурным контекстом, что в итоге помогает понять глубинные значения произведения и его место в культуре. Представляется возможным рассмотрение качества работы алгоритмов извлечения системы ключевых слов и формирования образов с точки зрения их структурной организации. Генеративные алгоритмы создают имитативную реальность, а имманентность художественной ценности искусства определяет уникальность и смысл созданного образного мира. Статья будет полезна всем, кто интересуется сущностью и спецификой цифрового искусства, взаимосвязью технологических инноваций и социокультурного контекста, созданием и визуализацией художественных образов в генеративных моделях ИИ, их осмыслением и интерпретацией.

Ключевые слова

цифровое искусство; генеративные модели; нейросети; художественный образ; визуализация; социокультурный контекст; промпт; DALL·E; Stable Diffusion; Kandinsky

Это произведение доступно по лицензии Creative Commons “Attribution” («Атрибуция») 4.0 Всемирная

1 Email: khandogin.ruslan[at]yandex.ru ORCID https://orcid.org/0000-0003-3280-7055
2 Email: kafedr.p[at]yandex.ru ORCID https://orcid.org/0000-0001-6482-8382

Введение

Наш мир стремительно меняется, мы живем в удивительную эпоху, когда искусственный интеллект способен создавать живописные и музыкальные произведения, фото- и видеоизображения, писать прозаические и поэтические сочинения. Все это, конечно, еще не позволяет нам говорить о создании общего искусственного интеллекта Artificial General Intelligence (AGI). Тем не менее, неспособность отличить человека от искусственного источника может вызвать некоторую путаницу и имеет серьезные последствия (Floridi & Chiriatti, 2020). Не случайно сегодня особое внимание уделяется вопросам, проблемам, распространённым концептам и интерпретациям понимания искусственного интеллекта, поскольку фиксируется рост недоверия и обеспокоенности (Taulli, 2023) в отношении современных генеративных систем, моделей и нейросетей.

Относительно недавно многие исследователи в сфере искусственного интеллекта в контексте разработок NLP (Natural Language Processing) обратили внимание и на проблему смыслообразования, в том числе и визуального. Подобные исследования кажутся и многообещающей, и сложной областью для работы. Так, например, создаваемые в последнее время специалистами модели пытаются извлечь неочевидную информацию из различных изображений (Khurana et al., 2023). Начиная с 2021 года, благодаря нейронным сетям, стал возможным перевод текстовой информации в графическую. На данном этапе различные диффузионные модели условного типа способны создавать реалистичные синтетические изображения, в том числе без явной опоры на устойчивые связи в своих лингвистических базах и связанных списках (Nichol et al., 2022) Такие модели способны к комплексному комбинированию разрозненных понятий, хотя и далеко не всегда могут уловить, зафиксировать и отразить некоторые языковые единицы, выражения и смысловые элементы, описывающие достаточно «необычные» и нетривиальные объекты, контексты и сценарии.

При этом лингвистический нарратив и изобразительный нарратив сталкиваются с одной и той же информацией, но мы их интерпретируем по-разному (Schlöder & Altshuler, 2023). Исследователи подчеркивают, что, несмотря на то что объединение графической и текстовой информации для обучения представлениям все еще остается недостаточно исследованной областью (Makarov et al., 2021), значительная часть современной аудитории получила доступ к таким сервисам, как Midjourney и Stable Diffusion. Кроме того, можно отметить, что научно-технические разработки последнего времени, связанные с технологиями глубокого машинного обучения (DeepLearning) и нейронных сетей (алгоритмы BERT и GPT), актуализировали проблему цифрового искусства и творчества (Миловидов, 2022), стратегии взаимодействия художника-человека и нечеловеческого агента, а также

характер этого взаимодействия (Фадеева, 2023), а цифровые медиа позволили сблизить разные виды искусства.

Отсутствие не только общего консенсуса в определении того, что есть искусство (еще Г. В. Ф. Гегель показал, что результаты художественной деятельности зависят от того, что именно понимается в ту или иную эпоху под искусством (Гегель, 2007)), но и фундаментальные расхождения в представлениях о значимости, осмысленности и обоснованности самого вопроса проблематизируют многочисленные и многомерные сущностные и феноменальные аспекты фундирования и восприятия в отношении цифрового искусства, особенно в контексте искусственного интеллекта. В современном приближении спор о природе, сущности и определении искусства актуализируется в рамках постоянного противостояния монистических теорий, плюралистических концептов и элиминативных принципов, а также взаимоотрицания или даже синтеза эссенциалистских и антиэссенциалистских подходов, что также ещё больше размывает предлагаемые и предполагаемые демаркационные линии (Bartel & Kwong, 2021). Цель данной работы заключается в том, чтобы показать и осмыслить способы конструирования новой художественной реальности искусственным интеллектом в контексте визуализации и образной репрезентации языковых описаний и запросов. Для этого в работе рассматриваются, анализируются и критически оцениваются примеры, специфические черты и особенности, конкретные результаты создания визуальных образов литературных героев и персонажей посредством генерации изображений по текстовым запросам и описаниям указанными моделями, системами и нейросетями. Можно предположить, что использование искусственным интеллектом определенных средств выразительности имеет социокультурную специфику и редуцировано с точки зрения существующих представлений о возможностях искусства.

Методы и методология исследования

Методологической основой исследования в рамках данной статьи является положение о взаимосвязи языка, мышления и культуры. Методология работы опирается на семиотический анализ текста, позволяющий раскрыть смысл и значения, закодированные в тексте. Художественный текст представляет собой наивысшую ступень в иерархии текстов культуры. Это многослойное семантическое поле, которое понимается как единство ключевых слов и образов, вступающее в сложные отношения с внешним контекстом. Образ представляет собой форму воспроизведения, осмысления и переживания явлений жизни путем создания эстетически воздействующих объектов. Ключевые слова – это слова, определяющие содержание текста и передающие его основной смысл. Характеристики ключевых слов проявляются на разных уровнях текста – морфологическом, лексическом, синтаксическом, прагматическом. Их распознавание генеративными системами подразумевает относи-

тельную сложность используемых методов и многоэтапность реализующих их алгоритмов (Ванюшкин & Гращенко, 2016). Довольно часто извлечение ключевых слов текста выполняется одновременно с классификацией текста. К числу распространенных задач, стоящих перед генеративными алгоритмами, относится процесс определения критических функций или атрибутов объекта в текстовых данных. Например, это понимание общей темы текста или определение характеристик объектов в тексте, таких как цвет или размер; выделение ключевых слов и фраз из текста, распознавание именованных сущностей (NER), также известное как извлечение сущностей или разбиение на фрагменты.

Распознавание образов — ещё одна важная операция; она представляет собой процесс выделения исходных данных из общей массы разнородных объектов, а затем их классификацию по тем или иным признакам. Для обучения компьютера распознаванию образов требуется множество размеченных данных, где объекты и классы уже известны.

В рамках данной работы генерация изображений на основе литературных образов проводилась в трёх моделях и системах программного обеспечения: DALL·E, Stable Diffusion и Kandinsky.

Для создания изображений были использованы бесплатные общедоступные сервисы, версии и приложения указанных проектов. Инструментарий Stable Diffusion использовался через NMKD Stable Diffusion GUI – бесплатный клиент для Windows, предоставляющий базовые возможности нейросети, а также через WEB-сервис[1] – онлайн-версию с ограниченным функционалом. Доступ к DALL·E (версия 2 и версия 3) осуществлялся прежде всего посредством Bing Image Creator, а также промежуточных онлайн-сервисов[2]. Работа с Kandinsky (версий 2.1 и 3.0) осуществлялась через онлайн-сервис нейросети Fusion Brain[3], ссылку на который предоставляет сам разработчик[4], телеграм-бот Kandinsky by Sber AI[5], а также приложение «СберБанк Онлайн» для Android (v 15.7.0). Также для создания некоторых вариантов описания, написания требуемых запросов использовался чат GPT версий 3.5 и 4.0[6].

Одним из отличий этих инструментов по сравнению с художественной деятельностью человека является непроясненность критериев выбора генеративной моделью тех или иных средств выразительности: линия, цвет, пластическая форма, пространственно-композиционные отношения. Так, например, художник в процессе творчества создаёт самые разнообразные эскизы, что позволяет в короткое время опробовать и реализовать множество вариантов образно-пластического решения той или иной художественной задачи.

1 https://stablediffusionweb.com/ru/app/image-generator
2 https://dalle.com.ru, https://www.dalle3.org/ru и др.
3 https://fusionbrain.ai/?utm_source=kandinsky
4 https://www.sberbank.com/promo/kandinsky/
5 https://t.me/kandinsky21_bot
6 https://chat.openai.com

Результаты

Спекулятивный характер современного искусства, стремление исключительно рациональным путем постичь смысл вещей, лежащий за границами эмпирии, приводит культуру к вопросу о том можно ли искусство сводить к технике, к инструменту, к сюжету или оно конституируется уникальным видением художника. Для осмысления этого тезиса сравним фрагменты текстов, а также результаты генерации изображений по текстовым запросам.

Сегодня написание эффективных запросов к различным языковым моделям, системам ИИ и соответствующим нейросетям становится уже самостоятельной (и достаточно высокооплачиваемой) профессией, зачастую называемой “Prompt Engineer” (Kutela et al., 2023). Стоит уточнить, что в рамках данной работы использовались два типа текстовых запросов. Во-первых, самостоятельно составленные краткие и развернутые запросы в прямом соответствии с конкретной целевой задачей. Под кратким запросом понимается сам персонаж, например Чеширский кот. Под развернутым — его описание, составленное на основе литературных первоисточников, оригинальных авторских текстов, переводов, научных и искусствоведческих размышлений и рефлексий, личного представления и естественной для нас интерпретации культурных традиций и аспектов символической реальности. Во-вторых, был использован самый популярный и доступный чат-бот с генеративным искусственным интеллектом с целью попытаться понять и оценить, даст ли он лучшие, по сравнению с «нашими», результаты, будет ли «понятнее», «ближе» и содержательнее для нейросети запрос, созданный искусственным интеллектом, а не человеческим, и будет ли в принципе существенное различие между визуализациями разного генеза. Тем более, что одна из рассматриваемых моделей – DALL·E – является продуктом компании OpenAI – разработчиков преслову-того чата GPT.

Первый образ, взятый для анализа, это образ Чеширского кота (Кэрролл, 2019), который в английской лингвокультуре символизирует нечто неуловимое, внезапно возникающее и бесследно исчезающее. Во времена написания романа «Алиса в Стране Чудес» часто использовалась поговорка — «улыбается, как чеширский кот». Для работы с образом, текстовым описанием и визуальной репрезентацией была привлечена генеративная система ИИ – Чат GPT 3.5. Чату GPT был задан запрос о том, как он сам сформулирует промпт для DALL·E – «родственной» нейросети, также разработанной OpenAI. В результате Чат GPT предлагает следующее описание для нейросети:

> «Представьте себе Чеширского кота, сидящего на ветке дерева среди густого тумана. Его тело мерцает и исчезает, оставляя только его широкий улыбающийся череп с длинными изогнутыми усами, которые выглядят как три белых ветви. Его глаза сверкают во мраке, создавая загадочное впечатление. Всё это олицетворяет загадку и мистику, которая окружает Чеширского кота в мире Алисы в стране чудес».

По данному описанию DALL·E сгенерировала следующие изображения (Рисунок 1):



Рисунок 1. Визуализация описания Чеширского Кота от Чата GPT 3.5 в DallE (A-D)

Figure 1. Cheshire Cat. Visualisation of the description by Chat GPT 3.5 in DallE (A-D)

Как можно заметить, в описании, предложенном для создания запроса чатом GPT, в явном виде указан сам персонаж, причём в тексте подразумевается, что это знакомый, устойчивый образ – «представьте себе Чеширского кота» – на который уже «накладываются» характеристики и описание положения и окружения («сидящий на ветке», «среди густого тумана»), а также отдельных элементов целостного образа (черепа, усов, глаз). При этом заметно, что чат GPT в своём описании придерживается не определённого, сухого, чёткого стиля и формата описания, но напротив – более поэтичного, образного, «мистического» и вместе с тем – неопределённого, предполагающего

творческий подход к интерпретации. Сам запрос, кажется, больше походит на обращение «писателя» к «художнику», чем на взаимодействие и коммуникацию двух ипостасей родственного «машинного разума». Визуальные репрезентации, предлагаемые DALL·E, являются под стать описанию стилевыми и метафоричными (изображения **A** и **B**). При рассмотрении сгенерированных изображений можно отметить и выделить некоторые специфичные аспекты и проявления. Так, например, данная в начале описания позиция «сидящего на ветке» явно и определённо выражена на всех изображениях, но уже следующая часть описания «среди густого тумана» раскрывается довольно своеобразно: на трёх из четырёх предложенных изображений Кот не столько находится среди тумана, сколько сам соткан из него, что особенно явно видно на изображениях **C** и**D**.

Результат, выданный Web-версией Stable Diffusion **(Рисунок 2)**, в этом отношении более точно следует исходному описанию, там сам кот имеет явно выраженную «телесную» форму, а туман является в первую очередь элементом окружения.



Рисунок 2. Визуализация описания Чеширского Кота от Чата GPT 3.5 в Stable Diffusion Web (A-B)

Figure 2. Cheshire Cat. Visualisation of the description by Chat GPT 3.5 in Stable Diffusion Web (A-B)

«Количественная» характеристика («как три белых ветви») выражена достаточно неопределённо и вариативно, причём во всех сгенерированных образах. На некоторых из них можно попытаться выделить именно три элемента, связанных с усами, но на большинстве их количество и отношение имеет крайне многообразный характер несмотря на то, что именно эта часть описания является наиболее формально определённой. Стоит отме-

тить, что общее впечатление передано более точно: предложенные картины и действительно можно интерпретировать и «ощутить» как загадочные и мистические.

На основании текста (Кэрролл, 2019) рассматриваемого произведения было также выделено совокупное описание Чеширского кота:

> «Кот крупнее обычных котов, а усы его были очень длинные; длинные так, что он поддерживал их одной лапой, как настоящий джентльмен, очень широкие усы, и когда он улыбался, усы расползались во все стороны, как три белые ветви, а глаза его сверкали во мраке как две светлячки, у него был смазанный мордочкой вид, и вообще он выглядел как животное, которое, внимательно поглядев на кого-то, непременно скажет что-нибудь оскорбительное».



Рисунок 3. Визуализация описания Чеширского Кота в DallE Bing Creator

Figure 3. Cheshire Cat. Visualisation of the description in DallE (by Bing Creator)

В данном описании, как можно заметить, уже отсутствует явное именование персонажа, поэтому прямое соотнесение с изначально имеющимся в наборе обучения классом, объектом или элементом исходной базы невозможно. В генерациях DALL·E от Bing (Рисунок 3) сразу бросается в глаза смена визуального стиля – от мистического, «опасного и пугающего», сверхъестественного, к «мягкому и доброму», немного загадочному, спокойному и мудрому. При добавлении в содержание запроса параметров абсурдистского и сказочного стиля интерпретации изображения принципиально не изменяются, добавляется текст автоматического перевода (и, как можно заметить в нём есть небольшие различия, несмотря на использование одного и того же инструментария), тон становится более контрастным или приглушённым, появляются некоторые дополнительные атрибуты, например свеча, которая, по всей видимости, должна дать источник света и добавить сказочности атмосфере. При этом отдельные элементы описания полностью игнорируются или проявляются и обнаруживаются с большим трудом. К примеру, ни на одном из четырёх изображений кот не поддерживает усы лапой, несмотря на явное указание на это в тексте запроса, да и оскорбительного выражения и вида заметить не удаётся, напротив, предлагаемые облик и мордочки котов можно охарактеризовать как добродушные и умилительные.

Web-версия Stable Diffusion (Рисунок 4) в аспекте визуализации предлагаемого описания радикально расходится с анимационно-фентезийным стилем DALL·E и реализует генерации в реалистической манере. Как можно заметить, предлагаемые коты – «настоящие», они реализованы в полноцветной палитре и трёхмерном пространственном ракурсе. При этом исходная образная связь с Чеширским котом разрывается ещё больше, а элементы описания от усов до лап имеют ещё меньше значения и практически не реализуются в предлагаемых вариациях.

В целом можно отметить значимые для образа Чеширского Кота элементы:

1. **Внешний вид:**
   - Крупный, полосатый кот с яркими фиолетовыми и розовыми полосами.
   - Большие, сверкающие глаза, часто меняющие цвет.
   - Широкая, загадочная улыбка.
   - Способен исчезать и появляться частично или полностью.
2. **Одежда и атрибуты:**
   - Не носит одежды и обуви.
   - Основной атрибут — его способность к внезапному исчезновению и появлению.

3. **Характер и поведение:**
   - Загадочный, хитрый.
   - Ироничный, саркастичный.
   - Непредсказуем, любит смущать и путать.
   - Получает удовольствие от своей загадочности и способности запутывать других.
   - Выражение лица обычно загадочное и насмешливое.

Улыбка его может выглядеть и дружелюбной, и слегка зловещей одновременно. Эта улыбка является его самой узнаваемой чертой.



Рисунок 4. Визуализация описания Чеширского Кота в Stable Diffusion Web (A-B)

Figure 4. Cheshire Cat. Visualisation of the description in Stable Diffusion Web (A-B)

Впрочем, для того, чтобы по достоинству оценить и сам роман, и образы его героев, нужно знать многое из того, что находится за его пределами. Необходимо, чтобы была сформирована определенная общность социальной истории, которая в дальнейшем отражается в знаниях об окружающем мире. Это фоновые знания – имплицитные смыслы-знания. Сам роман был адресован английским читателям другого века. «Более того, некоторые из шуток Кэрролла были понятны лишь тем, кто жил в Оксфорде; другие предназначались еще более узкому кругу – одним лишь прелестным дочерям ректора Лидделла», – отмечал известный американский математик и писатель, комментатор Л. Кэрролла М. Гарднер (Нефёдов & Чигрина, 2016).

Если подходить к художественному произведению с точки зрения его перекодирования из одной системы в другую, то неизбежно приходится выделять два плана – общий и единичный. Точный перевод ставит во главу угла

передачу общего, поэтому при нем происходит только замена на уровне языковых (знаковых, словарных) значений, а то, что является единичным и уникальным, остается непередаваемым. Именно поэтому на современном этапе развития искусственного интеллекта задача передачи того смыслового образа, который автор хотел создать в сознании своих читателей, посредством текста считается практически неразрешимой.

Еще одним знаковым персонажем, но уже отечественной культуры, является Кот Бегемот (Булгаков, 2016). Первоначально были проведены генерации без подробного описания образа, основывающиеся только на имени персонажа с помощью Web-версии Stable Diffusion по умолчанию **(A-B)** и в фотореалистичном стиле **(C-D) (Рисунки 5-6)**.



Рисунок 5. Генерация изображений по запросу “Cat Behemoth” в Stable Diffusion Web в режимах «по умолчанию» (A-B) и фотореалистичном стиле (C-D).

Figure 5. “Cat Behemoth” prompt image generation in Stable Diffusion Web by default (A-B) Photo style (C-D) specifications

Стоит отметить, что содержание и даже само построение романа задаёт явные аллюзии и связи с библейскими образами, символами, смыслами. «Изнутри может вырасти тема, идея, его смысл… Интерпретация темы может быть шагом к нашему внутреннему мироощущению, а может не быть, наталкиваясь на сопротивление нашего мышления, нашего опыта» (Кривцун, 2019, с. 20). Являясь представителем тёмных сил, сподвижником дьявола, демоническая сущность воспринимается отнюдь не как представление абсолютного зла, но как персонаж, обладающий не только харизмой и обаянием, но даже некоторой «человечностью» и сентиментальностью. Обращает на себя внимание то, что при стандартных генерациях без явных уточнений, многочисленных исправлений, привлечения дополнительного инструментария образ Бегемота является слишком «плоским», простым и односторонним, лишённым всей своей глубины, многогранности, идейного наполнения и культурного бэкграунда: это зачастую или «просто кот» (причём, как видно, далеко не обязательно чёрный) или некоторый «котодемон» (где, справедливости ради стоит отметить, «демоничность» выражается не только отдельными элементами вроде когтей, клыков, ушей и хвостов, но некоторой попыткой фиксации целостного образа, хоть и не всегда выходящей за такой набор исходных типизированных параметров). При этом предлагаемые нейросетями реалистичные изображения совершенно не отражают ни мистических, ни демонических, ни потусторонних, ни культуросообразных черт. Данные генерации представляют собой «обычные» изображения котов на некотором, чаще всего размытом, контрастно-бликующем боке-фоне.

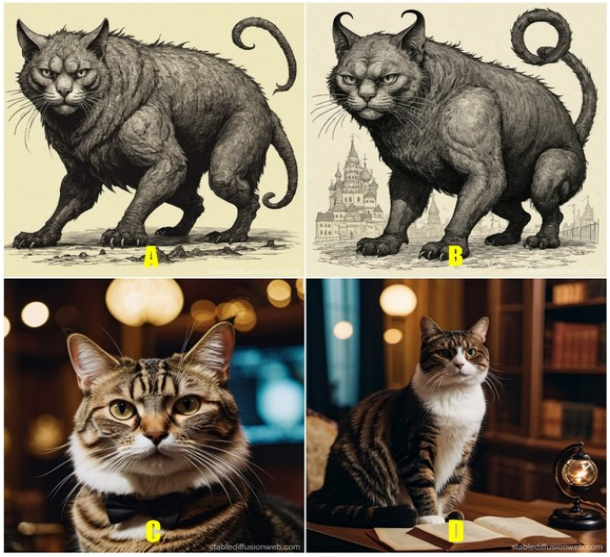

Рисунок 6. Генерация изображений по уточнённому запросу “Cat Behemoth…” в Stable Diffusion Web в режимах «по умолчанию» (A-B) и фотореалистичном стиле (C-D)

Figure 6. “Cat Behemoth…” clarified prompt image generation in Stable Diffusion Web by default (A-B) Photo style (C-D) specifications

Даже явно прописанное в промпте уточнение “character from the novel The Master and Margarita by the Russian writer Mikhail Bulgakov” (Рисунок 7) принципиально не меняет ситуацию. Веб-версия Stable Diffusion, например, скорее игнорирует эту часть запроса, выдавая того же кота на том же фоне, а за вышеуказанное отвечает галстук-бабочка, который, видимо, и делает кота персонажем романа Булгакова, а в других тот же кот, правда, уже без бабочки, находится перед книгой, а фон меняется с просто боке на размытое.

Первичная генерация изображений Кота Бегемота в Kandinsky с применением различных стилей и настроек (Рисунок 7) позволяет отметить некоторое эстетическое своеобразие художественной интерпретации **(A, B, D)**, предлагающей многообразные стилевые, форматные, цветовые черты и свойства, обнаруживая при этом определённое композиционное, предметное, эмоциональное единство (например, форма морды, типизированный «демонический» взгляд, острые уши, напоминающие рога, структура шерсти, форма и статика самой фигуры и т. д.) искомого образа. Несколько особняком здесь стоит генерация в реалистичной манере **(C)**, представляющая скорее изображение «обычного» кота и принципиально отличающееся во многих аспектах от рядоположенных версий, однако, при желании, определённые образные характеристики можно обнаружить и в предложенном варианте.



Рисунок 7. Генерация изображений по запросу «Кот Бегемот» в Kandinsky через интерфейс Fusion Brain в режимах «по умолчанию» (A), «реалистичное изображение» (C) и с применением стилей “Digital Art” (B), «Свой стиль» (D)

Figure 7. “Cat Behemoth” prompt image generation in Kandinsky (Fusion Brain) by default (A), realistic picture (C) and “Digital Art” (B), “No style” (D) specifications

Для генерации визуального образа в Kandinsky был предложен следующий промпт-описание на русском языке:

> «Кот, громадный, как боров, черный, как сажа или грач, и с отчаянными кавалерийскими усами, жутких размеров черный кот с пухлыми лапами, острыми ушами и круглой головой» (Булгаков, 2016) (Рисунок 8).



Рисунок 8. Визуализация описания Кота Бегемота в Kandinsky через интерфейс Fusion Brain (A) и Sber AI bot (B)

Figure 8. Cat Behemoth. Visualisation of the description in Kandinsky by Fusion Brain (A) and Sber AI bot (B)

Одним из «дальних литературных предков» Бегемота может являться гофмановский кот Мурр — именно от последнего член свиты мессира унаследовал «своё забавное самодовольство» (Чудакова, 2011).

Образ спутника Воланда в романе весьма многогранен и противоречив, в нём сталкиваются и находят отражение самые разные, зачастую противоположные, формы, аспекты, черты внешности и характера, модели поведения, речевые схемы. Внешность персонажа также отражает и его многомерную глубину, и радикальные трансформации от толстого кота-оборотня до худенького юного пажа, от жестокого потустороннего зверя до мещанского дамского угодника.

Образы отражают восприятие реальности человеком и ее воплощение в художественном тексте. При этом информация кодируется и обрабатывается человеческим сознанием по-разному: на одном уровне при помощи образов, на более глубинном при помощи концептов, которые можно рассматривать как словесный образ, связывающий человеческое мышление с языковой картиной мира.

Отдельно стоит заметить, что концепт признается базовой единицей культуры, ее концентратом. Концептуальная система носителя языка является системой его мнений и знаний о мире, отражающих познавательный опыт на доязыковом и языковом этапах и уровнях. Изучение культурных концептов позволяет констатировать наличие в определенной культуре чего-то существенного, так как в каждой культуре имеются свои уникальные обычаи, традиции. Формирование концептуальной системы проходит в несколько этапов. Невербальный способ образования концептуальной структуры можно отнести к первому этапу. На этом этапе человек знакомится с объектами, доступными его непосредственному восприятию. По каналам чувственного восприятия индивид получает определенную информацию, что формирует его систему представлений о мире. На этом же этапе осуществляется формирование некоторых «первичных» концептов, которые возникают в процессе ознакомления с окружающим миром, в результате перцептуального опыта. «Первичные» концепты являются необходимым условием для построения концептуальной системы. На языковом уровне культурные концепты сохраняются в коллективном языковом сознании, в памяти народа. При этом образ как аналоговая репрезентация предполагает синтетичность. Концепт формируется за счет аналитического подхода.



Рисунок 9. Генерация изображений по запросу «Создай изображение Кота Бегемота» в DallE 3 (A-B)

Figure 9. “Create an image of Cat Behemoth” prompt image generation in DallE 3 (A-B)

С одной стороны, образы Котов имеют стереотипный пласт фундирования и восприятия, обеспеченный фольклорным, социально-психологическим, обыденно-житейским, национальным и прочими базисами. С другой, эти

же основания находят принципиально разное, непредсказуемое, многогранное проявления в самых различных ракурсах, ситуативных моделях и конкретных феноменах.

Так, например, при запросе создания Кота Бегемота DALL·E (Рисунок 9) предлагает явно «детские», «мультяшные» визуализации. Несмотря на то, что в запросе «Кот» находится на первом месте, интерпретатор определяет ключевым основанием создания визуального образа именно бегемота. И сами предлагаемые образы принципиально расходятся и противоречат сущностному и культурно закреплённому образу булгаковского героя, что явно показывает несовпадение и отсутствие внутренней корреляции рассматриваемых лексических единиц и фиксируемых концептуализаций генеративной системы.



Рисунок 10. Генерация изображений по запросу «Создай изображение Кота Бегемота из Мастера и Маргариты Булгакова» в DallE 3 (A-D)
Figure 10. “Create an image of Cat Behemoth…” clarified prompt image generation in DallE 3 (A-D)

Генерация новых метафор — мощная форма лингвистических инноваций (Kenett et al., 2018), объединяющая отдаленно связанные или слабо связанные

компоненты в значимые уникальные концепции (Heruti & Mashal, 2023); однако кажется, что проявляющаяся концептуализация носит случайный, спорадический характер, и концепт фундируется не замыслом и идеей ИИ, а спецификой нашего осмысления и анализа предлагаемых репрезентаций.

При уточнении запроса создания изображения Кота Бегемота именно как героя «Мастера и Маргариты» Булгакова (Рисунок 10) получаются принципиально иные результаты, причём двух явно различных видов. Первый представляет собой именно процесс создания изображения – рисование художником на полотне, с находящимися рядом кистями, красками, палитрой эстетически оформленного в условно волшебном или мистическом окружении кота, впрочем, мало напоминающего именно булгаковского персонажа. Второй тип предлагаемых визуализаций уже гораздо ближе к искомому или ожидаемому как с точки зрения запечатлённых образов, композиционных решений, так и со стороны вариативности, атмосферности, «интересности» и большей претензии на творческую репрезентацию.

С целью написания промпта чату GPT 4.0 был сделан запрос описания Кота Бегемота для систем ИИ и нейросетей. В результате чат GPT предоставил следующие аспекты описания для дальнейшей генерации:

1. **Внешний вид:**
   - Крупный, черный кот с блестящей, гладкой шерстью.
   - Зеленые, сверкающие глаза.
   - Длинные, острые когти и густые усы.
   - Ходит на задних лапах.
2. **Одежда и атрибуты:**
   - В старомодном мужском костюме и шляпе.
   - Иногда держит бокал вина или шашку.
3. **Характер и поведение:**
   - Остроумный, саркастичный.
   - Любит создавать хаос и веселье.
   - Выражение лица часто насмешливое или хитрое.

И на основании данных характеристик по дополнительному запросу предоставил следующий промпт:

> «Создайте изображение Кота Бегемота из романа «Мастер и Маргарита» Михаила Булгакова. Кот Бегемот должен быть крупным, черным котом с блестящей, гладкой шерстью и сверкающими зелеными глазами. У него должны быть длинные, острые когти и густые усы. Кот должен ходить на задних лапах, одет в старомодный мужской костюм и шляпу. В одной лапе он держит

бокал вина или шашку. Выражение лица должно быть насмешливым или хитрым, передавая его остроумный и саркастичный характер».

Для генерации по данному запросу было использован телеграм-бот YesAi с доступом к DALL·E 3 (Рисунок 11).



Рисунок 11. Визуализация описания Кота Бегемота в DallE 3 YeSbot (A-D)

Figure 11. Cat Behemoth. Visualisation of the description in DallE 3 by YeS bot (A-D)

В данных результатах по запросу на русском языке явно видно вполне ожидаемое расхождение: на большинстве генераций представлен именно бегемот, то есть гиппопотам. При этом даже в рамках одной серии генераций система предлагает оба типа визуализаций: и кота, и гиппопотама. Предложенные персонажи держат бокал вина, как и сказано в установленном описании, хотя в некоторых вариациях напиток мало напоминает именно вино, а ёмкость с ним – бокал; а вот с другой описанной альтернативой ситу-

ация значительно сложнее. Во-первых, система не воспроизводит предполагаемую дизъюнкцию, а, скорее, предлагает конъюнкцию: во всех вариантах герой держит бокал в том или ином виде и (а не или) условную «шашку». Во-вторых, изображение «шашки» крайне вариативно и совсем не напоминает, собственно, шашку: в одной из версий это скорее трость, в другой – стилет, а в иных вообще длинный коготь, пучок волос или верёвки. То есть, в списке деталей и элементов описания, находящихся рядом, одни – считываемы, воспроизводимы и узнаваемы, а другие – случайны, неопределенны и крайне условны.

Еще одним образом, взятым для анализа, стал образ кота Мурра (Гофман, 2013). По запросу «*Кот Мурр – персонаж романа Гофмана Житейские воззрения кота Мурра*» с применением некоторых основных доступных инструментов, таких как настройка ArtGPT и установка стиля DigitalArt, Kandinsky сформировал соответствующие визуальные образы (Рисунок 12).



Рисунок 12. Генерация изображений в Kandinsky через интерфейс Fusion Brain по запросу «Кот Мурр» (A) и уточнённому запросу «Кот Мурр– …» в режимах «по умолчанию» (B), с применением настройки Art GPT (C) и с применением стиля “ Digital Art” (D)

Figure 12 “Tomcat Mur” prompt (A) and “Tomcat Murr…” clarified prompt image generation in Kandinsky (Fusion Brain) by default (B), realistic picture (C), ART GPT (B) and “Digital Art” style (D) specifications

Образы, созданные данной генеративной системой, достаточно вариативны и разнообразны. Среди довольно обычных «стандартизованных» примеров, коими достаточно полна всемирная паутина, выделяется результат «по умолчанию», полученный, как ни странно, без применения дополнительного инструментария: и своей стилистикой, и манерой исполнения, и определённой претензией на художественность (которую, впрочем, имела скорее не нейросеть, а мы сами, в соответствии со своими представлениями, в которых сочетаются личные воззрения, культурные доминанты, устоявшиеся паттерны и образцы художественной парадигмы). И это неудивительно, ведь в течение развития изобразительного искусства в работах многих художников изображения кошек становятся максимально реалистичными, достигают почти, или даже гипер-, фотографического сходства. Отчасти поэтому многие мастера решаются отказаться от соревнования с оптическими технологиями фиксации реальности, и одомашненные представители семейства кошачьих представляются то антропоморфными, то кубическими, то абстрактными, то мистическими, предполагая именно оригинальность воплощения, раскрытие (или сокрытие, что также свойственно образу кошки) сюжетной глубины, рефлексивные или социосообразные аспекты.

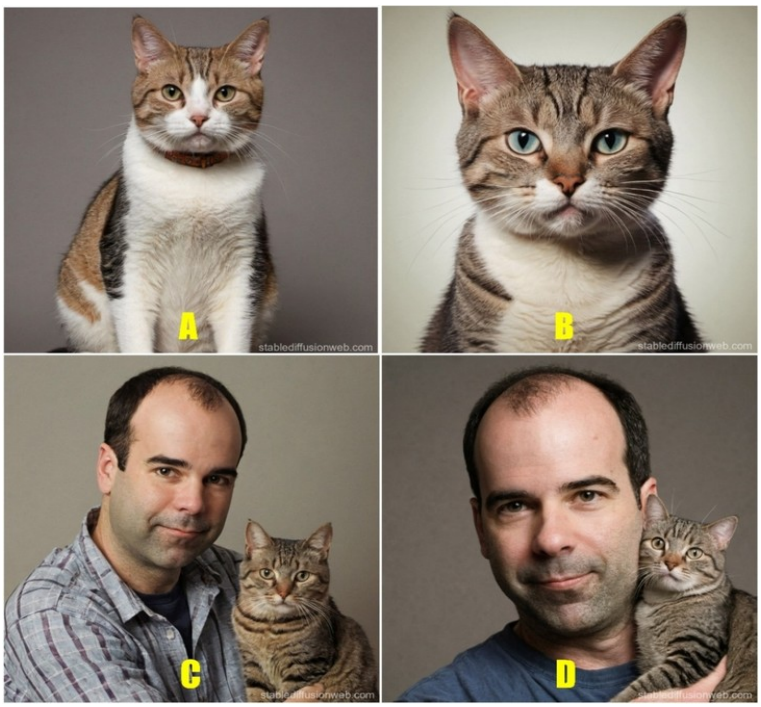

Рисунок 13. Генерация изображений по запросу “Tomcat Murr is a character ...” (A-B) и с уточнением “Hoffman's novel” (C-D) в Stable Diffusion Web

Figure 13. “Tomcat Murr is a character...” prompt (A-B) and with the addition “Hoffman’s novel” (C-D) prompt image generation in Stable Diffusion Web

По запросу *"Tomcat Murr Is a character of novel The Life and Opinions of the Tomcat Murr"* Веб-версия Stable Diffusion представила следующие результаты (Рисунок 13) (**A-B**), а при уточнении «Hoffman's novel» предложила условно портретные генерации, в которых появляется и оказывается в фокусе внимания уже человек (**C-D**).

> Кот мудрый, находчивый, остроумный, поэтически чувствующий, необыкновенной красоты. Серые и черные полоски, идущие вдоль спины, сходятся между ушами на темени и составляли на лбу какую-то живописную надпись иероглифами. Длинный красивый хвост также покрыт полосами и изгибается с необычайной энергией. Шерсть его пестрит и светится под лучами солнца до такой степени, что между черным и серым цветом можно было рассмотреть еще узенькие, золотисто-желтые полоски. В его физиономии есть что-то особенное, незаурядное, что голова его достаточно толста, а его борода уже теперь, в юности, бела и достаточно длинна (Гофман, 2013).

По данному описанию DALL·E 2 предложила следующие вариации визуализации (Рисунок 14):



Рисунок 14. Визуализация описания Кота Мурра в DallE 2

Figure 14. Tomcat Murr. Visualisation of the description in DallE 2

Следующее поколение данной модели по запросу «*Кот Мурр – персонаж романа Гофмана Житейские воззрения кота Мурра*» представило довольно вариативные художественные интерпретации (Рисунок 15).



Рисунок 15. Генерация изображений по запросу «Кот Мурр – персонаж …» в DallE 3

Figure 15. "Tomcat Murr is a character…" prompt image generation in DallE 3

Помимо этого, в рамках написания промтов для создания изображений и визуальных образов Кота Мурра мы обратились к Чату GPT для создания подходящего описания и требуемого запроса с точки зрения и ракурса представления «самого» искусственного интеллекта. Чат GPT 4.0 предложил такой вариант промпта для визуализации искомого образа посредством нейросетевой генерации:

> «Создай изображение Кота Мурра из романа Э. Т. А. Гофмана "Житейские воззрения кота Мурр". Кот Мурр — это серый кот с умными глазами, который сидит за письменным столом. На столе лежат книги, чернильница и перо, как символы его начитанности и способности писать. Кот выглядит самодовольно и уверенно, с легкой улыбкой на лице, возможно, носит маленькие очки для чтения. Фон должен быть уютным и классическим, с книжными полками и старыми картинами на стенах, чтобы передать атмосферу старинного кабинета ученого».

При запросе более подробного описания внешности ИИ предложил следующую версию:

«Создай изображение Кота Мурра из романа Э. Т. А. Гофмана «Житейские воззрения кота Мурра». Кот Мурр — это элегантный серый кот с густой, блестящей шерстью и большими, умными янтарными глазами. Он сидит за старинным деревянным письменным столом, покрытым книгами, свитками и чернильницей с пером. Кот носит маленькие круглые очки на носу и выглядит очень самодовольно и уверенно. Его лапы аккуратно сложены перед ним, а одна лапа лежит на открытой книге, словно он только что писал или читал. На заднем плане — уютный кабинет с книжными полками, старинными картинами и мягким освещением от масляной лампы, создающим теплую и интеллектуальную атмосферу. Важно передать его высокомерие и чувство собственного достоинства, возможно, добавив легкую улыбку на мордочке».

На основе данного промпта от чата GPT 4.0 были реализованы визуальные генерации в «родственной» ему нейросети и модели DALL·E 3 через различные платформы, интерфейсы и способы доступа (Рисунок 16).



Рисунок 16. Визуализация описания Кота Мурра от Чата GPT 4.0 в DallE 3 от Yes-bot (A), dalle3.org (B) и Bing Creator (C)

Figure 16. Tomcat Murr. Visualisation of the description by Chat GPT 4.0 in DallE 3 by Yes-bot (A), dalle3.org (B) и Bing Creator (C)

Сразу можно отметить, что посредством различных сервисов и интерфейсов были получены близкие или практически идентичные результаты, и BingCreator, DALL·E 3.org и YeS-bot создали изображение в одной стилистике, цветовой схеме, сходном типизированном окружении, практически тождественные как в общем отношении, так и частных деталях.

По данному описанию результаты генерации в Kandinsky и Stable Diffusion (Рисунок 17):



Рисунок17. Визуализация описания Кота Мурра от Чата GPT 4.0 в Kandinsky (A), Stable Diffusion Web (B) и Stable Diffusion GUI (C)

Figure 17. Tomcat Murr. Visualisation of the description by Chat GPT 4.0 in Kandinsky (A), Stable Diffusion Web (B) и Stable Diffusion GUI (C)

Таким образом, можно реконструировать следующие характеристики, присущие Коту Мурру:

1. **Внешний вид:**
   - Обычный домашний кот среднего размера.
   - Выразительное кошачье лицо с умным и проницательным взглядом.
   - Зелёные глаза, которые сверкают и излучают интеллигентность.
   - Имеет пушистую и ухоженную шерсть серо-полосатого цвета.
   - Длинные усы и острые когти.
2. **Одежда и атрибуты:**
   - Как типичный кот, Мурр не носит одежды и обуви.
   - В своей литературной деятельности Мурр использует чернила и бумагу.
   - Основной атрибут – способность к письму и литературным размышлениям.
3. **Характер и поведение:**
   - Остроумный, самоуверенный и интеллектуальный.
   - Независимый и самодовольный.
   - Склонен к философским размышлениям и самовыражению через письмо.
   - Стремится к самопознанию и самовыражению через творчество.
   - Относится к людям и другим животным с некой снисходительностью, ощущая своё превосходство в интеллекте и мудрости.

Можно отметить, что выявленные характеристики всех рассмотренных персонажей и соотнесение их с визуальными репрезентациями показывают

сложный механизм формирования образов, объединяющих связанные или изначально несвязанные компоненты в значимое целое.

Обсуждение

Под цифровым искусством или искусством новых медиа зачастую понимаются любые художественные практики, связанные с использованием цифровых технологий. Потенциал “text-to-image” моделей сегодня активно рассматривается в самых разных областях творческого и креативного мышления, связанного с различными аспектами визуализации. Одновременно стараясь утвердить их перспективы и раскрывая существующие или потенциальные проблемы, пороги и узкие места в ситуации всё нарастающего их применения, исследователи пытаются дать различные оценки таким феноменам и практикам с точки зрения субъектного осмысления и рефлексии (Paananen et al., 2023).

Несомненно, что искусственный интеллект способен создавать убедительный высокодетализированный визуальный «продукт», но означает ли это, что он способен создавать произведения искусства? Дело в том, что эстетичность – это не единственная их характеристика. Так, в современной философии искусства обсуждается целый ряд вопросов, связанных с символической, художественной и социокультурной ценностью произведений, созданных искусственным интеллектом. Большинство людей сходятся во мнении, что изображения, сгенерированные из текста, — это эстетика (Rasrichai et al., 2023).

DALL·E представляет собой модель генерации изображений, разработанную и реализованную OpenAI – создателями чата GPT, ставшего не только флагманом среди чат-ботов, систем генеративного ИИ и виртуальных собеседников, но прежде всего самой популярной, распространённой технологией искусственного интеллекта и практически синонимом или нарицательным именованием ИИ для массового пользователя. DALL·E основана на технологии GPT (GenerativePre-trainedTransformer) и способна генерировать изображения на основе текстового описания. Первая версия, DALL·E 1, была представлена в январе 2021 года. Этот проект впервые показал потенциал использования нейронных сетей для создания изображений, основанных на текстовом описании. DALL·E использует комбинацию техник, включая трансформеры и вариационные автокодировщики, для сгенерированных изображений. Следует отметить, что DALL·E способен генерировать изображения с учетом множества параметров, таких как форма, текстура, расположение объектов и так далее, основываясь на описании, предоставленном пользователем. Дальнейшие итерации (номерные версии) последовательно увеличивали объём данных, на которых проводится обучение, усложняли и уточняли алгоритмы и схемы работы и расширяли предоставляемый пользователю функционал. Так, DALL·E 2 использовала модель преобразователя GPT-3 с более чем

10 миллиардами параметров для создания изображений из текстовых описаний. Интерпретируя вводимые данные на естественном языке, каждая следующая версия генерирует изображения со значительно большим разрешением и повышенной реалистичностью, чем предшественник.

Stable Diffusion была выпущена в августе 2022 года, а уже в ноябре того же года вышла улучшенная версия 2.0, которая могла генерировать изображения с большим разрешением, а также была адаптирована для создания цифрового арта. Stable Diffusion использует модель скрытой диффузии вместо стандартной диффузионной модели. Она способна не только генерировать изображения по текстовому описанию, в том числе в стиле определённом авторском стиле, но и редактировать и дополнять уже готовые работы или скетчи, дорисовывая объекты или изменяя фон, а также данная нейросеть может быть встроена в популярный игровой движок Unreal Engine для создания локаций, ассетов, окружения и генерации внутриигровых объектов. Stable Diffusion представляет собой модель с открытым исходным кодом, доступную для загрузки, изменения и автономной работы, в отличие от многих других нейросетей, систем и сервисов ИИ, реализующих проприетарную модель с доступом посредством облачного сервиса, сайта или онлайн-платформы.

Kandinsky представляет собой крупномасштабную диффузионную мультиязычную модель генерации изображений из поступающего текстового запроса, которая также может создавать изображения на основе русскоязычного текста, что кажется вполне естественным и даже необходимым, так как является отечественной разработкой, связанной с одной из крупнейших российских корпораций. Прежде всего с точки зрения функционала, по сравнению с предыдущими версиями, в Kandinsky 3.0 была добавлена inpainting-модель, с помощью которой можно редактировать полученные изображения: менять отдельные объекты и целые области (inpainting) или расширять границы картинки, создавая панорамы (outpainting).

Вне зависимости от того в каких моделях и системах – DALL·E, Stable Diffusion, Kandinsky, – проводилась генерация изображений на основе литературных образов, существуют следующие характеристики алгоритмов извлечения ключевых слов: точность (Precision), полнота (Recall) и F-мера (F-measure), т. е. характеристика, которая позволяет дать оценку одновременно и по точности и полноте, их сбалансированности. Стоит отметить, что в настоящее время 100%-точность алгоритмов не может быть достигнута, поскольку результат зависит от многих факторов, в том числе, например, и от количества извлеченных ключевых слов. Основной недостаток здесь видится в том, что во многих алгоритмах их количество не зависит от размера текста. Следствием этого является необходимость разработки объективного и универсального критерия оценки качества работы алгоритмов извлечения ключевых слов.

Качество любой системы определяется исходя из составляющих ее элементов, а также способов их взаимосвязи, т. е. структурой. Структурный

уровень качества раскрывается на основе выделения свойств и отношений любой системы, в том числе и художественной.

Художественный образ всегда в большей или меньшей степени отражает реальность. Их тождественность реализуется во всей полноте только в процессе восприятия субъекта. Именно тогда полностью раскрывается и начинает функционировать уникальный мир, свернутый художником в акте создания произведения искусства. Образ предполагает наличие объективной или субъективной реальности, давшей толчок процессу художественного отображения. Затем авторский замысел, авторская идея трансформируется в творческом акте в реальность самого произведения, а после происходит еще один процесс трансформации. Таким образом, возникает новый образ, зачастую очень далекий от исходного, но сохраняющий в себе нечто единое, присущее всей системе образного воплощения реальности. Нельзя не отметить, что в цифровой среде символизм образа кота представлен чрезвычайно разнообразно, что связывается исследователями с широким спектром антропоморфизации кошачьего поведения (Тихонова, 2023, с. 157).

Социальные и философские аспекты и смыслы художественной литературы, связанной или фундированной образами животных, находят своё проявление не только в порождающих их традициях, но и в более широком культурном контексте. Например, образ Кота Бегемота находят своё проявление, отражение и прямые параллели в творчестве современных авторов, причём принадлежащих совершенно иной, нежели Михаил Булгаков, языковой, культурной и художественной традиции, таких как, например, Рабих Аламеддин – ливано-американский художник и писатель (Tabačková, 2018, p. 167). Так, рассматриваемое в рамках данной работы, произведение «Житейские воззрения кота Мурра» воспринимается восточным сознанием как острая сатира, ироничные нападки на социальную и бытовую пошлость и поверхностность, осуждение лицемерия и восхваление подлинной мудрости и любви (Zhang, 2023). Образ же Чеширского Кота выходит даже за пределы широкого социокультурного пространства, находя отражение не только в гуманитарном знании, художественных нарративах и бытовых практиках, но и в такой области как квантовая механика (Aharonov et al., 2021), именуя и описывая парадоксальное явление отчуждения объекта от своих собственных свойств.

Резкий качественный скачок в развитии и применении генеративных моделей и систем радикальным образом в самый короткий срок изменил архитектуру всех пространств существования и проявления визуальной культуры, на своём последнем этапе выражаясь не только общими концептуальными схемами, но прежде всего (по остроумной характеристике некоторых исследователей) «маленькими хитроумными уловками» (Steinfeld, 2023).

Создание промпта для генеративной системы в виде описания или характеристики определённого персонажа определяется сквозь призму исходных намерений пользователя и реализации исходно сформированного образа

желаемого результата. Итеративность действий стремится к постепенному, пошаговому приближению к воплощению изначального, но при этом и отчасти динамически меняющегося в зависимости от предложенных альтернатив, образа. Пользователь создаёт запросы, задаёт описания и уточнения, чтобы получить желаемый образ в текстовой форме естественного языка, но результаты действий, успешность или неудовлетворительность процесса оценивается визуально, а не лингвистически. Можно отметить, что создание описания для промпта — это довольно непростой процесс, требующий не только выделения определенных ключевых слов и элементов описания, касающихся как самого образа, так и стиля и эстетики, но и специфики устройства, принципов и схем работы самой генеративной системы.

Различные варианты использования генеративных нейросетей иллюстрируют большой потенциал инструментов ИИ в цифровом искусстве, дизайне, архитектуре, маркетинге, образовании и многих других сферах. В современной техногенной среде и технологической реальности проявляются и трансформируются не только традиционные отношения «животное – человек», но появляются и развиваются иные типы интеракций и пространств существования и взаимодействия (Schneiders E. et al., 2023).

Однако системы искусственного интеллекта по-прежнему испытывают значительные трудности визуализациии тех или иных образов, обычно из-за отсутствия семантического понимания и осмысленной интерпретации содержания изображения. Изображение — не просто набор пикселей или отдельных объектов, они несут семантическую и контекстную информацию, выражают определённые взгляды, ценности, идеи или идеалы, предлагая новые эффективные возможности для экспериментов и самовыражения.

Результаты исследования также показывают, что в среднем рассматриваемые модели показывают схожую эффективность; в зависимости от задач, примеров, специфики промпта и других факторов более адекватными, эстетичными, оригинальными и культуросообразными оказываются визуальные репрезентации различных нейросетей. При этом можно отметить, что Stable Diffusion является более универсальным, чем другие используемые инструменты; при использовании определённых запросов и описаний она генерирует изображения, соответствующие предполагаемым образным преставлениям в социальном и культурном аспектах. DALL·E способен создавать более глубокие и контекстуально определённые образы, а также можно быть более эффективным и эстетичным в создании артов в определённых стилях, например, сюрреалистической манере и аниме-стиле. Kandinsky также способен к созданию нетривиальных визуальных интерпретаций художественных образов, предлагая определённое разнообразие, но не всегда соотносясь с содержанием, контекстом и установленным описанием. Стоит учесть и тот факт, что рассматриваемые системы генерации поддерживают введение

большего числа параметров и потенциально способны к более комплексной и гибкой настройке.

Заключение

Генеративные модели и системы при помощи алгоритмов создают имитативную реальность, подражая существующим произведениям искусства. При этом форма выражения цифрового искусства еще находится в процессе становления, поскольку для его создателей в большей степени важна лишь содержательная сторона. Художественная ценность произведения искусства определяется уникальностью и неповторимостью в плане выражения и смысловой наполненностью содержания созданного художником образного мира. Как показывает проведённое исследование, результаты работы нейросетей и систем ИИ в сфере генерации изображений, визуализации различным образом сформированных запросов и описаний, репрезентации визуальных образов вполне способны рассматриваться и оцениваться в категориальных схемах искусства и понятиях, критериях эстетического, как со стороны воспринимающего субъекта, так и в объективно-теоретическом ракурсе.

Да, полученные изображения эстетичны, т. е. оцениваются как что-либо красиво-изящное, но для оценки их как произведения искусства этого все же недостаточно. Дело в том, что сущность художественного мышления проявляется через содержание образа, отражающего реальность. Художественная идея, представляя собой концептуальное мышление, не только интерпретирует типичные явления объективной действительности через художественный образ и формализируется в понятиях, но и отражает эмоционально-чувственное отношение субъекта к объективной действительности. Идея не сводится к простой совокупности образов. Она отражает не только сущее, т. е. существующее реально, но и должное, понимаемое как нравственно ценное, то, к чему необходимо стремиться, тем самым являясь основой для создания художественного произведения.

Проведенное исследование показывает непосредственную взаимосвязь языка описания, специфики языковой фундированности художественного образа, детерминирующих компонент промпта и соответствующей визуальной генерации, как, собственно, в предлагаемом изображении, так и в сущностных основаниях, интерпретативных смысловых формах и непосредственном интуитивном или эмоционально-чувственном восприятии.

Используя в качестве предмета и исходного материала устоявшиеся, литературно и мифологически определённые, социокультурно оформленные в личном и общественном сознании образы «культовых» представителей семейства «кошачьих», нами проведена визуализация образов посредством генерации изображений тремя нейросетевыми моделями на основе авторских, а также созданных с привлечением ИИ запросов и промптов. Визуальные объекты, генерируемые нейросетями и системами ИИ, воспринимающие субъ-

екты, вкупе с самим процессом генерации, отталкивающимся от предполагаемой исходной идеи и потенциального образа, а также языкового описания и сформулированного запроса, образуют единое неравновесное пространство идейного и эстетического взаимодействия.

Цифровая визуальность представляет собой быстро развивающуюся систему. При этом текст, будучи сам знаком и объединяя совокупность знаков, представляет собой обширное поле для семиозиса — процесса наделения знаков новыми значениями. В ходе интерпретации текст рассматривается как сложное семантическое целое, что подразумевает сопоставление и учет всех средств художественного изображения внутри текста с точки зрения их структурной организации, а также исследования параметров экстралингвистического характера. Ключевые понятия семиотики – знак и его интерпретация – также применимы к художественному тексту и его визуализации.

Список литературы

Aharonov, Y., Cohen, E., & Popescu, S. (2021). A dynamical quantum Cheshire Cat effect and implications for counterfactual communication. *Nature Communications*, 12(1), 4770. https://doi.org/10.1038/s41467-021-24933-9

Bartel, C., & Kwong, J. M. C. (2021). Pluralism, Eliminativism, and the Definition of Art. *Estetika: The European Journal of Aesthetics*, LVIII/XIV(2), 100–113. https://doi.org/10.33134/eeja.213

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Heruti, V., & Mashal, N. (2023). Effects of an Art Intervention Program Using Ambiguous Image-Text Interactions on Creative Thinking. *Empirical Studies of the Arts*, 02762374231215736. https://doi.org/10.1177/02762374231215736

Kenett, Y. N., Gold, R., & Faust, M. (2018). Metaphor Comprehension in Low and High Creative Individuals. *Frontiers in Psychology*, 9, 482. https://doi.org/10.3389/fpsyg.2018.00482

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kutela, B., Novat, N., Novat, N., Herman, J., Kinero, A., & Lyimo, S. (2023). The Rise of Prompt Engineering Jobs: A Review of the Nature of the Job, Qualifications, and Compensations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4625139

Makarov, I., Makarov, M., & Kiselev, D. (2021). Fusion of text and graph information for machine learning problems on networks. *PeerJ Computer Science*, 7, e526. https://doi.org/10.7717/peerj-cs.526

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., & Chen, M. (2021). *GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models*. arXiv. https://doi.org/10.48550/ARXIV.2112.10741

Paananen, V., Oppenlaender, J., & Visuri, A. (2024). Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing*, *22*(3), 458–474. https://doi.org/10.1177/14780771231222783

Rasrichai, K., Chantarutai, T., & Kerdvibulvech, C. (2023). Recent Roles of Artificial Intelligence Artists in Art Circulation. *Digital Society*, *2*(2), 15. https://doi.org/10.1007/s44206-023-00044-4

Schlöder, J. J., & Altshuler, D. (2023). Super Pragmatics of (linguistic-)pictorial discourse. *Linguistics and Philosophy*, *46*(4), 693–746. https://doi.org/10.1007/s10988-022-09374-x

Schneiders, E., Chamberlain, A., Fischer, J. E., Benford, S., Castle-Green, S., Ngo, V., Kucukyilmaz, A., Barnard, P., Row Farr, J., Adams, M., Tandavanitj, N., Devlin, K., Mancini, C., & Mills, D. (2023). TAS for Cats: An Artist-led Exploration of Trustworthy Autonomous Systems for Companion Animals. *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 1–5. https://doi.org/10.1145/3597512.3597517

Steinfeld, K. (2023). Clever little tricks: A socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, *21*(2), 211–241. https://doi.org/10.1177/14780771231168230

Tabačková, Z. (2018). Narrative odyssey of a congenital immigrant: The art of storytelling in Rabihala-meddine's the Angel of History. *Silesian Studies in English 2018. Proceedings of the 5th International Conference of English and American Studies 6th – 7th September 2018*, 167–175.

Taulli, T. (2023). *ChatGPT and Bard for Business Automation: Achieving AI-Driven Growth*. Apress. https://doi.org/10.1007/978-1-4842-9852-7

Zhang, B. (2023). From "Bildungsroman" to "World Classic Fairy Tale": The Translation and Reception of "Bambi" in Modern China. *Journal of Social Science Humanities and Literature*, *6*(6), 129–138. https://doi.org/10.53469/jsshl.2023.06(06).22

Булгаков, М. А. (2016). *Мастер и Маргарита*. Эксмо.

Ванюшкин, А. С., & Гращенко, Л. А. (2016). Методы и алгоритмы извлечения ключевых слов. *Новые информационные технологии в автоматизированных системах*, 19, 85–93.

Гегель, Г. В. Ф. (2007). *Лекции по эстетике (тома 1-2)*. Наука.

Гофман, Э. (2013). *Житейские воззрения кота Мурра*. Азбука.

Кривцун, О. А. (2019). Эволюция языка искусства: Культурные и художественные влияния. *Художественная культура*, 2, 2–25. https://doi.org/10.24411/2226-0072-2019-00012

Кэрролл, Л. (2019). *Приключения Алисы в Стране Чудес*. РОСМЭН.

Миловидов, S. V. (2022). Художественные особенности произведений компьютерного искусства, созданных с использованием технологий машинного обучения. *Артикульт*, 4, 36–48. https://doi.org/10.28995/2227-6165-2022-4-36-48

Нефёдов, И. В., & Чигрина, А. В. (2016). Отражение языковой игры в переводах сказки Льюиса Кэрролла «Алиса в стране чудес». *Актуальные проблемы гуманитарных и естественных наук*, 6–4, 55–60.

Тихонова, S. V. (2023). Сентиментальный визуальный контент новой нормальности: Как цифровые котики одомашнивают кризисы. *Вестник Санкт-Петербургского университета. Социология*, *16*(2), 149–167. https://doi.org/10.21638/spbu12.2023.203

Фадеева, Т. Е. (2023). «Союз» художника с нечеловеческим агентом—Утопия или рабочая модель художественного производства? *Известия Самарского научного центра Российской*

академии наук. Социальные, гуманитарные, медико-биологические науки, 25(88), 108–115. https://doi.org/10.37313/2413-9645-2023-25-88-108-115

Чудакова, М. О. (2011). *Не для взрослых. Время читать. Полка третья*. Время.

References

Aharonov, Y., Cohen, E., & Popescu, S. (2021). A dynamical quantum Cheshire Cat effect and implications for counterfactual communication. *Nature Communications*, 12(1), 4770. https://doi.org/10.1038/s41467-021-24933-9

Bartel, C., & Kwong, J. M. C. (2021). Pluralism, Eliminativism, and the Definition of Art. *Estetika: The European Journal of Aesthetics*, LVIII/XIV(2), 100–113. https://doi.org/10.33134/eeja.213

Bulgakov, M. A. (2016). *The Master and Margarita*. Eksmo Publ. (In Russian).

Carroll, L. (2019). *Alice's Adventures in Wonderland*. ROSMEN Publ. (In Russian).

Chudakova, M. O. (2011). *Not for adults. Time to read. Shelf three*. Time. (In Russian).

Fadeeva, T. E. (2023). "Union" of an Artist with a Non-Human Agent: Utopia or a Working Model of Artistic Production? *Izvestiya of the Samara Science Centre of the Russian Academy of Sciences. Social, Humanitarian, Medicobiological Sciences*, 25(88), 108–115. https://doi.org/10.37313/2413-9645-2023-25-88-108-115 (In Russian).

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Hegel, G. W. F. (2007). *Lectures on Aesthetics* (*Volumes* 1-2). Nauka Publ. (In Russian).

Heruti, V., & Mashal, N. (2023). Effects of an Art Intervention Program Using Ambiguous Image-Text Interactions on Creative Thinking. *Empirical Studies of the Arts*, 02762374231215736. https://doi.org/10.1177/02762374231215736

Hoffmann, E. (2013). *The Life and Opinions of the Tomcat Murr*. Azbuka Publ. (In Russian).

Kenett, Y. N., Gold, R., & Faust, M. (2018). Metaphor Comprehension in Low and High Creative Individuals. *Frontiers in Psychology*, 9, 482. https://doi.org/10.3389/fpsyg.2018.00482

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Krivtsun, O. A. (2019). The Evolution of the Language of Art: Cultural and Artistic Influences. *Art & Culture Sudies*, 2, 2–25. https://doi.org/10.24411/2226-0072-2019-00012 (In Russian).

Kutela, B., Novat, N., Novat, N., Herman, J., Kinero, A., & Lyimo, S. (2023). The Rise of Prompt Engineering Jobs: A Review of the Nature of the Job, Qualifications, and Compensations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4625139

Makarov, I., Makarov, M., & Kiselev, D. (2021). Fusion of text and graph information for machine learning problems on networks. *PeerJ Computer Science*, 7, e526. https://doi.org/10.7717/peerj-cs.526

Milovidov, S. V. (2022). Artistic Features of Computer Artworks Creating with Machine Learning Technology. *Articult*, 4, 36–48. https://doi.org/10.28995/2227-6165-2022-4-36-48 (In Russian).

Nefyodov, I. V., & Chigrina, A. V. (2016). Reflections of language play in translations of Lewis Carroll's tale "Alice in Wonderland". *Aktualnie problemi gumanitarnikh i estestvennikh nauk*, 6–4, 55–60. (In Russian).

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., & Chen, M. (2021). *GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models*. arXiv. https://doi.org/10.48550/ARXIV.2112.10741

Paananen, V., Oppenlaender, J., & Visuri, A. (2024). Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing*, *22*(3), 458–474. https://doi.org/10.1177/14780771231222783

Rasrichai, K., Chantarutai, T., & Kerdvibulvech, C. (2023). Recent Roles of Artificial Intelligence Artists in Art Circulation. *Digital Society*, *2*(2), 15. https://doi.org/10.1007/s44206-023-00044-4

Saratov State University, & Tikhonova, S. V. (2023). Sentimental visual content of the new normality: How digital cats domesticate crises. *Vestnik of Saint Petersburg University. Sociology*, *16*(2), 149–167. https://doi.org/10.21638/spbu12.2023.203 (In Russian).

Schlöder, J. J., & Altshuler, D. (2023). Super Pragmatics of (linguistic-)pictorial discourse. *Linguistics and Philosophy*, *46*(4), 693–746. https://doi.org/10.1007/s10988-022-09374-x

Schneiders, E., Chamberlain, A., Fischer, J. E., Benford, S., Castle-Green, S., Ngo, V., Kucukyilmaz, A., Barnard, P., Row Farr, J., Adams, M., Tandavanitj, N., Devlin, K., Mancini, C., & Mills, D. (2023). TAS for Cats: An Artist-led Exploration of Trustworthy Autonomous Systems for Companion Animals. *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 1–5. https://doi.org/10.1145/3597512.3597517

Steinfeld, K. (2023). Clever little tricks: A socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, *21*(2), 211–241. https://doi.org/10.1177/14780771231168230

Tabačková, Z. (2018). Narrative odyssey of a congenital immigrant: The art of storytelling in Rabihala-meddine's the Angel of History. *Silesian Studies in English 2018. Proceedings of the 5th International Conference of English and American Studies 6th – 7th September 2018*, 167–175.

Taulli, T. (2023). *ChatGPT and Bard for Business Automation: Achieving AI-Driven Growth*. Apress. https://doi.org/10.1007/978-1-4842-9852-7 (In Russian).

Vanyushkin, A. S., & Graschenko, L. A. (2016). Keyword extraction methods and algorithms. *Novie informatsionnie tekhnologii v avtomatizirovannikh sistemakh*, 19, 85–93. (In Russian).

Zhang, B. (2023). From "Bildungsroman" to "World Classic Fairy Tale": The Translation and Reception of "Bambi" in Modern China. *Journal of Social Science Humanities and Literature*, *6*(6), 129–138. https://doi.org/10.53469/jsshl.2023.06(06).22 (In Russian).