

Litera

Правильная ссылка на статью:

Жикулина К.П., Костромина В.В. Вычислительная креативность генеративной сети Midjourney в полимодальном пространстве // Litera. 2024. № 6. DOI: 10.25136/2409-8698.2024.6.70890 EDN: COCFNP URL: https://nbpublish.com/library_read_article.php?id=70890

Вычислительная креативность генеративной сети Midjourney в полимодальном пространстве

Жикулина Кристина Петровна

ORCID: 0000-0003-2488-4616

ассистент; кафедра общего и русского языкоznания; Российский университет дружбы народов имени Патриса Лумумбы

117198, Россия, г. Москва, ул. Миклухо-Маклая, 6, кафедра, Миклухо-Маклая, 12

✉ christina.zhikulina@gmail.com



Костромина Виктория Владимировна

магистр; кафедра общего и русского языкоznания; Российский университет дружбы народов имени Патриса Лумумбы

117198, Россия, Москва, г. Москва, ул. Миклухо-Маклая, дом 10, корпус 2

✉ kostromina_vv@pfur.ru



[Статья из рубрики "Автоматическая обработка языка"](#)

DOI:

10.25136/2409-8698.2024.6.70890

EDN:

COFCNP

Дата направления статьи в редакцию:

25-05-2024

Дата публикации:

01-06-2024

Аннотация: В данной статье исследуется полимодальное пространство в области вычислительной креативности у нейронных сетей. Объектом нашего исследования является полимодальная среда, которая объединяет в себе ряды разнородных кодов для выражения общей идеи, а предметом – возможность создания полимодального

цифрового искусства с помощью текстового и голосового промта в генеративной сети Midjourney. Цель исследования – доказать, что вычислительная креативность у машин может быть обнаружена и описана по результатам итераций в процессе создания изображений, что в свою очередь позволит говорить о сложной полимодальной системе как об отдельной цифровой категории полимодальности. Задачи, которые решались в рамках данного исследования: 1) обоснование использования термина полимодальный в контексте вычислительной креативности в цифровом искусстве нейронных сетей; 2) проведение серии тестов с генеративной сетью Midjourney для возможности описания вербально-визуального контента; 3) анализ и выведение алгоритма создания изображения в условиях суммаризации данных и трансформационной креативности. Нами был использован метод сплошной выборки при сборе языковых единиц по мере их встречаемости; контекстный анализ для систематического описания присутствующих вербальных и невербальных компонентах. Необходимо было провести эксперимент с генеративной сетью Midjourney для выявления закономерностей при создании графического пространства, а далее сравнить и сопоставить результаты итераций с оригинальным изображением. Научная новизна заключается в отсутствии исследований о полимодальном пространстве в контексте нейронных сетей и их генеративной способности. В ходе проведённого эксперимента нами были получены следующие результаты: термин «полимодальность» в контексте генеративной сети Midjourney и её «цифрового искусства» обусловлен наличием трёх каналов: верbalного, визуального и голосового (звукового); тесты показали, что способность нейронной сети к созданию изображений посредством промта находится на высоком уровне, однако прослеживаются грубые технические ошибки в системе, которые не позволяют пользователям в полной мере приблизиться к желаемому результату при генерации изображения; суммаризация данных позволяет говорить о наличии черт вычислительной креативности у генеративных сетей, так как в ходе создания изображения возникают новые образы (идеи), ранее не предполагаемые пользователем.

Ключевые слова:

искусственный интеллект, вычислительная креативность, трансформационная креативность, нейронная сеть, Миджорни, полимодальное пространство, полимодальный текст, итерация, промт, суммаризация

Финансирование. Благодарности:

Публикация выполнена в рамках грантовой поддержки научных проектов РУДН 124022500238-1 «Модель мультиязычного терминологического словаря».

Financing. Acknowledgments:

The publication has been supported by the RUDN University Scientific Projects Grant System, project No 124022500238-1 "Multilingual terminological dictionary model".

Введение

По словам мыслителя эпохи Просвещения Джонатана Свифта, воображение – это «искусство видеть вещи невидимые для окружающих» [\[20, с.11\]](#). Идея поиска новых мотивов на границе графики, науки, когнитивной психологии и математики применима сегодня ко многим произведениям искусства, созданным не только посредством красок и холста, но и с помощью компьютеров, алгоритмов, нейронных сетей и других форм искусственного интеллекта (далее – ИИ). Американские исследователи в области ИИ

называют такую возможность у машин – вычислительной креативностью (*computational creativity*).

Английский ученый Саймон Колтон предлагает следующее определение вычислительной креативности: «философия, наука и инженерия вычислительных систем, принимая на себя определённые обязанности, демонстрируют поведение, которое беспристрастные наблюдатели сочли бы творческим» [17]. В области компьютерного моделирования особое влияние оказывает подход профессора Маргарет Боден, в котором она выделяет исследовательскую креативность, где новая идея является продолжением традиций для предметной парадигмы, а творчество заключается в исследовании границ, содержания и потенциала креативного пространства [1, с. 116]. Также исследователь описывает комбинаторную креативность как создание нового с помощью комбинаций известных ранее идей (ассоциации) [15, с. 350] и трансформационную креативность, которая включает в себя несколько фундаментальных областей, формирующих пространство для появления ранее неизвестных идей [16, с. 362]. Хотя идеи, связанные с вычислительной креативностью машин не новы, настоящее развитие они получают только в наше время, когда суммаризация материалов и информации становится не только необходимостью, но и проблемой для многих направлений в исследованиях.

В 1968 году британский художник Гарольд Коэн создал компьютерную программу для рисования AARON, которая самостоятельно создавала картины. К более поздним примерам цифрового искусства относится совместное творчество пользователей с программой компьютерного зрения DeepDream, созданной в 2015 году разработчиком компании Google Александром Мордвинцевым. Программа использует искусственную нейросеть, которая ищет и усиливает закономерности в изображениях.

Нейросетевые технологии сегодня могут обучаться классификации и распознаванию объектов на основе множества тренировочных изображений или с помощью подробного текстового описания: осуществляют поиск неких паттернов и гиперболизируют их – «примерно так же, как поступает человек, когда смотрит на облака и пытается увидеть в них очертания животных» [19, с. 195]. Подобные программы способны помочь учёным лучше разобраться в том, как искусственные нейросети соотносятся с реальными сетями нейронов в зрительной коре мозга. Технологии такого типа помогают понять и то, как человеческий мозг ищет закономерности и смысл между несколькими компонентами [Там же].

С точки зрения Ю.А. Евграфовой, в современной лингвистике текст понимается как «единое структурно-смысловое целое (система линейного и нелинейного пространств), которое создаётся контаминацией элементов всех уровней, функционирующее в определённом прагматическом контексте, определяющим его восприятие и понимание» [3, с. 48]. Можно отметить, что исследователь выделяет «контаминацию элементов всех уровней» как ведущий фактор в понимании текста. В то же время А.П. Гусева подчёркивает, что в результате сочетания верbalного и неверbalного полимодальная коммуникация не является прямой, поскольку «план содержания высказывания, выражаемый значениями его компонентов, не совпадает с итоговым коммуникативным смыслом» [2, с. 100]. В одной из своих работ О.И. Максименко отмечает, что значительную долю знаний о мире человек получает благодаря зрению, «то есть визуальные знаки, к которым относятся как рисунки (в общем понимании этого слова), так и печатный текст, несут максимальную информацию» [6, с. 93]. Другими словами, создание изображения с нейронной сетью посредством текстового описания в условиях полимодальности тоже

должно включать в себя смысловые категории, требующие дополнительных усилий для истолкования. Подобный вывод можно сделать благодаря акценту на психологические особенности человека, но в случае с ИИ вопрос психологического восприятия мира пока остаётся гипотетическим.

Актуальность данной работы заключается в том, что вопросы и проблематика полимодального текста описываются в лингвистике почти пятьдесят лет, а нейронные сети начали активно использоваться для генерации «единого графического пространства» [\[11, с. 117\]](#), в котором соединяются вербальные и невербальные компоненты в пределах одного или нескольких изображений, относительно недавно. Многие пользователи жалуются на неверное истолкование запроса у генеративных сетей, и как результат этого – не точную или частичную генерацию запрашиваемого изображения. Алгоритмы обработки текстовых входных данных остаются неизученными, неописанными и интуитивными как для исследователей, так и для пользователей.

Новизной является то, что *полимодальный текст* и его создание с помощью ИИ исследуются впервые. Хотя «наше общество ревниво оберегает всё, что связано с творчеством», учёные из вычислительного подразделения Голдсмитского колледжа Саймон Колтон и Герейнт Уиггинс считают, что *вычислительная креативность* (или *машинное творчество*) – «вероятно, последний предел возможностей ИИ» [\[18, с. 26\]](#). Термин *вычислительная креативность* (или *машинное творчество*) можно рассматривать с нескольких сторон: 1) область разработки ИИ, где творчество моделируется посредством компьютера или другой техники (планшет, телефон); 2) программы, которые расширяют возможности человека и являются инструментами в творческом процессе.

Теоретическая значимость работы заключается в том, что у лингвистов появилась возможность исследовать языковые единицы и явления и их функционирование в новом формате *цифрового текста*; сопоставлять результаты с данными, которые были получены при исследованиях традиционных языковых форм; сделать вклад в развитие перспективного направления лингвистического исследования – «анализ влияния речевых практик в цифровой среде на систему языка и языковых компонентов» [\[9, с. 192\]](#). Также, в связи с тем, что многие исследователи отмечают высокую частотность в Интернете полимодальных текстов и простоту, с которой интернет-платформы и программы позволяют интегрировать и создавать изобразительные произведения (не только в коммуникации, но и в литературе, бизнесе и других сферах) [\[11, с. 115\]](#), полимодальные тексты и полимодальный вид коммуникации открывается для исследований с новой, неизученной стороны – развитие в *цифровой среде* с помощью технологий нейронных сетей.

Практическая значимость исследования заключается в том, что полученные материалы и результаты можно использовать для лекционных и семинарских занятий по «Общему языкознанию» в разделах, которые развиваются как теоретические, так и практические навыки в рамках дисциплины «Интернет-лингвистика». Примеры и их описания станут актуальными в таких темах как «Роль лингвистики в изучении Интернет-коммуникации», «Вербальная специфика коммуникации в Интернете», «Полимодальные тексты в Интернете» (разделы «Инфографика», «Ситуативные полимодальные произведения» и «Характеристика изобразительной составляющей полимодальных текстов»).

Материалы исследования

Материалами послужили минималистичные иллюстрации со вкладышей из популярной турецкой жевательной резинки ‘Love is...’ (Рис.1). Изображения подобного типа выбраны

не случайно, так как на картинках не прорисовываются мельчайшие детали, лица героев, пейзажи и т.д. Важно и то, что вербальный компонент (надпись-послание) на вкладыше содержит предложение, которое представлено с помощью двух автономных графических частей: «подлежащего 'Love is...' и предикативной части высказывания» [13, с. 152].

Рисунок 1/ Picture 1

Примеры вкладышей из турецкой жевательной резинки «Love is...» /

Examples of cartoon stuffers from Turkish chewing gum 'Love is...'



Источник: Коллекция вкладышей популярной жвачки 'Love is...'. URL: <https://www.liveinternet.ru/users/zimuka/post354225218/> (дата обращения: 01.05.2024) /

Source: The collection of cartoon stuffers of popular gum 'Love is...'. URL: <https://www.liveinternet.ru/users/zimuka/post354225218/> (accessed: 01.05.2024).

Соединение вербальных и невербальных компонентов в 'Love is' позволяют проанализировать полученные результаты со стороны полимодального текста, «связи его компонентов, использования/неиспользования конкретного семиотического кода, особенности декодирования авторской интенции и наличию в нём интертекста» [8, с. 299].

Результаты и обсуждения

В ходе эксперимента было сгенерировано и проанализировано около 20 изображений. Для обсуждения мы выбрали процесс генерации на примере только одного рисунка, так как этапы создания оказались долгими, а алгоритм сборки и обработки вербальных и невербальных компонентов, посредством нейронной сети Midjourney, оказался одинаковым для изображений любого типа.

Нейронная сеть Midjourney должна была воспроизвести образы персонажей со вкладыша с помощью текстового запроса, голосовых оценок пользователя и генеративных возможностей обработки данных.

Для описания тестов с нейронной сетью Midjourney нами был выбран вкладыш с сюжетом, в котором главные герои мини-комикса постарели (Рис.2). Образными признаками русского концепта «старость» могут «быть седой и покрытый морщинами» [10, с. 1228]. «Седой» как образный признак концепта «старость» может рассматриваться одинаково как в русской, так и в других культурах. Предполагается, что образный признак становится универсальным для вычислительной креативности у машин.

На вкладыше (Рис.2) мы видим две сидящие фигуры, которые изображены на белом фоне. Белый фон в разработке мини-комиксов 'Love is' является классическим компонентом изображений и позволяет сделать акцент на ключевых образах и их атрибутах. Кроме того, белый фон может указывать на представление ситуации вне времени и пространства – отсутствие хронотопа. Условными невербальными знаками выступают скамейка без спинки и лужайка с травой и цветами. Лужайка указывает на сезон, благодаря чему можно предположить в какой конкретный период разворачиваются события: весной, летом или ранней осенью. Данные указатели могут служить ассоциацией с ключевым вербальным элементом «любовь», а также символом жизни в связи с глаголом «прожить», использованном в части вербального компонента на изображении. Другой условный компонент позволяет раскрыть кинесику героев, поскольку скамейка со спинкой могла бы скрыть важную форму физической близости – объятия.

Фигуры двух людей – мужская (слева) и женская (справа), сидящих на скамейке, представляют собой иконический компонент, демонстрирующий семейную пару, которая прожила «вместе до старости». На гендерные признаки указывают классические представления о мужчине и женщине – у мужской фигуры короткие волосы, а в сравнении с женской мужская фигура выше по росту; у женской фигуры длинные волосы, рост ниже фигуры слева.

Отражение пожилого возраста пары обуславливается вышеописанным универсальным образным признаком – сединой (бледным цветом волос как у мужского персонажа, так и у женского), а также вещественным атрибутом – тростью. Важно, что седина смешивается с общим цветом головы: так, мы видим, что волосы не совсем седые, а светло-серые у мальчика и светло-желтые у девочки. Трость присутствует у обоих героев в идентичной вариации и уравнивает позицию персонажей, что указывает на объединяющий признак по возрасту. Стоит отметить, что мужская фигура одета в синий пиджак, а женская – в розовое платье, что подчеркивает гендерную идентичность героев.

В связи с тем, что пара изображается спиной к читателю, то иные визуальные признаки старости нам недоступны. Однако комплектация фигур не указывает на иные признаки старости: мы видим, что пара стандартной комплекции и повторяет графические образы иных комиксов 'Love is', намекая на детский или молодой возраст персонажей (как в предшествующих сериях вкладышей). Сформированный классический образ героев комиксов 'Love is' не позволяет говорить о героях как о пожилых людях, но может указывать на их временную роль в данном облике.

Слоган в качестве вербального компонента «Любовь это... прожить вместе до старости» актуализирует образ пожилого человека, но не закрепляет в качестве постоянной характеристики героев. При отсутствии вербального компонента возраст героев ставится под сомнение, в связи с чем при описании персонажей будет наиболее корректна формулировка «мальчик и девочка».

Рисунок 2/ Picture 2

Сюжет из вкладышей к жевательной резинке «Love is...» /

The cartoon stuffers from chewing gum 'Love is...'

Любовь это...



...прожить вместе до старости

Источник: Социальный интернет-сервис, фотохостинг «Pinterest». URL: <https://www.pinterest.ca/pin/2251868556772924/> (дата обращения: 01.05.2024) /

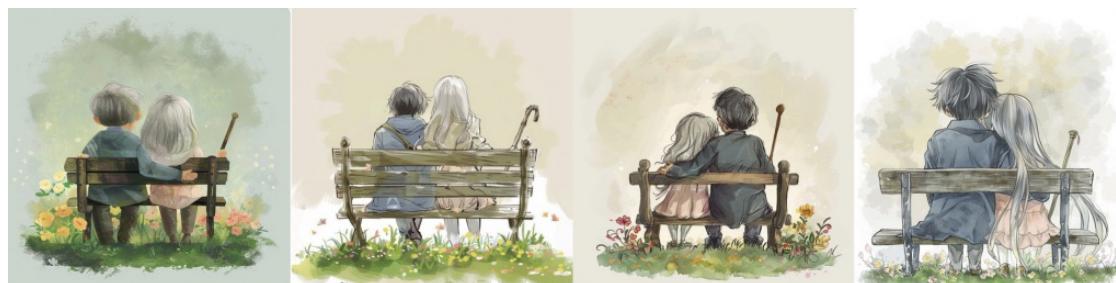
Source: An image sharing and social media service 'Pinterest'. URL: <https://www.pinterest.ca/pin/2251868556772924/> (accessed: 01.05.2024).

Рассмотрим подробно полученные результаты.

Рисунок 3/ Picture 3

Результат итерации в блоке №1 по созданию изображений в Midjourney /

The result of iteration in block №1 by creation images in Midjourney.



Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Девочка одета в светло-розовое платье, а мальчик в синий пиджак.

Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Девочка одета в светло-розовое платье, а мальчик в синий пиджак.

Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Девочка одета в светло-розовое платье, а мальчик в синий пиджак.

Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Девочка одета в светло-розовое платье, а мальчик в синий пиджак.

Источник: Телеграм чат-бот 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Имя пользователя: @GPT4Telegrambot (дата обращения: 12.05.2024) /

Source: Telegram chat-bot 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Username: @GPT4Telegrambot (accessed: 12.05.2024).

Блок 1

Мы видим, что один и тот же текст пользователя, заданный в чат-бот Midjourney, содержится под четырьмя вариациями картинок (Рис.3). Система обработки промтa (*prompt*), то есть запроса, подсказки, входных данных или инструкций, конвертирует сразу четыре разных изобразительных варианта для одного и того же текста. Предполагается, что по описательным характеристикам в промте генеративная сеть не всегда может подобрать точный стиль рисовки, а также создать детализацию образов на изображении с первого раза.

Некоторые пользователи считают, что одного подробного текста-описания достаточно для нейронной сети, чтобы она сгенерировала простое изображение. Однако большинство сетей сегодня находятся на «переходной» стадии развития. Это означает, что на классификацию запросов, обработку входных данных и расстановку акцентов нейронная сеть тратит большие ресурсы и время. Например, после отправки промтa Midjourney сообщает пользователю в чате, что «обработка запроса может занять 1-3 минуты» [\[14\]](#). В реальности один запрос может обрабатываться 10-15 минут. Более того, моделирование изображения происходит в несколько этапов (в нашем исследовании – блоков).

Также важно отметить, что типы речи, включая описание, которое мы используем в промте, представляют собой «мыслительный комплекс, выражающий связи одновременности, последовательности или причинно-следственной зависимости между явлениями» [\[7, с. 24\]](#), что в свою очередь влияет и на создание полимодального пространства: входные данные (текстовый запрос) + обработка запроса (нейронная сеть) = результат итерации (изображение). В данной схеме становится понятно, что в генерации картинок ведущим оказывается текст – текстовое описание – поиск и подборка информации, образов по вербальным-невербальным компонентам, и, самое важное, результат, объединяющий в себе мысль человека в вербальной форме и искусственную «мысль» в невербальном формате.

В первом промте мы вводили недостаточно качественное описание запроса, так как было необходимо протестировать взаимодействие двух каналов в Midjourney: текстового и изобразительного. Нами были умышленно допущены речевые или фактические ошибки, которые имитируют промт с точки зрения обычного пользователя, не знающего алгоритмов работы при запросе к нейронной сети. Так, в первом промте блока 1 оказывается, что «мальчик и девочка в старости сидят на скамейке», «сидят спиной» и «держат по палочке пенсионеров». Если оценивать данный текст с лингвистической точки зрения, то он больше напоминает набор слов. Однако для генеративной сети количество допущенных ошибок не является препятствием для создания изображения. Обращая внимание на невербальную семантику картинок, полученных нами (Рис.3), видно, что нейронная сеть изображает людей в старости, больше похожих телосложением на детей, но, одновременно, и на людей в старости за счёт седого цвета волос, что соответствует оригинальному отображению героев на вкладыше. Также сеть буквально понимает запрос в промте «сидят спиной», и мы видим, что представленные на рисунках герои сидят повернутыми спиной относительно смотрящего на изображение. «Палочку пенсионеров» Midjourney определяет по-разному: как посох, как клюку или как трость. Описания в промте «обнимают друг друга сзади» оказывается недостаточно: на трёх изображениях мальчик обнимает девочку одной рукой (Рис.3, картинки 1,3,4), на одном – объятия отсутствуют (Рис.3, картинка 2). Помимо всего прочего, на одном из изображений девочка и мальчик меняются местами (лево-право) на скамейке (Рис.2,

картинка 3).

Промт «зеленая лужайка с цветами» была воспринята нейронной сетью по-разному. Мы видим, что лужайка на указывает на поздний весенний период в связи с пышностью цветов (Рис.3, картинка 1); намекает на середину лета (Рис.3, картинка 2); на конец летнего сезона (Рис.3, картинка 3); изображён осенний период, поскольку цветы выглядят пожухшими и сухими (Рис.3, картинка 4). Невербальная характеристика лужайки на представленных результатах варьируется, поскольку дешифровать сезон в оригинал также не представляется возможным.

Вероятно, что в первом промте не хватило описательных характеристик для палочки, посадки и положения рук, а также условных компонентах (вид лавочки, характер фона), поэтому нейронная сеть предоставляет разные варианты в результатах итерации. Так как данные подробности отсутствовали в нашем описании, мы не можем говорить о неверном истолковании пользовательского запроса.

В следующем блоке (Рис.4) мы уточнили ряд компонентов в тексте и получили новые результаты.

Рисунок 4/ Picture 4

Результаты итераций в блоке №2 по созданию изображений в Midjourney /

The results of iterations in block №2 by creation images in Midjourney.



Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Герои выглядят как в жевательной резинке Love is... Девочка одета в светло-розовое платье, а мальчик в синий пиджак. Общий фон белый. На картинке надпись: «Любовь это...» (вверху), «... прожить вместе до старости» (внизу).

Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Герои выглядят как в жевательной резинке Love is... Девочка одета в светло-розовое платье, а мальчик в синий пиджак. Общий фон белый. На картинке надпись: «Любовь это...» (вверху), «... прожить

вместе до старости» (внизу). 16:11

Prompt: мальчик и девочка в старости сидят на скамейке, у мальчика седые черные волосы, короткие. У девочки светло-желтые седые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, а другими руками держат по палочке для пенсионеров. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Стиль картинки – минималистичный. Пример: как в жевательной резинке Love is... Девочка одета в светло-розовое платье, а мальчик в синий пиджак. Общий фон белый. На картинке надпись: «Любовь это...» (вверху), «... прожить вместе до старости» (внизу).

17:06

Источник: Телеграм чат-бот 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Имя пользователя: @GPT4Telegrambot (дата обращения: 12.05.2024) /

Source: Telegram chat-bot 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Username: @GPT4Telegrambot (accessed: 12.05.2024).

Блок 2

В данном блоке второй промт задаётся заново в чат-бот Midjourney с добавлением описательных характеристик. Невозможно не обратить внимание на то, что изображения в блоке 1 (Рис.3) далеки от оригинального изображения на вкладыше жевательной

резинки 'Love is' (Рис.2) – минималистичность отсутствует. В запросе мы дописали, что «герои выглядят как в жевательной резинке Love is...», а также: «общий фон белый» и «на картинке надпись: «Любовь это...» (вверху), «...прожить вместе до старости» (внизу)».

В результатах итерации мы видим, что на двух изображениях фон стал белым (Рис.4, картинки 1 и 4), «мальчик и девочка» обнимают друг друга сзади на всех четырёх вариантах, везде добавлена надпись-послание. Midjourney воспринимает положение надписи (вверху, внизу) по-своему: 1) когда подлежащее 'Love is' вверху изображения, а предикативная часть – внизу изображения (Рис.4, картинка 1); 2) когда подлежащее 'Love is' и предикативная часть находятся в одном месте, но 'Love is' – выше остального текста (Рис.4, картинка 4); 3) когда подлежащее 'Love is' и предикативная часть находятся на одном уровне (параллельно друг другу), но не сверху и снизу изображения, а справа и слева друг от друга (Рис.4, картинки 2 и 3). Важно также отметить, что текст не генерируется на русском языке: нейронная сеть переводит его на английский язык. Как показал эксперимент, при вынесении текста на изображения случаются задвоения слов типа «love is...is» или «together together» или «old...old».

В остальном можно отметить, что после введения дополнительного описания в промт «герои выглядят как в жевательной резинке Love is...», несколько вариаций рисунков Midjourney стали более приближенными к оригинальному изображению на выбранном нами вкладыше. Изменился внешний вид «девочки» – волосы стали короткими (Рис.4, картинки 1 и 3), хотя характеристика с длинными волосами не корректировалась нами во втором промте.

Промт «зеленая лужайка с цветами» в данной серии результатов выглядит наиболее однородно, наглядно и точно не указывает на сезон. Все компоненты носят универсальный характер – так же, как и в оригинальном вкладыше 'Love is'.

Далее мы следовали другому алгоритму: выбрали изображение (Рис.4, картинка 3), так как оно показалось нам наиболее приближенным к рисунку на вкладыше жевательной резинки, и продолжили создавать промты с дополненным описанием уже относительно этого изображения.

Рисунок 5/ Picture 5

Результаты итераций в блоках №3 и №4 по созданию изображений в Midjourney /

The results of iterations in blocks №3 and №4 by creation images in Midjourney.



Prompt: мальчик и девочка в старости сидят на скамейке без спинки. У мальчика короткие седые черные волосы. У девочки длинные светло-желтые волосы. Они сидят спиной, обнимают друг друга сзади, у каждого из них есть в руках трость. Под скамейкой, на которой они сидят, зелёная лужайка с цветами. Минималистичный стиль картинки. Девочка одета в светло-розовое платье, а мальчик в синем пиджаке. Общий фон картинки – белый. На картинке есть надпись: «Любовь это прожить вместе до старости».

Prompt: мальчик и девочка в старости сидят на скамейке без спинки. У мальчика короткие седые черные волосы. У девочки длинные светло-желтые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, у каждого из них есть в руках трость. Под скамейкой, на которой они сидят, зелёная лужайка с цветами.

Prompt: мальчик и девочка в старости сидят на скамейке без спинки. У мальчика короткие седые черные волосы. У девочки длинные светло-желтые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, у каждого из них есть в руках трость. Под скамейкой, на которой они сидят, зелёная лужайка с цветами.

Prompt: мальчик и девочка в старости сидят на скамейке без спинки. У мальчика короткие седые черные волосы. У девочки длинные светло-желтые волосы, длинные. Они сидят спиной, обнимают друг друга сзади, у каждого из них есть в руках трость. Под скамейкой, на которой они сидят, зелёная лужайка с цветами.

Источник: Телеграм чат-бот 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Имя пользователя: @GPT4Telegrambot (дата обращения: 12.05.2024) /

Source: Telegram chat-bot 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Username: @GPT4Telegrambot (accessed: 12.05.2024).

Блоки 3 и 4

В данных блоках детализация и неточности прорабатывались с помощью добавления третьего канала: голосовой промт. Посредством голосовых сообщений мы изменили в описании «палочку для пенсионеров» на «трость», убрали строчку с пояснением «герои выглядят как в жевательной резинке Love is...» и добавили описание для скамейки, на которой сидят герои. Нами было замечено, что два блока подряд скамейка изображалась Midjourney как объект со спинкой. Однако на оригинальном изображении, которое мы взяли за основу, скамейка нарисована без спинки (Рис.2).

В блоке 3 (Рис.1, картинка 1) сохраняется ввод текстового послания на рисунке посредством русского языка (Рис.5, картинка 1). Также, нейронная сеть повторяет серию задвоений слов «is...is» и «old old», дублирует одно и то же послание два раза (Рис.5, картинки 1-3). Скамейка остаётся со спинкой, однако впервые появляются трости с закруглённой ручкой у обоих героев на всех изображениях.

В блоке 4 (Рис.5, картинки 2-4) мы вносим письменные изменения в форму голосового промта из блока 3: переводим надпись-послание с русского на английский язык. Midjourney выдаёт серию изображений, приближенных к источнику, на который мы опирались, но с некоторыми неточностями:

1. – спинка у скамейки отсутствует только на одном изображении (Рис.5, картинка 3);
2. – трость со стороны «мальчика» находится поодаль (Рис.5, картинки 2,3);
3. – надпись-послание, введенное на английском языке, снова содержит задвоение слов (Рис.5, картинка 2);
4. – волосы у «девочки» не светло-жёлтые седые (как во всех промтах), а просто седые;
5. – цвет платья «девочки» меняется со светло-розового на жёлтое (Рис.5, картинка 1);

6. – волосы у «девочки» – короткие, а в промт введено описание, что длинные (на всех картинках);

7. – белый фон сохраняется только на одном изображении (Рис.5, картинка 4).

По нашему мнению, самыми приближенными к оригиналу оказываются несколько изображений (Рис.6), сгенерированных Midjourney. Продолжать генерацию по блокам с исправлением промта можно до бесконечности, однако на 4-5 блоках генеративная сеть начинает «ходить по кругу» в изображениях и их вариациях, изменяя только цвета и положения героев в пространстве.

Рисунок 6/ Picture 6

Итоговые результаты итераций в блоках по созданию изображений в Midjourney /

The final results of iterations in blocks by creation images in Midjourney.



Источники: Социальный интернет-сервис, фотохостинг «Pinterest». URL: <https://www.pinterest.ca/pin/2251868556772924/> (дата обращения: 01.05.2024); Телеграм чат-бот 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Имя пользователя: @GPT4Telegrambot (дата обращения: 12.05.2024) /

Sources: An image sharing and social media service 'Pinterest'. URL: <https://www.pinterest.ca/pin/2251868556772924/> (accessed: 01.05.2024); Telegram chat-bot 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Username: @GPT4Telegrambot (accessed: 12.05.2024).

Можно сделать вывод, что в совместном создании изображения человека и Midjourney самым важным элементом оказывается текст – промт. У профессиональных пользователей выработана система, по которой они выстраивают текст для промта, чтобы приблизиться к необходимому результату. Промт оказывается определённым видом

искусства или, по-другому, ключевым модулем при создании изображения.

Существуют также ошибки или неточности в генерации, которые не способен исправить даже хорошо написанный промт: (по результатам нашего исследования) посадка героев, подборка цвета, корректный перенос верbalного компонента – надписей-посланий на изображении.

В программировании термин *итерация* применяют к специальной обработке данных, при которой происходит многократное повторение действий, при этом рекурсия, то есть вызов функции (процедуры) из неё же самой – не реализуется. Другими словами, рекурсивная программа позволяет описывать повторяющиеся или бесконечные действия (вычисления) без явных повторов предшествующих частей программы и использования циклов. Мы считаем, что данный термин также подходит для описания действий при генерации изображения с помощью нейронной сети в условиях полимодальности. Более того, поэтапная система формирования полимодальной среды в цифровом мире может быть обозначена как суммаризация вербальных, невербальных и голосовых компонентов. Подобную систему уже описывали О.Ю. Коломийцева и А.Н. Москалёва в статье, затрагивающей вопросы реализации категории полимодальности в Инстаграм-дискурсе. Они отметили, что текст и изображение образуют «новую вербально-визуальную форму» [\[4, с. 115\]](#), которая интересна тем, что по отдельности визуальные и вербальные составляющие могут не нести глубинных смыслов [Там же]. К их системе мы хотели бы добавить и голосовой канал в полимодальное пространство Midjourney, что оправдывало бы использование термина *полимодальный* в контексте цифрового искусства с генеративными сетями.

Заключение

Приемлемость применения термина *полимодальный* в контексте вычислительной креативности (или машинного творчества) в цифровом искусстве нейронных сетей обусловлена многоканальностью процесса создания изображения. С одной стороны, происходит взаимодействие вербальных и невербальных компонентов посредством промта и визуализации его генеративной сетью. С другой стороны, для быстрых исправлений неточностей в описательных характеристиках промта можно использовать голосовые входные данные. Во взаимодействии всех частей возникает полимодальная среда, в которой создаётся полимодальное пространство.

В ходе эксперимента удалось протестировать и описать работу Midjourney. Вербально-визуальные изображения, сгенерированные с помощью данной нейронной сети, частично совпадают с оригинальным сюжетом, изображенным на вкладыше турецкой жевательной резинки 'Love is'. Мы предполагаем, что Midjourney обучена рисовать слишком хорошо, поэтому она запрограммирована работать по методу индукции, а не дедукции. Нейронная сеть не обучена минимизировать, но обучена – улучшать качество. Это доказывают результаты итерации в блоке 1, где изображения выглядят детализировано и более реалистично. Однако сюжет «мальчик и девочка в старости», изображённый на вкладыше 'Love is', удаётся сохранить на каждом сгенерированном изображении.

Взаимодействие всех блоков при создании изображения происходит на высоком уровне, однако, генеративная сеть достаточно часто некорректно итерирует запрос пользователя. Ошибочное видение запросов не всегда зависит от качественно написанного промта. Сегодня мы можем только предполагать, что данные возможности в нейронных сетях будут доработаны разработчиками.

Таким образом, вычислительная креативность (или машинное творчество) наблюдается у Midjourney хотя бы потому, что в блоках создания присутствуют варианты изображений, которые мало соответствуют характеристикам, заданным пользователем (особенно – первый промт). Даже если новое виденье изображения нейронной сетью является результатом выведения среднего арифметического из множественной суммаризации, это означает, что генеративная сеть работает по принципу, описанному Маргарет Боден, трансформационной креативности, которая предполагает создание ранее неизвестных идей с опорой на фундаментальные знания в заданной области.

Библиография

1. Белова С.С. Творчество: психологические и компьютерные модели / Психология. Журнал Высшей школы экономики. 2008. Т.5, №4. С. 112-119
2. Гусева А.П. Семиотически гетерогенный художественный текст как содержательно осложненная коммуникация // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2018. №18 (816). С. 98-109.
3. Евграфова Ю.А. Лингвосемиотика экрана: моделирование реальности в экранных текстах (на материале текстов кино, телевидения и сети Интернет) : дисс. на соискание ученой степени доктора филологических наук. (10.02.19). / МГОУ. – Москва, 2020.
4. Коломийцева О.Ю., Москалева А.Н. Способы реализации категории полимодальности в англоязычном инстаграм-дискурсе. Вестник Международного института рынка. 2021. №2. С. 115-125.
5. Коллекция вкладышей популярной жвачки 'Love is...'. Режим доступа: <https://www.liveinternet.ru/users/zimuka/post354225218/> (дата обращения: 01.05.2024).
6. Максименко О.И. Поликодовый vs. креолизованный текст: проблема терминологии. Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика, №. 2, 2012, С. 93-102.
7. Нечаева О.А. Функционально-смысловые типы речи: (Описание, повествование, рассуждение) : Автореферат дис. на соискание ученой степени доктора филологических наук. (10.02.01) / Моск. обл. пед. ин-т им. Н. К. Крупской. – Москва : [б. и.], 1975.
8. Новоспасская Н.В., Дугалич Н.М. Терминосистема теории поликодовых текстов // Русистика. 2022. Т. 20. № 3. С. 298-311. <http://doi.org/10.22363/2618-8163-2022-20-3-298-311>
9. Полонский А.В. Медиалект: язык в контексте медийной культуры // Современный дискурс-анализ. 2018. № 3 (20). Т.1. С. 189-199.
10. Сафаралиева Л.А., Перфильева Н.В. Моделирование многомерного лингвокультурного концепта на материале концепта СТАРОСТЬ // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2023. Т. 14. № 4. С. 1217-1234. <https://doi.org/10.22363/2313-2299-2023-14-4-1217-1234>
11. Сонин А.Г. Экспериментальное исследование поликодовых текстов : основные направления // Вопросы языкоznания. 2005. № 6. С. 115-123.
12. Социальный интернет-сервис, фотохостинг «Pinterest». Режим доступа: <https://www.pinterest.ca/pin/2251868556772924/> (дата обращения: 01.05.2024).
13. Степанова И. В. (2013). Креолизованный текст как средство презентации концепта Love (на материале комиксов Love is). Вестник Челябинского государственного университета, (24 (315)), С. 152-156.
14. Телеграм чат-бот 'ChatGPT | Midjourney | Claude | Suno AI — GPT4Telegrambot Inc.'. Имя пользователя: @GPT4Telegrambot (дата обращения: 12.05.2024).
15. Boden M.A. Creativity and artificial intelligence // Artificial Intelligence. 1998.103. Pp. 347-356.
16. Boden M.A. Computer models of creativity// Handbook of Creativity. R.J. Sternberg

- (ed.). Cambridge University Press, 1999. Pp. 351-372.
17. Colton S. From Computational Creativity to Creative AI and Back Again. *Interalia Magazine*. 2019. Режим доступа: <https://www.interaliamag.org/articles/simon-colton/> (дата обращения: 04.05.2024).
18. Colton S. & Wiggins, G. A. Computational creativity: The final frontier? In *ECAI 2012-20th European Conference on Artificial Intelligence, 27-31 August 2012, Montpellier, France-Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstration*. Frontiers in Artificial Intelligence and Applications. Vol. 242. 2012. IOS Press. Pp. 21-26. <https://doi.org/10.3233/978-1-61499-098-7-21>
19. Pickover A. Clifford. *Artificial Intelligence: An Illustrated History: From Medieval Robots to Neural Networks*. Sterling Publishing Co., Inc. (USA) via Alexander Korzhenevski Agency (Russia), 2021.
20. Santini C. *Kinttsugi: Finding Strength in Imperfection*. Andrews McMeel Publishing LLC. 2019.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Автор рецензируемой статьи обращает внимание на вычислительную креативность генеративной сети Midjourney в полимодальном пространстве. Предметная область работы достаточно понятна, в принципе обозначен и ценз актуальности вопроса. Следовательно, материал соотносится с одной из рубрик журнала, он, так или иначе, актуален. Формальные грейды исследования выдержаны: манифестация основных позиций дна – «Идея поиска новых мотивов на границе графики, науки, когнитивной психологии и математики применима сегодня ко многим произведениям искусства, созданным не только посредством красок и холста, но и с помощью компьютеров, алгоритмов, нейронных сетей и других форм искусственного интеллекта (далее – ИИ). Американские исследователи в области ИИ называют такую возможность у машин – вычислительной креативностью (computational creativity)», или «Нейросетевые технологии сегодня могут обучаться классификации и распознаванию объектов на основе множества тренировочных изображений или с помощью подробного текстового описания: осуществляют поиск неких паттернов и гиперболизируют их – «примерно также, как поступает человек, когда смотрит на облака и пытается увидеть в них очертания животных», или «Актуальность данной работы заключается в том, что вопросы и проблематика полимодального текста описываются в лингвистике почти пятьдесят лет, а нейронные сети начали активно использоваться для генерации «единого графического пространства», в котором соединяются вербальные и невербальные компоненты в пределах одного или нескольких изображений, относительно недавно». Вектор оценки нейросетей на данный момент субъективен, даются только варианты оценки, общего же описания нет. Автора пытается подчеркнуть новизну исследования, однако, и она слабо конкретна: «Новизной является то, что полимодальный текст и его создание с помощью ИИ исследуются впервые. Хотя «наше общество ревниво оберегает всё, что связано с творчеством», учёные из вычислительного подразделения Голдсмитского колледжа Саймон Колтон и Герейнт Уиггинс считают, что вычислительная креативность (или машинное творчество) – «вероятно, последний предел возможностей ИИ...». Считаю, что частный случай может заинтересовать, но он не является обобщением: «Материалами послужили минималистичные иллюстрации со вкладышей из популярной турецкой жевательной резинки 'Love is...' (Рис.1). Изображения подобного типа выбраны не

случайно, так как на картинках не прорисовываются мельчайшие детали, лица героев, пейзажи и т.д. Важно и то, что вербальный компонент (надпись-послание) на вкладыше содержит предложение, которое представлено с помощью двух автономных графических частей: «подлежащего 'Love is...' и предикативной части высказывания». Статистика / данные эксперимента введены верно, фактических нарушений нет: «В ходе эксперимента было сгенерировано и проанализировано около 20 изображений. Для обсуждения мы выбрали процесс генерации на примере только одного рисунка, так как этапы создания оказались долгими, а алгоритм сборки и обработки вербальных и невербальных компонентов, посредством нейронной сети Midjourney, оказался одинаковым для изображений любого типа». Оценка «визуала» дана объективно, автор старается прибегнуть к современной методологии анализа: «Фигуры двух людей – мужская (слева) и женская (справа), сидящих на скамейке, представляют собой иконический компонент, демонстрирующий семейную пару, которая прожила «вместе до старости». На гендерные признаки указывают классические представления о мужчине и женщине – у мужской фигуры короткие волосы, а в сравнении с женской мужская фигура выше по росту; у женской фигуры длинные волосы, рост ниже фигуры слева». Формально значимые моменты описываются, но должный диалог мнений остается: «мы видим, что один и тот же текст пользователя, заданный в чат-бот Midjourney, содержится под четырьмя вариациями картинок (Рис.3). Система обработки промта (prompt), то есть запроса, подсказки, входных данных или инструкций, конвертирует сразу четыре разных изобразительных варианта для одного и того же текста. Предполагается, что по описательным характеристикам в промте генеративная сеть не всегда может подобрать точный стиль рисовки, а также создать детализацию образов на изображении с первого раза». Удачно в режиме сопоставлений вводится авторский комментарий: «В данных блоках детализация и неточности прорабатывались с помощью добавления третьего канала: голосовой промт. Посредством голосовых сообщений мы изменили в описании «палочку для пенсионеров» на «трость», убрали строчку с пояснением «герои выглядят как в жевательной резинке Love is...» и добавили описание для скамейки, на которой сидят герои. Нами было замечено, что два блока подряд скамейка изображалась Midjourney как объект со спинкой. Однако на оригинальном изображении, которое мы взяли за основу, скамейка нарисована без спинки (Рис.2)». Вариация оценки полярна, что и важно, точки зрения вариативны. Итог работы соотносится с основным блоком, фактических нарушений не выявлено. Стоит согласиться, что «в ходе эксперимента удалось протестировать и описать работу Midjourney. Вербально-визуальные изображения, сгенерированные с помощью данной нейронной сети, частично совпадают с оригинальным сюжетом, изображенном на вкладыше турецкой жевательной резинки 'Love is'. Мы предполагаем, что Midjourney обучена рисовать слишком хорошо, поэтому она запрограммирована работать по методу индукции, а не дедукции. Нейронная сеть не обучена минимизировать, но обучена – улучшать качество. Это доказывают результаты итерации в блоке 1, где изображения выглядят детализировано и более реалистично. Однако сюжет «мальчик и девочка в старости», изображённый на вкладыше 'Love is', удаётся сохранить на каждом сгенерированном изображении». Импонирует в работе вариант нетривиальной оценки генеративной сети, ее особенностей и специфики. Основные требования издания учтены, текст не нуждается в принципиальной правке и коррективе. Рекомендую статью «Вычислительная креативность генеративной сети Midjourney в полимодальном пространстве» к публикации журнале «Litera».