

УДК 004.8

**ОБЗОР DALL-E, STABILITY AI И MIDJOURNEY ДЛЯ МУЗЫКАЛЬНЫХ
ОБЛОЖЕК****Серпинский Роман Эдуардович**

Студент группы ИУК5-61Б

Калужский филиал Московского государственного технического университета имени Н.Э.

Баумана

serpinskiyrea@student.bmstu.ru

Алехина Елена Дмитриевна

Старший преподаватель кафедры ИУК5 «Системы обработки информации»

Калужский филиал Московского государственного технического университета имени Н.Э.

Баумана

Alena-gur@bmstu.ru

Романовский Илья Олегович

Студент группы ИУК5-61Б

Калужский филиал Московского государственного технического университета имени Н.Э.

Баумана

romanovskiyio@student.bmstu.ru

Аннотация

В статье рассмотрены возможности трёх современных генеративных моделей изображений: DALL-E 3, MidJourney 5.2 и Stability AI. Выполнялись различные тестовые запросы, направленные на проверку способностей ИИ. Результаты показали, что каждая из моделей имеет свои уникальные сильные стороны. Данное исследование подчеркивает важность выбора подходящей модели для конкретных задач и открывает перспективы для дальнейших исследований и улучшений в области генеративных моделей изображений.

Ключевые слова: запрос, генерация, нейросеть, изображение**COMPARATIVE ANALYSIS OF CORPORATE INFORMATION NETWORK
MONITORING SYSTEMS****Serpinskiy Roman Eduardovich**

Student of the IUK5-61B group

Kaluga Branch of the Bauman Moscow State Technical University

serpinskiyrea@student.bmstu.ru

Alyokhina Elena Dmitrievna

Senior Lecturer at the Department of IUK5 "Information Processing Systems"

Kaluga Branch of the Bauman Moscow State Technical University
Alena-gur@bmstu.ru

Romanovsky Ilya Olegovich

Student of the IUK5-61B group

Kaluga Branch of the Bauman Moscow State Technical University
romanovskiyio@student.bmstu.ru

ABSTRACT

The article considers the possibilities of three modern generative image models: DALL·E 3, MidJourney 5.2 and Stability AI. Various test queries aimed at testing the AI's abilities were performed. The results showed that each of the models has its own unique strengths. This study highlights the importance of choosing an appropriate model for specific tasks and opens up prospects for further research and improvements in the field of generative image models.

Keywords: query, generation, neural network, image

Введение

С развитием технологий искусственного интеллекта, генерация изображений на основе текстовых описаний стала важной областью исследований, находящей своё применение в различных сферах, включая музыку. Обложки музыкальных альбомов играют ключевую роль в визуальной идентичности музыки, привлекая внимание слушателей и отражая суть музыкального произведения. Среди наиболее продвинутых и востребованных моделей, способных создавать уникальные и креативные обложки для музыкальных альбомов, выделяются DALL·E 3, Stability AI и MidJourney 5.2.

DALL·E является одной из первых моделей, которые продемонстрировали способность генерировать сложные изображения по текстовым запросам, включая абстрактные и сюрреалистические обложки. Stability AI фокусируется на стабильности и качестве создаваемых изображений, что важно для воспроизведения последовательного визуального стиля. MidJourney, в свою очередь, известна своей художественной выразительностью и креативностью, что делает её популярной среди музыкантов и дизайнеров.

DALL·E 3

Dalle [1], разработанный OpenAI, представляет собой значительный скачок в области генерации изображений ИИ. Этот инструмент быстро привлек внимание художников, дизайнеров и энтузиастов технологии по всему миру. Его главная привлекательность заключается в его революционном подходе к сочетанию искусственного интеллекта с художественным творчеством, что позволяет создавать потрясающие, высококачественные изображения.

Расширенное преобразование текста в изображение: выдающей особенностью Dalle является его способность интерпретировать и преобразовывать текстовые описания в визуально привлекательные изображения. Этот сложный процесс включает в себя алгоритмы глубокого обучения, которые понимают и визуализируют сложные концепции и абстрактные идеи, передаваемые в тексте.

Творческая универсальность: инструмент предлагает беспрецедентную универсальность, позволяя пользователям создавать изображения от реалистичных

представлений до сюрреалистического искусства. Эта гибкость делает его ценным активом для различных приложений, включая цифровое искусство, дизайн продукта и творческий мозговой штурм. [2]

Выходы с высоким разрешением: Dalle производит изображения с замечательной четкостью и детализацией, устанавливая высокий стандарт качества в искусстве, созданном искусственным интеллектом. Эта возможность особенно полезна для профессиональных художников и дизайнеров, которым для работы требуются изображения с высоким разрешением.

Stability AI

Stability AI [3] стал серьезным инструментом в генерации изображений на основе искусственного интеллекта, обслуживающим широкий круг пользователей от новичков до опытных профессионалов. Он известен своими надежными алгоритмами и пользовательским интерфейсом, который уравнивает простоту с функциональностью, делая передовую технологию искусственного интеллекта доступной для более широкой аудитории.

Стабильная и эффективная генерация изображений: в основе ИИ стабильности лежит его способность давать последовательные и надежные результаты. Алгоритмы платформы оптимизированы для эффективности, обеспечивая быструю генерацию изображений без ущерба для качества.

Удобный интерфейс: признавая важность пользовательского опыта, Stability AI имеет интуитивно понятный интерфейс, который упрощает процесс создания изображений, сгенерированных искусственным интеллектом. Эта философия дизайна делает его доступным для тех, кто новичок в искусстве искусственного интеллекта, при этом предлагая глубину, необходимую профессионалам. [4]

Настройка и управление: пользователи имеют значительный контроль над процессом генерации изображений, что позволяет настраивать его в соответствии с конкретными потребностями и предпочтениями. Этот уровень контроля особенно привлекателен для пользователей, которым требуются точные результаты для своих проектов.

Midjourney 5.2

Midjourney [5] – это генеративная модель искусственного интеллекта, разработанная для создания изображений на основе текстовых описаний. Запущенная в 2022 году, она быстро завоевала популярность благодаря своей способности создавать высококачественные, реалистичные и художественно выразительные изображения.

Художественные и абстрактные изображения: Midjourney специализируется на создании изображений, которые склоняются к художественному и абстрактному. Этот уникальный фокус отличает его, что делает его предпочтительным выбором для художников, графических дизайнеров и творческих профессионалов, ищущих вдохновение и новые визуальные выражения.

Интуитивно понятный дизайн: дизайн платформы подчеркивает простоту использования и интуитивно понятный пользовательский интерфейс. Даже те, у кого минимальный опыт в области искусственного интеллекта или цифрового искусства, могут эффективно ориентироваться и использовать его функции, способствуя творчеству и экспериментам. [6]

Сотрудничество с сообществом: Midjourney делает сильный акцент на сотрудничестве и участии сообщества. Он предлагает пользователям инструменты и пространства для обмена своими творениями, сбора отзывов и сотрудничества в проектах, способствуя созданию яркого и поддерживающего сообщества энтузиастов искусства ИИ.

Сравнение нейросетей

Было сделано несколько запросов для каждой нейросети, чтобы определить какая из них лучше всего подойдет для создания обложек. Включены следующие критерии, по которым будет осуществляться отбор:

- дизайн существ
- передача эмоций
- сюрреалистическая генерация
- точность выполнения запроса
- работа с фоном

1. Дизайн существ

Запрос: «An image of a 'thunderbird,' a mythical creature that controls storms, depicted in its natural habitat during a fierce thunderstorm.»

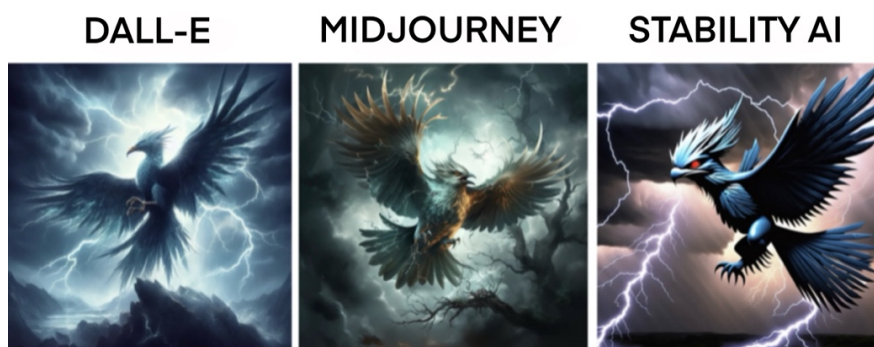


Рис. 1. Результат запроса «Дизайн существ»

Этот запрос требует творческого подхода к созданию несуществующих вещей, проверяя границы возможностей ИИ. Кроме того, в задании используются элементы окружающей среды и погоды, добавляются слои сложности, и можно увидеть, как ИИ генерирует специальные эффекты, такие как молния.

- DALL E: изображение создано в мрачных тонах с акцентом на молнии и грозовую атмосферу. Однако, само существо выглядит менее детализированным по сравнению с конкурентами, и в нем отсутствуют некоторые элементы драматичности и фантазии.

- MidJourney: этот ИИ продемонстрировал впечатляющую визуализацию. «Thunderbird» выглядит эпически, как мифическое существо, управляющее бурей. Детали, такие как перья, молнии и общее настроение изображения, переданы на высшем уровне, что делает его наиболее соответствующим запросу.

- Stability AI: здесь «thunderbird» выглядит не так впечатляюще и объемно, тоже самое касается фона, детализация и реализм уступают MidJourney. Существо выглядит мультяшным, что снижает общую эффектность изображения.

2. Передача эмоций

Запрос: «A portrait of a woman holding a wilted rose, her expression one of profound sadness and longing.»



Рис. 2. Результат запроса «Передача эмоций»

Может ли нейросеть передать чувства и эмоции персонажа? Ведь иногда дело не только во внешнем виде, но и в настроении. Этот запрос проверяет способность ИИ передавать эмоции с помощью мимики и символических элементов, позволяя понять нюансы человеческих эмоций. Лучше всего с заданием справилась Stability AI, которая корректно передала эмоции персонажа и атмосферу, в отличие от остальных нейросетей.

- DALL·E: изображение женщины с увядшей розой в руке выглядит тщательно выполненным, но выражение лица персонажа не полностью передает глубину чувств, таких как печаль и тоска. Символика увядшей розы присутствует, но эмоциональный подтекст недостаточно ярко выражен.

- MidJourney: здесь женщина изображена с более детализированным лицом и увядшей розой. Однако, хотя технически изображение выполнено хорошо, эмоции на лице персонажа кажутся несколько искусственными и менее естественными, что снижает общее впечатление от работы.

- Stability AI: Stability AI действительно передала эмоции персонажа наиболее удачно. Выражение лица женщины, её поза и сама увядшая роза создают сильное впечатление грусти и тоски. Атмосфера изображения поддерживает эмоциональный фон, делая его наиболее убедительным и трогательным среди всех представленных вариантов.

3. Сюрреалистическая генерация

Запрос: «Ancient Egyptian pharaoh using a modern smartphone while riding a dinosaur.»

DALL-E

MIDJOURNEY

STABILITY AI



Рис. 3. Результат запроса «Сюрреалистическая генерация»

В этом запросе объединяются разные временные периоды и проверяется, насколько ИИ владеет исторической достоверностью, чтобы не допустить анахронизма. Хороший генератор изображений должен обеспечить сочетание, которое при этом будет выглядеть правдоподобно.

- DALL·E: смогла показать наиболее сбалансированное и правдоподобное изображение, где древнеегипетский фараон органично взаимодействует с современным смартфоном и динозавром. Все элементы сцены хорошо сочетаются, и не возникает чувства разрыва между эпохами. Такой подход показывает высокую способность ИИ интегрировать различные временные слои и культурные элементы в одно целое, сохраняя при этом историческую достоверность и реалистичность.

- MidJourney: результат MidJourney выглядит наиболее качественно и детализировано, с высоким уровнем визуального исполнения. Тем не менее, изображение страдает от недостаточной гармонизации всех элементов и точности генерации. Несмотря на высокое качество, MidJourney не удалось достигнуть того уровня согласованности, который демонстрирует DALL·E.

- Stability AI: также представила интересное изображение, но оно менее впечатляющее по сравнению с DALL·E и MidJourney. Фараон и смартфон кажутся добавленными к сцене без особого внимания к деталям или историческому контексту. В результате изображение теряет в правдоподобности и выглядит несколько искусственно.

4. Точность выполнения запроса (согласованность)

Запрос: «An astronaut riding a steel horse on the moon. The astronaut is wearing a medieval armor with a party hat and a green sword.»

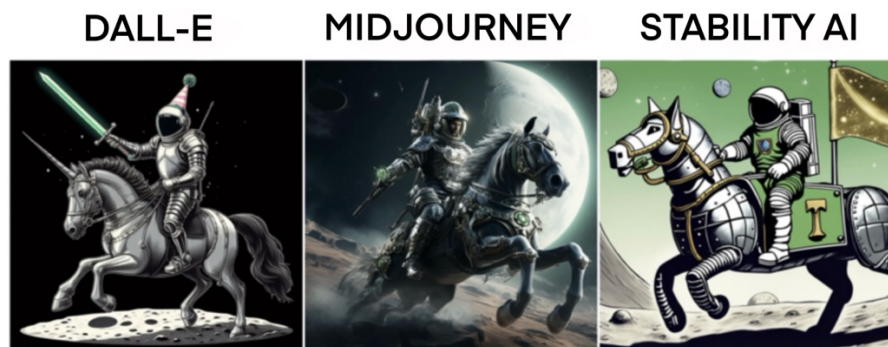


Рис. 4. Результат запроса «Точность выполнения запроса»

Точность выполнения запроса очень важна в генерации изображений, так как из-за её отсутствия теряется вся идейность работы.

- DALL-E: на изображении четко видны все элементы запроса. Астронавт действительно сидит на стальном коне на Луне, он одет в средневековую броню, на нём праздничная шляпа, а в руке он держит зелёный меч. Эта согласованность деталей и точность выполнения запроса делают результат работы DALL-E наиболее точным и завершённым.

- MidJourney: изображение выглядит визуально впечатляющим и художественно выразительным, но не содержит всех требуемых элементов. Хотя астронавт на стальном коне присутствует, не видно ни праздничной шляпы, ни зелёного меча, что снижает точность выполнения запроса.

- Stability AI: эта модель также сгенерировала изображение, где астронавт сидит на стальном коне на Луне, но элементы, такие как праздничная шляпа и зелёный меч, отсутствуют. Это показывает недостаточную согласованность в выполнении конкретных деталей запроса.

5. Работа с фоном

Запрос: «A dreamy landscape where clouds are made of cotton candy and rivers flow with liquid gold.»



Рис. 5. Результат запроса «Работа с фоном»

Данный запрос отвечает на вопрос, смогут ли эти модели искусственного интеллекта внести свое разнообразие и красочность изображению. Идея в том, чтобы получить цельную картину из несовместимых элементов. Задавая сочетание узнаваемых элементов в непривычном виде, можно оценить, насколько хорошо каждый ИИ понимает и строит сцены с нуля.

- DALL-E: справился с задачей лучше всех. Облака из сахарной ваты и реки с жидким золотом выглядят не только узнаваемо, но и эстетично. Пейзаж, созданный DALL-E 3, действительно впечатляет своей фантазийной красотой и гармонией, что подтверждает его

способность создавать сложные и необычные сцены, сохраняя при этом высокую визуальную привлекательность.

- MidJourney: также представил интересный и детализированный пейзаж, однако ему не удалось достичь нужного уровня, как у DALL·E. Облака и реки выглядят хорошо, но ощущение сказочности несколько уступает по сравнению с работой DALL·E. Несмотря на это, MidJourney продемонстрировал высокое качество визуализации и детализацию, которые могут впечатлить зрителя.

- Stability AI: предоставил неплохую попытку, но ей не хватило той же фантазийной красоты и детализации, что у DALL·E и MidJourney. Хотя элементы сцены — облака из сахарной ваты и реки с жидким золотом — присутствуют, их визуальное исполнение выглядит несколько плоско и менее захватывающе. Тем не менее, Stability AI показала, что способна справляться с творческими заданиями, хотя и нуждается в улучшении.

Вывод

В ходе данного исследования был проведен обзор и сравнительная оценка возможностей трёх передовых генеративных моделей изображений: DALL·E, MidJourney и Stability AI. Для анализа использовались различные тестовые запросы, направленные на проверку разнообразных аспектов работы этих систем, таких как дизайн существ, передача эмоций, историческая фантастика, когерентность и работа с фоном.

Дизайн существ. Этот тест показал, что MidJourney продемонстрировала лучшую способность к созданию детализированных и выразительных изображений мифических существ. Хотя DALL·E также достигла высоких результатов, изображение MidJourney было более эпическим и детализированным, создавая целое настроение. Это указывает на то, что MidJourney обладает выдающимися возможностями для генерации сложных и фантастических образов, лучше передавая динамику и атмосферу сцены.

Передача эмоций. В этом тесте Stability AI оказалась на высоте, лучше всех передав глубокие человеческие эмоции. Это свидетельствует о способности модели улавливать и воспроизводить тонкие нюансы мимики и символизма, что важно для задач, требующих эмоционального отклика. MidJourney и DALL·E также показали хорошие результаты, но им не удалось достичь такой же глубины эмоционального выражения, как у Stability AI.

Историческая фантастика. DALL·E продемонстрировала лучшие результаты в создании правдоподобных и когерентных изображений, объединяющих элементы из разных временных периодов. MidJourney, хотя и создала качественное изображение, не смогла гармонично вписать древнеегипетского фараона в сцену. Это подчеркивает, что DALL·E обладает лучшими способностями к созданию сложных и смешанных временных образов без потери исторической достоверности.

Точность выполнения запроса. DALL·E 3 продемонстрировала наилучшую когерентность, точно выполняя все аспекты запроса, включая такие мелкие детали, как зелёный меч и праздничная шляпа. MidJourney и Stability AI не смогли полностью передать заданные элементы, что указывает на важность детализации и внимательности к запросу, особенно для задач, требующих высокой точности исполнения.

Работа с фоном. Здесь снова DALL·E 3 показала лучшие результаты, создавая эстетически приятный и гармоничный пейзаж. Этот тест подчеркивает способность DALL·E 3 генерировать сложные и фантастические сцены, при этом сохраняя визуальную привлекательность и правдоподобие. MidJourney и Stability AI также показали хорошие результаты, но не достигли той же степени фантазийной красоты и детализации.

Заключение

Исследование показало, что каждая из моделей имеет свои сильные стороны и особенности. MidJourney проявила себя как сильный инструмент для создания

детализированных и выразительных изображений, особенно в области фантастических и мифических существ. Stability AI показала превосходные результаты в передаче эмоций, демонстрируя глубокое понимание мимики. DALL·E 3 оказалась наиболее универсальной и способной моделью, отлично справляющейся с заданиями на когерентность, историческую фантастику и работу с фоном.

Эти результаты подчеркивают важность выбора подходящей модели для конкретных задач дизайнера, создающего музыкальные обложки и указывают на потенциал дальнейшего развития и улучшения генеративных моделей изображений. В будущем целесообразно проводить более глубокие исследования, включающие дополнительные аспекты и разнообразные тестовые запросы, чтобы ещё точнее оценить возможности и ограничения каждой из моделей.

Список литературы:

1. DALL·E: Создание изображений с помощью текста. [Электронный ресурс] URL: <https://openai.com/>
2. Как работает DALL-E. [Электронный ресурс] URL: <https://habr.com/ru/companies/ruvds/articles/687508/>
3. Stability AI. (2023). Введение в модели Stability AI. [Электронный ресурс] URL: <https://www.stability.ai/>
4. Как пользоваться нейросетями Stable Diffusion и Stable Doodle от Stability AI. [Электронный ресурс] URL: <https://cpamonstro.com/neural-networks-ai/stability-ai-kak-polzovatsa-neirosetyami/?ysclid=lx8vru028n591141797>
5. Midjourney [Электронный ресурс] URL: <https://www.midjourney.com/>
6. MidJourney Documentation. (2022). Руководство пользователя MidJourney. [Электронный ресурс] URL: <https://www.midjourney.com/docs>

References:

1. DALL·E: Create images using text. [Electronic resource] URL: <https://openai.com/>
2. How does DALL-E work? [Electronic resource] URL: <https://habr.com/ru/companies/ruvds/articles/687508/>
3. Stability AI. (2023). An introduction to the Stability AI model. [Electronic resource] URL: <https://www.stability.ai/>
4. How to use Stable Diffusion and Stable Doodle neural networks from Stability AI. [Electronic resource] URL: <https://cpamonstro.com/neural-networks-ai/stability-ai-kak-polzovatsa-neirosetyami/?ysclid=lx8vru028n591141797>
5. Midjourney [Electronic resource] URL: <https://www.midjourney.com/>
6. MidJourney Documentation. (2022). MidJourney User's Guide. [Electronic resource] URL: <https://www.midjourney.com/docs>