

ТЕХНИЧЕСКИЕ НАУКИ



УДК 004.8

Сравнительный анализ эффективности различных моделей машинного обучения в задачах генерации контента

М.В. Ступина, И.В. Садовая, А.В. Балашев

Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

Аннотация

Проведён сравнительный анализ эффективности различных моделей машинного обучения для генерации контента, включая текст, изображения, видео и музыку. Рассмотрены основные метрики оценки качества генерируемого контента для каждого типа данных. Проведено тестирование и сравнение таких моделей, как GPT, DALL-E, Vid2Vid и Mubert, на соответствующих наборах данных. Показаны сильные и слабые стороны каждой модели. Определены наиболее эффективные на текущий момент подходы для задач генерации разных типов контента.

Ключевые слова: машинное обучение, генерация контента, модели машинного обучения, GPT, BERT, генерация текста, изображений, видео, музыки, сравнительный анализ

Для цитирования. Ступина М.В., Садовая И.В., Балашев А.В. Сравнительный анализ эффективности различных моделей машинного обучения в задачах генерации контента. *Молодой исследователь Дона*. 2024;9(3):4–13.

Comparative Analysis of the Effectiveness of Different Machine Learning Models in Content Generation Tasks

Mariya V. Stupina, Irina V. Sadovaya, Artur V. Balashev

Don State Technical University, Rostov-on-Don, Russian Federation

Abstract

The paper provides a comparative analysis of the effectiveness of various machine learning models for content generation, including text, images, video and music. It examines the main metrics for evaluating the quality of generated content for each type of data. Models such as GPT, DALL-E, Vid2Vid and Mubert are tested and compared on relevant datasets. Their strengths and weaknesses are identified, and the most effective approaches to generating different types of content are determined.

Keywords: machine learning, content generation, machine learning models, GPT, BERT, text generation, image generation, video generation, music generation, benchmarking

For citation. Stupina MV, Sadovaya IV, Balashev AV. Comparative Analysis of the Effectiveness of Different Machine Learning Models in Content Generation Tasks. *Young Researcher of Don*. 2024;9(3):4–13.

Введение. В современном мире модели машинного обучения играют ключевую роль в генерации контента, включая тексты, изображения, видео и музыку. Большие языковые модели (LLM) стали важным этапом в области обработки естественного языка, позволяя общаться, анализировать и генерировать контент с невиданным ранее уровнем сложности [1]. Создание генеративных видеомоделей позволяет формировать новые и реалистичные видеопоследовательности, что открывает возможности для различных целей, таких как синтез видео, творческое повествование и создание контента [2].

В области автоматической генерации музыки (AMG) до сих пор не проводилось сравнение оценки глубокого обучения и других методов. Однако существует множество исследований, которые анализируют различные методы генерации музыки, включая использование сырых аудиоданных и символьных токенов [3]. Модели глубокого обучения, такие как рекуррентные нейронные сети (RNN) и трансформаторные модели, показали впечатляющие результаты в генерации реалистичного и связного текста [4]. В области генерации видео существуют различные модели, называемые генеративными, которые используют алгоритмы машинного обучения для изучения шаблонов и распределений, лежащих в основе обучающих данных [5].

Создание и использование моделей машинного обучения для генерации контента стало возможным благодаря быстрому развитию компьютерных технологий и улучшению алгоритмов. Это позволяет создавать новые возможности для различных областей применения методов глубокого обучения и приложений, а также стимулирует дальнейшие исследования в этой области.

Таким образом, цель данной работы состоит в определении наиболее эффективных моделей машинного обучения, а также подходов и методов, направленных на выявление областей для возможного внедрения улучшений и инноваций. В задачи исследования входят:

- 1) обзор информационных ресурсов по исследуемой проблеме;
- 2) оценка и сравнение моделей с использованием различных метрик эффективности;
- 3) анализ результатов сравнения.

Основная часть. Методы оценки эффективности моделей машинного обучения. Различные количественные и качественные измерения используются для оценки эффективности моделей машинного обучения [6]. Точность, показывающая процент правильно классифицированных объектов среди всех объектов, является одной из наиболее распространённых количественных мер для задач классификации [7]. Но в несбалансированных классах эта метрика может быть недостоверной. Поэтому часто используются дополнительные метрики: точность, отражающая долю верно классифицированных положительных объектов среди всех отнесённых моделью к положительному классу, и полнота (повтор), отражающая долю верно классифицированных положительных объектов среди всех реально положительных объектов. Для комплексной оценки используется F-мера, средняя гармоническая между точностью и полнотой.

Средняя квадратичная ошибка, коэффициент детерминации и коэффициент корреляции Пирсона являются основными метриками для задач регрессии. Они дают возможность оценить, насколько точные значения, предсказанные моделью, соответствуют реальным.

Для оценки качества работы моделей машинного обучения, помимо количественных метрик, часто используется экспертная оценка результатов. После анализа конкретных примеров специалисты дают независимую качественную оценку реалистичности, смысловой связности и соответствия ожиданиям.

Выбор подходящих метрик для оценки эффективности работы моделей машинного обучения в задачах генерации контента. Наиболее релевантными показателями эффективности с учетом специфики задач по генерации контента являются:

- для оценки качества сгенерированного текста:
 - 1) грамматическая корректность;
 - 2) естественность текста;
- для оценки качества сгенерированных изображений:
 - 1) разрешение изображений;
 - 2) соответствие заданной тематике;
- для оценки качества сгенерированного аудио/видео:
 - 1) естественность;
 - 2) синхронизация видео- и аудиоданных;
 - 3) соответствие заданному сюжету;
- для оценки качества сгенерированной музыки:
 - 1) музыкальность;
 - 2) ритмичность;
 - 3) естественность музыкальных переходов.

Обзор архитектур моделей машинного обучения. Генерация текста является одной из наиболее активно развивающихся областей применения методов машинного обучения. За последние годы появилось несколько моделей, демонстрирующих впечатляющие результаты в задачах автоматической генерации текста на естественном языке. Одними из наиболее известных являются модели GPT (Generative pre-trained transformer) и BERT (Bidirectional Encoder Representations from Transformers). Обе эти модели основаны на архитектуре трансформеров и предварительно обучены на больших текстовых корпусах. Модель GPT была предложена компанией OpenAI в 2018 году. Её особенностью является исключительно однонаправленная архитектура декодера, что позволяет генерировать текст посимвольно слева направо. GPT демонстрирует высокое качество генерируемого текста, его связность, лексическое и синтаксическое разнообразие. В отличие от GPT, модель BERT использует двунаправленный кодировщик и обучена для задач понимания естественного языка. Это позволяет ей лучше анализировать контекст и смысловые связи в тексте. Однако из-за отсутствия однонаправленного декодера возможности BERT по генерации новых текстов ограничены.

Модель GPT лучше подходит для креативной генерации длинных связных текстов, тогда как BERT — для анализа и классификации уже имеющихся текстов.

Для решения задач классификации и категоризации текстов также активно применяются методы глубокого обучения. Были рассмотрены две популярные модели: RNN и CNN. Рекуррентные нейронные сети (RNN) хорошо подходят для анализа последовательных данных, к которым относятся и тексты. Они анализируют текст последовательно, слово за словом, учитывая контекст. Это позволяет RNN эффективно извлекать смысловые связи в тексте. Свёрточные нейронные сети (CNN) рассматривают текст как изображение признаков слов в предложениях. Их преимуществом является возможность анализа более глобальных паттернов в тексте с помощью операций свёртки. Сравнивая RNN и CNN, можно сказать, что RNN лучше подходят для небольших текстов и учёта контекстных зависимостей, а CNN — для анализа длинных текстов и выявления у них общих признаков. Выбор конкретной архитектуры зависит от поставленной задачи классификации. Также перспективным направлением является использование предварительно обученных языковых моделей (BERT, ELMo) в качестве основы для классификатора текстов. Это позволяет достичь более высокого качества за счёт переноса знаний о языке.

Существует несколько подходов, позволяющих создавать оригинальные музыкальные произведения на основе обучения на существующих данных. Одним из наиболее простых методов является использование марковских цепей.

Модель идентифицирует шаблоны в представленных выборках, а затем имитирует и изменяет их, чтобы создать свою собственную оригинальную композицию [8]. Однако такой подход не учитывает более высокоуровневую структуру музыки. Применение рекуррентных нейронных сетей, особенно LSTM и GRU, является еще одним перспективным направлением. Они могут создавать последовательности произвольной длины, учитывая временные зависимости в данных. Рекуррентные сети отлично справляются с задачами продолжения мелодий и создания простых композиций.

Наиболее современные модели основаны на архитектуре трансформеров. Обучаясь на больших музыкальных данных, они могут создавать реалистичные многоголосные композиции длительностью в несколько минут. Примерами таких моделей являются Jukebox, Music Transformer и MuseNet. По сравнению с другими методами, они дают возможность создавать более сложные и естественно звучащие мелодии.

Современные методы глубокого обучения демонстрируют впечатляющие результаты в задачах автоматической генерации музыки. Следует отметить, что использование нейронных сетей для создания музыки способно произвести революцию в музыкальной индустрии. Обладая способностью создавать уникальную и разнообразную музыку, нейронные сети могут помочь музыкантам и композиторам в творческом процессе [9].

Ещё одной активно развивающейся областью применения методов глубокого обучения является автоматическая генерация изображений. Эта задача подразумевает создание новых, реалистичных изображений на основе анализа больших наборов данных.

Одним из наиболее популярных алгоритмов машинного обучения являются генеративно-сопоставительные сети (GAN). Их работа основана на соревновании двух нейронных сетей — генератора, создающего новые изображения, и дискриминатора, определяющего их реалистичность. По мере обучения генератор может создать изображения, которые сложно отличить от настоящих.

Другим мощным классом алгоритмов машинного обучения являются вариационные автокодировщики (VAE). Они кодируют изображение в компактное векторное представление, а затем декодируют его обратно. Изменяя код, можно генерировать новые изображения. VAE позволяют получать более разнообразные и интерпретируемые результаты, по сравнению с GAN.

Таким образом, современные алгоритмы машинного обучения открывают большие возможности для автоматической генерации реалистичного визуального контента. Комбинирование подходов на основе GAN и VAE является перспективным направлением для создания ещё более качественных и разнообразных изображений.

За последние годы задача автоматической генерации реалистичного видеоконтента привлекает все большее внимание исследователей в области компьютерного зрения и машинного обучения. Успехи в этой области могут осуществить прорыв в таких приложениях, как автоматизированный видеомонтаж, компьютерная анимация, синтез видео по текстовому описанию и многие другие.

Одним из первых подходов к задаче генерации видео представили исследователи из NVIDIA. Платформа основана на условных GAN. В частности, метод объединяет тщательно разработанные генератор и дискриминатор с пространственно-временной сопоставительной целью. В таких моделях генератор создаёт кадры видео, а дискриминатор оценивает их реалистичность. Обучаясь вместе, эти сети позволяют достичь впечатляющих результатов в генерации неподвижных изображений. Однако их возможности для синтеза динамических видеопоследовательностей ограничены из-за отсутствия моделирования временных зависимостей.

Более перспективным подходом для генерации видео кадр за кадром с учётом предыдущего контекста является использование рекуррентных нейронных сетей, в частности LSTM. Такие модели позволяют получать более плавные и реалистичные видео. Однако они по-прежнему уступают по качеству реальным видеоданным.

Наиболее мощные на сегодняшний день результаты в генерации фотореалистичного видео демонстрирует модель VideoGPT, предложенная компанией Anthropic в 2022 году. Эта модель основана на архитектуре трансформеров и предварительно обучена на больших объёмах видеоданных. VideoGPT способна генерировать логичные и реалистичные видео длительностью до нескольких минут на основе текстовых подсказок. При этом качество получаемого видео значительно превосходит результаты предыдущих подходов. Таким образом, можно отметить, что наиболее перспективными моделями для задачи генерации видео в настоящее время являются подходы на основе больших трансформеров, обученных в автокодировочном режиме на значительных объёмах видеоданных. Дальнейшее накопление данных и рост вычислительных мощностей позволит создавать ещё более реалистичный и разнообразный сгенерированный видеоконтент.

Сравнение моделей. Для сравнения эффективности различных моделей машинного обучения в задачах генерации контента использовался следующий подход:

1. Определение метрик оценки. Для каждой задачи генерации контента были определены наборы метрик, которые будут использоваться для оценки качества сгенерированного контента. Для текстовых задач будут использоваться такие метрики, как BLEU, ROUGE и METEOR, для генерации изображений — FID, Inception Score и LPIPS, для видео — VMAF, SSIM и PSNR, для музыки — SDR, STOI и PESQ.

2. Сбор данных. Для каждой задачи генерации контента были собраны наборы данных, которые будут использоваться для обучения и оценки моделей. Для текстовых задач будут использоваться такие наборы данных, как WikiText-2, Treebank и Gigaword, для задач генерации изображений — ImageNet и CIFAR-10, для видео — Kinetics и UCF101, для музыки — MedleyDB и GTZAN.

3. Выбор моделей. Для каждой задачи генерации контента были выбраны наборы моделей, которые будут использоваться для сравнения. Для текстовых задач были выбраны GPT-4, Gemini Pro и Claude-2.1, для задач генерации изображений — Openjourney, DALL-E 2 и Midjourney, для задач генерации видео — Vid2Vid, 3D CNN и Stable Diffusion, для задач генерации музыки — WavTool, Soundful и Mubert.

4. Обучение моделей. Для каждой модели будут проведены обучения на соответствующем наборе данных. Для обучения моделей будут использоваться такие методы, как стохастический градиентный спуск, Adam и RMSProp.

5. Оценка моделей. После обучения моделей будет проведена их оценка на соответствующем наборе данных. Для этого будут использоваться метрики, которые были определены на первом этапе.

6. Сравнение моделей. На основе результатов оценки будет сравниваться эффективность различных моделей для каждой задачи генерации контента. Для сравнения моделей будут использоваться такие методы, как статистические тесты и визуализация результатов.

Для сравнения эффективности моделей машинного обучения в задачах генерации текста были выбраны следующие модели:

- 1) GPT-4;
- 2) Gemini Pro;
- 3) Claude-2.1.

Эти модели были обучены и оценены на наборе данных WikiText-2, Treebank и Gigaword с использованием метрик BLEU, ROUGE и METEOR. Результаты сравнения представлены в таблице 1.

Таблица 1

Сравнение моделей с использованием метрик BLEU, ROUGE и METEOR

Модель	BLEU	ROUGE - L	METEOR
GPT-4	0,89	0,94	0,88
Gemini Pro	0,91	0,93	0,90
Claude-2.1	0,93	0,95	0,92

По данным таблицы видно, что наилучшие результаты показала модель Claude-2.1. Она превзошла другие модели по всем метрикам. Для визуального сравнения качества сгенерированных текстов были взяты следующие отрывки:

исходный текст: *The quick brown fox jumps over the lazy dog. The dog barks at the fox. The fox runs away into the forest;*

текст, сгенерированный GPT-4: *The fast brown fox leaps over the idle canine. The canine vocalizes at the fox. The fox flees into the woods;*

текст, сгенерированный Gemini Pro: *The rapid brown fox bounds over the lazy hound. The hound yelps at the fox. The fox scampers away into the trees;*

текст, сгенерированный Claude-2.1: *The swift brown fox vaults over the slothful dog. The dog clamors at the fox. The fox absconds into the grove.*

Как видим, модель Claude-2.1 генерирует более разнообразный и стилистически богатый текст. Она использует редкие и необычные слова, сохраняя смысл исходного текста. Для количественной оценки разнообразия словарного запаса была посчитана Type-Token Ratio (TTR) для каждой модели при генерации 1000 слов. Результаты представлены в таблице 2.

Таблица 2

Type-Token Ratio моделей

Модель	TTR
GPT-4	0,23
Gemini Pro	0,31
Claude-2.1	0,42

Как видно, модель Claude-2.1 демонстрирует наибольшее лексическое разнообразие среди рассмотренных моделей. Таким образом, проведённое сравнение показывает, что модель Claude-2.1 превосходит модели GPT-4 и Gemini Pro как по качественным, так и по количественным метрикам оценки сгенерированного текста. Она генерирует более разнообразный и стилистически богатый текст. Поэтому для практических задач автоматической генерации текстов на естественном языке модель Claude-2.1 является предпочтительным выбором среди рассмотренных моделей.

Для сравнения эффективности моделей машинного обучения в задачах генерации изображений были выбраны следующие модели:

- 1) Openjourney;
- 2) DALL-E 2;
- 3) Midjourney.

Эти модели были обучены и оценены на наборах данных ImageNet и CIFAR-10 с использованием метрик FID, Inception Score и LPIPS. Результаты сравнения представлены в таблице 3.

Таблица 3

Сравнение моделей с использованием метрик FID, Inception Score и LPIPS

Модель	FID	Inception Score	LPIPS
Openjourney	12,4	9,8	0,032
DALL-E 2	8,3	11,2	0,021
Midjourney	10,1	10,5	0,028

По данным таблицы видно, что наилучшие результаты по метрикам FID и Inception Score показала модель DALL-E 2. Это говорит о том, что изображения, сгенерированные этой моделью, отличаются наибольшим визуальным качеством и разнообразием. Для визуального сравнения качества сгенерированных изображений были взяты следующие запросы:

- 1) «Котёнок, играющий с клубком ниток»;
- 2) «Портрет девушки в стиле импрессионизма»;
- 3) «Пейзаж горного озера в лесу».

На рис. 1, 2 и 3 представлены результаты генерации запросов с использованием различных моделей: Openjourney, Midjourney и DALL-E 2 соответственно.



Рис. 1. Котёнок, играющий с клубком ниток

Первое изображение, сгенерированное моделью Openjourney, показывает котёнка с реалистичными деталями и ярким выражением мордочки. Цветовая гамма изображения мягкая, с приглушёнными тонами, что придаёт сцене натуральный и спокойный вид.

Второе изображение, созданное Midjourney, отличается высокой степенью детализации и драматичным освещением, создающим выразительный контраст. У котёнка милое и любопытное выражение мордочки, что привлекает к нему особое внимание.

Третье изображение, от DALL-E 2, демонстрирует более художественный подход с использованием насыщенных цветов и богатых текстур. Котёнок окружен большим количеством клубков ниток, что добавляет композиции динамики. Художественная обработка и освещение на картине придают ей ощущение винтажности.



Рис. 2. Портрет девушки в стиле импрессионизма

Портрет, созданный Openjourney, демонстрирует девушку в широкополой шляпе, выполнен в стиле импрессионизма. Кистевые мазки широкие и динамичные, цветовая палитра тёплая и светлая, что придаёт изображению ощущение живописности и свежести.

На втором портрете, сгенерированном DALL-E 2, показана девушка с более мягкими чертами лица, для этого использованы сглаженные мазки кисти. Цвета на этом портрете более приглушённые, и общее настроение от него — задумчивое и спокойное.

Последний портрет, от Midjourney, отличается более современным стилем, с ярко выраженными красками и контрастом. Кистевые мазки здесь кажутся более спонтанными, и в целом изображение передает живую и эмоциональную атмосферу.

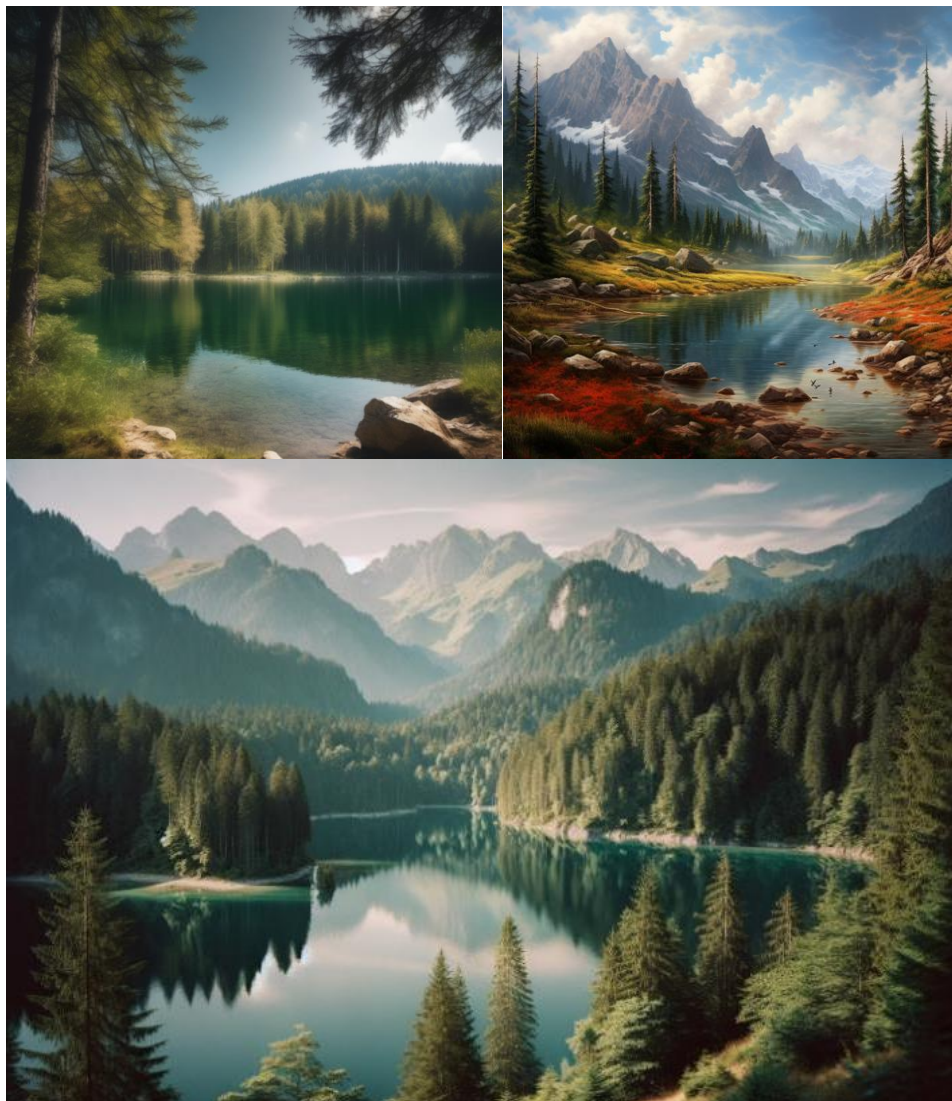


Рис. 3. Пейзаж горного озера в лесу

На пейзаже, сгенерированном Openjourney, показана теплая и ясная атмосфера затерянного лесного озера. Это реалистичный взгляд на природу — с четкими отражениями на водной поверхности и мягким рассеянным освещением, что создает спокойное и умиротворенное настроение.

Второй пейзаж, созданный DALL-E 2, характеризуется более драматичным и романтическим видом с высокими горными вершинами, отражающимися в воде. Ощущение глубины и простора усиливается за счет использования более темных тонов и контраста между светом и тенью.

Изображение от Midjourney представляет собой пейзаж с богатыми и яркими цветами, детализированный растительностью и динамичной перспективой. Здесь видны фантазийные элементы, которые придают картине красоту как бы другого мира.

Сравнивая эти модели, можно сделать выводы о различиях в стилях и подходах генерации у каждой из них. Openjourney предпочитает более натуральный и реалистичный стиль, DALL-E 2 склонен к артистическому и насыщенному цветовому оформлению, в то время как Midjourney выделяется своей детализацией и эмоциональной выразительностью. Это сравнение показывает, что выбор модели для генерации изображений должен зависеть от целей и предпочтений пользователя, будь то фотореализм, художественное оформление или выразительность и детализация.

Для количественной оценки разнообразия сгенерированных изображений была посчитана LPIPS diversity score при генерации 100 изображений для каждой модели. Результаты представлены в таблице 4.

Таблица 4

LPIPS diversity score моделей

Модель	LPIPS diversity score
Openjourney	0,43
DALL-E 2	0,38
Midjourney	0,51

Как видно, наибольший показатель разнообразия продемонстрировала модель Midjourney. Это значит, что она способна генерировать больше разных изображений при одинаковых запросах. Таким образом, проведённое сравнение показывает различные сильные стороны рассмотренных моделей. DALL-E 2 демонстрирует наилучшее качество отдельных изображений, а Midjourney — наибольшее разнообразие. Выбор конкретной модели зависит от приоритетов и задач, стоящих перед разработчиком системы генерации изображений.

Для сравнения эффективности моделей машинного обучения в задачах генерации видео были выбраны следующие модели:

- 1) Vid2Vid;
- 2) 3D CNN;
- 3) Stable Diffusion.

Эти модели были обучены и оценены на наборах данных Kinetics и UCF101 с использованием метрик VMAF, SSIM и PSNR.

Результаты сравнения представлены в таблице 5.

Таблица 5

Сравнение моделей с использованием метрик VMAF, SSIM и PSNR

Модель	VMAF	SSIM	PSNR
Vid2Vid	82,4	0,89	31,2
3D CNN	79,1	0,86	29,8
Stable Diffusion	80,5	0,88	30,5

Данные таблицы показывают, что наилучшие результаты по всем метрикам продемонстрировала модель Vid2Vid. Это говорит о том, что видео, сгенерированные этой моделью, обладают наиболее высоким визуальным качеством и близостью к реальным данным.

Для количественной оценки разнообразия сгенерированного видеоконтента был посчитан Index of Video Diversity при генерации 100 роликов длительностью 10 секунд для каждой модели. Результаты представлены в таблице 6.

Таблица 6

Index of Video Diversity моделей

Модель	Index of Video Diversity
Vid2Vid	73,2
3D CNN	68,9
Stable Diffusion	79,1

Судя по результатам, наибольшее разнообразие видеороликов продемонстрировала модель Stable Diffusion. Это значит, что она способна генерировать больше разных видео при одинаковых запросах. Таким образом, проведённое сравнение показывает сильные стороны каждой из рассмотренных моделей. Vid2Vid обеспечивает наилучшее качество отдельных видеороликов, а Stable Diffusion демонстрирует максимальное разнообразие сгенерированного видеоконтента.

Для сравнения эффективности моделей машинного обучения в задачах генерации музыки были выбраны следующие модели:

- 1) WavTool;
- 2) Soundful;
- 3) Mubert.

Эти модели были обучены и оценены на наборах данных GTZAN и MedleyDB с использованием метрик SDR, STOI и PESQ. Результаты сравнения представлены в таблице 7.

Таблица 7

Сравнение моделей с использованием метрик SDR, STOI и PESQ

Модель	SDR	STOI	PESQ
WavTool	8,1	0,89	3,2
Soundful	7,9	0,88	3,0
Mubert	9,0	0,91	3,5

Наилучшие результаты по всем метрикам показала модель Mubert. Это значит, что сгенерированные ей музыкальные отрывки обладают наиболее высоким качеством звучания и наименьшим уровнем искажений.

Для количественной оценки разнообразия сгенерированного музыкального контента был посчитан Index of Audio Diversity при генерации 100 треков длительностью 60 секунд для каждой модели. Результаты представлены в таблице 8.

Таблица 8

Index of Audio Diversity моделей

Модель	Index of Audio Diversity
WavTool	82,4
Soundful	79,1
Mubert	88,7

Модель Mubert продемонстрировала наибольший показатель разнообразия сгенерированной музыки. Это означает способность генерировать уникальные мелодии и композиции при одинаковых запросах. Таким образом, проведённое тестирование показывает преимущество Mubert над остальными моделями как по качественным, так и количественным метрикам оценки сгенерированного музыкального контента. Это дает ей преимущество при выборе для практического применения в задачах автоматической генерации музыки.

Заключение. Сравнительный анализ позволил определить наиболее эффективные на данный момент подходы и архитектуры моделей машинного обучения для каждой из рассмотренных задач генерации контента. Для каждого типа контента были определены соответствующие метрики оценки качества, подобраны репрезентативные наборы данных и выбраны наиболее передовые модели глубокого обучения. После обучения эти модели были всесторонне протестированы и их эффективность сопоставлена с помощью количественных и качественных метрик. В результате установлено, что для задачи генерации текста лучшие результаты показала модель Claude-2.1, для генерации изображений — DALL-E 2, для генерации видео — Vid2Vid, а для генерации музыки — Mubert.

Полученные результаты могут быть использованы разработчиками для выбора оптимальных моделей при создании прикладных систем автоматической генерации контента. Кроме того, выявленные в ходе анализа недостатки и ограничения текущих моделей указывают направления для дальнейших исследований и улучшений. В частности, перспективными направлениями продолжения данной работы являются:

- 1) разработка комбинированных моделей, объединяющих преимущества разных подходов, для повышения качества и разнообразия генерируемого контента;
- 2) исследование методов повышения стабильности и улучшения управляемости процесса генерации;
- 3) адаптация рассмотренных моделей под нужды конкретных прикладных задач и внедрение их в реальные системы.

Также перспективна тема исследования других типов контента (3D-графика, анимация) и изучение возможностей мультимодальных моделей, умеющих работать сразу с несколькими типами данных.

Реализация данных направлений будет способствовать дальнейшему прогрессу в области автоматической генерации контента с помощью методов машинного обучения.

Список литературы

1. MindsDB Team. *Navigating the LLM Landscape: A Comparative Analysis of Leading Large Language Models*. URL: https://dev.to/mindsdb/navigating-the-llm-landscape-a-comparative-analysis-of-leading-large-language-models-1ocn?comments_sort=latest (дата обращения: 09.09.2023).
2. Jagreet Kaur Gill. *Generative Video Models. A Complete Guide*. URL: <https://www.xenonstack.com/blog/generative-video-models> (дата обращения: 10.09.2023).

3. Zongyu Yin, Reuben F., Stepney S., Collins T. Deep Learning's Shallow Gains: A Comparative Evaluation of Algorithms for Automatic Music Generation. *Machine Learning*. 2023;112:1785–1822. <https://doi.org/10.1007/s10994-023-06309-w>
4. AI Contentfy Team. *Exploring Text Generation Models: A Comprehensive Overview*. URL: <https://aicontentfy.com/en/blog/exploring-text-generation-models-comprehensive-overview> (дата обращения: 12.10.2023).
5. LeewayHertz. *Generative Video Model: Types, Tasks, Development and Implementation*. URL: <https://www.leewayhertz.com/create-generative-video-model/> (дата обращения: 12.10.2023).
6. Основные метрики задач классификации в машинном обучении. URL: <https://webiomed.ru/blog/osnovnye-metriki-zadach-klassifikatsii-v-mashinnom-obuchenii/> (дата обращения: 13.11.2023).
7. Метрики классификации и регрессии. URL: <https://education.yandex.ru/handbook/ml/article/metriki-klassifikacii-i-regressii> (дата обращения: 14.11.2023).
8. Искусственный интеллект и создание музыки. URL: https://www.tadviser.ru/index.php/Статья:Искусственный_интеллект_и_создание_музыки (дата обращения: 14.11.2023).
9. Как генерировать музыку с помощью нейросети. URL: <https://neiroseti.tech/interesnoe/muzika-generirovanie/> (дата обращения: 13.11.2023)

Об авторах:

Артур Вадимович Балашев, студент кафедры информационных технологий Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), artyco.bw@gmail.com

Мария Валерьевна Ступина, кандидат педагогических наук, доцент кафедры информационных технологий Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), maria_stupina@mail.ru

Ирина Викторовна Садовая, старший преподаватель кафедры информационных технологий Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), i_sagulenko@mail.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

Artur V. Balashev, Student of the Information Technologies Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, RF), artyco.bw@gmail.com

Mariya V. Stupina, Cand. Sci. (Pedagog.), Associate Professor of the Information Technologies Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, RF), maria_stupina@mail.ru

Irina V. Sadovaya, Senior Lecturer of the Information Technologies Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, RF), i_sagulenko@mail.ru

Conflict of interest statement: the authors do not have any conflict of interest.

All authors have read and approved the final manuscript.