Известия Воронежского государственного педагогического университета. 2026. № 3 (312). С. 227–238.
Izvestia Voronezh State Pedagogical University. 2026. (3), 227–238.

Научная статья
УДК 81'32
DOI 10.47438/2309-7078_2026_3_227.

ЛИНГВИСТИЧЕСКИЙ ПРОФИЛЬ ИИ-СГЕНЕРИРОВАННЫХ ТЕКСТОВ ОТНОСИТЕЛЬНО ЕСТЕСТВЕННЫХ ТЕКСТОВ С РАЗНЫМИ ОЦЕНКАМИ КАЧЕСТВА

Татьяна Александровна Литвинова[1], Светлана Геннадьевна Любова[2], Елена Сергеевна Котлярова[3]

Воронежский государственный педагогический университет[1, 2, 3]
Воронеж, Россия

[1]*Доктор филологических наук, профессор кафедры высшей математики ORCID ID: 0000-0002-6019-3700, тел.: +7-473-264-44-17, e-mail: centr_rus_yaz@mail.ru*
[2]*Кандидат филологических наук, доцент кафедры теории, истории и методики преподавания русского языка и литературы, научный сотрудник лаборатории культурной безопасности языковых моделей, ORCID ID: 0000-0002-2586-410X, тел.: +7-473-255-14-78, e-mail: svetlana_lyubova_2305@mail.ru*
[3]*Аспирант кафедры русского языка, современной русской и зарубежной литературы, стажер-исследователь лаборатории культурной безопасности языковых моделей, ORCID ID: 0000-0002-4162-5002, тел.: +7-473-255-14-78, e-mail: kray-voronezh@mail.ru*

Аннотация. Широкое распространение генеративного искусственного интеллекта обусловливает необходимость разработки методов сопоставления ИИ-сгенерированных и естественных текстов с независимо полученными оценками качества. Большинство современных исследований сосредоточено на автоматическом распознавании ИИ-текстов либо на их непосредственной оценке человеком. При этом остаётся недостаточно изученным положение ИИ-текстов в лингвистическом пространстве естественных текстов известного качества. В статье представлены результаты сопоставления русскоязычных ИИ-сгенерированных текстов трёх типов - описаний, повествований и рассуждений - с соответствующими естественными текстами. Корпус естественных текстов включал 199 текстов 66 авторов, каждый из которых имел две независимые экспертные оценки качества; корпус ИИ-текстов состоял из 90 ответов, полученных по тем же заданиям. Для всех текстов был рассчитан единый набор из 14 формальных лингвистических показателей. Анализ методом главных компонент проводился только на естественных текстах, после чего ИИ-тексты проецировались в зафиксированное пространство естественных текстов. Для последующего анализа использовались восемь компонент, объяснявших 82,72 % дисперсии признаков естественных текстов. Установлено, что более высокое качество естественных текстов связано прежде всего с лексическим разнообразием, большей средней длиной и вариативностью длины слов, объёмом текста и меньшей частотой форм первого лица; часть связей зависела от типа задания. ИИ-тексты воспроизводили преимущественно лексическую составляющую данного профиля, но отличались от высоко оценённых естественных текстов по ряду структурных и пунктуационных характеристик. В целом 56,7 % ИИ-текстов располагались ближе всего к центроидам групп высокого качества; наиболее однозначный результат был получен для описаний, среди которых эта доля составила 80 %. В пространстве восьми нормированных компонент суммарная вариативность ИИ-текстов составляла только 17–22 % от вариативности естественных текстов соответствующего типа. Этот результат сохранялся после уравнивания размеров групп и исключения показателя объёма текста. Полученные данные позволяют рассматривать ИИ-тексты при использованной процедуре генерации как относительно стандартизированный тип текста, воспроизводящий отдельные признаки высокого качества, но не полный профиль высоко оценённых естественных текстов.

Ключевые слова: генеративный искусственный интеллект, большие языковые модели, качество текста, лингвистические признаки, анализ главных компонент, естественный текст, ИИ-сгенерированный текст, тип текста, лексическое разнообразие.

© Литвинова Т.А., Любова С.Г., Котлярова Е.С., 2026

Для цитирования: *Литвинова Т.А., Любова С.Г., Котлярова Е.С.* Лингвистический профиль ИИ-сгенерированных текстов относительно естественных текстов с разными оценками качества // Известия Воронежского государственного педагогического университета. 2026. № 3. С. 227–238. DOI 10.47438/2309-7078_2026_3_227.

Введение

Стремительное развитие больших языковых моделей привело к появлению текстов, которые не только соответствуют основным грамматическим и стилистическим нормам, но и могут получать высокие экспертные оценки по показателям связности, аргументированности и общего качества. В связи с этим простое противопоставление «человеческого» и «машинного» текста становится недостаточным. Важным становится не только активно исследуемый (см., например, [9]) вопрос о том, можно ли определить происхождение текста (ИИ/человек) по формализуемым лингвистическим признакам, но и вопрос о том, как ИИ-сгенерированный текст (далее также – ИИ-тексты) соотносится с внутренней вариативностью человеческого письма: соответствует ли он текстам высокого качества, располагается около некоторого нормативного среднего или образует самостоятельный лингвистический профиль?

Как показывают исследования, современные языковые модели способны создавать тексты, получающие оценки, сопоставимые с оценками естественных текстов и даже иногда превосходящие их. Так, в масштабном исследовании аргументативных эссе тексты, созданные ChatGPT, были оценены экспертами выше, чем студенческие работы, в том числе по параметрам структуры, логики и языкового оформления. Вместе с тем сгенерированные эссе отличались от человеческих по ряду языковых характеристик: большим лексическим разнообразием и частотностью номинализаций при меньшем количестве дискурсивных и эпистемических маркеров [5]. Сходная неоднозначность обнаруживается и при изучении художественных текстов: ИИ-сгенерированные рассказы могут получать более высокие оценки качества и читательской вовлеченности, однако сами оценки зависят от того, кому участники приписывают авторство текста [13]. Следовательно, высокое воспринимаемое качество текста, сходство с человеческим письмом и фактическое происхождение текста представляют собой связанные, но не тождественные параметры.

Следует отметить, что сопоставление ИИ-сгенерированных и естественных текстов (далее также – ЕТ) по какому-то одному лингвистическому показателю часто приводит к противоположным выводам. Особенно отчетливо указанное наблюдение проявляется при анализе лексических характеристик текста. Результаты подобного сопоставления зависят как от версии модели, так и от характера коммуникативного задания. Так, на сопоставимом материале ChatGPT-3.5 демонстрировал более ограниченный словарь и более низкое лексическое разнообразие по сравнению с человеческими текстами, тогда как тексты, сгенерированные GPT-4, приближались к соответствующим значениям естественных текстов или превышали их по отдельным показателям [12]. В более новом исследовании текстов, сгенерированных четырьмя поколениями ChatGPT, ИИ-тексты отличались от естественных по многомерному профилю лексического разнообразия; при этом новые модели не обязательно порождали более "человекоподобные" тексты, даже если демонстрировали высокие значения отдельных показателей данной группы [8]. Подобный факт означает, что лексическое разнообразие следует рассматривать не как самостоятельный эквивалент качества текстов, а как один из компонентов общего лингвистического профиля.

Различия между человеческими и машинными текстами исследуются также на уровне синтаксической организации, выраженности авторской позиции и характере взаимодействия с читателем. При сравнении естественных новостных текстов с результатами работы шести языковых моделей выяснилось, что ЕТ имеют более вариативную длину предложений и иные закономерности синтаксической организации; при этом различия между ЕТ и ИИ-текстами были выражены сильнее, чем различия между текстами, сгенерированными отдельными моделями [11]. В аргументативных эссе студентов было выявлено большее количество и разнообразие средств взаимодействия с адресатом, включая вопросы и авторские отступления, тогда как тексты, сгенерированные ChatGPT, характеризовались менее интерактивным способом построения аргументации [7]. Подобные результаты подтверждают необходимость учитывать не один лингвистический показатель, а совокупность структурных, лексических, пунктуационных и дискурсивных характеристик текста.

Корпусные исследования также показывают, что различия между ЕТ и ИИ-сгенерированными текстами имеют многомерный характер. В исследовании [1] ИИ-тексты сопоставлялись с естественной письменной и устной речью по признакам, предложенным в работе [2]. Было установлено, что ИИ-тексты не воспроизводят полностью профиль соответствующих ЕТ и занимают отличающиеся от них позиции по совокупности лингвистических признаков [1]. В работе [10] было показано, что сгенерированные тексты могут характеризоваться повышенной информационной плотностью, но одновременно производить впечатление содержательно недостаточных. На русскоязычном материале парных описаний произведений искусства также были выявлены устойчивые различия между ЕТ и ИИ-сгенерированными текстами на лексическом, морфосинтаксическом и семантическом уровнях: ИИ-тексты отличались более высокой лексической и синтаксической сложностью, большей абстрактностью и меньшей вариативностью, тогда как человеческие описания были в большей степени связаны с конкретным зрительным и телесным опытом [9].

Отдельное направление исследований связано с возможным снижением вариативности текстов, созданных с использованием генеративного ИИ. Экспериментальные данные показывают [4], что помощь языковой модели при создании текстов может повышать воспринимаемые качество текста, новизну и интерес к отдельному тексту, особенно у авторов с изначально более низкими показателями креативности. Однако

созданные с такой помощью тексты становятся более похожими друг на друга [4]. Это позволяет предположить, что ИИ способен воспроизводить относительно успешный нормативный профиль, одновременно сокращая диапазон индивидуальных способов его реализации. В таком случае близость ИИ-текстов к высокому среднему уровню качества ЕТ может сочетаться с их меньшей вариативностью или с расхождением с ЕТ по отдельным показателям.

При этом субъективная узнаваемость ИИ-текста не может служить надежной заменой его лингвистическому анализу. Так, в экспериментах, описанных в работе [6], участники определяли происхождение текстов практически на уровне случайного выбора и опирались на ошибочные эвристики, связывая с человеческим авторством формы первого лица, разговорные сокращения и упоминания семьи. Следовательно, воспринимаемая «ИИ-подобность» текста может не совпадать с его объективными языковыми характеристиками. Для последующего изучения восприятия ИИ-текстов необходим предварительный объективный слой, описывающий положение таких текстов относительно вариативности и качества ЕТ.

Несмотря на быстро растущее число исследований в области сравнительного анализа ИИ-текстов и ЕТ, в большинстве из них названные тексты рассматриваются как две однородные группы. Основной задачей работ становится либо выявление групповых различий, либо классификация текстов по источнику. При таком подходе внутренняя неоднородность человеческого письма, в частности различия между текстами разного качества, обычно не включается в аналитическую модель. Кроме того, значительная часть существующих исследований выполнена на материале английского языка и посвящена одному типу текста. Поэтому остается открытым вопрос о том, как ИИ-сгенерированные тексты соотносятся не просто с усредненной выборкой ЕТ, а с ЕТ с разным уровнем экспертно оцененного языкового качества.

Цель настоящего исследования состоит в определении положения ИИ-сгенерированных русскоязычных текстов в многомерном лингвистическом пространстве ЕТ с известными экспертными оценками качества. В отличие от существующих работ, в данном исследовании пространство признаков строится исключительно на корпусе ЕТ, после чего ИИ-тексты, созданные по тем же коммуникативным заданиям, проецируются в уже зафиксированную систему координат. Такой дизайн позволяет, во-первых, установить, какие измерения человеческой языковой вариативности связаны с качеством текста; во-вторых, определить близость ИИ-текстов к низкому, среднему и высокому уровням качества ЕТ; в-третьих, выявить признаки, по которым ИИ совпадает с качественными человеческими текстами или отличается от них; в-четвертых, проверить зависимость получаемого профиля от типа коммуникативного задания. Тем самым исследование создает объективную основу для дальнейшего анализа того, как положение текста в лингвистическом пространстве связано с воспринимаемой ИИ-подобностью текста, доверием к тексту и оценкой его коммуникативного качества.

Материал и методы исследования

Материал исследования

Естественные тексты были отобраны из экспериментального корпуса русскоязычной письменной речи, собранного ранее для исследования процесса порождения текста [14]. Исходно тексты создавались с использованием академического кейлоггера GenoGraphiX-Log, регистрировавшего процесс письма и сохранявшего итоговую версию текста. В настоящем исследовании анализировались только финальные тексты; сведения о процессе их порождения не использовались.

Итоговая выборка составила 199 текстов 66 авторов из двух экспериментальных серий. В первую серию вошли 90 текстов 30 авторов, во вторую — 109 текстов 36 авторов.

Участникам предлагалось выполнить три письменных задания:

1. Описание: «Опишите подробно вид из окна. Составьте текст-описание».
2. Повествование: «Поделитесь впечатлениями о произошедшем событии (книга, фильм, концерт или под.), которое Вас поразило. Составьте текст-повествование».
3. Рассуждение: «Порассуждайте о проблеме, которая Вас волнует (текст должен иметь проблемный характер). Пример: “Что такое счастье?”. Составьте текст-рассуждение».

Распределение текстов по типам заданий представлено в табл. 1.

Таблица 1. Состав исследовательского корпуса

Корпус	**Описание**	**Повествование**	**Рассуждение**	**Всего**
ЕТ	64	71	64	199
ИИ-тексты	30	30	30	90
Всего	**94**	**101**	**94**	**289**

Для формирования сопоставительной выборки ИИ-текстов использовались те же три задания. Генерация выполнялась с помощью модели GPT-5.6-sol при установленном уровне рассуждения medium. Для каждого задания было выполнено 30 независимых запросов, в результате чего было получено 90 текстов. В системной инструкции указывалось, что ответ должен содержать только текст на русском языке, без заголовка, комментариев и пояснений. Человеческие тексты или иные примеры модели не предоставлялись; история предыдущих генераций между запросами не сохранялась.

Максимальный объём ответа был технически ограничен значением max_output_tokens = 2500. Это значение задавало только верхнюю границу ответа и не использовалось как инструкция о требуемой длине

текста. Температура, top_p и другие параметры выборки явно не устанавливались; использовались настройки API по умолчанию. Фиксированное значение случайного начального состояния (seed) также не задавалось.

Из каждого ответа автоматически извлекался текстовый блок типа output_text. В корпус включался первый успешно полученный непустой ответ. Повторная генерация на основании содержания, качества или стилистических особенностей текста не проводилась. Ответы не подвергались ручному редактированию, сокращению, исправлению ошибок или иной постобработке. Таким образом, сопоставление проводилось между естественными и ИИ-сгенерированными текстами, созданными в ответ на одинаковые коммуникативные задания. Общие характеристики корпуса представлены в табл. 2.

Таблица 2. Характеристики исследовательского корпуса

Тип текста	**ЕТ, длина в словах (среднее ± SD)**	**ИИ-тексты, длина в словах (среднее ± SD)**
Описание	237,5 ± 55,4	197,3 ± 21,2
Повествование	285,7 ± 132,4	189,8 ± 20,6
Рассуждение	273,2 ± 128,0	211,9 ± 25,7

ИИ-сгенерированные тексты были в среднем короче ЕТ: различие на логарифмической шкале составляло от 0,60 до 0,83 стандартного отклонения ЕТ. ИИ-тексты также характеризовались значительно меньшей вариативностью объёма.

Процедура экспертного оценивания

Каждый ЕТ независимо оценивался двумя экспертами по десятибалльной шкале, где более высокое значение соответствовало более высокому общему качеству текста. Один эксперт имел практический опыт регулярной проверки учебных сочинений по критериям, близким к критериям государственной итоговой аттестации; второй обладал профессиональным редакторским опытом. Оценивание проводилось раздельно, без обсуждения конкретных текстов и последующего согласования выставленных баллов.

Общая оценка имела интегральный характер и отражала качество текста в целом, включая его языковое оформление, логичность и структурную организацию. В составе исходной разметки также фиксировались орфографические, пунктуационные, грамматические и речевые ошибки, описки, а также нарушения логики и структуры. Эти показатели в настоящей работе не использовались. Основным показателем качества являлось среднее арифметическое двух общих экспертных оценок.

Согласованность оценок экспертов оценивалась с помощью линейной корреляции Пирсона, ранговой корреляции Спирмена и коэффициента внутриклассовой корреляции абсолютного согласия. Корреляция оценок составила $r=0,746$ по Пирсону и $\rho=0,705$ по Спирмену ($p<0,001$ в обоих случаях). Коэффициент внутриклассовой корреляции для одной оценки был равен $ICC(A,1)=0,701$, а для среднего двух оценок - $ICC(A,2)=0,824$. Полное совпадение оценок наблюдалось в 39,7 % случаев, а расхождение не более чем на один балл — в 77,4 % случаев. Полученные значения подтверждают возможность использования среднего двух оценок как основного показателя качества.

Вместе с тем оценки экспертов различались по уровню: первый эксперт в среднем выставлял на 0,603 балла больше второго. Это систематическое различие учитывалось при интерпретации результатов и рассматривается как ограничение экспертной процедуры. Усреднение оценок позволило частично компенсировать индивидуальные особенности использования шкалы. В итоговой человеческой выборке средняя оценка качества составила 7,17 балла при стандартном отклонении 1,69; медиана была равна 7, диапазон значений – от 1 до 10 баллов.

В основном статистическом анализе качество рассматривалось как непрерывная переменная. Деление на уровни применялось только для визуализации пространства и вычисления расстояний до групповых центроидов. Границы групп соответствовали квартилям распределения: к низкому уровню относились тексты со средней оценкой не выше 6,25, к высокому - тексты с оценкой не ниже 8,50; остальные тексты составляли среднюю группу. ИИ-тексты экспертами непосредственно не оценивались: их положение определялось только посредством проекции в лингвистическое пространство ЕТ с известными оценками качества.

Лингвистические признаки

Набор лингвистических признаков был зафиксирован до анализа положения ИИ-текстов в пространстве ЕТ. Для естественных и ИИ-сгенерированных текстов применялись одни и те же функции предварительной обработки и расчета показателей. Набор включал 14 формальных признаков, характеризующих объём и структуру текста, длину предложений, лексическую вариативность, пунктуацию, выраженность авторской позиции и использование эксплицитных средств связи. При их извлечении не применялись внешние синтаксические анализаторы, морфологические модели, словари частотности или семантические представления текста. Такой выбор обеспечивал прозрачность процедуры и исключал зависимость результатов от качества автоматической морфологической или синтаксической разметки.

Предварительная обработка включала перевод буквенных последовательностей в нижний регистр и унификацию букв **ё** и **е**. Словесными токенами (словами) считались последовательности букв Unicode, которые могли включать внутренний дефис или апостроф. Таким образом, написания с дефисом внутри слова рассматривались как одно слово. При вычислении длины слова внутренний дефис исключался из подсчёта символов.

Предложения выделялись посредством разделения текста по точкам, вопросительным и восклицательным знакам, а также многоточию; учитывались только полученные фрагменты, содержавшие буквенные символы. Абзацы определялись по границам строк, при этом пустые фрагменты и фрагменты без буквенных символов исключались.

Состав и способы расчета признаков представлены в табл. 3.

Таблица 3. Лингвистические признаки текстов

Блок	**Показатель**	**Способ расчёта**
Объём	Логарифм числа слов	Натуральный логарифм ln(1+N), где N — число слов
Структура	Абзацы на 100 слов	Число абзацев, делённое на число слов и умноженное на 100
Длина предложений	Средняя длина предложения	Среднее число слов в предложении
Длина предложений	Коэффициент вариации длины предложений	Отношение стандартного отклонения длины предложений к их средней длине
Лексика	Средняя длина слова	Среднее число буквенных символов в слове
Лексика	Вариативность длины слов	Стандартное отклонение длины слов
Лексика	MATTR-50	Среднее значение type-token ratio в последовательно перемещаемых окнах по 50 слов [3].
Лексика	Hapax legomena	Число словоформ, встретившихся в тексте один раз, на 100 слов
Пунктуация	Запятые	Число запятых на 100 слов
Пунктуация	Тире	Число тире и изолированных дефисов на 100 слов
Пунктуация	Двоеточия и точки с запятой	Суммарное число двоеточий и точек с запятой на 100 слов
Пунктуация	Экспрессивные знаки	Суммарное число вопросительных и восклицательных знаков на 100 слов
Авторская позиция	Формы первого лица	Число выбранных личных и притяжательных местоимений первого лица на 100 слов
Средства связи	Эксплицитные коннекторы	Число словоформ из заранее заданного списка средств связи на 100 слов

Объём текста преобразовывался с использованием логарифма, что уменьшало влияние наиболее длинных текстов. Показатель числа абзацев нормировался на 100 слов, поскольку абсолютное число абзацев непосредственно зависит от длины текста.

Средняя длина предложения рассчитывалась в словах. Вариативность длины предложений оценивалась посредством коэффициента вариации. Использование относительного, а не абсолютного показателя позволяло сопоставлять тексты с различной средней длиной предложения.

Для оценки лексического разнообразия применялся показатель MATTR с окном 50 слов [3]. Для каждого последовательного окна вычислялось отношение числа различных словоформ к общему числу слов, после чего значения усреднялись по всем окнам текста. В отличие от обычного type-token ratio, MATTR в меньшей степени зависит от общего объёма текста [3]. Если текст содержал не более 50 слов, рассчитывался стандартный type-token ratio для всего текста. Частота hapax legomena определялась как число нормализованных словоформ, встретившихся ровно один раз, в расчёте на 100 слов.

Пунктуационные показатели рассчитывались непосредственно по символам исходного текста. В категорию тире включались длинное тире, короткое тире и дефис, не находившийся внутри буквенной последовательности. Поэтому дефисы в составе сложных слов не учитывались как самостоятельные пунктуационные знаки. Все пунктуационные показатели нормировались на 100 слов.

Показатель форм первого лица рассчитывался по закрытому списку, включавшему личные местоимения **я**, **мы** и их падежные формы, а также притяжательные местоимения **мой** и **наш** со словоизменительными вариантами. Поскольку морфологический анализ не применялся, показатель не охватывал глагольные формы первого лица без эксплицитно выраженного местоимения и рассматривался как формальный индикатор авторской позиции.

Для оценки использования эксплицитных средств связи применялся заранее определённый список словоформ: **и, а, но, или, либо, однако, зато, поэтому, поскольку, если, когда, хотя, чтобы, также, причём, следовательно, например, наоборот, иначе, ведь, словно, будто, пока**. Показатель отражает частоту выбранных средств связи, но не является полной оценкой когезии или логической связности текста. Его интерпретация ограничивается степенью эксплицитности отношений, выраженных перечисленными единицами.

Статистический анализ

Перед проведением анализа методом главных компонент (Principal Component Analysis, PCA) все 14 признаков центрировались и стандартизировались по выборке ЕТ. Модель PCA строилась только на 199 ЕТ. Для последующего анализа сохранялись первые восемь компонент, суммарно объяснявшие 82,72 % дисперсии. На предварительном этапе связь качества с каждой из всех 14 полученных компонент оценивалась с помощью коэффициента ранговой корреляции Спирмена. Поскольку одновременно проверялись несколько компонент, значения p корректировались методом Бенджамини — Хохберга.

Основная линейная модель включала восемь компонент и тип задания. Поскольку один автор мог быть представлен несколькими текстами, при расчёте статистической значимости учитывалась зависимость наблюдений внутри автора: стандартные ошибки корректировались с кластеризацией по автору. Поправка Бенджамини – Хохберга применялась к результатам проверки восьми компонент. Дополнительно проверялись взаимодействия компонент с типом задания. Возможная зависимость связи компонент с качеством от типа задания дополнительно проверялась в модели, включавшей взаимодействия компонент с типом задания; значения p для восьми таких проверок также корректировались методом Бенджамини — Хохберга.

Для содержательной интерпретации пространства отдельно анализировалась связь каждого из 14 лингвистических признаков с качеством текста. Каждый признак включался в отдельную линейную модель вместе с типом задания; признак предварительно стандартизировался, поэтому коэффициент отражал изменение оценки качества при увеличении его значения на одно стандартное отклонение. В этих моделях также учитывалась кластеризация текстов по автору, а значения p корректировались методом Бенджамини — Хохберга для 14 признаков.

Чтобы установить, различается ли связь признака с качеством между типами заданий, для каждого признака дополнительно строилась модель с его взаимодействием с типом задания. Для каждого признака рассчитывался общий тест жанровых различий, а поправка Бенджамини — Хохберга применялась к 14 полученным значениям p. Кроме того, связи признаков с качеством оценивались отдельно для описаний, повествований и рассуждений. В этих моделях признаки стандартизировались внутри соответствующего типа задания, а поправка на множественные сравнения применялась отдельно к 14 признакам в каждом типе.

ИИ-тексты не использовались при оценке связей лингвистических показателей с качеством. Различия средних значений между ИИ-сгенерированными и естественными текстами по отдельным признакам и компонентам выражались в единицах стандартного отклонения естественных текстов соответствующего типа. Объём текстов дополнительно описывался с помощью среднего числа слов, стандартного отклонения и коэффициента вариации.

Близость ИИ-текстов к низкому, среднему и высокому уровням качества ЕТ определялась по евклидову расстоянию до центроидов соответствующих групп ЕТ внутри того же типа задания. Расстояния рассчитывались по восьми сохранённым компонентам. Чтобы компоненты с большей исходной дисперсией не оказывали непропорционального влияния на результат, координаты по каждой компоненте делились на её стандартное отклонение в выборке ЕТ.

Для сопоставления однородности человеческих и ИИ-сгенерированных текстов рассчитывался общий внутригрупповой разброс в пространстве восьми нормированных компонент. Он определялся как сумма дисперсий по всем восьми компонентам: чем меньше значение этого показателя, тем более компактную и однородную область пространства занимает группа текстов. Математически данный показатель соответствует следу ковариационной матрицы.

Поскольку каждый тип ИИ-текстов включал 30 наблюдений, а группы ЕТ были больше, устойчивость результата к неодинаковому размеру выборок проверялась с помощью повторных подвыборок. Для каждого типа задания из естественных текстов 10 000 раз случайным образом отбиралось по 30 наблюдений, после чего заново рассчитывалось отношение вариативности ИИ-текстов к вариативности ЕТ. В качестве итогового показателя использовалась медиана полученных отношений. В дополнительном анализе PCA и оценка внутригрупповой вариативности повторялись после исключения показателя объёма текста, что позволило проверить, не объясняется ли выявленная компактность ИИ-текстов преимущественно различиями в их длине. В этом анализе пространство также строилось исключительно по ЕТ. Все расчёты выполнялись в R 4.6.0. Статистические проверки были двусторонними, уровень значимости составлял 0,05.

РЕЗУЛЬТАТЫ

Пространство естественных текстов и связь лингвистических признаков с качеством текста

Первые восемь главных компонент объясняли 82,72 % вариативности лингвистических признаков ЕТ.

Корреляционный анализ показал, что качество ЕТ положительно связано с PC1 (ρ = 0,304; pВН < 0,001) и PC6 (ρ = 0,226; pВН = 0,006) и отрицательно - с PC2 (ρ = 0,347; pВН < 0,001). Эти связи сохранились в линейной модели, одновременно учитывавшей восемь компонент и тип задания: для PC1 коэффициент составил 0,325 (pВН = 0,002), для PC2 - 0,388 (pВН < 0,001), для PC6 - 0,356 (pВН = 0,030). Модель объясняла 30,2 % вариативности экспертных оценок качества; скорректированный R^2 составил 0,265. Таким образом, качество текста оказалось связано не с одним показателем, а с сочетанием нескольких относительно независимых лингвистических признаков, прежде всего с лексическим разнообразием, структурно-пунктуационной организацией и объёмом текста.

PC1 характеризует преимущественно лексическое разнообразие и особенности лексической организации текста. На её положительном полюсе располагались тексты с большей средней длиной слова, более высокими значениями MATTR, большей долей слов, встречающихся один раз, и более выраженной вариативностью длины слов. Отрицательные нагрузки имели частота форм первого лица, вариативность и средняя длина предложений, а также объём текста. Положительная связь PC1 с качеством показывает, что более

высокие оценки получали прежде всего тексты с более разнообразной лексикой и более длинными словами.

PC2 противопоставляет дробное абзацное членение и выраженную авторскую позицию более развёрнутой и пунктуационно вариативной организации текста. Положительные нагрузки имели количество абзацев на 100 слов и частота форм первого лица; отрицательные - частота тире, вариативность длины предложений, объём текста, экспрессивная пунктуация и частота запятых. Отрицательная связь PC2 с оценками означает, что более высокому качеству ЕТ соответствует движение к отрицательному полюсу этой оси, то есть к более развёрнутой и пунктуационно разнообразной организации текста.

PC6 имеет смешанную структуру. Её положительный полюс связан прежде всего с объёмом текста и употреблением двоеточий и точек с запятой, а отрицательный - с частотой экспрессивной пунктуации, средств связи и запятых. Положительная связь PC6 с качеством показывает, что более высокие оценки по этой оси соответствуют прежде всего большей развёрнутости текста и более частому использованию двоеточий и точек с запятой.

Положение ИИ-текстов в пространстве естественных текстов

Проекция ИИ-текстов в пространство ЕТ показала (см. табл. 4), что ИИ-тексты располагаются преимущественно в области ЕТ среднего и высокого качества, однако не совпадают с ними по всем признакам. Средняя прогнозируемая позиция ИИ-текстов на шкале качества ЕТ составила 7,52 балла ($SD = 0{,}43$). Из 90 текстов 51 (56,7 %) оказался ближе всего к центроиду группы высокого качества, 23 (25,6 %) - среднего и 16 (17,8 %) - низкого качества.

Таблица 4. Положение ИИ-текстов в пространстве естественных текстов

Тип текста	Оценка качества естественных текстов	Прогноз качества ИИ-текстов (среднее± SD)	Число ИИ-текстов ближе к низкому качеству	Число ИИ-текстов ближе к среднему качеству	Число ИИ-текстов ближе к высокому качеству
Описание	7,30	7,56 ± 0,29	3 (10,0%)	3 (10,0%)	24 (80,0%)
Повествование	7,18	7,20 ± 0,36	11 (36,7%)	6 (20,0%)	13 (43,3%)
Рассуждение	7,01	7,79 ± 0,40	2 (6,7%)	14 (46,7%)	14 (46,7%)
Всего	7,17	7,52 ± 0,43	16 (17,8%)	23 (25,6%)	51 (56,7%)

Во всех трёх типах текстов средние координаты ИИ-текстов по PC1 превышали соответствующие средние значения ЕТ на 1,54-1,80 стандартного отклонения. Это смещение к положительному полюсу PC1, связанному в выборке ЕТ с более высокими оценками качества, отражает прежде всего большую среднюю длину слова и высокое лексическое разнообразие ИИ-текстов. Средние координаты ИИ-текстов по PC2 также были смещены в положительную сторону - на 0,38–0,72 стандартного отклонения. Поскольку более высокие оценки ЕТ были связаны, напротив, с отрицательными значениями PC2, это смещение указывает на расхождение между ИИ- и высоко оценёнными ЕТ по структурно-пунктуационной организации.

В выборке ЕТ более высокие оценки качества были связаны прежде всего с лексическим разнообразием, большей средней длиной и вариативностью длины слов, более частым употреблением тире и бóльшим объёмом текста. Отрицательные связи были обнаружены для частоты форм первого лица и средств связи. ИИ-тексты воспроизводили только часть этого профиля: они характеризовались особенно высоким лексическим разнообразием, большей средней длиной слова и меньшей частотой форм первого лица. Вместе с тем они не отличались устойчиво повышенной вариативностью длины слов, имели меньший объём и иной пунктуационный профиль, а частота средств связи зависела от типа задания. Следовательно, положение ИИ-текстов в области, соответствующей относительно высоким оценкам ЕТ, не означает полного совпадения с лингвистическим профилем высоко оценённых ЕТ.

Результаты проекции ИИ-текстов в пространство естественных текстов с учетом типа текстов

Связь отдельных лингвистических признаков с качеством оказалась частично зависимой от типа текста. При раздельном включении признаков в модели с контролем типа текста более высокие оценки были связаны с лексическим разнообразием MATTR ($b = 0{,}778$; pВН $< 0{,}001$), большей средней ($b = 0{,}571$; pВН $< 0{,}001$) и стандартной ($b = 0{,}483$; pВН $= 0{,}017$) длиной слова, более частым употреблением тире ($b = 0{,}385$; pВН $= 0{,}017$) и бóльшим объёмом текста ($b = 0{,}396$; pВН $= 0{,}024$). Отрицательные связи были получены для форм первого лица ($b = 0{,}375$; pВН $= 0{,}032$) и частотности средств связи ($b = 0{,}308$; pВН $= 0{,}034$). Вместе с тем влияние объёма текста, абзацного членения и употребления двоеточий и точек с запятой значимо различалось между типами заданий: соответственно $F(2, 65) = 7{,}84$, pВН $= 0{,}005$; $F(2, 65) = 7{,}57$, pВН $= 0{,}005$ и $F(2, 65) = 10{,}59$, pВН $= 0{,}001$. Поэтому положение ИИ-текстов в пространстве ЕТ далее рассматривалось отдельно для описаний, повествований и рассуждений (рис. 1).

Совместная проверка взаимодействий восьми компонент с типом задания также показала, что конфигурация связей между лингвистическим пространством и качеством в целом различалась в зависимости от задания: $F(16, 65) = 4{,}83$, $p < 0{,}001$. Вместе с тем после поправки на множественные сравнения взаимодействие ни одной отдельно взятой компоненты с типом задания не достигло статистической значимости. В частности, направления связей PC1 и PC2 с качеством оставались относительно устойчивыми.

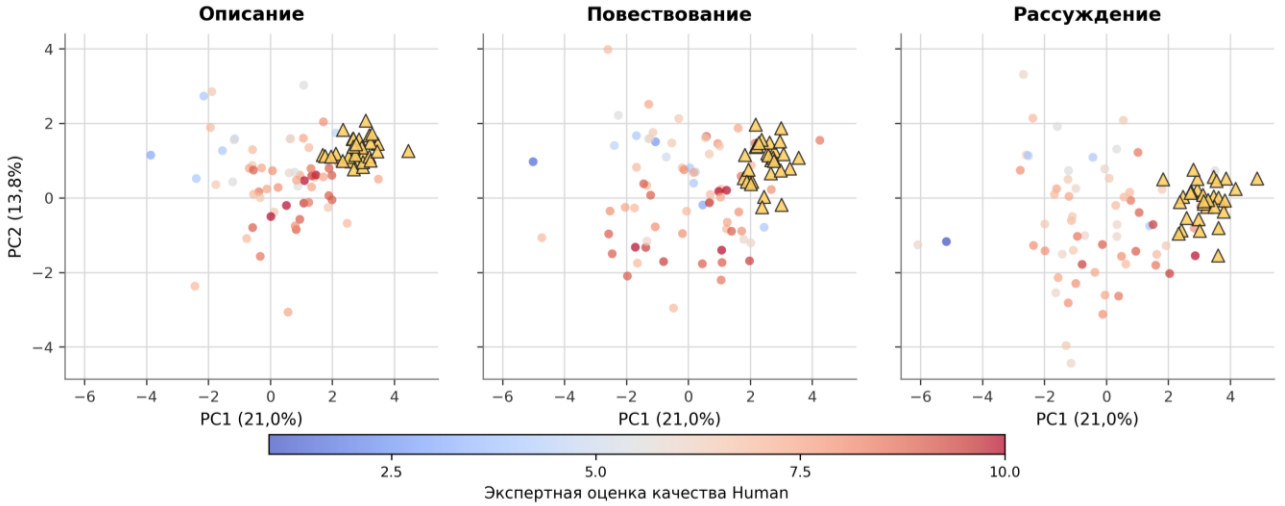


Рис. 1. Проекция ИИ-текстов в пространство, построенное на естественных текстах (ЕТ). Кругами обозначены ЕТ; их цвет отражает среднюю экспертную оценку качества. Жёлтыми треугольниками обозначены ИИ-тексты. На рисунке представлены первые две главные компоненты; расстояния до центроидов групп качества рассчитывались в пространстве восьми нормированных компонент

Наиболее выраженная близость к области высокого качества ЕТ была обнаружена для описаний. Средняя прогнозируемая оценка ИИ-текстов этого типа составила 7,56 балла (SD = 0,29). Из 30 ИИ-описаний 24 оказались ближе всего к центроиду ЕТ высокого качества, три - среднего и три - низкого качества. Средняя координата ИИ-описаний по PC1 превышала среднюю координату соответствующих ЕТ на 1,66 стандартного отклонения. По PC2 среднее положительное смещение составило 0,72 стандартного отклонения. Поскольку более высокие оценки ЕТ были связаны с отрицательными значениями PC2, последнее смещение было направлено в сторону, противоположную общей связи этой компоненты с качеством.

В описаниях из выборки ЕТ наиболее выраженную положительную связь с качеством имело лексическое разнообразие MATTR: увеличение показателя на одно внутригрупповое стандартное отклонение сопровождалось повышением оценки в среднем на 0,819 балла (pBH < 0,001). Положительные связи также были выявлены для средней длины слова (b = 0,732; pBH = 0,003) и доли уникальных слов (b = 0,638; pBH = 0,036). Напротив, более частое употребление форм первого лица (b = 0,573; pBH = 0,036) и средств связи (b = 0,552; pBH = 0,036) было связано с более низкими оценками качества. Таким образом, в описаниях качество определялось прежде всего разнообразием и характеристиками лексики, а также меньшей эксплицитностью авторской позиции.

Профиль ИИ-описаний в значительной степени соответствовал этому направлению. По сравнению с описаниями из группы ЕТ MATTR был выше на 1,15 стандартного отклонения, средняя длина слова - на 0,99, а доля однократных слов - на 1,97 стандартного отклонения. Частотность форм первого лица, напротив, была ниже на 1,38 стандартного отклонения. Эти особенности объясняют преимущественную близость ИИ-описаний к группе ЕТ высокого качества. Вместе с тем они ИИ-тексты данного типа короче ЕТ на 0,74 стандартного отклонения, отличаются меньшей вариативностью длины предложений (−0,54) и более редким употреблением тире (−0,57). Следовательно, высокая позиция ИИ-описаний связана прежде всего с лексическим профилем, тогда как их структурно-пунктуационная организация не полностью совпадает с организацией высоко оценённых ЕТ данной группы.

Для повествований была обнаружена иная картина. Средняя прогнозируемая оценка качества ИИ-текстов составила 7,20 балла (SD = 0,36), что является наименьшим значением среди трёх анализируемых типов текстов. Распределение относительно центроидов групп качества также было наиболее неоднородным: 13 текстов оказались ближе к группе высокого качества, шесть — среднего и 11 — низкого качества. При этом смещение по PC1 оставалось выраженным и составляло 1,54 стандартного отклонения, а положительное смещение по PC2 достигало 0,62 стандартного отклонения.

В повествованиях группы ЕТ качество было положительно связано с тремя показателями: лексическим разнообразием MATTR (b = 0,866; pBH = 0,005), объёмом текста (b = 0,717; pBH < 0,001) и частотностью двоеточий и точек с запятой (b = 0,660; pBH < 0,001). Следовательно, высокое лексическое разнообразие не исчерпывало профиль качественного повествования: более высокие оценки получали также развёрнутые тексты с более частым употреблением двоеточий и точек с запятой.

ИИ-повествования превосходили соответствующие ЕТ по MATTR на 1,66 стандартного отклонения, средней длине слова — на 1,12, доле однократных слов — на 1,86 стандартного отклонения. Частотность двоеточий и точек с запятой также была несколько выше среднего для ЕТ — на 0,36 стандартного отклонения. Однако объём ИИ-повествований был ниже на 0,83 стандартного отклонения, а вариативность длины предложений — на 0,73. Кроме того, ИИ-повествования

содержали больше абзацев в пересчёте на 100 слов (+0,81) и больше эксплицитных средств связи (+0,88). Таким образом, ИИ-тексты данного типа демонстрировали более высокое лексическое разнообразие, но не воспроизводили развёрнутость повествования, которая в выборке ЕТ была одним из основных признаков качества. Сочетание этих разнонаправленных характеристик согласуется со смешанным распределением ИИ-повествований между областями текстов высокого, среднего и низкого качества в выборке ЕТ.

ИИ-рассуждения получили наиболее высокую среднюю прогнозируемую оценку качества — 7,79 балла (*SD* = 0,40). Однако относительно центроидов они не концентрировались исключительно в области высокого качества: 14 текстов были ближе к высокой группе, 14 — к средней и только два — к низкой. Смещение относительно рассуждений из группы ЕТ по PC1 было максимальным среди трёх типов текстов и составляло 1,80 стандартного отклонения. Положительное смещение по PC2 было менее выраженным, чем в двух других типах текста, и составляло 0,38 стандартного отклонения.

В рассуждениях из группы ЕТ после поправки на множественные сравнения значимым оказался показатель абзацного членения: увеличение количества абзацев на 100 слов на одно стандартное отклонение сопровождалось снижением оценки на 0,767 балла (*p*ВН = 0,002). Для средней длины слова (*b* = 0,551; *p*ВН = 0,055), MATTR (*b* = 0,523; *p*ВН = 0,056) и частотности двоеточий и точек с запятой (*b* = 0,349; *p*ВН = 0,056) были выявлены близкие к порогу значимости тенденции. При этом коэффициенты для средней длины слова и MATTR были положительными, а для двоеточий и точек с запятой – отрицательным. В повествованиях последний показатель, напротив, был положительно связан с качеством. Различие направлений соответствует выявленному взаимодействию признака с типом задания и показывает, что значение одного и того же формального показателя зависит от коммуникативной задачи.

Лексические характеристики ИИ-рассуждений значительно превышали показатели человеческих текстов: сдвиг составлял 1,99 стандартного отклонения для MATTR, 1,13 - для средней длины слова и 2,35 - для доли уникальных слов. Одновременно ИИ-рассуждения были короче на 0,60 стандартного отклонения и отличались особенно низкой вариативностью длины предложений — на 1,21 стандартного отклонения ниже среднего для ЕТ. Количество абзацев на 100 слов было выше на 0,41 стандартного отклонения, то есть ИИ смещался в направлении, которое в рассуждениях группы ЕТ было связано со снижением оценки. Также отмечались более редкое употребление тире (–0,68) и средств связи (–0,34), но более частое употребление двоеточий и точек с запятой (+0,34). Таким образом, выраженное лексическое разнообразие соответствовало высокой прогнозируемой позиции ИИ-рассуждений, тогда как особенности абзацного членения и структурно-пунктуационной организации сближали значительную часть текстов не с высокой, а со средней группой качества ЕТ.

Таким образом, вопрос о соответствии ИИ-текстов определённому уровню качества ЕТ не может рассматриваться без учёта типа текста. ИИ-описания наиболее последовательно воспроизводили лексические признаки высоко оценённых ЕТ. В повествованиях выраженное лексическое разнообразие сочеталось с меньшей развёрнутостью текста, что согласуется с наиболее неоднородным распределением относительно центроидов качества ЕТ. В рассуждениях была получена наиболее высокая средняя прогнозируемая оценка, однако лексическое преимущество сочеталось с более дробным абзацным членением, связанным в выборке ЕТ со снижением качества, и существенно меньшей вариативностью длины предложений. Общим для трёх типов заданий было высокое лексическое разнообразие ИИ-текстов, тогда как их структурно-пунктуационные характеристики зависели от типа текста и не всегда соответствовали профилю высоко оценённых ЕТ.

Отдельно была оценена внутригрупповая вариативность текстов в пространстве восьми нормированных главных компонент. Суммарный разброс, рассчитанный как след ковариационной матрицы, во всех трёх типах текстов был существенно меньше у ИИ-сгенерированных текстов. Для описаний значение этого показателя составило 0,98 у ИИ-текстов и 5,79 у ЕТ, то есть 16,9 % уровня ЕТ; для повествований — соответственно 1,65 и 7,35, или 22,4 %; для рассуждений — 1,97 и 9,39, или 21,0 %. Следовательно, независимо от типа задания ИИ-тексты занимали значительно более компактную область лингвистического пространства, чем ЕТ.

Этот результат сохранялся после уравнивания размеров групп. При повторном расчёте по 10 000 случайных подвыборок, включавших по 30 ЕТ каждого типа, медианные значения отношения вариативности ИИ-текстов к вариативности ЕТ составили 16,9 % для описаний, 22,5 % для повествований и 21,0 % для рассуждений. В дополнительном анализе после исключения показателя объёма текста и повторного построения PCA только по ЕТ различия также сохранились: вариативность ИИ-текстов составляла 17,8 % человеческого значения для описаний, 24,1 % для повествований и 22,4 % для рассуждений.

Обсуждение

Полученные результаты показывают, что положение ИИ-текстов в пространстве качества ЕТ нельзя свести к однозначному выводу об их соответствии «хорошим» или «средним» текстам. Более половины ИИ-текстов были ближе всего к центроидам групп высокого качества ЕТ, а их средняя прогнозируемая позиция составила 7,52 балла. Однако это соответствие было связано прежде всего с лексическими характеристиками: ИИ-тексты отличались высоким лексическим разнообразием, большей средней длиной слова и высокой долей уникальных слов. По структурным и пунктуационным характеристикам совпадение с высоко оценёнными ЕТ было менее последовательным. Таким образом, ИИ воспроизводит отдельные измерения качества ЕТ, но не его единый целостный профиль.

Существенной оказалась зависимость результатов от типа текста. ИИ-описания наиболее последовательно приближались к области высокого качества ЕТ, поскольку именно лексическое разнообразие и характеристики слов были наиболее тесно связаны с оценками ЕТ данного типа. В повествованиях лексическое

преимущество ИИ сочеталось с меньшим объёмом и более однородной организацией предложений, тогда как качество ЕТ было связано также с их развёрнутостью. В рассуждениях высокая прогнозируемая позиция ИИ-текстов сопровождалась отличиями в абзацном членении и вариативности длины предложений. Следовательно, один и тот же набор характеристик ИИ-текстов имеет неодинаковое значение в разных коммуникативных задачах. Это подтверждает необходимость рассматривать соответствие ИИ-текстов качеству ЕТ как результат взаимодействия источника текста, типа задания и конкретного набора языковых характеристик.

Отдельного внимания заслуживает существенно меньшая внутригрупповая вариативность ИИ-текстов. Во всех трёх типах заданий их суммарный разброс составлял лишь 17–22 % от разброса соответствующих ЕТ. Этот результат сохранялся после уравнивания размеров групп и исключения показателя объёма текста. Следовательно, ИИ не просто располагается около среднего нормативного центра или области высокого качества ЕТ, а формирует более компактный и стандартизированный тип текста. Высокие лексические показатели сочетаются в нём с ограниченным разнообразием способов текстовой организации. Это позволяет предположить, что одним из отличительных признаков ИИ-текста может быть не отдельная языковая характеристика, а необычно устойчивое сочетание таких характеристик.

Предложенная процедура создаёт основу для дальнейшего изучения восприятия ИИ-текстов. Зафиксированное пространство ЕТ позволяет отделить измеряемый лингвистический профиль текста от последующих читательских суждений о его происхождении, качестве и достоверности. В дальнейших исследованиях координаты текста и его расстояние до центроидов групп качества ЕТ могут быть сопоставлены с воспринимаемой ИИ-похожестью текста, доверием к нему и оценкой достоверности представленной в тексте информации, что позволит установить, какие свойства перемещают ИИ-текст в область высокого качества и одновременно делают его более распознаваемым как сгенерированный.

Исследование имеет ряд **ограничений**. Лингвистическое пространство строилось по 14 сравнительно простым и легко извлекаемым признакам, не охватывающим семантическую релевантность, глубину аргументации, композиционную целостность и сложные синтаксические отношения. Поэтому его следует рассматривать как одну из возможных операционализаций качества текста, а не как исчерпывающую модель. Вместе с тем компактный набор признаков обеспечил идентичность обработки ЕТ и ИИ-текстов и прозрачность интерпретации результатов.

ИИ-тексты генерировались одной моделью с использованием минимального запроса, поэтому перенос результатов на другие модели и режимы генерации требует дополнительной проверки. Выборка из 90 ИИ-текстов достаточна для пилотной проекции, но не отражает всего разнообразия возможных ответов модели. Кроме того, прогнозируемая позиция ИИ-текстов основана на модели человеческих экспертных оценок и не является результатом их непосредственного оценивания. Дальнейшие исследования могут включать независимую оценку ИИ-текстов, расширенный набор признаков и сопоставление нескольких моделей. Указанные ограничения не отменяют основного результата: положение и профиль ИИ-текстов в человеческом пространстве качества систематически зависят от типа задания.

Выводы

Проведённое исследование позволило определить положение ИИ-сгенерированных текстов в лингвистическом пространстве, построенном исключительно на материале ЕТ с независимо полученными оценками качества. ИИ-тексты не участвовали в формировании пространства, а были спроецированы в него после фиксации структуры главных компонент. Такой подход позволил сопоставить тексты человека и ИИ в единой системе координат, не подстраивая параметры модели под особенности машинной генерации.

Результаты показали, что ИИ-тексты воспроизводят прежде всего лексическую составляющую ЕТ высокого качества. Они характеризуются повышенным лексическим разнообразием, большей средней длиной слова и меньшей частотой форм первого лица. Вместе с тем их профиль не совпадает полностью с профилем высоко оценённых ЕТ. Расхождения обнаружены по ряду структурных характеристик, включая объём текста, абзацное членение и употребление отдельных знаков препинания; их выраженность и направление зависят от типа текста.

Наиболее близкими к области высокого качества оказались ИИ-описания: 80 % из них располагались ближе всего к центроиду группы ЕТ высокого качества. Для повествований соответствие было менее однозначным: высокое лексическое разнообразие сочеталось с меньшей развёрнутостью текста, вследствие чего ИИ-повествования распределились между всеми тремя областями качества, причём доли текстов, ближайших к центроидам высокого и низкого качества, были сопоставимы. В ИИ-тексты типа "рассуждение" также воспроизводили лексические признаки ЕТ высокого качества, но характеризовались более дробным абзацным членением и особой пунктуационной организацией.

Во всех трёх типах заданий при использованной модели и процедуре генерации ИИ-тексты образовали существенно более компактные области пространства: их суммарная внутригрупповая вариативность составляла приблизительно 17–22 % от вариативности соответствующих ЕТ. Данное обстоятельство указывает на ИИ как на более стандартизированный тип текста, в котором отдельные признаки высокого качества выражены особенно сильно, тогда как межтекстовая и структурная вариативность существенно ограничена. Полученное пространство текстов ЕТ разного качества создаёт основу для дальнейшего изучения связи лингвистического профиля текста с воспринимаемой ИИ-похожестью, доверием к тексту и оценкой его достоверности.

Источник финансирования

Исследование выполнено за счет гранта Российского научного фонда No 26-78-10113, https://rscf.ru/project/26-78-10113

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Библиографический список

1. Berber Sardinha T. AI-generated vs human-authored texts: A multidimensional comparison // Applied Corpus Linguistics. 2024. Vol. 4, no. 1. Article 100083. DOI: 10.1016/j.acorp.2023.100083.
2. Biber D. Variation across Speech and Writing. Cambridge : Cambridge University Press, 1988. 299 p. DOI: 10.1017/CBO9780511621024.
3. Covington M.A., McFall J.D. Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR) // Journal of Quantitative Linguistics. 2010. Vol. 17, no. 2. P. 94–100. DOI: 10.1080/09296171003643098.
4. Doshi A.R., Hauser O.P. Generative AI enhances individual creativity but reduces the collective diversity of novel content // Science Advances. 2024. Vol. 10, no. 28. Article eadn5290. DOI: 10.1126/sciadv.adn5290.
5. Herbold S., Hautli-Janisz A., Heuer U., Kikteva Z., Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays // Scientific Reports. 2023. Vol. 13. Article 18617. DOI: 10.1038/s41598-023-45644-9.
6. Jakesch M., Hancock J.T., Naaman M. Human heuristics for AI-generated language are flawed // Proceedings of the National Academy of Sciences. 2023. Vol. 120, no. 11. Article e2208839120. DOI: 10.1073/pnas.2208839120.
7. Jiang F.K., Hyland K. Does ChatGPT Write Like a Student? Engagement Markers in Argumentative Essays // Written Communication. 2025. Vol. 42, no. 3. P. 463–492. DOI: 10.1177/07410883251328311.
8. Kendro K., Maloney J., Jarvis S. Do Large Language Models Produce Texts With “Human-Like” Lexical Diversity? Evidence From Four ChatGPT Models // International Journal of Applied Linguistics. 2026. Vol. 36, no. 3. P. 2822–2834. DOI: 10.1111/ijal.70115.
9. Litvinova T.A., Molchanova V.A., Dekhnich O.V., Seredin P.V. Describing Images: Linguistic Signatures of AI and Humans // 2025 38th Conference of Open Innovations Association (FRUCT). 2025. P. 198–208. DOI: 10.23919/FRUCT67853.2025.11239274.
10. Markey B., Brown D.W., Laudenbach M., Kohler A. Dense and Disconnected: Analyzing the Sedimented Style of ChatGPT-Generated Text at Scale // Written Communication. 2024. Vol. 41, no. 4. P. 571–600. DOI: 10.1177/07410883241263528.
11. Mucoz-Ortiz A., Gymez-Rodrнguez C., Vilares D. Contrasting Linguistic Patterns in Human and LLM-Generated News Text // Artificial Intelligence Review. 2024. Vol. 57. Article 265. DOI: 10.1007/s10462-024-10903-2.
12. Reviriego P., Conde J., Merino-Gymez E., Martнnez G., Hernбndez J.A. Playing with words: Comparing the vocabulary and lexical diversity of ChatGPT and humans // Machine Learning with Applications. 2024. Vol. 18. Article 100602. DOI: 10.1016/j.mlwa.2024.100602.
13. Sears S., Weisberg D.S. Bot or not: Can people tell the difference between stories written by a human or by an AI system? // Judgment and Decision Making. 2026. Vol. 21. Article e21. DOI: 10.1017/jdm.2026.10042
14. Литвинова Т. А., Молчанова В. А. Исследование межсловных пауз в текстах на русском языке по данным кейлоггера с использованием моделей гауссовой смеси // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10, № 4. С. 147-166. DOI: 10.18413/2313-8912-2024-10-4-0-8

References

1. Berber Sardinha, T. (2024) AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*. 4 (1), 100083. Available from: doi: 10.1016/j.acorp.2023.100083.
2. Biber, D. (1988) *Variation across Speech and Writing*. Cambridge. Cambridge University Press. 299 p. Available from: doi: 10.1017/CBO9780511621024.
3. Covington, M.A., McFall, J.D. (2010) Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*. 17 (2), 94–100. Available from: doi: 10.1080/09296171003643098.
4. Doshi, A.R., Hauser, O.P. (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*. 10 (28), eadn5290. Available from: doi: 10.1126/sciadv.adn5290.
5. Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., Trautsch, A. (2023) A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific Reports. 13, 18617. Available from: doi: 10.1038/s41598-023-45644-9.
6. Jakesch, M., Hancock, J.T., Naaman, M. (2023) Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences*. 120 (11), e2208839120. Available from: doi: 10.1073/pnas.2208839120.
7. Jiang, F.K., Hyland, K. (2025) Does ChatGPT Write Like a Student? Engagement Markers in Argumentative Essays. *Written Communication*. 42 (3), 463–492. Available from: doi: 10.1177/07410883251328311.
8. Kendro, K., Maloney, J., Jarvis, S. (2026) Do Large Language Models Produce Texts With “Human-Like” Lexical Diversity? Evidence From Four ChatGPT Models. *International Journal of Applied Linguistics*. 36 (3), 2822–2834. Available from: doi: 10.1111/ijal.70115.
9. Litvinova, T.A., Molchanova, V.A., Dekhnich, O.V., Seredin, P.V. (2025) Describing Images: Linguistic Signatures of AI and Humans. In: *38th Conference of Open Innovations Association (FRUCT)*, pp. 198–208. Available from: doi: 10.23919/FRUCT67853.2025.11239274.
10. Markey, B., Brown, D.W., Laudenbach, M., Kohler, A. (2024) Dense and Disconnected: Analyzing the Sedimented Style of ChatGPT-Generated Text at Scale. *Written Communication*. 41 (4), 571–600. Available from: doi: 10.1177/07410883241263528.
11. Mucoz-Ortiz, A., Gymez-Rodrнguez, C., Vilares, D. (2024) Contrasting Linguistic Patterns in Human and LLM-Generated News Text. *Artificial Intelligence Review*. (57), 265. Available from: doi: 10.1007/s10462-024-10903-2.

12. Reviriego, P., Conde, J., Merino-Gymez, E., Martнnez, G., Hernбndez, J.A. (2024) Playing with words: Comparing the vocabulary and lexical diversity of ChatGPT and humans. *Machine Learning with Applications*. 18, 100602. Available from: doi: 10.1016/j.mlwa.2024.100602.

13. Sears, S., Weisberg, D.S. (2026) Bot or not: Can people tell the difference between stories written by a human or by an AI system? *Judgment and Decision Making*. 21, e21. Available from: doi: 10.1017/jdm.2026.10042.

14. Litvinova, T.A., Molchanova, V.A. (2024) Issledovanie mezhslovnykh pauz v tekstakh na russkom yazyke po dannym keiloggera s ispol'zovaniem modelei gaussovoi smesi [Investigating between-word pause duration in Russian typed texts using mixture modeling based on keystroke data]. *Nauchnyi rezul'tat. Voprosy teoreticheskoi i prikladnoi lingvistiki*. 10 (4), 147–166. Available from: doi: 10.18413/2313-8912-2024-10-4-0-8. (In Russian)

Поступила в редакцию 01.07.2026
Подписана в печать 30.09.2026

Original Article
UDC 81'32
DOI: 10.47438/2309-7078_2026_3_227.

LINGUISTIC PROFILE OF AI-GENERATED TEXTS RELATIVE TO HUMAN-AUTHORED TEXTS WITH DIFFERENT QUALITY RATINGS

Tatiana A. Litvinova[1], Svetlana G. Lyubova[2], Elena S. Kotlyarova[3]

Voronezh State Pedagogical University[1, 2, 3]
Voronezh, Russia

[1]*Dr. Philol. Sci., Professor of the Department of Higher Mathematics*
ORCID ID: 0000-0002-6019-3700, tel.: +7-473-264-44-17, e-mail: centr_rus_yaz@mail.ru
[2]*Cand. Philol. Sci., Associate Professor of the Department of Theory, History and Methods of Teaching Russian Language and Literature, Researcher at the Laboratory for Cultural Safety of Language Models,*
ORCID ID: 0000-0002-2586-410X, tel.: +7-473-255-14-78,
e-mail: svetlana_lyubova_2305@mail.ru
[3]*Postgraduate Student of the Department of Russian Language, Modern Russian and Foreign Literature, Research Intern at the Laboratory for Cultural Safety of Language Models,*
ORCID ID: 0000-0002-4162-5002, tel.: +7-473-255-14-78, e-mail: kray-voronezh@mail.ru

Abstract. The widespread adoption of generative artificial intelligence has created a need for methods that compare AI-generated texts with human-authored texts whose quality has been assessed independently. Most current studies focus either on the automatic detection of AI-generated texts or on their direct evaluation by human raters. However, the position of AI-generated texts within a linguistic space derived from human-authored texts of known quality remains insufficiently understood. This article presents a comparison of three types of Russian-language AI-generated texts (descriptions, narratives, and argumentative texts) with corresponding human-authored texts. The human corpus comprised 199 texts written by 66 authors, with each text receiving two independent expert quality ratings; the AI corpus comprised 90 responses generated for the same writing tasks. The same set of 14 formal linguistic features was calculated for all texts. Principal component analysis was conducted exclusively on the human-authored texts, after which the AI-generated texts were projected into the fixed human reference space. Eight components explaining 82.72% of the variance in the linguistic features of the human-authored texts were retained for subsequent analysis. Higher quality in the human-authored texts was associated primarily with greater lexical diversity, greater mean word length and variability in word length, greater text length, and a lower frequency of first-person forms; some of these associations varied across writing tasks. AI-generated texts reproduced mainly the lexical dimension of this profile but differed from highly rated human-authored texts in several structural and punctuation-related characteristics. Overall, 56.7% of the AI-generated texts were closest to the centroids of the high-quality human groups. The clearest pattern was found for descriptions, 80% of which were closest to the high-quality centroid. In the space of eight normalized principal components, the total within-group variability of AI-generated texts amounted to only 17–22% of that observed in the corresponding human-authored texts. This result remained stable after equalizing group sizes and excluding text length from the feature set. Under the generation procedure used in this study, AI-generated texts can therefore be regarded as a relatively standardized text type that reproduces selected characteristics associated with high-quality human writing, but not the full profile of highly rated human-authored texts.

Key words: generative artificial intelligence, large language models, text quality, linguistic features, principal component analysis, human-authored text, AI-generated text, writing task, lexical diversity.

Cite as: Litvinova, T.A., Lyubova, S.G., Kotlyarova, E.S. (2026) Linguistic profile of AI-generated texts relative to human-authored texts with different quality ratings. *Izvestia Voronezh State Pedagogical University*. (3), 227–238. (In Russ., abstract in Eng.). DOI: 10.47438/2309-7078_2026_3_227.

Received 01.07.2026
Accepted 30.09.2026