



DOI: 10.22363/2313-1683-2025-22-1-144-174

EDN: UHNYYIW

УДК 372.881.1

Исследовательская статья

ChatGPT в англоязычном академическом письме: признаки машинной генерации текста в обзорах литературы студентов бакалавриата и магистратуры

В.А. Дугарцыренова 

Национальный исследовательский университет «Высшая школа экономики», Москва,
Российская Федерация
✉ vdugartsyrenova@hse.ru

Аннотация. В эпоху стремительного развития технологий искусственного интеллекта (ИИ) остро встает проблема недобросовестного использования студентами средств ИИ для генерации англоязычных научных работ или их фрагментов с последующим представлением их как авторского продукта. Эта проблема усугубляется отсутствием надежного инструментария для установления признаков машинного происхождения таких текстов. В настоящем исследовании на основе результатов предварительного качественного и количественного анализа 45 сгенерированных ChatGPT-3.5 англоязычных обзоров литературы к научным проектам (*research proposals*) предпринята попытка выявить отличительные характеристики таких текстов и определить их присутствие в англоязычных обзорах литературы, подготовленных студентами бакалавриата и магистратуры двух российских вузов. На первом этапе исследования в рамках анализа искусственно сгенерированных текстов и небольшой выборки работ магистрантов ($N = 12$) было установлено, что многие характеристики искусственно сгенерированных обзоров литературы в ярко выраженной форме проявились в студенческих текстах, что позволило рассматривать их в качестве признаков машинной генерации текста. Одним из таких признаков явилась аномально высокая повторяемость одних и тех же абстрактных слов и оборотов в искусственно сгенерированных и студенческих текстах. На этом этапе исследования с помощью корпусного менеджера AntConc был проведен количественный анализ частотности употребления таких оборотов в искусственно сгенерированных текстах и составлен список наиболее частотных слов и оборотов (обозначенных в работе как «обороты из языка ChatGPT»), указывающих на машинное происхождение текста. На втором этапе исследования первоначальные выводы о признаках машинной генерации текстов были подкреплены, уточнены и дополнены на основе качественного и количественного анализа объединенной выборки из 47 англоязычных обзоров литературы магистрантов и бакалавров. По результатам анализа работ было выявлено десять признаков машинной генерации текста, относящихся к структурно-содержательным и языковым аспектам подготовки англоязычных обзоров литературы, которые описываются и иллюстрируются в работе. Полученные результаты, выводы и рекомендации могут быть полезны широкому кругу чи-

© Дугарцыренова В.А., 2025



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

тателей, в особой мере педагогам и ученым, принимающим участие в обучении написанию англоязычных научных работ, экспертизе и рецензировании таких работ, а также осуществляющим руководство научными проектами на английском языке.

Ключевые слова: искусственный интеллект, ChatGPT-3.5, англоязычное академическое письмо, англоязычные обзоры литературы в проектах научного исследования, искусственно сгенерированные англоязычные обзоры литературы, признаки машинной генерации текста

Введение

В условиях интернационализации науки и образования, интенсификации сотрудничества между странами и академической мобильности, усиления конкуренции на глобальном рынке труда возрастают требования к уровню готовности современного специалиста к успешной профессиональной деятельности и взаимодействию в глобальном пространстве. В контексте этих изменений важным показателем профессиональной состоятельности работников науки и образования становится их международная видимость, выражающаяся, в частности, в наличии международных публикаций (чаще всего на английском языке). Данные тенденции не замедлили сказаться на образовательной политике российских вузов, усилившей акцент на профессиональной и научной составляющих иноязычной подготовки будущих специалистов. Так, в НИУ ВШЭ подготовка студентами всех образовательных программ письменного обоснования научного (дипломного) проекта (*research proposal*) на английском языке (АЯ) и его последующая защита (также на английском языке) становится обязательным условием завершения обучения уже на уровне бакалавриата.

В этих реалиях особую актуальность для участников образовательного процесса как в России, так и за ее пределами приобретают возможности нейросетей. Как показывает ряд работ (Арзютова, 2023; Воронин, Палёнова, 2024; Dalgıç, Yaşar, Demir, 2024; Deng, Lin, 2023; González et al., 2024; Huang, Tan, 2023; Lingard, 2023; Pavlik, 2023; Rudolph, Tan, Tan, 2023), на настоящий момент можно выделить несколько возможностей и функций нейросетей, которые позволяют решать широкий спектр задач в различных областях, включая задачи познавательной и научной деятельности. К таким возможностям относятся быстрая обработка и структурирование информации по запросу (например, плана научного исследования по ключевым словам, названию темы или исследовательским вопросам; упражнений и тестов разных видов и пр.); генерация и адаптация текстов разных жанров и на разных языках; редактирование, перевод и оформление текстов; математический и статистический анализ; моделирование и экспериментальная работа.

Неудивительно, что все большее количество авторов обращается к средствам ИИ для оптимизации процессов проведения научных исследований и подготовки научных текстов. Это подтверждается ростом опубликованных научных статей и материалов конференций, подготовленных с помощью ИИ. Так, авторы статьи Huang, Tan (2023) открыто заявляют, что на исходных ста-

дних подготовки статьи часть текста генерировалась ChatGPT, а затем редактировалась авторами. В отдельных публикациях (ChatGPT Generative Pre-Trained Transformer, Zhavoronkov, 2022; Pavlik, 2023) нейросеть ChatGPT даже указана в качестве соавтора, причем Zhavoronkov отводит ей роль первого автора за вклад в генерацию аргументов относительно возможностей и ограничений исследуемого препарата (рапамицина) на основе имеющихся медицинских данных (ChatGPT Generative Pre-Trained Transformer, Zhavoronkov, 2022).

В условиях усиления роли ИИ в научной сфере нельзя не отметить растущую обеспокоенность мирового академического сообщества (Else, 2023; Rahimi, Abadi, 2023; Sallam, 2023; Stokel-Walker, 2023) тем, что нейросети открывают широкие возможности для фальсификации научных данных и публикации «фейковых» исследований, распространения недостоверной информации в связи с феноменом так называемых «галлюцинаций» (“hallucinations”) — генерацией нейросетью ошибочных или вымышленных фактов из-за нехватки данных или неверной интерпретации запроса. По мере развития инструментов ИИ становится все труднее отличить научный текст, написанный человеком, от сгенерированного ИИ. Так, согласно результатам эксперимента, проведенного группой ученых из американских университетов (Gao et al., 2023), аннотации к 50 медицинским статьям, полностью сгенерированные нейросетью ChatGPT на основе оригинальных аннотаций, оказались настолько правдоподобными, что 34 % из них не смогли распознать как продукт машинной генерации ни специализированные программы для выявления ИИ, ни медицинские эксперты (32 %), которым эти тексты предъявлялись вместе с оригинальными аннотациями (Else, 2023). Система антиплагиата, использовавшаяся в эксперименте, подтвердила 100-процентную оригинальность всех текстов.

Данные опасения подтверждаются появлением опубликованных статей и препринтов, содержащих сфабрикованный контент. Так, нами было обнаружено англоязычное исследование, в котором почти все используемые источники оказались вымышленными (Rafiq, Nawaz, Afzal, 2025). В свете данных тенденций звучат призывы об установлении норм, регламентирующих использование средств ИИ в научной деятельности (Else, 2023; Sallam, 2023; Stokel-Walker, 2023). Журналы отдельных издательств, например, Elsevier¹, Science² и Springer³, уже ввели запрет на присвоение ИИ статуса соавтора и цитирование каких-либо работ, написанных в соавторстве с ИИ, и обязали авторов отразить в рукописи факт использования ИИ (например, в сопроводи-

¹ Elsevier. The use of generative AI and AI-assisted technologies in writing for Elsevier // Elsevier.com. 2023. URL: <https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier> (accessed: 21.03.2024)

² Science. Science journals: Editorial policies // Science.org. 2023. URL: https://www.science.org/content/page/science-journals-editorial-policies?adobe_mc=MCMID%3D79624516929185125304012049119377686396%7CMCORGID%3D242B6472541199F70A-4C98A6%2540AdobeOrg%7CTS%3D1711463628 (accessed: 21.03.2024)

³ Springer. Author guidelines. Guidance on the use of Large Language Models (LLM) // Link.springer.com. 2023. URL: <https://link.springer.com/journal/40616/updates/26032570> (accessed: 21.03.2024)

тельном письме или в разделе «Благодарности») в случаях, если они прибегали к средствам ИИ при проведении исследования или подготовке текста.

В российском образовательном контексте активное внедрение технологий ИИ на разных уровнях образования сопровождается все более широким использованием обучающимися нейросетей для решения различных учебно-познавательных задач, включая подготовку научных текстов. К сожалению, как показал наш опыт экспертизы англоязычных работ магистрантов и бакалавров в рамках курсов академического письма (АП), это все чаще приводит к недобросовестному применению ИИ — интеграции частично или полностью скопированного текста, сгенерированного ИИ. Это в особой мере характерно для студентов с низким уровнем владения иностранным языком (ИЯ), отсутствием необходимых компетенций и опыта в подготовке научных текстов на ИЯ и недостаточной уверенностью в своих возможностях. Данная проблема усугубляется тем, что специализированные детекторы ИИ (например, GPTZero, Content at Scale, Copyleaks, Crossplag, DetectGPT, GPT-2 Output Detector, Sapling, Turnitin, ZeroGPT и др.), призванные помочь в выявлении машинного происхождения текста, демонстрируют значительные погрешности в разграничении человеческого и искусственно сгенерированного текста, что подтверждается результатами исследований (Liang et al., 2023; Weber-Wulff et al., 2023). В частности, как показало исследование Liang et al. (2023), известные детекторы ИИ могут ошибочно определять тексты (англоязычные эссе и научные аннотации), написанные с ошибками авторами-носителями языка, как искусственно сгенерированные, а тексты носителей языка как человеческие, то есть низкое языковое качество текстов может определяться как признак машинной генерации. Более того, как утверждают авторы, эффективность распознавания сгенерированных текстов заметно снижается после их редактирования с помощью запросов улучшить качество языка текста (например, “Elevate the provided text by employing literary language” — «Сделай текст более литературным»), что еще более затрудняет выявление машинной генерации таких текстов.

Приходится констатировать, что ввиду стремительного развития средств ИИ и относительной новизны явления проблема недобросовестного применения ИИ студентами при подготовке иноязычных научных текстов на настоящий момент не получила должного освещения. В отечественной и зарубежной литературе отсутствуют исследования, посвященные анализу иноязычных научных работ студентов (в частности, подготовленных в рамках курсов академического письма) на предмет их машинной генерации. Лишь в отдельных отечественных исследованиях делаются попытки выявить и описать признаки машинного происхождения текстов на примере искусственно сгенерированных русскоязычных научных текстов. Например, в работе М.Н. Черкасовой и А.В. Тактаровой на основе анализа пяти научных текстов, созданных системой «Махтекст» (по серии запросов составить «Научный текст-диссертация на тему: основы прагмалингвистики, начала функциональной и скрытой прагматики»), выделяются семь признаков машинного происхождения таких текстов: ненаучность стиля, присутствие галлюцинаций ИИ, отсутствие

конкретики в ответе на поставленный вопрос, размытость высказывания при его большом объеме, нейтральность в оценочных суждениях, чрезмерное использование списков и отточенная орфография и пунктуация (Черкасова, Тактарова, 2024).

В свете изложенной проблемы в настоящем исследовании предпринята попытка выявить и описать признаки машинной генерации текста на основе анализа корпуса сгенерированных англоязычных обзоров литературы и работ аналогичного жанра, подготовленных магистрантами первого года обучения и студентами-бакалаврами двух российских вузов. Отправной точкой для проведения исследования послужили результаты анализа магистерских обзоров литературы (объемом 1300-1500+ слов) в рамках проводимого нами англоязычного курса по академическому письму, в ходе которого в восьми работах были обнаружены смысловые, языковые и структурно-логические аномалии, прямо или косвенно указывающие на машинное происхождение текстов или их частей: ссылки на несуществующие источники, безошибочное использование сложной академической лексики студентами с невысоким уровнем владения английским языком (АЯ), резкие различия в качестве языкового оформления текста в разных частях одной работы, а также включение в текст списков. Для проверки этих предположений было решено сформировать корпус искусственно сгенерированных англоязычных обзоров литературы по темам магистрантов ($N = 45$), изучение которого позволило бы выявить отличительные характеристики искусственно сгенерированных текстов данного жанра и установить наличие или отсутствие данных характеристик в студенческих работах, а затем проверить и дополнить наши выводы на основе анализа более широкой выборки работ.

Процедура и методы исследования

Исследование проводилось в два этапа. Задачи *первого этапа* включали:

1) создание корпуса искусственно сгенерированных англоязычных обзоров литературы по 12 темам работ магистрантов первого года обучения международной программы Communications and International Public Relations НИТУ МИСИС (информация о магистрантах представлена в табл. 1);

2) анализ искусственно сгенерированных текстов для выявления отличительных характеристик таких текстов как маркеров их машинного происхождения;

3) анализ обзоров литературы магистрантов на предмет проявления в них выявленных характеристик сгенерированных текстов.

Корпус искусственно сгенерированных англоязычных обзоров литературы был сформирован на основе серии запросов к нейросети ChatGPT-3.5. Первый запрос содержал установку на создание англоязычного обзора литературы для проекта магистерского исследования, в котором указывались тема исследования и требуемый объем текста в соответствии с заданием в курсе АП, а также делался акцент на включении только существующих источников, выверенных в библиографических базах данных и оформленных в формате цитирования APA.

Таблица 1 / Table 1

Информация о магистрантах НИТУ МИСИС / Characteristics of NUST MISIS students

Образовательная программа / Educational program	Страны / Countries	Уровни владения АЯ* (кол-во чел.) / Levels of English Proficiency (number of people)			
		B1	B2	C1	C2
Communications and International Public Relations (CIPR)	Россия	—	2	4	1
	Алжир	—	—	1	—
	Вьетнам	1	—	—	—
	Египет	—	—	1	—
	Нигерия	—	1	—	—
	Пакистан	1	—	—	—

*По Общеввропейской шкале уровней владения ИЯ.

*According to the Common European Framework of Reference for Languages.

Поскольку исходные версии сгенерированных текстов не соответствовали условиям запроса (например, отличались меньшим объемом и содержали несуществующие источники), формулировались дальнейшие запросы до получения версии текста, максимально отвечающей требованиям к заданию. Пример последовательности наших запросов по одной работе представлен в табл. 2. На каждом этапе цитируемые источники проверялись в библиографических базах данных. При обнаружении в тексте или списке литературы несуществующих источников в последующих запросах вновь указывалось на необходимость верификации этих источников в базах данных и корректировки текста с использованием только существующих источников. Всего итераций обзоров для одной работы получилось 3-4. В итоге было получено 45 сгенерированных текстов (включая все их итерации). Объем текстов варьировался от 643 до 1469 слов. Средний объем текстов составил 1276 слов.

Таблица 2 / Table 2

Последовательность запросов, используемых для генерации обзоров литературы по одной работе / Prompts used to generate literature reviews per topic

№	Формулировка запросов / Prompt
1	Can you write a detailed literature review of 1,300-1,500 words (without references) for a Master's level research proposal "Multi-Level Marketing Strategy: Concepts, Characteristics and Legal Development Experiences for Sales Companies in Vietnam"? Please include real references verified in scientific databases and search engines formatted in the APA style.
2	Many of the references are non-existent sources. Also, this review has numerous short subsections. Can you reorganize it into a more coherent text using real sources verified in scientific databases and search engines?
3	This version is much shorter than 1,300 words. Can you expand it into a text of at least 1,300 words using real sources verified in scientific databases and search engines?
4	You did not remove the non-existent sources. Please rewrite the text without those sources while keeping its length to at least 1,300 words.

После создания корпуса был проведен качественный и количественный анализ сгенерированных текстов. В качестве отправной точки для проведения анализа использовались критерии оценивания студенческих обзоров литературы, разработанные в рамках курса АП: содержание, организация текста, языковое оформление и оформление ссылок на источники по международному формату АРА (Приложение 1). В ходе анализа отдельные выводы формулировались индуктивно.

В рамках анализа особенностей языкового оформления сгенерированных обзоров литературы был также проведен анализ частотности употребления повторяющихся слов и оборотов из языка ChatGPT (например, абстрактных оборотов высокого уровня сложности, существительных во множественной форме, прилагательных, обозначающих важность явления), многократное использование которых могло бы служить признаком машинного происхождения текста. В этих целях был составлен исходный список таких слов и оборотов (*nuances, complexities, landscape, complex interplay, multifaceted nature, pivotal* и пр.). Исходный список далее уточнялся и дополнялся посредством количественного анализа сгенерированных текстов с использованием функций KWIC (Key Words in Context) и Cluster корпусного менеджера AntConc. До начала анализа из всех текстов были удалены списки литературы. Каждая единица из списка вводилась в поле поиска и анализировалась по частотности употребления во всем корпусе. Глаголы вводились в разных словоформах (например, *encompass, encompasses, encompassed, encompassing*). При подсчете частотности употребления таких единиц суммировались результаты поиска по всем словоформам. В ходе анализа выявлялись новые единицы с высокой частотностью употребления, которые не были включены в исходный список. Эти единицы добавлялись в список для последующего анализа и верификации. Среди таких единиц были существительные во множественном числе (например, *impacts, realities*) и прилагательные в сравнительной степени (*broader* и *further*), которые рассматривались нами как отдельные единицы ввиду частотности их употребления в самостоятельном значении именно в этих формах. В качестве минимальной частоты встречаемости единицы во всем корпусе было выбрано 10. По результатам анализа был сформирован список наиболее высокочастотных слов и оборотов из языка ChatGPT (Приложение 2). На рис. 1 иллюстрируются результаты поиска частотности употребления наречия *collectively* с помощью функции KWIC. Результаты отображены в виде конкорданса — структурированного списка всех употреблений данного слова в корпусе. Как можно видеть, наречие *collectively* встречается 11 раз, причем чаще всего в сочетании с глаголом *underscore*.

Наконец, на этом этапе исследования был проведен качественный и количественный анализ студенческих обзоров литературы, который позволил установить наличие в них выявленных характеристик сгенерированных текстов.

На втором этапе исследования ставились задачи уточнить и дополнить полученные выводы о признаках машинной генерации в англоязычных обзорах литературы на большей выборке работ — рассмотренных ранее 12 работ магистрантов и 35 новых работ — обзоров литературы (объемом 800–1000+) в проектах дипломных исследований студентов-бакалавров факультета мировой экономики и мировой политики НИУ ВШЭ, поступивших к нам в рамках принятой в НИУ ВШЭ ежегодной процедуры лингвистического рецензирования таких проектов. Все 35 авторов работ являлись гражданами России с уровнями владения АЯ В1+–С1, 18 из них обучались по образовательной программе «Востоковедение», 13 — «Экономика», 4 — «Международные отношения».

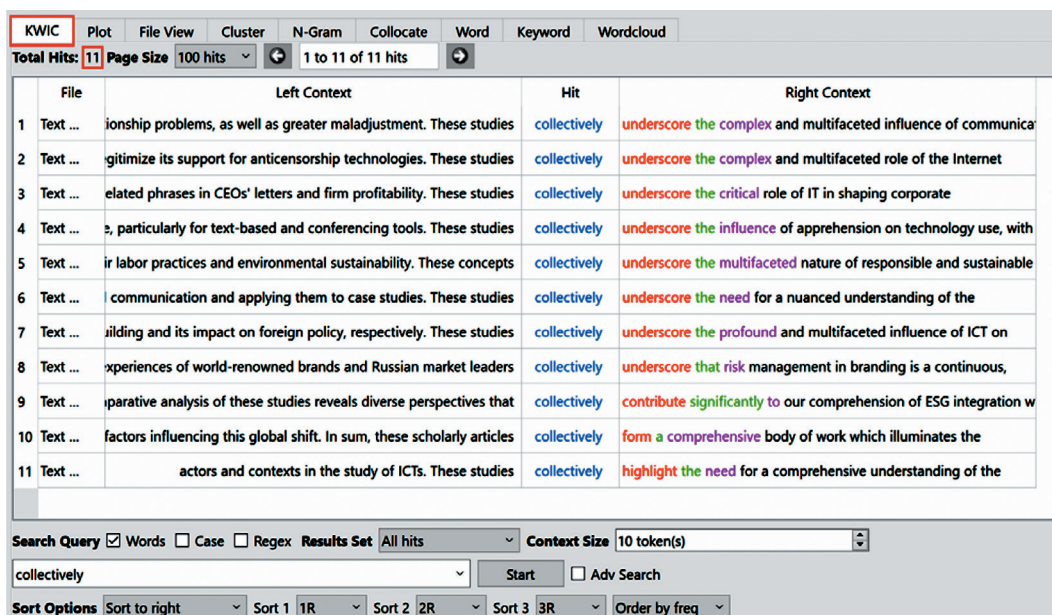


Рис. 1. Частотность употребления заданного слова *collectively*

Источник: скриншот экрана результатов анализа с использованием функции KWIC корпусного менеджера AntConc (выполнено В.А. Дугарцыреновой)

Figure 1. Frequency analysis for the key word *collectively*

Source: screenshot of the analysis results using the KWIC function of the AntConc corpus manager (performed by Vera A. Dugartsyrenova)

Все поступившие на рецензирование работы бакалавров успешно прошли независимую предварительную проверку на оригинальность, выполненную сотрудником факультета с помощью программы «Антиплагиат». 29 из них были подготовлены авторами самостоятельно на основе рекомендованного им учебного пособия и других вспомогательных ресурсов. Остальные 6 (17 %) были выполнены в рамках проводимого нами факультативного англоязычного курса «Академическое письмо».

В рамках этого этапа исследования все тексты были проанализированы на предмет проявления в них признаков машинной генерации текста, выявленных на первом этапе исследования, и других особенностей, характерных для искусственно сгенерированных обзоров литературы. Был проведен качественный и количественный анализ таких проявлений. В ходе анализа языкового оформления работ фиксировалось наличие в них частотных слов и оборотов из языка ChatGPT (Приложение 2), таких как *multifaceted* или *collectively underscore*. Затем с помощью функции KWIC в AntConc проводился подсчет встречаемости этих единиц в каждой работе, а также определялось общее и среднее количество их повторов в одной работе и всем корпусе. Подсчет частотности включал в себя как ключевые слова, используемые в отдельности (например, *nuanced*), так и все производные от них словосочетания (например, *nuanced challenges*), а также словоформы, образуемые от глаголов из списка (например, *navigate* — *navigation*). При этом из подсчета исключались

слова и обороты из списка, имевшие явную тематическую обусловленность в контексте конкретного исследования (например, *intricate* во фразе *intricate IT concepts* в работе о стратегиях объяснения сложных ИТ-понятий неспециалистам).

Применительно к признакам, связанным с безошибочным использованием сложных синтаксических и грамматических конструкций, артиклей и правил пунктуации, в рамках настоящего исследования анализ их проявлений был ограничен подсчетом количества и процентной доли работ, в которых эти признаки были зафиксированы хотя бы один раз.

Результаты исследования

Первый этап исследования: характеристики искусственно сгенерированных текстов

По результатам качественного и количественного анализа сгенерированных англоязычных обзоров литературы ($N = 45$) были выявлены следующие отличительные характеристики таких текстов, относящиеся к их *содержанию, организации, языковому оформлению*, а также *оформлению в них ссылок на источники в формате APA*.

Содержание

1. Информация охватывает разнообразные аспекты темы исследования, но разброс слишком обширен, что придает обзору справочный характер.

2. Подразделы, посвященные разным аспектам темы, включаются по принципу тематической релевантности.

3. Почти в каждом новом параграфе (при том, что параграфы состоят из 2–5 предложений) вводится новый аргумент, который не получает дальнейшего развития. Отсутствует узкоспециальная терминология.

4. Аргументы в отдельных параграфах подкрепляются максимум 1–3 ссылками на источники. Некоторые параграфы не содержат ссылок на источники, хотя в тексте присутствуют отсылки к неким исследованиям (*Several studies highlight ...*).

5. Часть цитируемых источников в исходной версии и последующих итерациях всех сгенерированных текстов являются вымышленными.

6. Свыше 70 % источников в работах по малоизученным темам, а также исследованиям в неанглоязычных контекстах оказываются вымышленными (что может означать, что аргументы, опирающиеся на такие источники, также носят фиктивный характер). При запросе удалить такие ссылки (например, по исследованиям в неанглоязычных регионах) — в последующих версиях текстов они заменяются другими. Удаление вымышленных ссылок возможно при указании фамилий авторов. При удалении вымышленных источников и сохранении требуемого объема текста наблюдается почти полное отсутствие ссылок на источники.

7. Ссылки на источники, опубликованные до даты обучения нейросети, часто достоверны.

8. В определениях ключевых понятий часто отсутствуют ссылки на источники.

9. Во всех текстах наблюдается отсутствие прямых цитат.

10. Информация о существующих источниках в открытом доступе может содержать неточности (например, в описании контекста, методов и результатов исследований, статистических цифрах).

11. Библиографические ссылки на существующие источники могут содержать ошибки в годах публикации источников (ошибки исправляются частично или не устраниваются).

12. DOI в библиографических ссылках часто ошибочны, что в большинстве случаев указывает на фантомный характер ссылок.

13. Внутритекстовые ссылки могут не совпадать с затекстовыми.

Организация текста

1. 76 % текстов начинаются со вводного параграфа, в 18 % из них присутствует заголовок “Introduction”. Такой параграф может содержать список с пунктами плана раздела (3 случая).

2. В 53 % текстов вводный параграф завершается формулировкой цели обзора с помощью типичных оборотов *This literature review synthesizes research on ... / critically examines ...*

3. 49 % текстов дробятся на 3–12 подразделов с 1–3 краткими параграфами и претенциозными заголовками (“Conceptualizing Inclusive Marketing”, “Theoretical Perspectives on ...”, “Theoretical and Practical Implications”). В подразделах встречаются нумерованные и/или маркированные списки из 4–5 пунктов с заголовками (11 % текстов).

4. При делении текста только на два подраздела каждый из них завершается заключительным параграфом, который в общих чертах обобщает содержание подраздела (11 % случаев).

5. 84 % текстов содержат краткий заключительный параграф с максимально обобщенными формулировками (в 74 % из них параграф не имеет заголовка и идентифицируется по вводным оборотам *In conclusion, in sum, in summary*; в остальных случаях параграф имеет заголовок “Conclusion”).

6. В 53 % текстах в заключительном параграфе приводятся рекомендации для проведения будущих исследований, которые можно опознать по формулировкам *Future research should continue to explore/could explore ...*

7. Тексты могут включать отдельный параграф с рекомендациями для проведения будущих исследований (например, с заголовками “Future Directions and Research Gaps”, “Recommendations for Future Research”) (два случая) или параграф с практическими рекомендациями (три случая).

8. Максимальное количество предложений в параграфах во всех текстах не превышает пяти.

Языковое оформление

1. Во всех текстах выдерживается нейтральный академический стиль.

2. Все тексты демонстрируют безошибочное использование сложных синтаксических и грамматических конструкций.

3. В текстах наблюдается высокая повторяемость низкочастотных академических слов и оборотов (*nexus, foundational, necessitate, elucidate*).

4. Для всех текстов характерно частотное использование абстрактных слов и оборотов (*complexities, multifaceted nature, navigate ... landscape, collectively underscore*), передающих отвлеченные, многоаспектные понятия и предполагающих широкую интерпретируемость.

5. Во всех текстах присутствуют существительные во множественном числе, обозначающие собирательные категории и абстрактные процессы (*global impacts, operational efficiencies, vulnerabilities, challenges and opportunities*).

6. С высокой частотностью встречаются прилагательные, подчеркивающие важность или остроту явления (*critical, pivotal, essential, crucial, central, vital*).

7. Тексты демонстрируют безупречное владение правилами употребления артиклей и пунктуации.

Оформление ссылок по требованиям формата APA

1. Внутритекстовые и затекстовые ссылки оформляются в целом корректно в соответствии с требованиями формата APA.

2. В затекстовых ссылках наблюдаются небольшие погрешности в оформлении названий некоторых видов источников, например, связанные с использованием заглавных букв во всех словах названий книг и научных статей.

Последующий анализ студенческих обзоров литературы на предмет проявления в них выявленных характеристик сгенерированных текстов на этом этапе исследования показал, что данные характеристики в явно выраженном виде присутствовали в 8 из 12 (67 %) рассмотренных работ магистрантов из России (3), Алжира (1), Вьетнама (1), Египта (1), Нигерии (1) и Пакистана (1). Так, в пяти работах было отмечено присутствие несуществующих источников и их описаний, причем две из них (предполагающие проведение исследования во Вьетнаме и Пакистане) были целиком или на 80 % составлены из соединенных между собой описаний несуществующих источников. При относительно небольшом объеме текстов (до 1500 слов) пять работ демонстрировали избыточное деление текста на множество кратких подразделов (вплоть до 12), в четырех из данных работ в подразделах использовались нумерованные или маркированные списки. В двух работах присутствовал краткий заключительный параграф с абстрактными формулировками выводов. В восьми работах весь текст или его фрагменты отличались аномально частотным и безошибочным использованием слов и оборотов из языка ChatGPT (Приложение 2), сложных синтаксических и грамматических структур, артиклей и пунктуации. Такие «совпадения» дали нам основания рассматривать их в качестве признаков машинной генерации текста.

Таким образом, на данном этапе мы выделили **семь признаков машинной генерации текстов** или их частей:

1) цитирование в тексте и списке литературы несуществующих источников;

- 2) избыточное дробление текста на множество кратких подразделов;
- 3) представление информации в виде списков;
- 4) включение обобщающего заключительного параграфа в формате ChatGPT;
- 5) частотное использование слов и оборотов из языка ChatGPT;
- 6) безошибочное использование сложных синтаксических и грамматических конструкций и артиклей;
- 7) безошибочное соблюдение правил англоязычной пунктуации.

Второй этап исследования: анализ корпуса студенческих работ

В ходе анализа объединенного корпуса студенческих работ ($N = 47$) были выявлены новые признаки машинной генерации текста. Так, в девяти работах бакалавров в описаниях отдельных существующих источников (23 случая) было отмечено присутствие абстрактных формулировок, насыщенных шаблонными оборотами из языка ChatGPT. Проверка содержания этих источников показала, что часть информации в описаниях недостоверна. При генерации искусственных описаний к этим источникам с помощью ChatGPT-3.5 были обнаружены сходства в содержательных и языковых характеристиках этих описаний: бессодержательность формулировок, неточность информации и использование лексики, характерной для языка ChatGPT. Это позволило нам выделить наличие таких описаний как признак машинной генерации текста. Кроме того, было зафиксировано включение авторами-бакалаврами вводных параграфов, содержащих типичные для искусственно сгенерированных текстов размытые формулировки и сообщающих в последнем предложении цель раздела (*This literature review synthesizes/critically examines ...*), а также завершение обзора литературы формулировкой рекомендаций для будущих исследователей (*Future research should explore ...*).

В целом по результатам анализа было выявлено **десять признаков машинной генерации текста в англоязычных обзорах литературы** студентов магистратуры и бакалавриата на уровне содержания текста, его организации и языкового оформления. Далее следует информация об особенностях и частотности проявления каждого признака, которая иллюстрируется примерами из работ (с пометами «бакал.» и «магистр.»).

На уровне содержания

1. Цитирование в тексте и списке литературы несуществующих источников. Описания несуществующих источников (научных статей, диссертаций, глав сборников и пр.), как правило, включают в себя вымышленный контекст исследований, методы их реализации, результаты и выводы; имена выдуманных или реальных авторов (чаще всего экспертов в области проводимого исследования); названия существующих или вымышленных журналов; вымышленный диапазон страниц в статьях и главах сборников. Названия работ в библиографических ссылках часто «склеиваются» из ключевых слов по теме исследования. Ссылки на источники оформляются в целом корректно в запрошенном пользователем формате цитирования.

Данный признак был обнаружен в 12 работах (5 магистрантов и 7 бакалавров, 26 % выборки); общее число его проявлений составило 46. Так, в работах трех международных магистрантов количество вымышленных источников варьировалось от 5 до 9 источников. **Пример (1)** содержит ссылки на несуществующую научную статью и ее описание из работы на тему “Multi-Level Marketing Strategy: Concepts, Characteristics and Legal Development Experiences for Sales Companies in Vietnam” («Многоуровневый маркетинг: концепции, особенности и опыт правового регулирования компаний по продажам во Вьетнаме»). В данном случае реальным является только название журнала. Всего в работе было семь ссылок, все из которых относились к вьетнамскому контексту исследования (что подтверждается фамилиями авторов) и *все* были вымышленными.

(1)

Nguyen, H., Do, T. H., & Dang, H. T. (2023). The impact of multi-level marketing participation on the psychological well-being of Vietnamese distributors. *Journal of Business Economics and Management*, 25(5), 609-619.

Nguyen, H. et al (2023) examine the relationship between MLM participation and psychological well-being among Vietnamese distributors. They find that MLM participation is associated with lower levels of life satisfaction, happiness, and self-esteem. Their findings underscore the importance of recognizing the psychological aspects of MLM participation and addressing the potential harms it may pose to the well-being of individuals involved. The article highlights the need for policymakers, MLM companies, and distributors themselves to prioritize measures that promote psychological well-being, such as providing adequate support systems, promoting financial literacy, and fostering a more balanced and sustainable approach to MLM activities. They suggest that this may be due to the stress and pressure associated with MLM, as well as the financial instability and social isolation that some distributors experience. (*магистр.*)

Примечательно, что сгенерированные на эту тему тексты также содержали не менее половины подобных фантомных источников и их описаний. При этом в двух последующих версиях, несмотря на запрос удалить несуществующие источники (с указанием конкретных имен авторов), нейросеть продолжала генерировать новые фантомные ссылки. Только четвертая версия текста не содержала фантомных ссылок, поскольку в запросе было четко указано исключить все фантомные ссылки и анализировать только реальные источники. Однако в этой версии было рассмотрено только три источника при сохранении исходно заданного объема текста (не менее 1300 слов).

2. Включение малосодержательных и/или неточных описаний существующих источников. Подобные описания носят обобщенный, поверхностный характер, некоторые могут собираться из ключевых слов по теме исследования. Часть информации (например, про методы, результаты или выводы исследования) может быть недостоверной.

Проявление этого признака было отмечено в девяти работах студентов-бакалавров (19 %) (всего 23 случая). **Пример (2)** наглядно демонстрирует размытость формулировок в описаниях двух цитируемых источников, включающих обороты с поэтическим окрасом (*harmonious blend of ... changes*) и обороты из языка ChatGPT (Приложение 2), такие как *complexities, transcend*

и *underscore* (подчеркнуты и выделены курсивом). В этом примере источник *Berkhout et al. (2012)* обозначен как “article” (статья), хотя при ознакомлении с его содержанием становится понятно, что это не вполне исследовательская статья, а вступительное слово редакторов (editorial) к специальному выпуску журнала *Sustainability Science*.

(2)

Frans Berkhout, Peter Marcotullio, and Tatsuya Hanaoka’s 2012 article offers a *comprehensive* exploration into the *complexities* of energy transition. They delve into the concept of ‘Transition Studies,’ *highlighting the need for a shift to renewable energy sources that transcends mere technological advancements*. The authors argue that this shift requires a *harmonious blend of economic, political, institutional, and socio-cultural changes*. Their discussion *underscores the importance of adopting an ethical approach that seamlessly integrates technology with sustainability*. This perspective *echoes the sentiments of Jean-Louis Armand, who advocates for long-range responsibility in sustainability science*. The authors provide a detailed review of the current state and the various challenges involved in transitioning to renewable energy. They emphasize the increasing focus on renewable energy sources, spurred by concerns about the sustainability and acceptability of nuclear power and carbon capture and storage technologies. (бакал.)

Для сравнения сгенерированное ChatGPT-3.5 этого же источника объемом 289 слов (рис. 2) содержит схожие абстрактные формулировки, включающие слова и обороты из языка ChatGPT (в частности, такие как *complexities*, *dynamics* и *comprehensive*, повторяются несколько раз): *delve into the complexities¹ of energy transitions; crucial for achieving sustainability goals; the importance of understanding the dynamics¹ ... of energy transitions; discuss various dimensions of energy transitions; stress the need for comprehensive¹ approaches; take into account complex interactions; address the complexities² of energy transitions; provide a ... comprehensive² analysis of the complexities³ and dynamics² of energy transitions*.

На уровне организации текста

1. Включение вводного параграфа. Вводный параграф (может иметь подзаголовок *Introduction*) задает контекст исследования. В последнем предложении такого параграфа часто сообщается цель написания раздела (*This review synthesizes research/insights/findings from ...*).

Данный признак проявился в десяти работах бакалавров (21 %). В **примере (3)** представлен пример вводного параграфа, в котором также присутствуют размытые формулировки с использованием шаблонных оборотов и существительных во множественном числе из языка ChatGPT (подчеркнуты и выделены курсивом), и в конце приводится цель работы (*This review synthesizes insights from multiple studies ...*).

(3)

The COVID-19 pandemic emphasizes the *crucial* role of sustainable financing in enhancing economic *resilience* in times when global crises arise. This is an approach in which investors consider not only the financial return, but also the social, environmental and governance *impacts* of their investments. Such investments cover a wide range of projects, from renewable energy to clean technology and social enterprises. *This review*

synthesizes insights from multiple studies to evaluate how sustainable financial strategies influenced the vulnerability of national economies to market fluctuations during the pandemic. (бакал.)

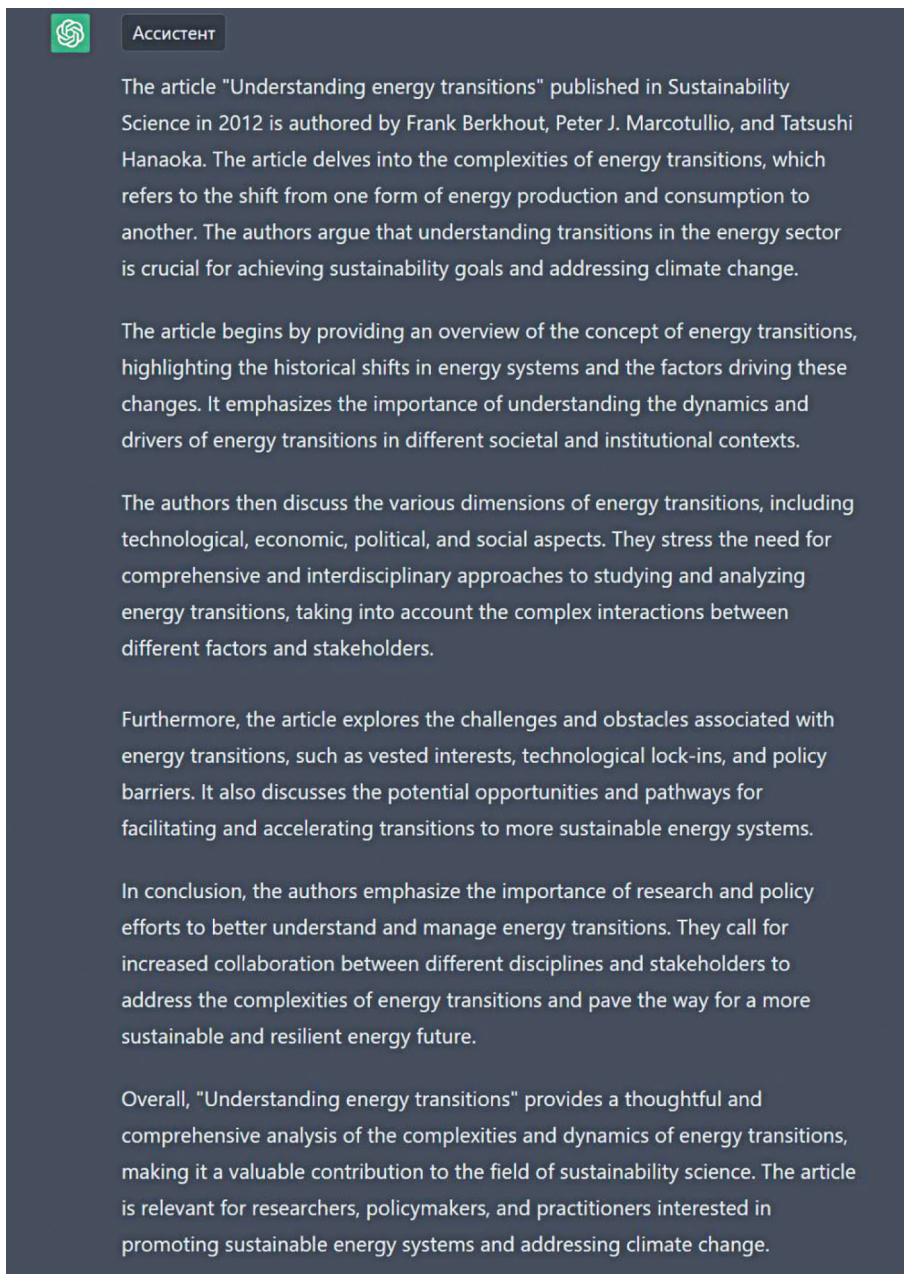


Рис. 2. Сгенерированное ChatGPT-3.5 описание источника
Источник: скриншот ответа ChatGPT-3.5 (выполнено В.А. Дугарцыреновой)

Figure 2. ChatGPT-3.5 generated description of the source
Source: screenshot of ChatGPT-3.5 response (performed by Vera A. Dugartsyrenova)

2. Деление текста на множество кратких подразделов. Текст неестественным образом делится на множество кратких подразделов, часто с многообещающими заголовками (“...: A Conceptual Overview”, “... Theoretical Foun-

dations”). Подразделы далее дробятся на небольшие параграфы (длиной 1–3 предложения). Информация в подразделах носит краткий реферативный характер и не раскрывает аспекта темы. Количество цитируемых в подразделах источников минимально (1–2). Ссылки на исследования в неанглоязычных контекстах часто оказываются вымышленными. Взаимосвязь между подразделами в тексте не обозначается (и не прослеживается).

Этот признак был зафиксирован в пяти работах магистрантов и семи бакалавров (26 %). В **примере (4)** представлен наиболее яркий случай проявления этого признака. В работе раздел включает в себя 12 кратких подразделов, в свою очередь состоящих из множества параграфов объемом 1–2 предложения. В данном примере мы приводим часть параграфов (первые три) в первых двух подразделах. Между подразделами не прослеживается четкая логическая связь, отсутствует анализ каких-либо источников. Последующие подразделы имеют схожую структуру и содержание.

(4)

The Role of Language Socialization in Child and Literacy Development: A Review of Literature

Introduction

Literacy skills are foundational for both academic success and lifelong learning, essential to a child’s growth. Reading, writing, and comprehension are integral components of literacy development, enabling children to navigate the complexities of today’s world.

Language socialization emerges as a crucial determinant shaping children’s literacy trajectories within the broader framework of child development. This term encompasses the social and cultural processes individuals undergo within their communities to acquire and refine language skills.

Language acquisition, language use, and language ideologies are influenced by a myriad of interactions, practices, and experiences occurring within family, peer, and educational settings. ...

Definition and Conceptual Framework:

This subsection aims to define language socialization and underscore its significance in shaping children’s literacy abilities.

Language socialization refers to the process through which individuals acquire language within the context of their culture and social networks. It encompasses a myriad of interactions, practices, and experiences within peer groups, families, schools, and communities.

Through language socialization, children not only develop linguistic proficiency but also learn the cultural and social norms surrounding language use ... (*магистр.*)

3. Представление информации в виде списков. В подразделах встречаются нумерованные или маркированные списки из 4–5 пунктов с вводным предложением. Пункты часто включают краткие заголовки, выделенные жирным шрифтом и содержащие двоеточие (**1. High awareness of general measures: ...**). Каждый пункт состоит из 1–2 предложений или кратких параграфов, включающих ссылки на источники (зачастую несуществующие), но не во всех случаях. При интеграции в текст списки могут не соотноситься с общим контекстом и стилем повествования в работе.

13 работ (4 магистрантов и 9 бакалавров) (28 %) включали информацию, представленную в виде списков, причем в четырех работах такие списки предъявлялись дважды. Всего было выявлено 15 таких проявлений в корпусе. В **примере (5)**, дав определение понятия «бренд», автор переходит к рассмотрению процессов формирования и продвижения брендов и включает краткую справочную выжимку в виде списка из пяти пунктов с заголовками. В выжимке вкратце обобщаются сходства и различия между Google и Yandex. Как можно видеть, информация носит поверхностный характер, связь приводимых фактов с процессами формирования и продвижения брендов не поясняется. Отсутствует взаимосвязь с первым параграфом, где вводится понятие «бренд». Выжимка не содержит ссылок на источники.

(5)

Many Russian scholars defined the concept of “brand” as the trade name of a competitive product, service, or organization that was popular for a long time among a specific group of consumers, satisfying their expectations. This approach was also followed by Philip Kotler, Thomas Neilson, Malcolm McDonald, Geoffrey R. Hodgson, and Anatoly Volkov (2021).

Let’s consider brand formation and promotion by comparing two companies, Google and Yandex:

(далее следует справочная вставка, в которой вкратце обобщаются сходства и различия между Google и Yandex):

Uniqueness and quality of the product or service: Both brands offer innovative technologies and products that solve users’ problems and improve their lives. Google is known for its search engines, cloud services, and devices such as Google Search, YouTube, Gmail, Maps, Drive, Play, News, Hangouts, Translate, Photos, Books, Docs, Blogger, Custom Search, Earth, Calendar, Keeper, Analytics, AdSense, and App Engine. Yandex provides services such as Alisa (virtual assistant), ya.ru, Yandex.Afisha, Yandex Browser, Yandex Disk, Yandex Direct, Yandex Food, Yandex Mail, Yandex Maps, Yandex Market, Yandex Metrica, Yandex Music, Yandex News, Yandex Taxi, and Yandex Translator.

User-friendliness: Both brands strive to make their products and services as simple and convenient to use as possible in order to attract and retain users.

Innovation and technological progress: As global leaders in their industry, Google and Yandex constantly invest in research and development of new technologies. They implement advanced search algorithms, artificial intelligence, and machine learning.

Social responsibility: Both brands carry out charitable programs, support scientific research, and invest in the development of education and information literacy.

Marketing strategies: Google and Yandex use various channels and tools such as online advertising, television, events, sponsorship, and others to attract attention and convince potential users of the advantages of their products.

Both brands have a strong international reputation and enjoy the trust of users. They continue to develop and improve their brands to remain innovative and successful in the market. *(магистр.)*

Пример (6) иллюстрирует фрагмент с нумерованным списком из четырех пунктов (мы приводим только два). Примечательно, что, хотя в каждом пункте цитируются источники, часть из них (*Connolly & Hanson, 2016; Korhonen et al., 2018*), отсутствуют в списке литературы. Заголовки демонстрируют безошибочное использование заглавных букв в ключевых словах (в соответ-

ствии с нормами международного научного дискурса), что нехарактерно для русскоязычных авторов.

(6)

The identification of research gaps has shown that small business is underrepresented as an object of systemic analysis. It is necessary to focus on the following aspects:

1. Impact on Financial Access. Sanctions targeting Russia's financial institutions have constrained the availability of foreign investment and credit for small enterprises (Åslund, 2015). The disconnection from SWIFT and the freeze on foreign assets have led to increased borrowing costs and limited funding options for small businesses. Studies indicate that while large corporations can leverage internal reserves or state-backed loans, small firms face heightened financial uncertainty (Korhonen et al., 2018).

2. Operational and Supply Chain Challenges. International restrictions on technology transfers and imports have disrupted operational capacities, especially for firms reliant on foreign equipment and materials (Dreger et al., 2016). Supply chain disruptions have forced companies to seek alternative suppliers from Asia and Latin America, often at higher costs and lower quality (Belkin, 2021). Small manufacturers, for instance, have struggled with procuring essential components, leading to production delays and cost inflation (Connolly & Hanson, 2016). (бакал.)

4. Включение обобщающего заключительного параграфа. В заключение обзора источников по одному аспекту темы либо в конце всего раздела включается краткий параграф из 1–2 предложений, который обобщает содержание источников или всего раздела. Параграф открывается типичными оборотами *in summary*, *in sum*, *in conclusion*. Выводы носят максимально обобщенный характер.

Данный признак проявился в восьми работах (6 бакалавров и 2 магистрантов) (17 %). В **примере (7)** представлен параграф, который автор приводит после анализа ряда источников по одному из аспектов исследования на тему “ESG Narrative in the North American Energy Industry” («ESG-повестка в энергетической отрасли Северной Америки»). Как можно видеть, данное обобщение содержит абстрактные формулировки (*non-insignificant intersection*, *evolving landscape*, *emerging expectations*, *nuanced conclusions* и *this critical domain*), которые не раскрывают конкретных фактов и выводов по рассмотренным источникам, а лишь упоминают общие аспекты темы — некие ожидания, вызовы и роль инвестиционных консультантов. Такое обобщение скорее направлено на создание впечатления о способности автора оценивать вклад источников в разработку проблемы. Текст изобилует шаблонными словами и оборотами из языка ChatGPT (подчеркнуты и выделены курсивом).

(7)

In conclusion, there exists a non-insignificant intersection between fiduciary duty and responsible investment practices. This intersection underscores the evolving landscape, emerging expectations for fiduciaries, persistent implementation challenges, and the crucial role of investment consultants in shaping responsible investment practices. The nuanced conclusions drawn from these insights contribute to a comprehensive understanding of the dynamics within this critical domain. (магистр.)

В **примере (8)** приведен обобщающий параграф, который автор поместил в заключение к обзору литературы. В нем перечисляются ранее рассмотренные

источники (Berkhout, Marcotullio, Hanaoka; Smith; Kühne; Rabbi). Однако, как и в предыдущем случае, формулировки остаются бессодержательными, что подчеркивается многократным повторением существительных во множественном числе, обозначающих абстрактные понятия и процессы (*various aspects and challenges, multiple dimensions, unique insights and perspectives, complexities and necessities*), и других частотных единиц из языка ChatGPT (подчеркнуты и выделены курсивом). Описания фокуса каждого источника максимально сжаты. В них воспроизводятся ключевые слова из названий данных источников в списке литературы.

(8)

In summary, these scholarly articles collectively form a comprehensive body of work that elucidates the various aspects and challenges of transitioning to renewable energy. From Berkhout, Marcotullio, and Hanaoka's integrated approach to energy transition, Smith's critical analysis of the MLP, Kühne's conflict resolution framework, Rabbi's focus on energy security in the context of carbon neutrality, to Pietrzak's detailed examination of energy transformation in the EU, this literature review provides a holistic understanding of the multiple dimensions involved in the global shift towards sustainable energy systems. Each article contributes unique insights and perspectives, enriching the overall understanding of the complexities and necessities of the energy transition. (бакал.)

5. Завершение обзора литературы формулировкой рекомендаций для будущих исследователей. Финальный параграф часто указывает на лакун(ы) в исследовании проблемы и завершается рекомендациями для будущих исследователей (*Future research should continue to explore/should focus on ...*) либо практическими рекомендациями. Рекомендации могут предлагаться списком в отдельном параграфе (например, под заголовком “Future Directions”).

Десять работ студентов-бакалавров (21 %) в корпусе содержали рекомендации для проведения будущих исследований в заключительном параграфе, что нетипично для традиционных текстов описываемого жанра, но было отмечено нами как отличительная характеристика организации текста в 53 % искусственно сгенерированных обзоров литературы. В примере (9) приведен пример такого параграфа из студенческой работы. Рекомендации приводятся в последнем предложении параграфа и включают типичные формулировки, присутствующие в таких параграфах в сгенерированных текстах (*Future studies could greatly benefit from ...*).

(9)

In sum, the existing literature on Turkey-Iran relations during Erdogan's presidency paints a comprehensive picture of a bilateral relationship marked by geopolitical maneuvering, economic collaborations, challenges posed by international sanctions, and the impact of domestic political dynamics. Despite the wealth of research available, there remains an imperative for further investigations that weave together these diverse threads to furnish a more holistic understanding of the forces shaping Turkey's diplomatic and economic engagements with Iran. Future studies could greatly benefit from adopting an interdisciplinary approach that merges perspectives from international relations, economics, and political science, thereby enriching our grasp of the multifaceted nature of Turkey's policies towards Iran. (бакал.)

На уровне языкового оформления текста

1. Частотное использование слов и оборотов из языка ChatGPT.

В тексте с несбалансированно высокой частотностью используется абстрактная академическая лексика продвинутого уровня сложности, которая подрывает уникальность авторского стиля письма. Использование такой лексики может резко контрастировать с реальным уровнем владения ИЯ автором, демонстрируемым им без опоры на вспомогательные материалы и технические средства. Могут часто встречаться слова/обороты в форме множественного числа, передающие обобщенные и многоаспектные понятия и явления (*global impacts, emerging landscapes, distinct emphases, linguistic diversities*), и прилагательные, подчеркивающие важность проблемы исследования и рассматриваемых понятий (*critical, pivotal, crucial, essential*). Отдельные слова и обороты могут иметь поэтический окрас (*contours, fabric, tapestry, finesse*), придающий академическому стилю повествования не всегда уместную выразительность.

Данный признак многократно проявился в 16 работах (34 %) из корпуса. В целом авторами было использовано 46 различных единиц из списка слов и оборотов ChatGPT (Приложение 2) с различной степенью частотности (от 1 до 7 раз). Так, наречие *collectively* было использовано одним автором 7 раз в сочетании с глаголами *underscore* (6) и *highlight* (1). Общее количество упоминаний этих единиц в 16 работах составило 412 (228 в работах магистрантов, 184 — бакалавров). Магистранты использовали от 8 до 27 разных единиц (в среднем 14 на человека), бакалавры — от 6 до 19 (в среднем 13 единиц). Самыми частотными единицами (с 10 и более упоминаниями) оказались: *underscore* (27 упоминаний), *nuanced* (23), *dynamics* (20), *critical* (19), *complex* (14), *landscape* (14), *navigate* (14), *multifaceted* (14), *unique* (13), *crucial* (12), *intricate* (12), *pivotal* (12), *encompass* (10), *evolving* (10) и *transitions* (10).

В примерах (10–11) встречаются единицы из языка ChatGPT (Приложение 2), включая слова, обозначающие остроту явления (подчеркнуты и выделены курсивом). Такие единицы пронумерованы для иллюстрации частотности их использования в небольшом сегменте текста. В первом фрагменте отмечено 11 единиц, причем *complex* встречается два раза. Во втором фрагменте таких единиц 8. Некоторые единицы (*against the backdrop of, deliberate equilibrium, delicate act of balancing between, mold diplomatic posture*) имеют поэтический окрас, не вполне соотносящийся с нормами академического стиля. При всей «выразительности», торжественности слога, достигаемой при использовании таких единиц, они передают минимум конкретики и требуют пояснений.

(10)

In the *complex*¹ socio-political *landscape*² of Balochistan, language plays a very important role, acting as a force that closely shapes the essence of Baloch youth identity. Faced with adversities and discriminatory forces, these resilient individuals wield language as a *multifaceted*³ tool of resistance, empowering them to assert their cultural heritage against the backdrop of oppression (Baloch, 2019). The significance of Balochi goes beyond linguistic expression; it acts as a symbol of empowerment, providing

a profound means for cultural affirmation and contributing significantly to the *complex*⁴ process of identity formation. Balochi becomes more than just a language; it serves as a cultural vault, a medium through which Baloch youth express their *unique*⁵ identity amidst a linguistic diversity. ...

...

Within this linguistic diversity, Baloch youth strategically go through the *nuances*⁶ of Balochi, Urdu, and English. Balochi remains a *vital*⁷ tool for cultural expression, while Urdu and English serve *broader*⁸ communicative purposes, creating a deliberate equilibrium in their linguistic preferences. This conscious *navigation*⁹ reflects the effort to *balance*¹⁰ tradition and modernity, heritage and progress, demonstrating the *complexity*¹¹ of their identity construction (Baloch, 2018). Balochi is not merely a language; it is a cultural anchor, a source of pride, and a mean of expressing identity in a rapidly changing world. (*магистр*.)

(11)

The ramifications of international sanctions on Iran, alongside Turkey's strategies to *navigate*¹ these constraints, constitute another *pivotal*² theme in the literature. Research by experts like Lindenstrauss (2018) *delves into*³ the intricacies of Turkey's stance as it seeks to harmonize its economic interests with Iran against its obligations under international sanctions regimes. These studies shed light on the diplomatic finesse Turkey employs to alleviate the negative *impacts*⁴ of sanctions on its bilateral ties with Iran, offering *critical*⁵ insights into the delicate act of balancing between adherence to global norms and pursuit of national interests.

Moreover, the influence of domestic politics in shaping Turkey's foreign policy towards Iran receives considerable attention in scholarly discussions. Works by entities such as the International Crisis Group (2016) dissect how Erdogan's political ideology and domestic policy imperatives mold Turkey's diplomatic posture towards Iran. These analyses reveal the profound *interplay*⁶ between Turkey's internal political *landscape*⁷ and its external diplomatic endeavors, *underscoring*⁸ how domestic considerations can significantly sway foreign policy orientation ... (*бакал.*)

Необходимо отметить, что, безусловно, данные слова и обороты (например, *nuanced understanding*, *complex interplay* и *underscore*) имеют полное право на существование и широко используются в англоязычных научных работах. Однако аномально частотное использование данных оборотов в пределах одного текста студентами (зачастую без понимания значений и смысловых нюансов их употребления) свидетельствует о недостаточном понимании авторами того, как такой текст воспринимается читателем. К сожалению, если преподаватель не знаком с реальным языковым потенциалом студента объективно выявить машинную генерацию текста только по этому признаку может быть сложно. Эту проблему может решить лишь последующая оценка иноязычной речи студента в условиях взаимодействия с ним при отсутствии у него доступа к вспомогательным материалам (например, во время защиты проекта). Наш опыт взаимодействия с авторами этих фрагментов в рамках курса АП (в случае с магистрантами) и на защитах проектов (у бакалавров) показал, что их академическая речь не соответствует тому уровню владения сложными языковыми средствами, который демонстрируют их тексты. В устных ответах студентов можно было отметить ограниченный запас академической лексики, обилие грамматических ошибок (в том числе в построении разных типов во-

просов), ошибки в словообразовании, подчас неверный выбор средств связности, а главное — незнание значений тех слов и оборотов из языка ChatGPT (*complexities, interplay, nuanced*), которые они с аномальной частотностью использовали в своих работах.

2. Безошибочное использование сложных синтаксических и грамматических конструкций и артиклей. Текст изобилует многосоставными (сложносочиненными и сложноподчиненными) предложениями, сложными грамматическими оборотами и конструкциями (определительными и причастными оборотами, оборотами с инфинитивом и герундием, расщепленными инфинитивами и пр.), а также демонстрирует безупречное владение артиклями (характерное только для образованных носителей или очень компетентных пользователей АЯ).

Проявления этого признака были отмечены в 16 работах — 8 магистрантов и 8 бакалавров (34 %). В **примерах (12–13)** даны два фрагмента (авторы которых не отличаются высоким уровнем владения АЯ), в которых демонстрируется безукоризненное использование сложных по синтаксической структуре предложений, включающих обороты с герундием и расщепленным инфинитивом (выделены курсивом), и артиклей (выделены жирным шрифтом, перед нулевыми артиклями добавлено нижнее подчеркивание).

(12)

The analysis of stock market behavior, particularly **the** impact of human behavior and technological advancements, provides **a** critical backdrop for understanding **the** intricate dynamics of financial markets. *Building upon this foundation, the* following research endeavors to explore **the** influence of private investor sentiment on **the** Russian stock market. *By utilizing* machine learning models to classify text data and *considering* various contributing factors, **the** study seeks to address **the** central question: how does private investor sentiment shape **the** landscape of **the** Russian stock market? (бакал.)

(13)

To comprehensively explore **the** factors contributing to literacy development, *it is imperative to understand* **the** intricate relationship between language socialization and children's literacy skills, *encompassing* reading, writing, and comprehension abilities. (магистр.)

3. Безошибочное соблюдение правил англоязычной пунктуации. 16 работ (8 магистрантов и 8 бакалавров) (34 %) демонстрировали безошибочное соблюдение правил англоязычной пунктуации — аспекта письменной речи, который вызывает затруднения у многих неносителей АЯ. **Примеры (14–16)** наглядно иллюстрируют, как умело авторы используют запятые после вводных фраз, состоящих из трех и более слов (*Within today's landscape, ...* — пример 14), выделяют запятыми уточняющие обороты (*a comparative analysis, particularly of chapters 8–10 of "The Old Book of Tang" and the corresponding chapters of "The New Book of Tang"*, — пример 15), применяют «оксфордскую запятую» перед последним элементом перечисления (*shaped, revised, and reinterpreted* — пример 15), а также ставят точку с запятой для разделения частей сложного предложения, где вторая часть поясняет первую (*The significance of Balochi goes beyond*

linguistic expression; it acts as a symbol of empowerment ... — пример 16). Эти и другие правила англоязычной пунктуации редко изучаются в российских программах обучения АЯ даже на вузовской ступени образования.

(14)

Within today's landscape, social media platforms and recommendation web services form an integral part of everyday life and serve as an effective tool for fashion retail and publishing entities. (бакал.)

(15)

By conducting a comparative analysis, particularly of chapters 8-10 of "The Old Book of Tang" and the corresponding chapters of "The New Book of Tang", this research aims to unravel the complexities of historiography and to gain a deeper understanding of how historical narratives are shaped, revised, and reinterpreted over time. (бакал.)

(16)

The significance of Balochi goes beyond linguistic expression; it acts as a symbol of empowerment, providing a profound means for cultural affirmation and contributing significantly to the complex process of identity formation. (магистр.)

Обсуждение результатов

Проведенное исследование позволило выявить десять признаков машинной генерации текстов в англоязычных обзорах литературы, подготовленных студентами бакалавриата и магистратуры. Некоторые из этих признаков (цитирование несуществующих источников, включение неточных описаний источников, представление информации в виде списков, безошибочное использование пунктуации) соотносятся с признаками машинной генерации текста, ранее зафиксированными в искусственно сгенерированных русскоязычных научных текстах (Черкасова и Тактарова, 2024). Присутствие каждого из этих признаков свидетельствует о некритичном отношении авторов к предлагаемому ИИ контенту и их готовности интегрировать искусственно сгенерированный текст без его критического осмысления и адаптации под индивидуальный стиль, уровень владения ИЯ и задачи исследования. Данные результаты подтверждают опасения относительно негативного влияния ИИ на процессы организации научной деятельности (Dergaa et al., 2023; Else, 2023; Rahime, Abadi, 2023) и рисков утраты студентами навыков самостоятельной работы, рефлексии и критического мышления (Лопанова, Савина, 2024; Choi et al., 2023). Они также подчеркивают необходимость пересмотра существующих подходов к обучению написанию научных текстов в сложившихся условиях.

В свете полученных результатов предлагается **ряд практических рекомендаций** для оптимизации подходов к подготовке студентов к написанию англоязычных научных работ (прежде всего обзоров литературы) в условиях масштабного внедрения ИИ.

1. Необходимо развивать у обучающихся умения критической оценки возможностей и ограничений ИИ, а также качества сгенерированного контента. Это предполагает позиционирование ИИ как вспомогательного инструмента, полезного на отдельных стадиях работы (например, на стадии генерации идей, составления плана текста, поиска потенциальных источников), но

не как источника готовых решений (в том числе полностью сгенерированных текстов), заменяющих самостоятельный поиск и анализ информации и подготовку авторского текста. Формированию умений критического отношения к ИИ может способствовать включение в учебный процесс таких заданий, как: 1) сопоставление обзоров литературы в высокорейтинговых статьях (опубликованных до появления генеративного ИИ) и сгенерированных обзоров литературы по темам этих статей (с точки зрения отбора источников по конкретным аспектам проблемы, глубины их анализа, аргументации и логической организации текста); 2) анализ достоинств (например, соблюдение академического стиля письма) и недостатков искусственно сгенерированных обзоров литературы на отобранных примерах (например, включение несуществующих источников, отсутствие глубины в анализе источников, чрезмерно дробное деление текста, нарушения взаимосвязи между частями текста); 3) разбор анонимизированных студенческих работ с признаками ИИ, нацеленный на выявление этих признаков. В этой связи особое значение приобретает развитие у студентов навыков проверки достоверности сгенерированной информации (особенно ссылок и приводимых фактов) с помощью научных и поисковых баз данных и оценки ее релевантности в контексте проводимого исследования. Этой цели могут служить аудиторные или внеаудиторные задания на перекрестную проверку студентами ссылок в работах друг друга, например, в формате взаимного рецензирования (peer review) работ до их сдачи преподавателю.

2. В контексте обучения ИЯ важно формировать у студентов умения критической оценки языкового качества сгенерированных текстов (кажущееся «идеальным» только по формальным признакам, но не являющееся таковым) и поощрять их к формированию индивидуального стиля письма. Это подразумевает целенаправленное обучение студентов распознаванию языковых аномалий в сгенерированных текстах (например, чрезмерно частотном использовании абстрактной лексики и существительных во множественной форме, сложных синтаксических и грамматических структур и пунктуации, которые могут не соответствовать уровню владения языком студентов), которые подрывают уникальность авторского стиля и могут свидетельствовать о машинном происхождении текстов. В этом процессе представляется полезным ознакомление студентов с представленным в данном исследовании списком слов и оборотов из языка ChatGPT, многократное повторение которых может не только производить негативное впечатление на читателя, но и выдавать машинное происхождение текстов. Включение заданий на выявление в сгенерированных текстах таких оборотов и замены их на альтернативные может способствовать расширению словарного запаса авторов. Студентам также может быть рекомендовано использовать возможности нейросетей для составления индивидуальных списков научных оборотов (с примерами их употребления в текстах в той или иной предметной области) с целью их дальнейшего осознанного применения в авторском тексте.

3. В целях развития осознанного подхода к познавательной деятельности представляется целесообразным поощрять студентов к ведению подробной документации и анализу собственного процесса подготовки научного текста на всех или отдельных его этапах — от генерации идей и составления

плана раздела (в том числе с использованием ИИ) до поиска и отбора релевантных источников по теме (с указанием используемых баз данных, ключевых слов и критериев поиска), написания комментариев к источникам и создания черновиков текста. Данный процесс может сопровождаться анализом степени и эффективности применения ИИ в решении тех или иных задач. Результаты такой саморефлексии могут фиксироваться в дневнике или портфолио и затем демонстрироваться в классе, например, с приложением скриншотов, подтверждающих процесс самостоятельного поиска литературы в библиографических базах, и текстов найденных публикаций.

4. Особое внимание следует уделять обсуждению этических аспектов использования ИИ при выполнении различных учебных и научных задач. Для достижения этой цели рекомендуются тренинги и задания, предполагающие устные обсуждения и дебаты относительно допустимых и недопустимых форм использования ИИ и последствий нарушения этических норм, а также анализ релевантности привлечения ИИ для решения конкретных задач. Например, студентам может быть предложено рассмотреть примеры работ с разным уровнем интеграции ИИ и обсудить, какие из них соответствуют этическим нормам, какие — нет. Кроме того, для студентов может представлять интерес практика совместного составления правил этического использования ИИ в рамках конкретного курса или работы над заданием (вместо «навязанных» преподавателем или университетом правил), которая может содействовать развитию личной ответственности за соблюдение этих правил.

5. В целях развития у студентов самостоятельного мышления и индивидуального стиля письма необходимо регулярно вовлекать их в практику самостоятельного создания текстов (например, в классе) без использования каких-либо вспомогательных средств, включая ИИ. В качестве примера задания можно предложить им написать черновой вариант фрагмента обзора литературы по определенному аспекту темы, в котором требуется синтезировать информацию из нескольких источников на основе ранее сделанных по ним заметкам.

6. Наконец, для педагогов крайне важно повышать квалификацию в области применения ИИ, что позволит им научиться адекватно оценивать возможности и ограничения разных средств ИИ в решении конкретных учебно-познавательных задач, освоить стратегии эффективного взаимодействия с ИИ, а также развить умения выбора нестандартных форм работы и заданий, снижающих риск злоупотребления ИИ студентами. Например, в этих целях могут быть организованы тренинги, посвященные обсуждению педагогических стратегий и методов, способствующих развитию у студентов критического и ответственного отношения к использованию ИИ в соответствии со стоящими перед ними задачами.

Ограничения исследования. Представленное в настоящем исследовании описание признаков машинной генерации призвано повысить осведомленность научных работников и педагогов (особенно тех, кто ведет дисциплину «академическое письмо» на АЯ) о существующих формах недобросовестного применения средств ИИ при подготовке англоязычных научных текстов. Вместе с тем следует признать, что описанные в работе признаки были выявлены на материале текстов только одного жанра научной прозы (обзоров литературы в англо-

язычных проектах научного исследования), на ограниченной выборке работ и с учетом текущих возможностей и ограничений нейросети ChatGPT-3.5. Хотя ChatGPT на настоящий момент является одной из наиболее распространенных и продвинутых языковых моделей, можно предположить, что представленные нами признаки машинной генерации не отражают все возможные формы недобросовестного применения ИИ при написании англоязычных научных работ. Требуются дальнейшие исследования, которые позволят расширить представления о данных формах как в рамках представленного жанра, так и других жанров научной прозы на больших выборках работ, в разных контекстах обучения и на примерах других нейросетей.

Заключение

Результаты настоящего исследования подчеркивают остроту проблемы недобросовестного применения ИИ студентами при подготовке англоязычных научных текстов. Выявленные нами признаки машинной генерации текстов свидетельствуют о недостаточно сформированных у некоторых авторов (магистрантов и бакалавров) умениях критического осмысления и объективной оценки предложенного ИИ контента (с характерными для него аномалиями) и низком уровне ответственности за качество собственных текстов, что, в частности, проявилось в слепом копировании чуждых для их уровня и индивидуального стиля языковых конструкций, использовании несуществующих источников и неоправданном дроблении небольшого по объему текста на многочисленные разделы.

Данные результаты указывают на необходимость переосмысления традиционных подходов к обучению написанию и оцениванию работ с целью предотвращения развития у студентов шаблонного мышления, избыточной зависимости от ИИ и снижения рисков академического мошенничества. Необходимо целенаправленно развивать у студентов ответственное и критическое отношение к использованию ИИ в учебной и научной деятельности, а также формировать методологическую и этическую грамотность, позволяющую различать допустимые и недопустимые формы использования ИИ, оценивать последствия нарушения этических норм и проводить четкую границу между авторским и сгенерированным текстом. Представленные нами признаки машинной генерации текста могут быть использованы в качестве критериев экспертной оценки оригинальности и качества англоязычных научных работ (в частности обзоров литературы в проектах научных исследований) в различных образовательных контекстах, что особенно актуально в условиях отсутствия надежных автоматизированных средств для выявления машинного происхождения текстов. Более того, характеристики искусственно сгенерированных англоязычных обзоров литературы и особенности их проявления в студенческих текстах, выявленные в ходе нашего исследования, могут способствовать более глубокому осмыслению и разработке проблемы недобросовестного применения ИИ в научно-познавательной деятельности, а также совершенствованию образовательных программ и методов преподавания иноязычного академического письма в эпоху ИИ.

Список литературы / References

- Арзютова С.Н. Использование ChatGPT в обучении английскому языку // Гуманитарные исследования. Психология и педагогика. 2023. № 16. С. 39–47. <https://doi.org/10.24412/2712-827X-2023-16-39-47> EDN: QQQVED
Arzyutova, S.N. (2023). ChatGPT using in English language teaching. *Humanitarian Studies. Pedagogy and Psychology*, (16), 39–47. (In Russ.) <https://doi.org/10.24412/2712-827X-2023-16-39-47>
- Воронин А.Н., Палёнова В.В. Будущее психологии: эффективное взаимодействие с ChatGPT возможно? // Вестник РУДН. Серия: Психология и педагогика. 2024. Т. 21. № 3. С. 831–857. <https://doi.org/10.22363/2313-1683-2024-21-3-831-857> EDN: GUPGVY
Voronin, A.N., & Palenova, V.V. (2024). The future of psychology: Is effective interaction with ChatGPT possible? *RUDN Journal of Psychology and Pedagogics*, 21(3), 831–857. (In Russ.) <https://doi.org/10.22363/2313-1683-2024-21-3-831-857>
- Лопанова Е.В., Савина Н.В. К проблеме использования нейросетей в учебной деятельности студентов // Пространство педагогических исследований. 2024. Т. 1. № 1. С. 23–40. <https://doi.org/10.23859/3034-1760.2024.33.52.002> EDN: XNOWWH
Lopanova E.V., & Savina, N.V. (2024). On the problem of using neural networks in students' educational activities. *Education Research Environment*, 1(1), 23–40. (In Russ.) <https://doi.org/10.23859/3034-1760.2024.33.52.002>
- Черкасова М.Н., Тактарова А.В. Искусственно сгенерированный академический текст (лингвопрагматический аспект) // Филологические науки. Вопросы теории и практики. 2024. Т. 17. № 7. С. 2551–2557. <https://doi.org/10.30853/phil20240363> EDN: YYWRDN
Cherkasova, M.N., & Taktarova, A.V. (2024). Artificially generated academic text (a linguopragmatic aspect). *Philology. Theory & Practice*, 17(7), 2551–2557. (In Russ.) <https://doi.org/10.30853/phil20240363>
- ChatGPT Generative Pre-trained Transformer, C., & Zhavoronkov, A. (2022). Rapamycin in the context of Pascal's Wager: Generative pre-trained transformer perspective. *Oncoscience*, 9, 82–84. <https://doi.org/10.18632/oncoscience.571>
- Choi, E.P.H., Lee, J.J., Ho, M.-H., Kwok, J.Y.Y., & Lok, K.Y.W. (2023). Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. *Nurse Education Today*, 125, 105796. <https://doi.org/10.1016/j.nedt.2023.105796>
- Dalgıç, A., Yaşar, E., & Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 34, 100481. <https://doi.org/10.1016/j.jhlste.2024.100481>
- Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615–622. <https://doi.org/10.5114/biolsport.2023.125623>
- Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, 613(7944), 423–423. <https://doi.org/10.1038/d41586-023-00056-7>
- Gao, C.A., Howard, F.M., Markov, N.S., Dyer, E.S., Ramesh, S., Luo, Y., & Alexander T. Pearson, A.T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine*, 6, 75. <https://doi.org/10.1038/s41746-023-00819-6>
- González, R., Poenaru, D., Woo, R., Trappey, A. F., Carter, S., Darcy, D., Encisco, E., Gulack, B., Miniati, D., Tombash, E., & Huang, E.Y. (2024). ChatGPT: What every pediatric surgeon should know about its potential uses and pitfalls. *Journal of Pediatric Surgery*, 59(5), 941–947. <https://doi.org/10.1016/j.jpedsurg.2024.01.007>
- Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research*, 13(4), 1148–1154.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education*, 12(1), 261–270. <https://doi.org/10.5334/pme.1072>

Pavlik, J.V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 78(1), 84–93. <https://doi.org/10.1177/10776958221149577>

Rafiq, S., Nawaz, A., & Afzal, A. (2025). The future of AI in academic writing: A case study of undergraduate and postgraduate assessments. *Dialogue Social Science Review*, 3(1), 1280–1297.

Rahimi, F., & Abadi, A.T.B. (2023). ChatGPT and publication ethics. *Archives of Medical Research*, 54(3), 272–274. <https://doi.org/10.1016/j.arcmed.2023.03.004>

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11(6), 887. <https://doi.org/10.3390/healthcare11060887>

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613(7945), 620–621. <https://doi.org/10.1038/d41586-023-00107-z>

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), 26. <https://doi.org/10.1007/s40979-023-00146-z>

ПРИЛОЖЕНИЕ 1 / APPENDIX 1

Критерии оценивания англоязычного обзора литературы /
Literature review assessment criteria

Content	Points
1. Discussion of the literature reflects an ability to draw critical connections among material from various sources in developing a line of argument (<i>through discussing findings from individual studies/studies as groups; comparing/contrasting diverse perspectives; highlighting limitations of some aspects of prior research where necessary, etc.</i>).	/20
2. The choice of sources is appropriate throughout (at least 10 relevant, current, and authoritative sources are used).	/10
3. To support an argument, evidence from sources is integrated effectively using quoting, paraphrasing, summarizing, and synthesizing strategies (if used, direct quotes are kept to a minimum of 1-3).	/10
Organization	
4. Paragraph division is effective and demonstrates a logical flow of ideas from one issue to the next.	/5
5. Each paragraph has a clear focus and demonstrates a logical progression of ideas.	/5
6. Appropriate (formal) cohesive devices are used to connect ideas within and between paragraphs.	/5
Language	
7. Move-specific grammar and vocabulary structures are varied and appropriate to the author's communicative intentions in the Literature Review section.	/10
8. The use of tenses and verb forms (active and passive) is aligned to specific communicative functions conveyed in this section.	/5
9. Hedging is used to counterbalance any biased or overly strong claims.	/5
10. The review is free of vocabulary, grammar, word order, spelling, style, and punctuation errors.	/15
Format	
11. All cited sources are acknowledged properly in the in-text references and the reference list following the APA style conventions.	/10
Total: ____/100 points	

ПРИЛОЖЕНИЕ 2 / APPENDIX 2

Список частотных оборотов из языка ChatGPT / List of frequently occurring ChatGPT lexical items

Ключевое слово / Key word	Частотность употребления / Frequency	Частые коллокации* / Frequent collocates*
Существительные / Nouns		
dynamics	52	complex, evolving, operational
landscape	44	evolving, complex, global
dimension	33	critical, various, holistic
complexity	29	of
resilience	24	demonstrate, build
impacts	23	broader, negative
interplay	20	dynamic, complex
complexities	15	navigate
realities	15	multifaceted, operational
transitions	12	smooth, sustainable
nuances	11	cultural, cognitive, contextual
Прилагательные / Adjectives		
critical	122	role, insights, dimension, area, perspectives
complex	65	interplay, dynamics, ways, landscape
essential	60	for, to; component, feature
crucial	54	for, to; role
comprehensive	51	understanding, analysis, overview, view
broader	49	analyses, factors, structures, discourse, literature
dynamic	43	nature, interplay, arena
evolving	42	nature, landscape, dynamics, framework
multifaceted	36	nature, approach, role, realities
unique	35	challenges, opportunities, factors
future	29	research, studies
nuanced	25	understanding, challenges, conclusions
pivotal	25	role, theme
vital	19	for, in; role
fundamental	15	to; element, shifts
holistic	13	understanding, approach, strategies
intricate	12	relationship, nature, interplay, issues
central	12	to; role, theme
balanced	11	approach, view
further	10	investigation(s), research, studies
persistent	10	challenges
tailored	10	approaches, strategies, solutions, choices
Глаголы / Verbs		
underscore	81	importance, the need
leverage	51	relationships, resources; X and X
navigate	44	landscape, challenges, nuances, complexities; between
balance	37	X with X; X and X
encompass	27	processes, practices, identities
delve into	22	dynamics, analysis, intricacies, dimensions
persist	13	challenges, divides
elucidate	10	aspects, shifts
transcend	10	mere X; barriers, divides, divisions
Наречия / Adverbs		
critically	21	examine, discuss, analyze
collectively	11	underscore

*Коллокации представлены по убывающей частотности их употребления в корпусе (слева направо).

*Collocates are presented in descending order of their frequency of use in the corpus (from left to right).

История статьи:

Поступила в редакцию 3 апреля 2024 г.

Доработана после рецензирования 25 июля 2024 г.

Принята к печати 29 июля 2024 г.

Для цитирования:

Дугарцыренова В.А. ChatGPT в англоязычном академическом письме: признаки машинной генерации текста в обзорах литературы студентов бакалавриата и магистратуры // Вестник Российского университета дружбы народов. Серия: Психология и педагогика. 2025. Т. 22. № 1. С. 144–174. <http://doi.org/10.22363/2313-1683-2025-22-1-144-174>

Заявление о конфликте интересов:

Автор заявляет об отсутствии конфликта интересов.

Сведения об авторе:

Дугарцыренова Вера Аркадьевна, кандидат педагогических наук, доцент Школы иностранных языков Национального исследовательского университета «Высшая школа экономики», Российская Федерация, 101000, Москва, ул. Мясницкая, д. 20. ORCID: 0000-0002-7476-7461; eLibrary SPIN-код: 9368-4699. E-mail: vdugartsyrenova@hse.ru

DOI: 10.22363/2313-1683-2025-22-1-144-174

EDN: UHNYIW

UDC 372.881.1

Research article

ChatGPT Presence in Academic Writing: Detecting AI-Generated Text in Undergraduate and Graduate Students' Research Proposal Literature Reviews

Vera A. Dugartsyrenova 

National Research University Higher School of Economics, Moscow, Russian Federation

[✉vdugartsyrenova@hse.ru](mailto:vdugartsyrenova@hse.ru)

Abstract. With the rapid development of artificial intelligence (AI) tools, concerns emerge regarding students' unethical uses of these tools to produce AI-generated research texts or their parts, and to present them as original writing. This issue is compounded by the lack of reliable tools for detecting machine-generated text. To address these concerns, the present study aimed to identify distinctive features of ChatGPT-generated research proposal literature reviews ($N = 45$) and investigate the presence of these features in English-language literature reviews produced by undergraduate and graduate students from two Russian universities. During the first stage, an analysis of AI-generated texts and a small sample of graduate students' ($N = 12$) literature reviews was conducted. Findings revealed that many features typical of AI-generated texts were clearly present in student texts suggesting that these features may serve as indicators of machine-generated writing. One such feature was the unusually high recurrence of lexical items (predominantly with abstract meanings) in both AI-generated and student texts. Drawing on these insights, a frequency analysis was performed using AntConc to explore the occurrence of these items in AI-generated texts and compile a list of the most frequent items indicative

of machine-generated writing (referred to in this study as “ChatGPT language”). At the second stage, findings on the initial indicators were validated, refined, and expanded based on an analysis of a larger sample of 47 English language literature reviews prepared by bachelor and master students. The study identified ten indicators of AI-generated writing pertaining to content, structure, and language use in literature reviews, which are detailed and illustrated in the paper. The study’s findings contribute valuable practical and research insights which may aid all those involved in teaching English language academic writing, reviewing students’ academic texts, and supervising research projects across diverse EAP contexts.

Key words: artificial intelligence, ChatGPT-3.5, L2 academic writing, English language literature reviews to research proposals, AI-generated literature reviews, indicators of machine-generated writing

Article history:

Received 3 April 2024

Revised 25 July 2024

Accepted 29 July 2024

For citation:

Dugartsyrenova, V.A. (2025). ChatGPT presence in academic writing: Detecting AI-generated text in undergraduate and graduate students’ research proposal literature reviews. *RUDN Journal of Psychology and Pedagogics*, 22(1), 144–174. (In Russ.) <http://doi.org/10.22363/2313-1683-2025-22-1-144-174>

Conflicts of interest:

The author declares that there is no conflict of interest.

Bio note:

Vera A. Dugartsyrenova, PhD in Pedagogics, is Associate Professor at the School of Foreign Languages, National Research University Higher School of Economics (20 Myasnitskaya St, Moscow, 101100, Russian Federation). ORCID: 0000-0002-7476-7461; eLibrary SPIN-code: 9368-4699. E-mail: vdugartsyrenova@hse.ru