

DOI: 10.31862/2500-2953-2025-3-81-102

УДК 372.881.111.1:(004.8+371.261)

**С.В. Боголепова, А.А. Греченко**

Национальный исследовательский университет  
«Высшая школа экономики»,  
101000 г. Москва, Российская Федерация

## Возможности генеративного искусственного интеллекта для проверки академических текстов на английском языке

Целью работы является определение потенциала генеративного искусственного интеллекта для оценки академических текстов на основе критериев, учитывающих специфику англоязычного академического дискурса. 20 текстов в формате Project Proposal были оценены как опытным преподавателем, так и генеративным искусственным интеллектом. Выставленные баллы были проанализированы статистически, а обратная связь – тематически. Выявлено, что искусственный интеллект способен выявлять сильные и слабые стороны студенческих работ и генерировать обратную связь на основе критериев; однако уровень согласия между баллами искусственного интеллекта и эксперта оказался низким, а форматы предоставления обратной связи заметно отличались. Внедрение искусственного интеллекта может помочь стандартизировать оценивание и провести учебную аналитику, однако для обеспечения надежности оценивания и индивидуализации обратной связи необходима экспертность преподавателя.

**Ключевые слова:** англоязычный академический дискурс, искусственный интеллект в обучении языкам, генеративный искусственный интеллект в образовании, оценочная обратная связь

© Боголепова С.В., Греченко А.А., 2025



Контент доступен по лицензии Creative Commons Attribution 4.0 International License  
The content is licensed under a Creative Commons Attribution 4.0 International License

ДЛЯ ЦИТИРОВАНИЯ: Боголепова С.В., Греченко А.А. Возможности генеративного искусственного интеллекта для проверки академических текстов на английском языке // Рема. Rhema. 2025. № 3. С. 81–102. DOI: 10.31862/2500-2953-2025-3-81-102

DOI: 10.31862/2500-2953-2025-3-81-102

**S. Bogolepova, A. Grechenko**

HSE University,  
Moscow, 101000, Russian Federation

## Using generative artificial intelligence for the evaluation of academic texts in English

The paper aims to assess the potential of generative AI for the evaluation of academic texts based on criteria that account for the features of the Anglo-Saxon academic discourse. Twenty texts in the “Project Proposal” format were evaluated both by an experienced instructor and by generative AI. The scores assigned were subject to statistical analysis, and the feedback was analysed thematically. It was found that AI can identify strengths and weaknesses in students’ works and generate feedback based on the criteria; however, the level of agreement between AI and expert scores was low, and the formats of feedback provision differed significantly. The integration of AI can help standardise assessments and facilitate educational analytics, but to ensure reliable evaluation and personalized feedback, the expertise of an instructor is necessary.

**Key words:** English-language academic discourse, artificial intelligence in language learning, generative artificial intelligence in education, evaluative feedback

FOR CITATION: Bogolepova S., Grechenko A. Using generative artificial intelligence for the evaluation of academic texts in English. *Rhema*. 2025. No. 3. Pp. 81–102. DOI: 10.31862/2500-2953-2025-3-81-102

## 1. Введение

Технологии искусственного интеллекта (ИИ) в настоящее время активно используются участниками образовательного процесса. Генеративный ИИ (генИИ) на основе больших языковых моделей обучен на огромных объемах текстовых данных, что определяет его ценность для изучения и преподавания иностранных языков. Преподаватели иностранных языков и смежных дисциплин используют потенциал генИИ для разработки и адаптации учебных материалов, поурочного и долгосрочного планирования обучения, организации работы обучающихся по развитию языковых навыков и речевых умений [Титова, 2024; Crompton et al., 2024; Mananay, 2024].

ГенИИ имеет несомненный потенциал для содействия преподавателю в проверке письменноречевой продукции обучающихся и обеспечения обратной связи (ОС) по ней [Боголепова, Жаркова, 2024; Сысоев и др., 2024б; Li, Liu, 2024; Mizumoto et al., 2024]. Исследования, продемонстрировавшие этот потенциал, использовали в качестве материала эссе на иностранном языке; при этом возможности генИИ для оценивания академических работ до сих пор остаются неизученными. Англоязычный академический дискурс обладает специфическими чертами, которые учитываются как при обучении написанию таких текстов, так и при их оценивании [Дугарцыренова, 2016].

Данная работа ставит целью изучить возможности генИИ для проверки работ в формате Project Proposal и обеспечения ОС по ним. Сопутствующей задачей является сравнение результатов оценки преподавателем и генИИ, а также формата и содержания ОС от двух проверяющих. Сделаны выводы о возможной интеграции усилий преподавателя и потенциала генИИ.

## 2. Анализ литературы

### 2.1. Особенности академического дискурса

Написание академического текста – сложный процесс, объединяющий умения выстраивать убедительную аргументацию, логично и эффективно организовывать информацию в тексте, адекватно интегрировать ссылки на научную литературу [Chauhan, 2022]. Академическое письмо выходит за рамки использования адекватных языковых средств или стилистических приемов, включая в себя социокультурные и когнитивные аспекты [Ahmed, 2022; Chauhan, 2022]. Исследователи характеризуют академический дискурс как форму коммуникации, целью которой является создание текста, обладающего научной ценностью и значимостью [Ahmed, 2022; Britton, 2023].

Академический язык включает в себя специфическую лексику и клише, что позволяет добиваться точности выражения смыслов и обеспечивать формальность дискурса, а также сигнализировать о принадлежности научному дискурсивному сообществу. Важной чертой академического письма является 'хеджирование' (*hedging*), реализуемое посредством определенных грамматических структур и лексических сочетаний [Nasiri, 2012].

Академический стиль требует последовательной организации текста и следования установленным стандартам [Li, Schmitt, 2009]. Письменный академический дискурс ориентирует читателя посредством дискурсивных маркеров и метатекстовых элементов, позволяющих сделать текст связным и логичным [Farahani, 2022]. В академическом письме абзац признается наименьшей самостоятельной дискурсивной единицей, в связи с этим четкое понимание структуры внутри текста и каждого абзаца имеет критическое значение [Wirantaka, 2016; Бакулев, Боголепова, 2024]. Такие структурные элементы, как введение с описанием проблематики и постановкой исследовательских задач, обзор литературы, интегрирующий работу в соответствующее исследовательское поле, описание результатов работы и соответствующие выводы, традиционно присутствуют в академическом тексте, что облегчает его восприятие и анализ [Ahmed, 2022].

Следование общепринятым стандартам оформления является важным аспектом академического письма, обеспечивающим единообразие и ясность представленной информации. Несмотря на важность этого аспекта, при написании академического текста многие студенты пренебрегают им, что впоследствии приводит к затруднению понимания и интерпретации академического текста [Epp et al., 2021; Бакулев, Боголепова, 2024]. Следование академическим нормам и правилам написания академического текста во многом зависит от научной области, а также от общего уровня подготовки учащихся [Samraj, 2004].

Таким образом, при оценивании письменных академических работ важно обращать внимание не только на языковые особенности дискурса, но и на соответствие требованиям к организации и оформлению текста, его когезию и когерентность, равно как и на соответствие содержания исследовательской работы предписаниям научной области.

## 2.2. Использование генерированного искусственного интеллекта для проверки письменных работ

В целом, генИИ может использоваться для проверки работ и обеспечения оценочной ОС [Сысоев и др., 2024a]. Хотя генИИ имеет потенциал для индивидуализации оценки и автоматизации проверки, все

еще существует ряд трудностей при эксплуатации ИИ для упомянутых целей [Khalifa, Albadawy, 2024; Боголепова, 2025]. Выявлено, что генИИ может достаточно результативно выявлять грамматические и лексические ошибки, орфографические неточности, проблемы со стилем и структурой текста [Solak, 2024]. Однако ключевой сложностью для ИИ все еще остается ограниченная способность к пониманию особенностей контекста и тонкостей формата той или иной письменной работы. В отличие от преподавателя, ИИ не может учитывать особенности авторского стиля, оценивать глубину или оригинальность мысли [Escalante et al., 2023; Jansen et al., 2024; Solak, 2024].

В равной степени важно рассмотреть не только результативность применения генИИ в оценивании, но и отклик обучающихся на такие нововведения. При получении ОС как от преподавателя, так и от ИИ, студенты отмечали более высокое качество ОС от преподавателя ввиду возможности личного взаимодействия и обсуждения [Jansen et al., 2024]. Однако был сделан вывод о том, что смешанный подход имеет потенциал для формулировки ОС, т.к. он интегрирует структурированность и быстроту ОС от ИИ и более глубокое знание контекста и поддержку от преподавателя [Escalante et al., 2023].

Были проанализированы возможности генИИ для рецензирования научных статей. Выявлена высокая степень согласия между оценками, выставленными генИИ и экспертами-рецензентами (75%) [Checco et al., 2021], что может свидетельствовать о применимости генИИ для автоматизированной проверки академических текстов. Авторы также делают вывод о том, что использование ИИ может ускорить процесс анализа и оценивания статей, при этом снижая риск влияния субъективных факторов на итоговую оценку. Более того, ИИ смог успешно идентифицировать серьезные недостатки в работах, например, бездоказательные утверждения или недостаточно глубокий анализ литературы [Tyser et al., 2024]. Вместе с тем исследователи заметили сложности в выявлении специфических проблем, связанных с соблюдением научных норм и стандартов. Для этого требуется более точное понимание специфики научной области и высокий уровень критического мышления, на что на данном этапе технологического развития ИИ не способен.

При работе с ИИ качество сгенерированного материала во многом зависит от четкости, контекстуальной полноты и детальности используемого промпта [Giray, 2023]. В промпте необходимо использовать конкретные глаголы действия, определить роль ИИ, точно задать ожидаемый результат, а также дать генИИ опору или шаблон [Ekin, 2023].

Разные цели требуют разных подходов к составлению промптов [Schulhoff et al., 2024]. В частности, для создания креативного контента

наиболее подходящим будет стратегия промпт-инжиниринга с использованием иллюстрирующих примеров (few-shot prompting). Стратегия формулировки промптов без примеров и иллюстраций (zero-shot prompting) может быть применена для стандартных и повторяющихся сценариев, но она становится менее эффективной, когда появляется необходимость в решении более комплексных задач. Тогда более релевантной будет стратегия использования цепочки рассуждений (chain-of-thought prompting), т.к. ее основная цель – предоставить ИИ пошаговую инструкцию действий для получения самого точного результата [Dang et al., 2022; Schulhoff et al., 2024].

Таким образом, академический дискурс характеризуется рядом особенностей, включая особый стиль изложения, использование специализированной лексики и клише, соблюдение установленных стандартов оформления, а также логическую связность и последовательность в структуре представляемого материала. В современном образовательном процессе активное внедрение технологий ИИ открывает новые возможности для автоматической проверки и оценки академических текстов, однако эти возможности до сих пор не изучены и не оценены.

### 3. Методы

#### 3.1. Материал исследования

Материалом исследования стали 20 студенческих работ в формате Project Proposal, написанные студентами, обучающимися по направлению «Лингвистика» в Национальном исследовательском университете «Высшая школа экономики» (НИУ ВШЭ), в 2023/2024 учебном году. Project Proposal – концепция выпускной квалификационной работы на английском языке, которая имеет определенную структуру и оформляется по стандарту APA. Работы были проверены преподавателем, в течение четырех последних лет ведущим дисциплину «Академическое письмо», по заданным критериям, включающим следующие категории:

1) содержание и использование источников (от 0 до 4 баллов) – в рамках категории оцениваются обоснованность исследовательской ниши пробела, актуальности и цели работы, умения аргументации с опорой на релевантные, актуальные и авторитетные источники, информация из которых интегрирована через цитирование, перифраз и синтез, способность четко описать методы и планируемые результаты;

2) организация и формат (от 0 до 2 баллов) – оценивается когезия и когерентность текста, структура абзаца, использование разнообразных

связующих средств, соответствующих контексту, оформление согласно стилю АРА;

3) языковое оформление (от 0 до 4 баллов) – оценивается использование языковых средств для реализации коммуникативных функций, их разнообразие, соблюдение правил пунктуации и академического (научного) стиля письма.

### 3.2. Оценка работ посредством искусственного интеллекта

Для оценки работ посредством гениев был разработан промпт на основе описанных выше критериев. На первичном этапе его разработки был применен инструмент для оптимизации исходного промпта (<https://promptperfect.jina.ai/interactive>). PromptPerfect не вносит серьезных корректировок, однако помогает избавиться от многословности при составлении промптов и улучшает качество запрашиваемых критериев с языковой точки зрения, что позволяет формулировать итоговый промпт более четко и структурно.

Для получения обратной связи был выбран гений на платформе Talk.AI (<https://talkai.info/chat/>). Пользователи платформы могут без регистрации и ограничений в доступе использовать чат-бот на основе архитектуры GPT, уже демонстрировавший потенциал для обеспечения ОС по студенческим эссе [Боголепова, Жаркова, 2024; Сысоев и др. 2024a].

Посредством разработанного промпта на платформе были оценены первые несколько работ, однако ИИ не обеспечивал достаточно качественную ОС по всем критериям одновременно, что приводило к поверхностным или избыточно детализированным оценкам в отдельных областях без конкретных примеров. По этой причине промпт был реструктурирован на отдельные компоненты, каждый из которых был сфокусирован на конкретном критерии. На следующем этапе разработки была добавлена информация о необходимости оценивания только заданных параметров при выставлении баллов. Например, при оценке критерия «Содержание» ИИ должен был полностью игнорировать лексические и грамматические ошибки, а также ошибки, связанные с логикой и структурой текста. Кроме того, в промпте было отмечено, что итоговый балл по каждому критерию должен быть выражен целым числом, а не дробным (рис. 1).

Благодаря фокусу на каждом критерии по очереди система смогла предоставлять более точные и релевантные комментарии по работе, приводя четкие примеры из академических текстов студентов. Такой подход позволил выявить конкретные сильные и слабые стороны по каждому критерию, что значительно снизило риск упущения важных аспектов и ошибок в анализируемых работах.

Imagine you are an academic writing instructor. Your students wrote a paper called "Project Proposal" in which they review academic literature on a topic stated in the heading and propose how they will conduct research or do a project to achieve the goal and objectives stated in the introduction. Assess the paper based on the criteria below, suggest a grade out of 4 points, and provide detailed feedback ONLY for the aspect "Content", listing all the examples from the paper that illustrate both the strong sides and drawbacks. All mistakes and non-compliance with the criterion "Content and source use" must be listed. Currently, grammatical and lexical mistakes are not analyzed. The score for a criterion must be a whole number; fractional assessments are not permitted.

**The criteria: <...>**

**Рис. 1.** Промпт для формулировки обратной связи на основе критериев (для оценки содержания работы)

**Fig. 1.** Prompt for formulating criteria-based feedback (for assessing the content of the work)

### 3.3. Анализ результатов

По каждой работе были сопоставлены баллы, выставленные преподавателем и ИИ как по отдельным категориям, так и итоговые. Количественный анализ был осуществлен с помощью онлайн-инструмента Source Table (<https://sourcetable.com/>). Были вычислены статистические коэффициенты, позволившие оценить согласованность оценок двух проверяющих.

Текстовые комментарии были проанализированы тематически. Первично выделив основные темы и подтемы с помощью инструмента генИИ (Talk.AI), авторы затем вручную проанализировали ОС, используя цветовую кодировку, в результате чего некоторые темы были добавлены или скорректированы. После этого авторы вторично проанализировали ОС и внесли финальные правки в список основных тем. Наконец, были посчитаны комментарии в соответствии с каждой из тем, что позволило выделить основные акценты в оценочной ОС от генИИ как по отдельным работам, так и по всей исследуемой выборке.

В свою очередь, комментарии от преподавателя по 20 проверенным работам также были собраны и тематически проанализированы. На первой стадии работы с ОС комментарии были вычитаны и распределены по аспектам, заданным в критериях оценивания. Далее был проведен качественный и количественный анализ ОС преподавателя, а также выполнена цветовая кодировка по ранее сформированным темам для дальнейшего сопоставления с комментариями, данными генИИ.



## 4. Результаты

### 4.1. Количественная оценка

Количественный анализ дал неоднозначные результаты. Только 25% работ были оценены одинаково преподавателем и ИИ, но 60% оценок различаются меньше, чем на один балл. Величина каппы Коэна (0,096) сигнализирует о крайне низкой согласованности результатов оценивания. Однако коэффициент Пирсона ( $r = 0,68$ ;  $p = 0,0005$ ) показывает, что в вопросе отнесения работы к определенной категории по качеству преподаватель и генИИ преимущественно совпадают, т.е. если преподаватель оценил работу высоко, генИИ также ставит высокую оценку. Преподаватель в целом имеет тенденцию к выставлению более высоких оценок ( $M = 7,1$ , что на 0,8 больше, чем в среднем выставляет ИИ), а также выставляет оценки в более широком диапазоне (от 4 до 10 баллов; ИИ – от 5 до 7 баллов).

Сравнив баллы по категориям, мы наблюдаем наибольшую согласованность (как систематическую, так и абсолютную) в оценке языкового оформления. По аспекту «Содержание и использование источников» наблюдается наибольшее различие по оценке.

### 4.2. Качественная оценка

ГенИИ формулировал оценочную ОС на основе заданных критериев, выделяя как сильные, так и слабые стороны в работах. ИИ структурировал ОС двумя разными способами: или вычленил отдельные аспекты в критериях оценки и давал комментарий по плюсам и минусам каждого из них, сопровождая утверждения примерами, или по очереди перечислял сильные и слабые стороны внутри каждой крупной категории, также иллюстрируя тезисы примерами (табл. 1 и 2).

Агрегированная ОС от ИИ позволяет определить, что сильными сторонами проанализированных работ являются четкая идентификация исследовательской ниши и постановка исследовательских задач, достаточная контекстуализация работы в исследовательском поле, разделение работы на разделы, каждый из которых выполняет свою функцию, общий академический стиль работы (эти аспекты выделены в более 50% работ).

ИИ, при этом, позволяет выделить и слабые стороны в академических работах студентов, такие как поверхностный анализ и слабость аргументации, перегруженность абзацев и предложений, отсутствие связности между элементами академического дискурса и однообразие средств связи, многочисленные отступления от стандартов оформления (табл. 2).

Таблица 1

Тематический анализ обратной связи от искусственного интеллекта:  
сильные стороны студенческих работ  
[Thematic analysis of feedback by AI: Strengths of students' works]

Аспект [Aspect]	Тематический анализ [Thematic analysis]	Пример [Example]
Обоснование [Justification]	Выявление исследовательской ниши ( $n = 15$ ) и четкая постановка исследовательских целей/задач (13); аргументация актуальности (7) [Identification of a research niche ( $n = 15$ ) and clear formulation of research goals/objectives (13); argumentation of relevance (7)]	The objectives of the research are clearly stated, particularly the focus on how emotional information and information structure are transmitted in simultaneous interpretation.
Контекстуализация [Contextualization]	Описание исследовательского поля и соотнесение своего исследования с ним (16), синтез информации из различных источников при анализе литературы (5), убедительная аргументация практической значимости работы (6) [Description of the research field and correlation of your research with it (16), synthesis of information from various sources in the analysis of literature (5), convincing argumentation of the practical significance of the work (6)]	The literature review is comprehensive, incorporating a variety of sources that discuss the evolution of taste as a sensory experience (e.g., Classen, 1993; Goossens, 2007) and the role of metaphors in food discourse.
Источники [Sources]	Авторитетные (7) / релевантные (10) / современные (5) [Authoritative (7) / relevant (10) / modern (5)]	The literature review includes a variety of sources that demonstrate the interconnection between language and culture, such as Kramsch (2014) and Melton (2010).
Методы исследования [Research methods]	Аргументация выбора и четкое описание методов (7) [Reasoning for choice and clear description of methods (7)]	The methods section outlines a clear strategy for the research, utilizing both content analysis and comparative analysis.

Организация текста [Organization of text]	Четкие разделы (введение, литературный обзор и т.д.) (16) и заголовки к ним (8); развитие идей от раздела к разделу (7) [Clear sections (introduction, literature review, etc.) (16) and headings to them (8); development of ideas from section to section (7)]	The paper generally follows a logical progression from the introduction through the literature review, methodology, and anticipated results.
Абзацы [Paragraphs]	Корректное разделение на абзацы (8), отдельный абзац развивает одну идею (7) [Correct division into paragraphs (8), each paragraph develops one idea (7)]	For instance, the literature review is divided into coherent paragraphs that discuss different aspects and studies related to CLIL.
Предложения [Sentences]	Сложные предложения (4), адекватное сочетание простых и сложных предложений (4) [Complex sentences (4), adequate combination of simple and complex sentences (4)]	The author employs complex sentences to convey sophisticated ideas, such as, "Owing to the achievements in the domain of neuroscience, at the present time the obstacles students face, while acquiring foreign languages, can be tackled."
Стиль [Style]	Формальный (академический) стиль (11), речевые клише, ему свойственные (5) [Formal (academic) style (11), speech cliches characteristic of it (5)]	The overall tone remains formal and academic, which is appropriate for the context. Phrases like "the primary aim of this paper" and "this study presupposes" maintain an academic style.
Лексика [Vocabulary]	Разнообразная (9), релевантная (7), академическая (9); используются специальная/тематическая лексика (4) и термины (3) [Diverse (9), relevant (7), academic (9); uses specialized/topical vocabulary (4) and terms (3)]	The author demonstrates a good command of academic vocabulary, using terms such as "morphemic," "etymological," "semantic," and "neologism" appropriately throughout the text.

Тематический анализ обратной связи от искусственного интеллекта: слабые стороны студенческих работ  
[Thematic analysis of feedback by AI: Weaknesses of students' works]

Таблица 2

Аспект [Aspect]	Тематический анализ [Thematic analysis]	Пример [Example]
Обоснование [Justification]	Недостаточно доказательной информации (7) и примеров (5), поверхностный (не критический) анализ (10); отсутствие поддержки основного тезиса на протяжении всей работы (5), недостаточная аргументация актуальности проблемы и выбора исследовательской ниши (8), нечеткая формулировка аргументов (7), слабая аргументация значимости исследования и практического применения результатов (7), аргументация не развита / слабая (5) [Insufficient supporting information (7) and examples (5), superficial (not critical) analysis (10); lack of support for the main thesis throughout the work (5), insufficient argumentation for the relevance of the problem and the choice of research niche (8), unclear formulation of arguments (7), weak argumentation for the significance of the study and the practical application of the results (7), argumentation is not developed / weak (5)]	While the gap is mentioned, the justification could be further strengthened by providing more specific examples or data illustrating the current state of social advertising in Russia. Some arguments lack depth or could be better supported with a more thorough analysis of the implications of the identified stereotypes.
Контекстуализация [Contextualization]	Перечисление идей из источников без интеграции (5); слабое соотнесение идей из источников и основного тезиса (9) [Listing ideas from sources without integration (5); weak correlation of ideas from sources and the main thesis (9)]	While the literature review mentions various studies, the connections between these studies and the central argument could be more explicitly drawn. The argument regarding the distinction between localization and intercultural communication is not always clearly articulated.

Источники [Sources]	Устаревшие (13) / однообразные (1) / нерелевантные (2) / неавторитетные (2) / поддерживающие только одну точку зрения (1) / ненаучные (4) [Outdated (13) / monotonous (1) / irrelevant (2) / unauthoritative (2) / one-sided (1) / unscientific (4)]	A few sources appear outdated or lack direct relevance to the specific focus on social advertising (e.g., Weber, 1968).
Абзацы [Paragraphs]	Слишком длинные и развивают несколько идей (18) [Too long and develop several ideas (18)]	Some paragraphs, especially in the Literature Review, feel packed with information, which can disrupt the flow. For instance, the discussion on the use of technologies in education combines multiple studies without clear segmentation.
Организация текста [Text organization]	Переходы между разделами (9), абзацами (2), предложениями (3) и идеями (12) слишком резкие или неочевидные; нет подразделов в анализе литературы (2), методах (1) и планируемых результатах (1) [Transitions between sections (9), paragraphs (2), sentences (3), and ideas (12) are too abrupt or unclear; there are no subsections in the literature review (2), methods (1), and planned results (1)]	However, within some sections, particularly the Literature Review, the flow of ideas can feel disjointed. For example, the transition between the definitions of neologisms and the discussion of their societal relevance could be smoother, as the abrupt shifts may confuse readers.
Связки [Links]	Повторяющиеся и однообразные (17), употреблены неверно (7) [Repetitive and monotonous (17), used incorrectly (7)]	However, some cohesive devices are used repetitively (“What is more” appears multiple times), which can lead to redundancy.
Форматирование [Formatting]	Недочеты в оформлении списка литературы (отсутствие капитализации, наклонного шрифта, номеров страниц, DOI; использование точек и запятых; неверное расположение элементов) (19), некорректное форматирование внутритекстовых ссылок (5),	Some in-text citations lack proper punctuation or formatting (e.g., missing commas or incorrect placement of years), which detracts from the overall professionalism of the paper.

Окончание табл. 2

Аспект [Aspect]	Тематический анализ [Thematic analysis]	Пример [Example]
	несоответствие внутритекстовых ссылок списку литературы (8) [Deficiencies in the formatting of the bibliography (lack of capitalization, italic font, page numbers, DOI; use of periods and commas; incorrect placement of elements) (19), incorrect formatting of in-text references (5), inconsistency between in-text references and the bibliography (8)]	The reference list itself contains minor formatting errors, such as inconsistent use of italics and punctuation.
Стиль [Style]	Многословность и избыточность (11), длинные (перегруженные) и сложные для понимания предложения (12), разговорный стиль / отклонения от академического стиля (8) [Verbosity and redundancy (11), long (overloaded) and difficult to understand sentences (12), colloquial style / deviations from academic style (8)]	Some sentences are overly complex or convoluted, making them hard to follow. For example, “The emergence of the study can be observed in the works of the prominent Soviet psychologists of the 20th century” could be more straightforwardly phrased as “The study emerged from the work of prominent Soviet psychologists in the 20th century.”
Грамматика [Grammar]	Непоследовательность в использовании времен (7), грамматические ошибки (5), ошибки в параллельных конструкциях (2) [Inconsistency in the use of tenses (7), grammatical errors (5), errors in parallel constructions (2)]	The paper shifts between present and past tense inconsistently. For example, “this study will help to understand” (future) and “this study confirms” (present) are used together without clear rationale.
Пунктуация [Punctuation]	Использование запятых: сложные предложения (2), причастные и другие обороты (4), списки (3), выделение связей (1); использование кавычек (3), точки с запятой (1) [Use of commas: complex sentences (2), participial and other phrases (4), lists (3), identifying connectives (1); use of quotation marks (3), semicolons (1)]	In the sentence “It means that texts become available to international audiences which, in turn, contributes to the transfer of knowledge,” there should be a comma before “which” to clarify the non-restrictive clause.

Преподаватель подходит к оценочной ОС иначе. Основная часть ОС дана в виде комментариев на полях, указывающих на недочеты в работе, количество которых значительно варьируется от работы к работе (от 0 до 48). Преподаватель дает намного более сжатую ОС: в среднем на работу в ОС от ИИ 1108 слов, от преподавателя – 124 слова. Большинство работ (17) также снабжены краткой итоговой ОС на основе критериев, в которой отмечаются как сильные, так и слабые стороны работы (см. пример, рис. 2).

**Content and source use:** Fully corresponds; within a broader context; support is appropriate; the gap is clearly stated; the choice of sources is appropriate.

**Language:** Organization is clear, some issues of a paragraph division are present, some inconsistencies in the APA style are apparent.

**Organization and format:** Effective communicative structures, which are accurately used, yet there are a few inaccuracies (mainly in punctuation and in articles), style is academic.

**Рис. 2.** Пример обратной связи преподавателя на основе критериев

**Fig. 2.** Instructor's feedback based on criteria

Комментарии преподавателя «на полях» акцентировали схожие с ИИ аспекты, однако преподносились в ином формате. Их формулировки указывали на тип ошибки, недостаточность аргументации или отсутствие содержательных элементов без непосредственного их исправления (табл. 3). Больше всех комментариев относились к организации и форматированию текста (51 и 45 соответственно, в совокупности 96 для всей категории).

Тематический анализ ОС от ИИ выявил ее недостатки. Иногда ИИ относит информацию к неверной категории. Например, обсуждая содержание работы, в пример платформа приводит недостатки организации текста: *the transition between discussing the historical perception of taste and its current status in hedonistic culture could be more seamless to enhance understanding*. В нескольких случаях ИИ считает ошибкой то, что ей не является, например, предлагая добавить запятые там, где они не требуются. ИИ не дает полноценного представления о том, насколько часто обнаруженный недостаток встречается в работе, для этого необходим отдельный анализ. ГенИИ не полностью осознает специфику академических текстов и требований к ним, считая, например, тавтологией повторения основной терминологии и предлагая перефразировать ее.

Таблица 3

**Тематический анализ комментариев преподавателя**  
**[Thematic analysis of instructor's commentaries]**

Категория [Category]	Темы [Topics]	Пример комментария [Example of a comment]
Содержание и использование источников (59 комментариев) [Content and use of sources (59 comments)]	Четкая аргументация; включение необходимых содержательных элементов в раздел; поддержка тезисов ссылками на источники; опора на современные и релевантные источники; разяснение используемой терминологии [Clear argumentation; inclusion of necessary substantive elements in the section; supporting thesis with references to sources; reliance on current and relevant sources; explanation of the terminology used]	Evidence/reference to prove that... Were the goals outlined? Anticipated results remain unclear concerning the results of the morphemic analysis, semantic analysis and etymological analysis Clarification on methodology, sampling and data processing is needed The authors of the statement remain unclear
Язык (42 комментария) [Language (42 comments)]	Грамотное использование языковых средств; четкое донесение смысла посредством языковых средств; использование речевых клише; академический стиль текста [Competent use of linguistic means; clear communication of meaning through linguistic means; use of speech clichés; academic writing style]	Questionable use of the grammar form meaning unclear Hedging is needed ineffective and bulky grammar structure, inappropriate word
Организация текста и форматирование (96 комментариев) [Text organization and formatting (96 comments)]	Структура работы; разделение на абзацы; использование связей; форматирование по APA [Structure of the work; division into paragraphs; use of transitions; APA formatting]	Ineffective paragraphing The paragraph is not connected to the previous idea A transition is needed Page number missing



## 5. Обсуждение результатов

Данное исследование посвящено исследованию потенциала гениИИ для оценивания академических работ в формате Project Proposal. Выявлено, что ИИ ориентируется на заданные критерии и способен аргументировать выставленную отметку, а также сформулировать ОС в соответствии с ними. Как и в предыдущих исследованиях (см. обзор [Боголепова, 2025]), были определены такие положительные черты ОС от ИИ, как информативность и ее конструктивный характер, но при этом для нее также характерны избыточность и шаблонность.

Хотя согласованность оценок преподавателя и гениИИ низка, оба проверяющих сходно оценивают качество работы. Наибольшая согласованность в оценке наблюдается в категории «Языковое оформление», наименьшая – в категории «Содержание и использование источников». Можно предположить, что опытный преподаватель лучше, чем ИИ, понимает специфические требования к содержанию работы, что было показано и на примере ОС на эссе [Steiss et al., 2024].

И гениИИ, и преподаватель ориентируются на критерии оценивания и акцентируют следующие аспекты, характеризующие качественный академический (научный) текст:

- интеграция исследования в соответствующий академический контекст;
- связность текста на уровне разделов, абзацев, предложений;
- абзацы, не перегруженные информацией и развивающие один основной тезис;
- академический стиль и соответствующие языковые средства и речевые клише, без повторений, избыточности и перегруженности предложений;
- корректное оформление работы и источников согласно стандарту APA;
- четкая аргументация и поддержка тезисов.

ОС от ИИ и комментарии преподавателя, однако, имеют свою специфику. ИИ чаще подмечает отдельные неточности в языковом и пунктуационном оформлении, ошибках в форматировании и несоответствиях на уровне предложения, в то время как преподаватель отмечает отдельные неточности в тексте, а в итоговой ОС суммирует сильные и слабые стороны на уровне всей работы. Кроме того, преподаватель формулирует комментарии, указывая на тип ошибки, задавая наводящие вопросы, но не предлагая однозначных исправлений. Склонность ИИ к выделению локальных аспектов и преподавателя – к балансу характеристик, а также развивающий характер комментариев преподавателя отмечают и в других исследованиях [Dai et al., 2023; Guo Wang, 2024].

ГенИИ не учитывает специфики контекста и не имеет информации об индивидуальных особенностях и потребностях отдельного студента, поэтому дает подробную ОС по всем критериям, что было отмечено и в исследованиях на другом текстовом материале [Almegren et al., 2024]. Преподаватель же, напротив, такой информацией обладает и может таргетировать ОС в соответствии с ней. В то же время опыт взаимодействия со студентом может являться и причиной необъективной оценки (так называемый «эффект гало» [Spear, 1996]). Первичная оценка работы с помощью генИИ может помочь минимизировать этот эффект и дать преподавателю основу, которую он будет редактировать в соответствии с контекстом.

Краткость комментариев преподавателя является несомненным плюсом в эпоху коротких форматов и сокращающегося времени концентрации внимания, но они могут быть недостаточно информативными для дальнейшей коррекции деятельности обучающегося [Сысоев и др., 2024б]. Хотя ИИ уступает опытным преподавателям по способности оценить содержание работы, информативная ОС от ИИ может стать элементом формирующего оценивания [Steiss et al., 2024]. Постоянная доступность генИИ не только для получения ОС, но и для уточнения неясных моментов и получения дополнительной информации делает его инструментом непрерывного развития и обучения, однако студентов необходимо обучать целенаправленному использованию ИИ для решения таких задач [Авраменко, Буланова, 2024].

ГенИИ, как было выявлено, дает возможность не только оценить сильные и слабые стороны отдельной работы, но и выделить типичные недочеты в работах группы студентов, что позволяет осуществить образовательную аналитику и скорректировать учебную программу в соответствии с нуждами студентов. Потенциал ИИ для индивидуализации и персонализации обучения является неоспоримым фактом [Сысоев, 2025]; в плане анализа письменноречевой продукции индивидуализация может осуществляться как при получении ОС от ИИ студентом, так в результате анализа агрегированной ОС преподавателем и последующей коррекции программы обучения и разработки учебных материалов.

## 6. Выводы

В ходе исследования было выявлено, что генИИ способен оценивать академические работы в формате Project Proposal и давать ОС, основанную на критериях. Однако наблюдается низкая степень согласованности оценок генИИ и преподавателя, что, по всей вероятности, обусловлено различиями в подходах к оцениванию и особенностями формирования

итоговых оценочных суждений: ИИ ориентируется на заданные критерии и дает более стандартные, шаблонные комментарии, в то время как преподаватель использует более субъективные и контекстуальные оценки, формируя итоговую ОС в виде комментариев по тексту и кратких выводов с акцентом на сильные и слабые стороны работы. В целом, генИИ может быть эффективным инструментом для формирования предварительной или формирующей оценки и выявления типичных ошибок.

## Библиографический список / References

Авраменко, Буланова, 2024 – Авраменко А.П., Буланова Е.Р. Перспективы развития самостоятельной работы студентов в контексте интеграции технологий искусственного интеллекта в иноязычное образование // Рема. Rhema. 2024. № 1. С. 79–91. DOI: 10.31862/2500-2953-2024-1-79-91. [Avramenko A.P., Bulanova E.R. Perspectives of developing students' independent work in the context of integrating artificial intelligence technologies into foreign language education. *Rhema*. 2024. No. 1. Pp. 79–91. (In Rus.). DOI: 10.31862/2500-2953-2024-1-79-91]

Бакулев, Боголепова, 2024 – Бакулев А.В., Боголепова С.В. Современные подходы к обучению иностранному языку в преподавании академического письма // Вестник Томского государственного университета. 2024. № 509. С. 113–120. DOI: 10.17223/15617793/509/11. [Bakulev A.V., Bogolepova S.V. Modern approaches to foreign language teaching in academic writing instruction. *Vestnik Tomskogo gosudarstvennogo universiteta*. 2024. No. 509. Pp. 113–120. (In Rus.). DOI: 10.17223/15617793/509/11]

Боголепова, 2025 – Боголепова С.В. Возможности инструментов искусственного интеллекта для проверки письменных работ и формулировки обратной связи // Дискурс профессиональной коммуникации. 2025. Т. 7. № 1. С. 70–88. DOI: 10.24833/2687-0126-2025-7-1-70-88 [Bogolepova S.V. The potential of AI tools for written work assessment and feedback formulation. *Diskurs professionalnoy kommunikatsii*. 2025. Vol. 7. No. 1. Pp. 70–88. (In Rus.). DOI: 10.24833/2687-0126-2025-7-1-70-88]

Боголепова, Жаркова, 2024 – Боголепова С.В., Жаркова М.Г. Исследование потенциала генеративных моделей для оценивания эссе и обеспечения обратной связи // Отечественная и зарубежная педагогика. 2024. Т. 1. № 5 (101). С. 123–137. DOI: 10.24412/2224-0772-2024-101-123-137 [Bogolepova S.V., Zharkova M.G. Exploring the potential of generative models for essay evaluation and feedback provision. *Otechestvennaya i zarubezhnaya pedagogika*. 2024. Vol. 1. No. 5 (101). Pp. 123–137. (In Rus.). DOI: 10.24412/2224-0772-2024-101-123-137]

Дугарцыренова, 2016 – Дугарцыренова В.А. Трудности обучения иноязычному академическому письму // Высшее образование в России. 2016. № 6. С. 106–112. [Dugartsyrenova V.A. Challenges in teaching foreign academic writing. *Vysshee obrazovanie v Rossii*. 2016. No. 6. Pp. 106–112.]

Сысоев, 2025 – Сысоев П.В. Персонализированное обучение на основе технологий искусственного интеллекта: насколько готовы современные студенты к новым возможностям получения образования // Высшее образование

в России. 2025. Т. 34. № 2. С. 51–71. DOI: 10.31992/0869-3617-2025-34-2-51-71 [Sysoev P.V. AI-based personalized learning: How ready are modern students for new educational opportunities? *Vysshee obrazovanie v Rossii*. 2025. Vol. 34. No. 2. Pp. 51–71. (In Rus.). DOI: 10.31992/0869-3617-2025-34-2-51-71]

Сысоев и др., 2024а – Сысоев П.В., Филатов Е.М., Сорокин Д.О. Обратная связь в обучении иностранному языку: от информационных технологий к искусственному интеллекту // *Язык и культура*. 2024. № 65. С. 242–261. DOI: 10.17223/19996195/65/11 [Sysoev P.V., Filatov E.M., Sorokin D.O. Feedback in foreign language learning: From information technologies to artificial intelligence. *Yazyk i kultura*. 2024. No. 65. Pp. 242–261. (In Rus.). DOI: 10.17223/19996195/65/11]

Сысоев и др., 2024б – Преподаватель vs искусственный интеллект: сравнение качества предоставляемой преподавателем и генеративным искусственным интеллектом обратной связи при оценке письменных творческих работ студентов / П.В. Сысоев, Е.М. Филатов, Н.И. Хмаренко, С.С. Мурунов // *Перспективы науки и образования*. 2024. № 5 (71). DOI: 10.32744/pse.2024.5.41 [Sysoev P.V., Filatov E.M., Khmarenko N.I., Murunov S.S. Teacher vs AI: Comparing the quality of feedback provided by instructors and generative AI in assessing students' written work. *Perspektivy nauki i obrazovaniya*. 2024. No. 5 (71). (In Rus.). DOI: 10.32744/pse.2024.5.41]

Титова, 2024 – Титова С.В. Технологические решения на базе искусственного интеллекта в обучении иностранным языкам // *Вестник Московского университета. Сер. 19: Лингвистика и межкультурная коммуникация*. 2024. Т. 27. № 2. С. 18–37. [Titova S.V. AI-based technological solutions in foreign language teaching. *Vestnik Moskovskogo universiteta. Ser. 19: Lingvistika i mezhkulturnaya kommunikatsiya*. 2024. Vol. 27. No. 2. Pp. 18–37. (In Rus.)]

Ahmed, 2022 – Ahmed A.A.M. Academic writing: Types, elements, and strategies. *International Journal of Language, Humanities, and Education*. 2022. Vol. 5. No. 2. Pp. 60–70.

Almegren et al., 2024 – Almegren A., Mahdi H. S., Hazaea A. N. et al. Evaluating the quality of AI feedback: A comparative study of AI and human essay grading. *Innovations in Education and Teaching International*. 2024. DOI: 10.1080/14703297.2024.2437122

Britton, 2025 – Britton C. Academic writing: Key components, structure, and avoiding plagiarism. *Journal of Perioperative Practice*. 2025. Pp. 207–212. DOI: 10.1177/17504589241310405

Chauhan, 2022 – Chauhan P. Fundamentals of academic writing: A literature review. *Journal of NELTA*. 2022. Vol. 27. No. 1-2. Pp. 161–180. DOI: 10.3126/nelta.v27i1-2.53201

Checco et al., 2021 – Checco A., Bracciale L., Loreti P. et al. AI-assisted peer review. *Humanities and Social Sciences Communications*. 2021. Vol. 8. No. 25. DOI: 10.1057/s41599-020-00703-8

Crompton et al., 2024 – Crompton H. et al. AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*. 2024. Vol. 55. No. 6. Pp. 2503–2529. DOI: 10.1111/bjet.13460

Dai et al., 2023 – Dai W., Lin J., Jin F. et al. Can large language models provide feedback to students? A case study on ChatGPT. *EdArXiv Preprint*. 2023. DOI: 10.35542/osf.io/hcgzj

Dang et al., 2022 – Dang H. et al. How to prompt? Opportunities and challenges of zero-and few-shot learning for human-AI interaction in creative applications of generative models. *arXiv preprint arXiv:2209.01390*. 2022.

Ekin, 2023 – Ekin S. Prompt engineering for ChatGPT: A quick guide to techniques, tips, and best practices. *Authorea Preprints*. 2023.

Escalante et al., 2023 – Escalante J., Pack A., Barrett A. AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*. 2023. Vol. 20. No. 1. Pp. 57. DOI: 10.1186/s41239-023-00425-2

Farahani, 2023 – Farahani M.V. Dynamicity of interaction in academic discourse: Evidence from a corpus-based study. *Forum for Linguistic Studies*. 2023. Vol. 5. No. 3. DOI: 10.59400/fls.v5i3.1895

Giray, 2023 – Giray L. Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*. 2023. Vol. 51. No. 12. Pp. 2629–2633. DOI: 10.1007/s10439-023-03272-4

Guo, Wang, 2024 – Guo K., Wang D. To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. 2024. Vol. 29. No. 7. Pp. 8435–8463. DOI: 10.1007/s10639-023-12146-0

Jansen et al., 2024 – Jansen T., Höft L., Bahr L. et al. Comparing generative AI and expert feedback to students' writing: Insights from student teachers. *Psychologie in Erziehung und Unterricht*. 2024. Vol. 71. No. 2. Pp. 80–92. DOI: 10.2378/peu2024.art08d

Khalifa, Albadawy, 2024 – Khalifa M., Albadawy M. Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*. 2024. Article number 100145. DOI: 10.1016/j.cmpbup.2024.100145

Li, Liu, 2024 – Li W., Liu H. Applying large language models for automated essay scoring for non-native Japanese. *Humanities and Social Sciences Communications*. 2024. Vol. 1. Article number 723. DOI: 10.1057/s41599-024-03209-9

Li, Schmitt, 2009 – Li J., Schmitt N. The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*. 2009. Vol. 18. No. 2. Pp. 85–102. DOI: 10.1016/j.jslw.2009.02.001

Mananay, 2024 – Mananay J.A. Integrating artificial intelligence (AI) in language teaching: Effectiveness, challenges, and strategies. *International Journal of Learning, Teaching and Educational Research*. 2024. Vol. 23. No. 9. Pp. 361–382. DOI: 10.26803/ijlter.23.9.19

Mizumoto et al., 2024 – Mizumoto A., Shintani N., Sasaki M., Feng Teng M. Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*. 2024. Vol. 3. No. 2. Article number 100116. DOI: 10.1016/j.rmal.2024.100116

Nasiri, 2012 – Nasiri S. Academic writing: The role of culture, language and identity in writing for community. *International Journal of Learning & Development*. 2012. Vol. 2. No. 3. Pp. 1–8. DOI: 10.5296/ijld.v2i3.1530

Samraj, 2004 – Samraj B. Discourse features of the student-produced academic research paper: Variations across disciplinary courses. *Journal of English for Academic Purposes*. 2004. DOI: 10.1016/S1475-1585(03)00053-5

Schulhoff et al., 2024 – Schulhoff S. et al. The prompt report: A systematic survey of prompting techniques. *arXiv preprint arXiv:2406.06608*. 2024. Vol. 5.

Solak, 2024 – Solak E. Examining writing feedback dynamics from ChatGPT AI and human eucators: A comparative study. *Pedagogika – Pedagogy*. 2024. Vol. 96. No. 7. Pp. 955–969. DOI: 10.53656/ped2024-7.05

Spear, 1996 – Spear M. The influence of halo effects upon teachers' assessments of written work. *Research in Education*. 1996. Vol. 56. No. 1. Pp. 85–87. DOI: 10.1177/003452379605600107

Steffen et al., 2021 – Epp S., Hoffmann M., Lell N. et al. Stereo: A pipeline for extracting experiment statistics, conditions, and topics from scientific papers. *The 23rd International Conference on Information Integration and Web Intelligence (iiWAS2021)*. 2021. P. 25. DOI: 10.1145/3487664.3487712

Steiss et al., 2024 – Steiss J., Tate T., Graham S. et al. Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*. 2024. Vol. 91. Art. 101894. DOI: 10.1016/j.learninstruc.2024.101894

Tyser et al., 2024 – Tyser K. et al. AI-driven review systems: Evaluating LLMs in scalable and bias-aware academic reviews. *arXiv preprint arXiv:2408.10365*. 2024.

Wirantaka, 2016 – Wirantaka A. Paragraph writing of academic texts in an EFL context. *Journal of Foreign Language Teaching and Learning*. 2016. Vol. 1. No. 2. Pp. 34–45.

Статья поступила в редакцию 01.05.2025

The article was received on 01.05.2025

## Сведения об авторах / About the authors

**Боголепова Светлана Викторовна** – кандидат филологических наук, доцент; доцент Школы иностранных языков, Национальный исследовательский университет «Высшая школа экономики», г. Москва

**Svetlana V. Bogolepova** – PhD in Philology; associate professor at the School of Foreign Languages, HSE University, Moscow

ORCID: <https://orcid.org/0000-0003-1050-9110>

E-mail: [sbogolepova@hse.ru](mailto:sbogolepova@hse.ru)

**Греченко Анастасия Александровна** – тьютор, приглашенный преподаватель Школы иностранных языков, Национальный исследовательский университет «Высшая школа экономики», г. Москва

**Anastasiia A. Grechenko** – tutor, visiting lecturer at the School of Foreign Languages, HSE University, Moscow

ORCID: <https://orcid.org/0009-0000-3173-6932>

E-mail: [agrechenko@hse.ru](mailto:agrechenko@hse.ru)

Все авторы прочитали и одобрили окончательный вариант рукописи

All authors have read and approved the final manuscript