



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

<https://doi.org/10.24833/2687-0126-2025-7-1-70-88>

ВОЗМОЖНОСТИ ИНСТРУМЕНТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОВЕРКИ ПИСЬМЕННЫХ РАБОТ И ФОРМУЛИРОВКИ ОБРАТНОЙ СВЯЗИ

С.В. Боголепова

*Национальный исследовательский университет «Высшая школа экономики»
sbogolepova@hse.ru*

Аннотация: Статья посвящена исследованию возможностей генеративного искусственного интеллекта (ИИ) для решения задач оценки письменных работ и предоставления обратной связи на них. Поставлена цель определить возможности и ограничения генеративного ИИ при его использовании для проверки письменноречевой продукции обучающихся и формулировки обратной связи на нее. Для достижения цели осуществлен систематический обзор двадцати двух оригинальных исследований, проведенных в российском и зарубежном контекстах, результаты которых были опубликованы в 2022-2025 годах. Выявлено, что результаты критериального оценивания генеративными моделями согласуются с оценками преподавателей, причем генеративный ИИ превосходит проверяющих в способности оценить язык и аргументацию. На надежность этой оценки, однако, негативное влияние оказывают нестабильность оценивания, галлюцинации генеративных моделей и их ограниченная способность учитывать особенности контекста. Несмотря на детальность и конструктивность обратной связи от генеративного ИИ, она недостаточно конкретна и избыточна, что может затруднить ее восприятие обучающимися. Обратная связь от генеративной модели таргетирует преимущественно локальные недостатки, в то время как проверяющие преподаватели обращают внимание на глобальные проблемы, такие как неполное соответствие содержания заданной теме или нарушения в логике изложения. В отличие от преподавателей, генеративный ИИ дает шаблонную обратную связь, не используя косвенные формулировки и наводящие вопросы, способствующие развитию умений саморегуляции. Тем не менее, эти недостатки можно скорректировать последующими запросами к генеративной модели. Также выявлено, что обучающиеся открыты к получению обратной связи от генеративного ИИ, однако предпочитают получать ее от преподавателей и сверстников. Результаты обсуждаются в контексте использования генеративных моделей для проверки письменных работ и формулировки обратной связи преподавателями иностранного языка. Выводы подчеркивают необходимость критического подхода к использованию генеративных моделей в оценивании письменных работ, а также важность подготовки преподавателей для эффективного взаимодействия с этими технологиями.

Ключевые слова: ИИ в образовании, генеративные модели в преподавании иностранного языка, автоматическая оценка текста, обратная связь на письменные работы, оценивание письменноречевой продукции.

Для цитирования: Боголепова, С.В. (2025). Возможности инструментов искусственного интеллекта для проверки письменных работ и формулировки обратной связи. *Дискурс профессиональной коммуникации*, 7(1), 70–88. <https://doi.org/10.24833/2687-0126-2025-7-1-70-88>

1. ВВЕДЕНИЕ

Инструменты искусственного интеллекта (ИИ) стали неотъемлемой частью жизни общества, и образование не является исключением. ИИ способен выполнять задачи, традиционно приписываемые человеческому интеллекту: обучаться, планировать, анализировать, делать выводы и т.д. Генеративный ИИ – это спектр инструментов ИИ, способных синтезировать/генерировать контент, почти неотличимый от того, что может создавать человек (Gupta et al., 2024). Генеративный ИИ, способный создавать тексты, имеет в основе большие языковые модели, т.е. нейронные сети с миллиардами параметров, неконтролируемо обученных на огромных объемах текстовых данных (Fan et al., 2024). Генеративный ИИ или генеративные модели (ГМ) коммуницируют с пользователем посредством чат-бота. Сейчас в мире более 300 миллионов человек используют возможности ГМ (AI tool user numbers worldwide from 2020-2030, 2024).

ГМ широко применяются участниками образовательного процесса. Обучающиеся используют их для информационных запросов, генерации идей и редактирования текстов, анализа и резюмирования информации (Almassaad et al., 2024; Chan & Hu, 2023). Изучающие иностранный язык могут выполнять тренировочные и тестовые задания, получая мгновенную обратную связь, а также коммуницировать с моделью (Annamalai et al., 2023). Преподаватели создают образовательный контент, формируют учебные планы и программы, выстраивают индивидуальные образовательные траектории для своих студентов (Константинова и др., 2023).

Автоматические системы уже более двадцати лет используются для оценки письменных текстов, однако только в последнее время революционное развитие систем ИИ позволило вывести оценку на новый уровень и сделать такие инструменты доступными широкому кругу пользователей (Li & Liu, 2024; Stahl et al., 2024). В то время как более ранние автоматизированные системы проверки умели распознавать только заданные лингвистические параметры и работали только с определенным типом текста, современные ГМ обучены на огромных объемах разнотипных текстов и могут работать с любыми критериями, заданными в запросе (Li & Liu, 2024; Hussein et al., 2019). Более того, помимо непосредственного оценивания современные системы способны давать развернутую обратную связь по заданным аспектам, чего предыдущие системы делать не могли.

Технологии ИИ помогают решить задачу объективного и валидного оценивания на разных этапах обучения (Титова, 2024). ГМ дают преподавателям возможность частично автоматизировать проверку письменных работ студентов, выделить как сильные, так и слабые стороны работ, таким образом делая обратную связь индивидуализированной и экономя время преподавателя (Kasneci et al., 2023). Автоматическая оценка посредством ГМ позволяет избежать эффекта гало, т.е. влияния субъективного отношения к автору текста со стороны проверяющего (Mizumoto & Eguchi, 2023). Обучающиеся сами могут обращаться к ГМ для получения обратной связи на свои тексты и последующей их доработки (Сысоев, Филатов, 2024).

Несмотря на наличие обзоров, анализирующих использование технологий генеративного ИИ в образовании (Bahroun et al., 2023; Crompton & Burke, 2023; Wang et al., 2024) и применение ИИ и, в частности, ГМ в обучении иностранному языку (Сысоев, 2023; Титова, 2024; Crompton et al., 2024; Li et al., 2024), автором не было обнаружено обзоров исследований, посвященных изучению практик использования генеративного ИИ для проверки письменных работ и формулировки обратной связи. Подобный обзор, интегрирующий как отечественные исследовательские работы, так и труды зарубежных исследователей, позволит понять, какой потенциал ГМ имеют для решения упомянутых задач, а также выявить их ограничения, которые необходимо учитывать при работе с ГМ. Соответственно, цель данного исследования — определить возможности и ограничения инструментов генеративного искусственного интеллекта для оценки письменных работ обучающихся и формулировки обратной связи по ним в образовательном процессе на основе обзора современных исследовательских статей.

В процессе достижения поставленной цели были найдены ответы на следующие исследовательские вопросы:

1. Насколько эффективно инструменты генеративного ИИ могут оценивать письменные работы по сравнению с оценкой преподавателей?
2. Каковы характеристики обратной связи, генерируемой инструментами ИИ, и как она соотносится с обратной связью от преподавателей?
3. Каковы сильные и слабые стороны использования генеративного ИИ для оценки письменных работ и предоставления обратной связи?
4. Как обучающиеся воспринимают и оценивают обратную связь, полученную от генеративного ИИ, в сравнении с обратной связью от преподавателей?

Автор обсуждает полученные результаты в контексте применения ГМ преподавателями иностранного языка.

2. МАТЕРИАЛЫ И МЕТОДЫ

Для отбора и анализа статей автор использовал методику *PRISMA*, подразумевающую следующие шаги (Crompton et al., 2024):

- автоматический и ручной поиск исследований в базах данных;
- удаление из подборки дубликатов и статей, к которым нет полнотекстового доступа;
- скрининг статей на предмет соответствия необходимому типу исследования, году публикации, принадлежности предметной области и соответствия исследовательской цели;
- качественный анализ отобранных исследований.

Поиск исследований проводился в базах данных «*Scopus*» («*Science Direct*»), «*ResearchGate*», «Киберленинка». Использовались ключевые слова:

- на русском языке: искусственный интеллект для оценки текста, генеративные модели в обучении иностранному языку, большие языковые модели в образовании, генеративный искусственный интеллект в образовании, автоматическая проверка текста, оценка/оценивание письменных работ/ письменной продукции, сгенерированная обратная связь;
- на английском языке: *Artificial Intelligence for Text Evaluation, Generative Models in Language Teaching, Large Language Models in Education, Generative Artificial Intelligence in Education, Automatic Text Scoring, Assessment/Evaluation of Written Production, AI-generated Feedback*.

Активное использование ГМ, таких как *ChatGPT*, началось в 2022 году (Yu & Guo, 2023), поэтому по ключевым словам были отобраны оригинальные исследования, опубликованные в 2022-2025 гг. Было принято решение включить в подборку и препринты, т.к. они дают возможность быстрой диссеминации научных результатов, что важно, когда речь идет о быстро развивающейся и недавно исследуемой технологии.

В результате поиска по базам данных было отобрано 198 статей. После скрининга в пуле осталось 22 статьи, соответствующие критериям отбора (тип, тематика, год публикации), что можно считать достаточным, учитывая короткий временной промежуток и узкую тематику исследования (Evans & Benefield, 2001). Отбирались оригинальные исследования использования ГМ для оценки письменных работ и обеспечения обратной связи, их результаты были систематизированы и частично описаны (Amobonye et al., 2024).

3. РЕЗУЛЬТАТЫ

3.1. Оценивание письменных работ

Благодаря доступности и простоте ГМ используются и для проверки письменных работ обучающихся, однако в исследованиях пока нет единогласия по вопросу надежности результатов такой проверки. Исследование на основе корпуса в 12100 эссе в формате международных экзаменов с помощью запроса, включающего критерии, показало, что проверка эссе ГМ точно соотносит уровень работы с международной шкалой (Mizumoto & Eguchi, 2023). При последовательной проверке одной и той же работы ГМ выдают результаты, отличающиеся друг от друга не более, чем на 2 балла (из девяти), т.е. результаты проверки отличаются внутренней согласованностью.

Исследование письменных экзаменационных ответов открытого типа также показало высокий уровень внутренней согласованности проверки работ ГМ *ChatGPT*; при этом оценки более, чем 70 % работ, выставленные преподавателем и ГМ, отличались не более, чем на один балл из шести (Jauhainen & Garagorry Guerra, 2024). Было замечено, что преподаватели оценивали языковую составляющую работы даже строже, чем ГМ *Chat GPT*.

Подобные описанным выше результаты были получены и при анализе результатов проверки студенческих эссе ГМ *ChatGPT-3.5* и *Perplexity* и их сравнении с оценкой эссе преподавателем (Боголепова, Жаркова, 2024). Хотя при сравнении баллов, выставленных преподавателем и ГМ, при допущении расхождения в 1 балл (из девяти) согласованность ранжировалась от 70 до 90%, межэкспертное соответствие оценок было низким. Сравнение оценок эссе по заданным критериям ГМ и практикующими преподавателями выявило, что оценки *ChatGPT-4* и преподавателей демонстрируют высокую согласованность, однако периодически ГМ галлюцинируют, приписывая студенческим работам свойства, им не присущие (Pack et al., 2024).

Мицумото и др. сравнили результаты проверки письменных работ преподавателем английского языка, специализированным инструментом *Grammarly* и ГМ *Chat GPT* (Mizumoto et al., 2024). Исследователи показали, что результаты ГМ *ChatGPT-4* по оценке языковой составляющей положительно коррелируют ($\rho = 0.79$) с баллами, выставленными преподавателем, и превосходят в этом результаты специализированного приложения. Подобные результаты были получены в отношении и общей оценки работы. Также было показано, что ГМ в сравнении со специализированными инструментами автоматической оценки более точно

определяют соответствие работы уровню владения различными иностранными языками (Li & Liu, 2024). Исследователи делают вывод о том, что ГМ подходят к анализу и оценке текста более холистически и более «по-человечески», чем такие инструменты.

Несмотря на согласованность в оценке, подходы к оцениванию ГМ и экспертом имеют свои особенности (Awidi, 2024). Хотя человек лучше подмечает нюансы, выставленная им оценка не всегда соответствует критериям и сопутствующим вербальным комментариям, а отдельные фрагменты работы или личность автора, импонирующие проверяющему, могут повлиять на объективность оценки. ГМ, в свою очередь, точно руководствуются критериями, но нечувствительны к контексту. Что касается аспектов оценивания, ГМ превосходят проверяющих в оценке языка и аргументации, в то время как люди лучше определяют соответствие работы заданной теме (Bouziane & Bouziane, 2024).

3.2. Формулировка обратной связи

Ряд исследований посвящен изучению качества обратной связи, даваемой ГМ, и ее сравнению с обратной связью, даваемой проверяющим преподавателем.

Сравнение обратной связи от преподавателя и ГМ *ChatGPT 4.0* на основе выборки в 350 студенческих эссе показал, что ГМ не уступают преподавателю в способности давать обратную связь на содержание работы, аргументацию и организацию и даже превосходят его в вопросе обратной связи по языку и оригинальности («нешаблонности мышления»), которые эксперты оценили выше, чем обратную связь от преподавателя (Сысоев и др., 2024). Способность ГМ к адекватной оценке оригинальности и аргументации на основе критериев отмечены и в других статьях (Awidi, 2024; Steiss et al., 2024). Сысоев и др. (2024) отметили, что ГМ *Chat GPT* дает более развернутые комментарии, чем преподаватель, и предлагает варианты исправления ошибок, а не только выделяет их.

В другом исследовании, в котором экспертами сравнивались обратная связь от ГМ и преподавателя на 50 студенческих эссе было выявлено, что ГМ *ChatGPT* дает более детальную, чем преподаватель, обратную связь, причем равномерно распределяет комментарии между всеми аспектами оценивания, в то время как преподаватели акцентируют содержание и язык (Guo & Wang, 2024). Отличительной чертой комментариев преподавателей было использование не прямых формулировок и вопросов, нацеленных на рефлекссию и самокоррекцию. Можно было бы предположить, что подобные формулировки культурно обусловлены, однако отсутствие саморегулирующей обратной связи от ГМ было отмечено и в исследовании в другом образовательном контексте (Dai et al., 2023).

Используя не прямые формулировки, преподаватель обращает внимание обучающегося на глобальные недостатки (в структуре предложения, аргументации), в то время как ГМ прямо исправляют локальные ошибки и предлагают переформулировки отдельных предложений (Lin & Crosthwaite, 2024). Последнее не всегда адекватно, когда реципиент обратной связи – изучающий иностранный язык, потому что качество его продукции вполне может соответствовать уровню коммуникативной компетенции обучающегося. Акцент на грамматической правильности в обратной связи от ГМ можно назвать избыточным (Wu et al., 2023).

Обратная связь от преподавателя и ГМ отличается и по другим параметрам. Сгенерированная обратная связь более четкая и информативная (Сысоев и др., 2024; Dai et al., 2023). Однако преподаватели в своей обратной связи более конкретно указывают на то, что необходимо исправить, и акцентируют важные для конкретного случая аспекты, причем делают это разными способами (выделение, подчеркивание, использование кода) (Steiss et al.,

2024). ГМ не только избыточно исправляют неточности, но находят ошибки там, где их нет, потому что соответствующих конструкций не было в обучающей выборке или потому что просто была поставлена задача найти ошибки (Jiang et al., 2023).

Боголепова и Жаркова (2024) также отмечают, что хотя обратная связь от ГМ достаточно подробна, она не всегда достаточно конкретна, чтобы студент смог внести корректировки в работу. Это такие комментарии, как *“The logical flow of ideas is generally smooth, but there are a few instances where the connections between ideas could be more explicit”* или *“However, there are a few instances where word choice could be improved for clarity and precision”*. В этих комментариях отмечены слабые стороны работы, однако не уточняется, какие фрагменты подлежат исправлению. В то же время исследователи отметили конструктивный характер обратной связи от ГМ, выделяющей как недостатки, так и сильные стороны работы. Действительно, ГМ могут содействовать в формулировке обратной связи, чтобы она была более позитивной и конструктивной (Hirunyasiri et al., 2023).

Исследователи акцентируют излишнюю информативность и избыточность обратной связи от ГМ (Боголепова, Жаркова, 2024; Lin & Crosthwaite, 2024; Steiss et al., 2024). Однако в контексте высшего образования детальность обратной связи от ГМ и ее фокус на индивидуальных особенностях работы студента являются мотивирующими факторами (Kinder et al., 2025). Студенты готовы использовать ГМ для улучшения письменных работ (Jansen et al., 2024) и демонстрируют высокий уровень мотивации и вовлеченности при работе с обратной связью от ГМ (Chan et al., 2024).

Другие исследования ставили задачей выяснить, какую обратную связь на эссе, полученную от ГМ или от преподавателя, студенты считают более качественной (Escalante et al., 2023; Jansen et al., 2024). По мнению студентов, обратная связь от ГМ уступает по качеству обратной связи со стороны преподавателя (студенты посчитали полезной 59% обратной связи от генеративной модели и 88% – от преподавателя (Jansen et al., 2024). В другом подобном исследовании мнения студентов по поводу предпочтения обратной связи от ГМ или преподавателя разделились пополам (Escalante et al., 2023). При этом было показано, что прирост в умениях письма не зависел от источника обратной связи.

Также продемонстрировано, что обратная связь на письменные работы, полученная от сверстников, ценится студентами выше обратной связи от ГМ (Saini et al., 2024). ГМ имеют тенденцию давать описательную обратную связь, в то время как обратная связь от сверстников побуждает реципиентов критически осмысливать свою работу, т.к. другие студенты, как и преподаватели, используют в обратной связи наводящие вопросы или привлекают внимание к непоследовательности в изложении (Banihashem et al., 2024).

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Обзор нацелен на изучение возможностей и ограничений ГМ при их применении для проверки студенческих работ и обеспечения обратной связи. Исследования показывают, что ГМ способны оценивать письменные работы наравне с преподавателем, но при этом имеют свои особенности в подходах к оцениванию и формулировке обратной связи. ГМ превосходят преподавателей в оценке языковых аспектов и аргументации, однако их обратная связь часто оказывается избыточной и неконкретной, сфокусированной на локальных недостатках. Несмотря на детальность и конструктивность обратной связи от ГМ, студенты предпочитают комментарии преподавателей и сверстников, т.к. они способствуют критиче-

скому осмыслению работ. Однако обучающиеся открыты к получению обратной связи от ГМ, и ее обстоятельность может быть плюсом, когда с ней работают студенты с высоким уровнем коммуникативной компетенции в иностранном языке.

4.1. Проверка и оценивание письменных работ

Хотя ГМ уже могут оценивать письменные работы наравне с преподавателем, *надежность* и *объективность* оценивания остаются под вопросом. ГМ, в отличие от человека, руководствуются только заданными критериями, а не эмоциями, но и модели «устают» и совершают ошибки. При последовательной оценке одной и той же работы модель может выдавать отличающиеся результаты. Хотя эта разница невелика и подобные отклонения допускаются при проверке даже письменной части экзаменов с высокими ставками, признать результаты надежными и полагаться только на них мы не можем. Более того, ГМ могут ошибаться в оценке уровня текста по «Общеввропейской шкале компетенций» (Benedetto et al., 2025), что особенно важно учитывать в контексте оценки письменноречевой продукции обучающихся на иностранном языке.

Так как ГМ учатся на определенном наборе данных, они могут считать неверным то, чего не было в данных, на которых модель обучалась (Bahroun et al., 2023). Искаженная интерпретация может касаться жанровых особенностей текста, стиля аргументации или использования специфического языка. Это может привести к предвзятости, дискриминирует отдельные группы обучающихся и также снижает надежность оценивания.

Еще одним недостатком оценки работ генеративными моделями можно назвать ее недостаточную *обоснованность*. ГМ работают по принципу «черного ящика»: известны только вводная информация и сгенерированные данные, а в результате каких операций происходит генерация, остается неизвестным (Lye & Lim, 2024; Iorliam & Ingio, 2024). По этой причине остается неясным, каким образом модель пришла к конечным выводам, и довольно размытые формулировки типа «*работа могла бы быть организована более логично*» в обратной связи, к сожалению, не проясняют картину. Частично проблему помогает решить использование цепочки запросов. В последующих запросах можно попросить ГМ пояснить оценку по отдельным категориям и привести примеры ошибок и варианты их исправления.

В то же время ГМ не уступают преподавателю в способности верно интерпретировать критерии оценивания и выставлять баллы, соответствующие уровню работы. Среди ГМ наилучшие результаты показывает модель последнего поколения *ChatGPT-4* (Pack et al., 2024), и мы можем ожидать дальнейшей эволюции языковых моделей.

4.2. Обратная связь на письменные работы

Проанализируем обратную связь от ГМ относительно характеристик эффективной обратной связи (Hattie & Timperley, 2007). Она, как уже отмечалось, не всегда дает точные указания на недостатки и определяет действия, которые можно предпринять, чтобы в последующих работах их не иметь. Другими словами, обратная связь от ГМ часто бывает *неконкретной*. Это ограничение преодолимо: последующие запросы, требующие уточнения формулировок и приведения примеров, а также добавление в запрос детальных критериев оценивания помогают конкретизировать обратную связь.

Обратная связь от генеративной модели *избыточна и недостаточно персонализирована*. Хотя она выделяет сильные стороны и недостатки работы, только преподаватель, знающий цели и задачи учебного курса, содержание обучения, образовательные потребности и личностные особенности студента может отобрать соответствующие комментарии. Более того, реципиенты обратной связи владеют иностранным языком на разном уровне, и чтобы обратная связь была доступной, преподавателю может понадобиться упростить формулировки.

Студенты ценят содержательные и развернутые комментарии по своим работам, и инструменты ИИ позволяют сгенерировать информативную обратную связь за короткое время или давать ее автоматически (Gombert et al., 2024). В то же время, слишком детализированная обратная связь может перегрузить и демотивировать студента, поэтому задачей преподавателя является отбор информации в соответствии с потребностями учащегося и планируемыми образовательными результатами (Kumar, 2023).

Нельзя не отметить такие свойства обратной связи от ГМ, даваемой на основе критериев, как *однообразие* и *шаблонность*. Согласимся с А. А. Кореневым, утверждающим, что «шаблонность обратной связи (например, использование из раза в раз одних и тех же фраз или структурных моделей) может привести к тому, что обучающиеся потеряют веру в оригинальность данной обратной связи и доверие к ее источнику» (Корнев, 2018, с. 124). Взаимодействие с чат-ботами студенты считают несколько механистическим и однообразным (Annamalai et al., 2023), но преподаватель может не только отбирать, но и видоизменять комментарии ГМ перед предоставлением студенту.

Эффективная обратная связь способствует развитию *умений саморегуляции*. Она может быть дана в форме вопросов, подталкивающих ученика к анализу и самостоятельному исправлению ошибки. Вместо исправлений преподаватели могут использовать заранее оговоренный шифр для обозначения разных типов ошибок, на основании которого впоследствии требуется переработка письменного продукта. Однако стоит отметить, что по отдельному запросу ГМ может сформулировать вопросы и предложения, способствующие развитию этих умений.

При всех недостатках обратная связь от ГМ носит *конструктивный характер*, позволяя обучающимся осознать сильные стороны работы и понять, каким образом она может быть улучшена. ГМ дают мотивирующие и позитивные комментарии, на формулировку которых у преподавателя не всегда хватает времени и энергии. Обратная связь от преподавателя не всегда *своевременна*, при этом ГМ способны выдавать ее мгновенно при условии доступа к технологиям.

Перечисленные выше потенциальные недостатки можно предвосхитить, грамотно сформулировав запрос. Для достижения поставленных целей в запрос рекомендуется включить задачу для модели в повелительном наклонении, дать ГМ шаблон или пример в качестве опоры, а также описать в запросе контекст и уточнить все необходимые детали (Sidorkin, 2024). С помощью последующих запросов возможно:

- получить список языковых и стилистических ошибок с возможными исправлениями;
- сформулировать обоснование для оценки по отдельным аспектам;
- уточнить, какие фрагменты работы необходимо улучшить в плане содержания и организации, и каким образом;
- упростить обратную связь, сделав формулировки доступными обучающимся, владеющим изучаемым языком на разном уровне;

- сформулировать вопросы и утверждения, побуждающие учащегося к самостоятельному улучшению отдельных аспектов работы и др.

Использование цепочки запросов, уточняющих различные аспекты контекста оценивания, помогает сделать обратную связь более конкретной, разнообразной и персонализированной (Stahl et al., 2024).

4.3. Использование ГМ преподавателями иностранного языка

Осознавая как возможности, так и ограничения ГМ, преподаватели иностранного языка могут использовать их потенциал для оптимизации выполнения профессиональных задач, относящихся к проверке письменных работ и формулировке обратной связи по ним. Оценка письменных работ студентов может стать “полуавтоматической” (Kasnesi et al., 2023): преподаватель может сверять результат своей проверки с результатом ГМ, а также использовать в обратной связи студенту актуальные с точки зрения задач обучения и потребностей учащегося комментарии, при необходимости меняя их формат.

У преподавателя может возникнуть искушение переложить обратную связь по письменным работам на ИИ полностью, ведь ГМ объективно оценивают письменные работы и дают развернутые конструктивные комментарии. Но в большей степени не сама обратная связь, а то, как получатель с ней работает, определяет ее эффективность и обучающий потенциал (Winstone et al., 2021). К сожалению, даже студенты с большим опытом изучения иностранных языков слабо взаимодействуют с обратной связью по письменным работам (Боголепова и др., 2023). Если преподаватель не выступает в роли медиатора между генеративной моделью и обучающимся, видоизменяя обратную связь таким образом, чтобы она стала понятной и доступной учащемуся, а также учитывала его индивидуальные особенности и специфику образовательного контекста, вероятность эффективного взаимодействия студента с обратной связью становится еще ниже. Задачей преподавателя также является проследить, что реципиенты проработали полученную обратную связь, например, переписав работу с учетом обратной связи после ее обсуждения в классе (Meyer et al., 2024).

Обучающиеся отмечают детальность и информативность обратной связи, сформулированной ГМ, но в то же время ценят обратную связь от преподавателя, особенно если она дается в форме устного обсуждения (Escalante et al., 2023). У будущих преподавателей должно быть сформировано понимание того, когда и как может использоваться потенциал ГМ в соответствии с методической целесообразностью их применения. Быстрое развитие технологий ИИ создает ситуацию, в которой необходима постоянная подготовка и переподготовка специалистов, способных использовать экспертные знания при работе с ГМ и принимать профессионально обоснованные решения по их использованию (Константинова и др., 2023).

4.4. Ограничения исследования

Проведенное исследование имеет некоторые ограничения. Анализ проведен на небольшой выборке, т.к. в нем рассматривались оригинальные исследования, проведенные в последние несколько лет (с 2022 до начала 2025 года). Дальнейшие исследования возможностей ГМ для проверки письменных работ и формулировки обратной связи в различных контекстах дадут возможность уточнить и расширить полученные результаты. Был проведен только качественный анализ исследований, потому что они значительно отличались по методам сбора и интерпретации итогов, что, однако, не помешало выявить сходные результаты.

5. ВЫВОДЫ

Оценивание письменных работ и формулировка обратной связи при помощи ГМ имеет как преимущества, так и серьезные недостатки. Преподавателям, использующим (или планирующим использовать) модели для решения данных профессиональных задач, нужно знать эти недостатки, чтобы понимать, по каким параметрам сгенерированную информацию нужно критически оценивать и дорабатывать. Необходимо осознавать слабые стороны ГМ, такие как ограниченная надежность оценки, избыточность и неконкретность обратной связи, галлюцинации, недоступность комментариев для студентов с низким уровнем владения языком и др.

Широкодоступные инструменты ИИ не были созданы для решения задач обучения и появились недавно, поэтому возможности их использования в педагогических целях могут быть оценены только экспериментально. Дальнейшие исследования могут быть проведены на больших текстовых выборках, анализировать результаты проверки текстов различных жанров, сравнивать результаты оценки различными ГМ.

Конфликт интересов:

Автор заявляет об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. 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:10.24412/2224-0772-2024-5-123-137
2. Боголепова С.В., Кирсанова М.А., Пивоварова А.А. Как аукнется, так и откликнется? Отношение студентов-лингвистов к взаимооцениванию и обратной связи // Вестник Томского государственного университета. 2023. № 487. С. 118–128. doi:10.17223/15617793/487/14
3. Константинова Л.В., Ворожихин В.В., Петров А.М., Титова Е.С., Штырно Д.А. Генеративный искусственный интеллект в образовании: дискуссии и прогнозы [Электронный ресурс] // Открытое образование. 2023. №2. С. 36-48. URL: <https://cyberleninka.ru/article/n/generativnyy-iskusstvennyy-intellekt-v-obrazovanii-diskussii-i-prognozy> (дата обращения: 18.11.2024).
4. Корнев А.А. Обратная связь в обучении и педагогическом общении [Электронный ресурс] // Rhema. Рема. 2018. №2. С. 112-127. URL: <https://cyberleninka.ru/article/n/obratnaya-svyaz-v-obuchenii-i-pedagogicheskom-obschenii> (дата обращения: 18.11.2024).
5. Сысоев П.В. Использование технологий искусственного интеллекта в обучении иностранному языку: тематика методических работ за 2023 год и перспективы дальнейших исследований // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2024. № 29(2). С. 294-308. <https://doi.org/10.20310/1810-0201-2024-29-2-294-308>
6. Сысоев П.В., Филатов Е.М. Методика обучения учащихся и студентов написанию эссе в триаде «обучающийся — преподаватель — искусственный интеллект» // Вестн. Моск. ун-та. Сер. 19. Лингвистика и межкультурная коммуникация. 2024. Т. 27. № 2. С. 38–54. doi:10.55959/MSU-2074-1588-19-27-2-3

7. Сысоев П.В., Филатов Е.М., Хмаренко Н.И., Мурунов С.С. Преподаватель vs искусственный интеллект: сравнение качества предоставляемой преподавателем и генеративным искусственным интеллектом обратной связи при оценке письменных творческих работ студентов // *Перспективы науки и образования*. 2024. № 5 (71). doi:10.32744/pse.2024.5.41
8. Титова С.В. Технологические решения на базе искусственного интеллекта в обучении иностранным языкам: аналитический обзор // *Вестник Московского университета. Сер. 19. Лингвистика и межкультурная коммуникация*. 2024. Т. 27. № 2. С. 18–37. doi: 10.55959/MSU-2074-1588-19-27-2-2
9. AI tool user numbers worldwide from 2020-2030 [Электронный ресурс] // Statista Research Department. 09.12.2024. URL: <https://www.statista.com/forecasts/1449844/ai-tool-users-worldwide> (дата обращения: 26.12.2024).
10. Almassaad A., Alajlan H., Alebaikan R. Student Perceptions of Generative Artificial Intelligence: Investigating Utilization, Benefits, and Challenges in Higher Education // *Systems*. 2024. Vol. 12. No. 10. Article 385. <https://doi.org/10.3390/systems12100385>
11. Amobonye A., Lalung J., Mheta G., Pillai S. Writing a Scientific Review Article: Comprehensive Insights for Beginners // *The Scientific World Journal*. 2024. Article 7822269. <https://doi.org/10.1155/2024/7822269>
12. Annamalai N., Ab Rashid R., Hashmi U., Mohamed M.H., Alqaryouti M., Sadeq A. Using chatbots for English language learning in higher education // *Computers and Education: Artificial Intelligence*. 2023. Vol. 5. Article 100153. <https://doi.org/10.1016/j.caeai.2023.100153>
13. Awidi I.T. Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool // *Computers and Education: Artificial Intelligence*. 2024. Vol. 6(3). Article 100226. <https://doi.org/10.1016/j.caeai.2024.100226>
14. Bahroun Z., Anane C., Ahmed V., Zacca A. Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis // *Sustainability*. 2023. Vol. 15(17). Article 12983. <https://doi.org/10.3390/su151712983>
15. Banihashem S. K., Kerman N. T., Noroozi O., Moon J., Drachsler H. Feedback sources in essay writing: peer-generated or AI-generated feedback? // *International Journal of Educational Technology in Higher Education*. 2024. Vol. 21. № 1. Article 23. <https://doi.org/10.1186/s41239-024-00455-4>
16. Benedetto L., Gaudeau G., Caines A., BATTERY P. Assessing how accurately large language models encode and apply the common European framework of reference for languages // *Computers and Education: Artificial Intelligence*. 2025. Vol. 8. Article 100353. <https://doi.org/10.1016/j.caeai.2024.100353>
17. Bouziane K., Bouziane A. AI versus human effectiveness in essay evaluation // *Discover Education*. 2024. Vol. 3(1). Article 201. <https://doi.org/10.1007/s44217-024-00320-6>
18. Chan C.K.Y., Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education // *International Journal of Educational Technology in Higher Education*. 2023. Vol. 20(1). Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
19. Chan S., Lo N., Wong A. Generative AI and Essay Writing: Impacts of Automated Feedback on Revision Performance and Engagement. *rEFlections*. 2024. Vol. 31. № 3. 1249-1284. <https://doi.org/10.61508/refl.v31i3.277514>
20. Crompton H., Burke D. Artificial intelligence in higher education: the state of the field // *International Journal of Educational Technology in Higher Education*. 2023. Vol. 20(1). Article 22. <https://doi.org/10.1186/s41239-023-00392-8>

21. Crompton H., Edmett A., Ichaporia N., Burke D. AI and English language teaching: Affordances and challenges // *British Journal of Educational Technology*. 2024. Vol. 55. № 6. P. 2503-2529. <https://doi.org/10.1111/bjet.13460>
22. Dai W., Lin J., Jin F., Li T., Tsai Y., Gasevic D., Chen G. Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT // *EdArXiv Preprint*. 2023. <https://doi.org/10.35542/osf.io/hcgzj>
23. 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. Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
24. Evans J., Benefield P. Systematic Reviews of Educational Research: Does the Medical Model Fit? // *British Educational Research Journal*. 2001. Vol. 27. № 5. P. 527-541. <https://doi.org/10.1080/01411920120095717>
25. Fan L., Li L., Ma Z., Lee S., Yu H., Hemphill L. A Bibliometric Review of Large Language Models Research from 2017 to 2023 // *ACM Transactions on Intelligent Systems and Technology*. 2024. Vol. 15. № 5. Article 91. P. 1-25. <https://doi.org/10.1145/3664930>
26. Gombert S., Fink A., Giorgashvili T., Jivet I., Di Mitri D., Yau J., Frey A., Drachsler H. From the Automated Assessment of Student Essay Content to Highly Informative Feedback: a Case Study // *International Journal of Artificial Intelligence in Education*. 2024. Volume 34. P. 1378-1416. <https://doi.org/10.1007/s40593-023-00387-6>
27. 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. <https://doi.org/10.1007/s10639-023-12146-0>
28. Gupta P., Ding B., Guan Ch., Ding D. Generative AI: A systematic review using topic modelling techniques // *Data and Information Management*. 2024. Vol. 8. № 2. Article 100066. <https://doi.org/10.1016/j.dim.2024.100066>
29. Hattie J., Timperley H. The Power of Feedback // *Review of Educational Research*. 2007. Vol. 77. № 1. Pp. 81–112. <https://doi.org/10.3102/003465430298487>
30. Hirunyasiri D., Thomas D.R., Lin J., Koedinger K.R., Alevan V. Comparative analysis of GPT-4 and human graders in evaluating praise given to students in synthetic dialogues // *EdArXiv Preprint*. 2023. <https://doi.org/10.48550/arXiv.2307.02018>
31. Hussein M.A., Hassan H., Nassef M. Automated language essay scoring systems: a literature review // *PeerJ Computer Science*. 2019. Vol. 5. Article e208. <http://doi.org/10.7717/peerj-cs.208>
32. Iorliam A., Ingio J.A. A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing // *Journal of Computing Theories and Applications*. 2024. Vol. 1. № 3. P. 311–325. <https://doi.org/10.62411/jcta.9447>
33. Jansen T., Höft L., Bahr L., Fleckenstein J., Möller J., Köller O., Meyer J. Comparing Generative AI and Expert Feedback to Students' Writing: Insights from Student Teachers // *Psychologie in Erziehung und Unterricht*. 2024. Vol. 71. No. 2. P. 80–92. <http://dx.doi.org/10.2378/peu2024.art08d>
34. Jauhainen J.S., Garagorry Guerra A. Generative AI in education: ChatGPT-4 in evaluating students' written responses // *Innovations in Education and Teaching International*. 2024. Pp. 1–18. <https://doi.org/10.1080/14703297.2024.2422337>
35. Jiang Z., Xu Z., Pan Z., He J., Xie K. Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning // *Languages*. 2023. Vol. 8. № 4. Article 247. <https://doi.org/10.3390/languages8040247>

36. Kasneci E. et al. ChatGPT for good? On opportunities and challenges of large language models for education // *Learning and Individual Differences*. 2023. Vol. 103. Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
37. Kinder A. et al. Effects of adaptive feedback generated by a large language model: A case study in teacher education // *Computers and Education: Artificial Intelligence*. 2025. Vol. 8. Article 100349. <https://doi.org/10.1016/j.caeai.2024.100349>
38. Kumar R. Faculty members' use of artificial intelligence to grade student papers: a case of implications // *International Journal for Educational Integrity*. 2023. Vol. 19. № 1. Article 9. <https://doi.org/10.1007/s40979-023-00130-7>
39. Li B., Lowell V.L., Wang Ch., Li X. A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching // *Computers and Education: Artificial Intelligence*. 2024. Volume 7. Article 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
40. Li W., Liu H. Applying large language models for automated essay scoring for non-native Japanese // *Humanities and Social Sciences Communications*. 2024. Vol. 11. Article 723. <https://doi.org/10.1057/s41599-024-03209-9>
41. Lin Sh., Crosthwaite P. The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback // *System*. 2024. Vol. 127. Article 103529. <https://doi.org/10.1016/j.system.2024.103529>
42. Lye Ch.Y., Lim L. Generative Artificial Intelligence in Tertiary Education: Assessment Redesign Principles and Considerations // *Education Sciences*. 2024. Vol. 14. № 6. Article 569. <https://doi.org/10.3390/educsci14060569>
43. Meyer J., Jansen T., Schiller R., Liebenow L.W., Steinbach M., Horbach A., Fleckenstein J. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions // *Computers and Education: Artificial Intelligence*. 2024. Vol. 6. Article 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
44. Mizumoto A., Eguchi M. Exploring the potential of using an AI language model for automated essay scoring // *Research Methods in Applied Linguistics*. 2023. Vol. 2. № 2. Article 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
45. 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. № 2. Article 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
46. Pack A., Barrett A., Escalante J. Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability // *Computers and Education: Artificial Intelligence*. 2024. Vol. 6. Article 100234. <https://doi.org/10.1016/j.caeai.2024.100234>
47. Saini A.K., Cope B., Kalantzis M., Zapata G.C. The Future of Feedback: Integrating Peer and Generative AI Reviews to Support Student Work // *EdArXiv preprint*. 2024. <https://doi.org/10.35542/osf.io/x3dct>
48. Sidorkin A.M. Embracing Chatbots in Higher Education: The Use of Artificial Intelligence in Teaching, Administration, and Scholarship. New York: Routledge, 2024. <https://doi.org/10.4324/9781032686028>
49. Stahl M., Biermann L., Nehring A., Wachsmuth H. Exploring LLM Prompting Strategies for Joint Essay Scoring and Feedback Generation [Электронный ресурс] // *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*. Mexico City, Mexico. Association for Computational Linguistics, 2024. Pp. 283–298. URL: <https://aclanthology.org/2024.bea-1.23> (дата обращения: 15.12.2024).

50. Steiss J., Tate T., Graham S., Cruz J., Hebert M., Wang J., Moon Y., Tseng W., Warschauer M., Olson C.B. Comparing the quality of human and ChatGPT feedback of students' writing // *Learning and Instruction*. 2024. Vol. 91. Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
51. Wang N., Wang X., Su Y.Sh. Critical analysis of the technological affordances, challenges and future directions of Generative AI in education: a systematic review // *Asia Pacific Journal of Education*. 2024. Vol. 44. № 1. P. 139–155. <https://doi.org/10.1080/02188791.2024.2305156>
52. Winstone N., Boud D., Dawson P., Heron M. From feedback-as-information to feedback-as-process: a linguistic analysis of the feedback literature // *Assessment & Evaluation in Higher Education*. 2021. Vol. 47. No. 2. Pp. 213–230. <https://doi.org/10.1080/02602938.2021.1902467>
53. Wu H., Wang W., Wan Y., Jiao W., Lyu M. ChatGPT or Grammarly? Evaluating ChatGPT on Grammatical Error Correction Benchmark // *ArXiv preprint*. 2023. arXiv:2303.13648. <https://doi.org/10.48550/arxiv.2303.13648>
54. Yu, H., Guo Y. Generative artificial intelligence empowers educational reform: current status, issues, and prospects // *Frontiers in Education*. 2023. Vol. 8. Article 1183162. <https://doi.org/10.3389/feduc.2023.1183162>

REFERENCES

1. Almassaad, A., Alajlan, H., & Alebaikan, R. (2024). Student Perceptions of Generative Artificial Intelligence: Investigating Utilization, Benefits, and Challenges in Higher Education. *Systems*, 12(10), 385. <https://doi.org/10.3390/systems12100385>
2. Amobonye, A., Lalung, J., Mheta, G., & Pillai, S. (2024). Writing a Scientific Review Article: Comprehensive Insights for Beginners. *The Scientific World Journal*, 7822269. <https://doi.org/10.1155/2024/7822269>
3. Annamalai, N., Rashid, R.A., Munir Hashmi, U., Mohamed, M.H., Alqaryouti, M., & Sad-
eq, A. (2023). Using chatbots for English language learning in higher education. *Computers and Education: Artificial Intelligence*, 5, 100153. <https://doi.org/10.1016/j.caeai.2023.100153>
4. Awidi, I.T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, 6(3), 100226. <https://doi.org/10.1016/j.caeai.2024.100226>
5. Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability*, 15(17), 12983. <https://doi.org/10.3390/su151712983>
6. Banihashem, S.K., Kerman, N.T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1), 23. <https://doi.org/10.1186/s41239-024-00455-4>
7. Benedetto, L., Gaudeau, G., Caines, A., & Buttery, P. (2025). Assessing how accurately large language models encode and apply the common European framework of reference for languages. *Computers and Education: Artificial Intelligence*, 8, 100353. <https://doi.org/10.1016/j.caeai.2024.100353>

8. Bogolepova, S.V., & Zharkova, M.G. (2024). Issledovanie potentsiala generativnykh modelei dlya otsenivaniya esse i obespecheniya obratnoy svyazi [Researching the potential of generative language models for essay evaluation and feedback provision]. *Domestic and Foreign Pedagogy*, 1(5), 123–137 (in Russian). <https://doi.org/10.24412/2224-0772-2024-101-123-137>
9. Bogolepova, S.V., Kirsanova, M.A., & Pivovarova, A.A. (2023). Kak auknetsya, tak i otkliknetsya? Otnoshenie studentov-lingvistov k vzaimootsenivaniyu i obratnoy svyazi [As you give, you will receive? Language students' attitudes to peer-assessment and feedback]. *Vestnik Tomskogo gosudarstvennogo universiteta [Tomsk State University Journal]*, 487, 118–128 (in Russian). <https://doi.org/10.17223/15617793/487/14>
10. Bouziane, K., & Bouziane, A. (2024). AI versus human effectiveness in essay evaluation. *Discover Education*, 3(1), 201. <https://doi.org/10.1007/s44217-024-00320-6>
11. Chan, C.K.Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>
12. Chan, S., Lo, N., & Wong, A. (2024). Generative AI and Essay Writing: Impacts of Automated Feedback on Revision Performance and Engagement. *rEFLections*, 31(3), 1249–1284. <https://doi.org/10.61508/refl.v31i3.277514>
13. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>
14. Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503–2529. <https://doi.org/10.1111/bjet.13460>
15. Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y., Gasevic, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. *EdArXiv Preprint*. <https://doi.org/10.35542/osf.io/hcgzj>
16. Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 57. <https://doi.org/10.1186/s41239-023-00425-2>
17. Evans, J., & Benefield, P. (2001). Systematic Reviews of Educational Research: Does the Medical Model Fit? *British Educational Research Journal*, 27(5), 527–541. <https://doi.org/10.1080/01411920120095717>
18. Fan, L., Li, L., Ma, Z., Lee, S., Yu, H., & Hemphill, L. (2024). A bibliometric review of large language models research from 2017 to 2023. *ACM Transactions on Intelligent Systems and Technology*, 15(5), 91, 1–25. <https://doi.org/10.1145/3664930>
19. Gombert, S., Fink, A., Giorgashvili, T., Jivet, I., Di Mitri, D., Yau, J., Frey, A., Drachsler, H. (2024). From the automated assessment of student essay content to highly informative feedback: A case study. *International Journal of Artificial Intelligence in Education*, 34, 1378–1416. <https://doi.org/10.1007/s40593-023-00387-6>
20. Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
21. Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. *Data and Information Management*, 8(2), 100066. <https://doi.org/10.1016/j.dim.2024.100066>

22. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
23. Hirunyasiri, D., Thomas, D.R., Lin, J., Koedinger, K.R., & Aleven, V. (2023). Comparative analysis of GPT-4 and human graders in evaluating praise given to students in synthetic dialogues. *EdArXiv Preprint*. <https://doi.org/10.48550/arXiv.2307.02018>
24. Hussein, M.A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science*, 5, e208. <http://doi.org/10.7717/peerj-cs.208>
25. Iorliam, A., & Ingio, J.A. (2024). A comparative analysis of generative artificial intelligence tools for natural language processing. *Journal of Computing Theories and Applications*, 1(3), 311–325. <https://doi.org/10.62411/jcta.9447>
26. Jansen, T., Höft, L., Bahr, L., Fleckenstein, J., Möller, J., Köller, O., & Meyer, J. (2024). Comparing generative AI and expert feedback to students' writing: Insights from student teachers. *Psychologie in Erziehung und Unterricht*, 71(2), 80–92. <https://doi.org/10.2378/peu2024.art08d>
27. Jauhainen, J.S., & Garagorry Guerra, A. (2024). Generative AI in education: ChatGPT-4 in evaluating students' written responses. *Innovations in Education and Teaching International*, 1-18. <https://doi.org/10.1080/14703297.2024.2422337>
28. Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages*, 8(4), 247. <https://doi.org/10.3390/languages8040247>
29. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
30. Kinder, A., Briese, F.J., Jacobs, M., Dern, N., Glodny, N., Jacobs, S., Leßmann, S. (2025). Effects of adaptive feedback generated by a large language model: A case study in teacher education. *Computers and Education: Artificial Intelligence*, 8, 100349. <https://doi.org/10.1016/j.caeai.2024.100349>
31. Konstantinova, L.V., Vorozhikhin, V.V., Petrov, A.M., Titova, E.S., & Shtykhno, D.A. (2023). Generativnyy iskusstvennyy intellekt v obrazovanii: diskussii i prognozy [Generative Artificial Intelligence in Education: Discussions and Forecasts]. *Open Education*, 27(2), 36-48. Retrieved 2024, November 18, from <https://cyberleninka.ru/article/n/generativnyy-iskusstvennyy-intellekt-v-obrazovanii-diskussii-i-prognozy> (in Russian).
32. Korenev, A.A. (2018). Obratnaya svyaz' v obuchenii i pedagogicheskom obshchenii [Feedback in learning, teaching and educational communication]. *Rhema*, 2, 112–127. Retrieved 2024, November 18, from <https://cyberleninka.ru/article/n/obratnaya-svyaz-v-obuchenii-i-pedagogicheskom-obshchenii/viewer> (in Russian).
33. Kumar, R. (2023). Faculty members' use of artificial intelligence to grade student papers: A case of implications. *International Journal for Educational Integrity*, 19(1), 9. <https://doi.org/10.1007/s40979-023-00130-7>
34. Li, B., Lowell, V.L., Wang, Ch., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
35. Li, W., & Liu, H. (2024). Applying large language models for automated essay scoring for non-native Japanese. *Humanities and Social Sciences Communications*, 1, 723. <https://doi.org/10.1057/s41599-024-03209-9>

36. Lin, Sh., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, 103529. <https://doi.org/10.1016/j.system.2024.103529>
37. Lye, C.Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences*, 14(6), 569. <https://doi.org/10.3390/educsci14060569>
38. Meyer, J., Jansen, T., Schiller, R., Liebenow, L.W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
39. Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
40. Mizumoto, A., Shintani, N., Sasaki, M., & Feng Teng, M. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
41. Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence*, 6, 100234. <https://doi.org/10.1016/j.caeai.2024.100234>
42. Saini, A.K., Cope, B., Kalantzis, M., & Zapata, G.C. (2024). The Future of Feedback: Integrating Peer and Generative AI Reviews to Support Student Work. *EdArXiv preprint*. <https://doi.org/10.35542/osf.io/x3dct>
43. Sidorkin, A.M. (2024). *Embracing chatbots in higher education: The use of artificial intelligence in teaching, administration, and scholarship*. New York: Routledge. <https://doi.org/10.4324/9781032686028>
44. Stahl, M., Biermann, L., Nehring, A., & Wachsmuth, H. (2024). Exploring LLM prompting strategies for joint essay scoring and feedback generation. In E. Kochmar et al. (Eds.), *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)* (pp. 283–298). Association for Computational Linguistics.
45. Statista Research Department. (2024, December 9). *AI tool user numbers worldwide from 2020-2030*. Retrieved 2024, December 26, from <https://www.statista.com/forecasts/1449844/ai-tool-users-worldwide>
46. Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C.B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
47. Sysoyev, P.V. (2024). The use of artificial intelligence technologies in foreign language teaching: the subject of methodological works for 2023 and prospects for further research. *Tambov University Review. Series: Humanities*, 29(2), 294-308 (in Russian). <https://doi.org/10.20310/1810-0201-2024-29-2-294-308>
48. Sysoyev, P.V., & Filatov, E.M. (2024). Metodika obucheniya uchaschikhsya i studentov napisaniyu esse v triade «obuchayushchiesya — prepodavatel' — iskusstvennyy intellekt» [A method for teaching foreign language creative writing to students within the framework “Learner — Teacher — Artificial Intelligence.”] *Lomonosov Linguistics and Intercultural Communication Journal*, 27(2), 38–54 (in Russian). <https://doi.org/10.55959/MSU-2074-1588-19-27-2-3>

49. Sysoyev, P.V., Filatov, E.M., Khmarenko, N.I., & Murunov, S.S. (2024). Prepodavatel' vs iskusstvennyy intellekt: sravnenie kachestva predostavlyаемой преподавателем i generativnym iskusstvennym intellektom obratnoy svyazi pri otsenke pismennykh tvorcheskikh rabot studentov [Teacher vs. Artificial Intelligence: A comparison of the quality of feedback provided by a teacher and generative artificial intelligence in assessing students' creative writing]. *Perspektivy nauki i obrazovaniya [Perspectives of Science and Education]*, 71(5), 694–712 (in Russian). <https://doi.org/10.32744/pse.2024.5.41>

50. Titova, S. V. (2024). Tekhnologicheskie resheniya na baze iskussvennogo intellekta v obuchenii inostrannym yazykam: analiticheskiy obzor [Technological solutions based on artificial intelligence in teaching foreign languages: An analytical review]. *Lomonosov Linguistics and Intercultural Communication Journal*, 27(2), 18–37 (in Russian). <https://doi.org/10.55959/MSU-2074-1588-19-27-2-2>

51. Wang, N., Wang, X., & Su, Y.S. (2024). Critical analysis of the technological affordances, challenges and future directions of generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 44(1), 139–155. <https://doi.org/10.1080/02188791.2024.2305156>

52. Winstone, N., Boud, D., Dawson, P., & Heron, M. (2021). From feedback-as-information to feedback-as-process: A linguistic analysis of the feedback literature. *Assessment & Evaluation in Higher Education*, 47(2), 213–230. <https://doi.org/10.1080/02602938.2021.1902467>

53. Wu, H., Wang, W., Wan, Y., Jiao, W., & Lyu, M. (2023). ChatGPT or Grammarly? Evaluating ChatGPT on Grammatical Error Correction Benchmark. *arXiv preprint*, arXiv:2303.13648. <https://doi.org/10.48550/arxiv.2303.13648>

54. Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: current status, issues, and prospects. *Frontiers in Education*, 8, 1183162. doi: 10.3389/fed-uc.2023.1183162

Информация об авторе:

Боголепова Светлана Викторовна – кандидат филологических наук, доцент, доцент школы иностранных языков Национального исследовательского университета «Высшая школа экономики», г. Москва. Научные интересы: ИКТ и искусственный интеллект в преподавании иностранных языков, разработка учебных курсов, академическое письмо. ORCID: 0000-0003-1050-9110

Получено: 20 ноября 2024 г.

Принято к публикации: 10 февраля 2025 г.

POTENTIAL OF ARTIFICIAL INTELLIGENCE TOOLS FOR TEXT EVALUATION AND FEEDBACK PROVISION

Svetlana V. Bogolepova

National Research University Higher School of Economics

sbogolepova@hse.ru

Abstract: The article aims to explore the potential of generative artificial intelligence (AI) for assessing written work and providing feedback on it. The goal of this research is to determine the possibilities and limitations of generative AI when used for evaluating students' written production and providing feedback. To accomplish the aim, a systematic review of twenty-two original studies was conducted. The selected studies were carried out in both Russian and international contexts, with results published between 2022 and 2025. It was found that the criteria-based assessments made by generative models align with those of instructors, and that generative AI surpasses human evaluators in its ability to assess language and argumentation. However, the reliability of this evaluation is negatively affected by the instability of sequential assessments, the hallucinations of generative models, and their limited ability to account for contextual nuances. Despite the detailisation and constructive nature of feedback from generative AI, it is often insufficiently specific and overly verbose, which can hinder student comprehension. Feedback from generative models primarily targets local deficiencies, while human evaluators pay attention to global issues, such as the incomplete alignment of content with the assigned topic. Unlike instructors, generative AI provides template-based feedback, avoiding indirect phrasing and leading questions contributing to the development of self-regulation skills. Nevertheless, these shortcomings can be addressed through subsequent queries to the generative model. It was also found that students are open to receiving feedback from generative AI; however, they prefer to receive it from instructors and peers. The results are discussed in the context of using generative models for evaluating written work and formulating feedback by foreign language instructors. The conclusion emphasises the necessity of a critical approach to using generative models in the assessment of written work and the importance of training instructors for effective interaction with these technologies.

Keywords: AI in education, generative models in language teaching, automatic text evaluation, feedback on written works, automated essay scoring.

How to cite this article: Bogolepova, S.V. (2025). Potential of Artificial Intelligence Tools for Text Evaluation and Feedback Provision. *Professional Discourse & Communication*, 7(1), 70–88 (in Russian). <https://doi.org/10.24833/2687-0126-2025-7-1-70-88>

About the author:

Svetlana V. Bogolepova, Cand. Sci. (Philology), is an Associate Professor at National Research University Higher School of Economics (Moscow, Russia). Her academic research interests lie in the areas of ICT and artificial intelligence in language teaching, course design, and academic writing. ORCID: 0000-0003-1050-9110.

Received: November 20, 2024.

Accepted: February 10, 2025.