

DOI: 10.25629/НС.2026.01.03

УДК: 37.026:004.9

ВАК: 5.8.2 Теория и методика обучения и воспитания

5.8.7 Методология и технология профессионального образования

2.3.8 Информатика и информационные процессы

## ОТДЕЛЬНЫЕ НЕГАТИВНЫЕ ПОСЛЕДСТВИЯ ПРИМЕНЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАЗОВАНИИ

Гарбук С.В.<sup>1</sup>, Мерецков О.В.<sup>2,3</sup>, Млякова М.Р.<sup>1</sup>

<sup>1</sup>Всероссийский институт научной и технической информации Российской академии наук

<sup>2</sup>Московский педагогический государственный университет

<sup>3</sup>Академия стандартизации, метрологии и сертификации (учебная)

### Аннотация

*В статье рассматриваются возможности и риски внедрения систем искусственного интеллекта в образование в контексте электронного обучения. Показано, что ИИ способен повысить вовлеченность обучающихся, поддерживать индивидуализацию обучения и снизить издержки, однако широкое применение в образовании получают преимущественно системы машинного обучения, чья работа часто недостаточно интерпретируема и может непредсказуемо деградировать в реальной эксплуатации. На основе анализа документированных кейсов и результатов тестирования больших языковых моделей при реферировании научных статей (200 публикаций ведущих журналов) выявляется ключевая угроза – «чрезмерное обобщение» с утратой существенных деталей и риском неверного толкования научных выводов. Дополнительно проанализированы примеры генерации исторически некорректного визуального контента, ложноположительного определения авторства студенческих работ средствами анти-плагиат/AI-detection, а также дискриминационных эффектов систем прокторинга в отношении обучающихся с ограниченными возможностями здоровья. Делается вывод о необходимости верификации ИИ-алгоритмов и/или СИИ перед их применением в образовательных процедурах, а также о недопустимости как безусловного доверия решениям ИИ, так и полного их игнорирования; ответственность за образовательные решения должна сопровождаться критической экспертной оценкой результатов работы ИИ.*

### Ключевые слова

*искусственный интеллект (ИИ), система искусственного интеллекта (СИИ), электронное обучение (ЭО), цифровой образовательный контент, доверенный искусственный интеллект, большие языковые модели (LLM)*

### Об авторах

**Гарбук Сергей Владимирович**, кандидат технических наук, Всероссийский институт научной и технической информации Российской академии наук, AuthorID: 598035, ORCID: 0000-0001-5385-3961, garbuk@list.ru

**Мерецков Олег Вадимович**, кандидат педагогических наук, Академия стандартизации, метрологии и сертификации (учебная), AuthorID: 1091906, ORCID: 0009-0007-2936-1730, oleg-jan@mail.ru

**Млякова Майя Рашидовна**, Всероссийский институт научной и технической информации Российской академии наук, m.mlyakova@viniti.ru

## Введение

В настоящее время технологии искусственного интеллекта (ИИ) активно интегрируются в отрасль образования [1]. Использование технологий ИИ открывает новые, уникальные возможности для образовательной системы, способствуя повышению вовлеченности учащегося в образовательный процесс, например, за счет игрофикации обучения на основе технологий дополненной реальности [2, с. 137], переносу фокуса преподавателя с рутинных задач, связанных с визуализацией учебного материала, составлением типовой документации, разработкой фонда оценочных средств на индивидуализацию обучения, снижению экономических затрат учебных заведений на лабораторные эксперименты, а также повышению уровня эффективности обучения [3, с. 274] и т.д.

Например, при сдаче «гаокао» – национального вступительного экзамена в университеты Китая, школьники, которые обучались с помощью системы Squirrel AI (формирует индивидуальную программу обучения на основе анализа уровня знаний учащегося), получают на 30–40% больше положительных оценок, чем те, кто не пользовался такой системой. Указанный экзамен является одним из самых конкурентных (в 2025 г. его сдали 13,35 млн человек) и влияет на карьеру и социальную мобильность в Китае.

В образовании преимущественно применяются системы искусственного интеллекта (СИИ), основанные на методах машинного обучения, которые принципиально не обладают полной интерпретируемостью для человека. Данный подход является весьма универсальным, но имеет существенный недостаток: качество работы алгоритмов машинного обучения зачастую непредсказуемо деградирует в ходе их реальной эксплуатации [4, с. 6].

Для рассмотрения возможностей применения подобных систем в процессе электронного обучения кратко перечислим возможные аспекты педагогической деятельности в соответствующем контексте:

- подготовка и проведение занятий;
- проверка результатов самостоятельной работы обучающихся (домашних заданий);
- промежуточная и итоговая аттестация [5-9].

## Методы

Методы исследования основаны на аналитическом обзоре, сравнительном анализе и case study, направленных на выявление рисков применения систем искусственного интеллекта в электронном обучении. Рассматривались СИИ, используемые для подготовки и проведения занятий, проверки самостоятельных работ, а также для промежуточной и итоговой аттестации, включая дистанционный прокторинг. Предметом анализа выступали проявления некорректной работы таких систем и их последствия для достоверности учебного содержания, справедливости оценивания и психологического состояния обучающихся.

Эмпирическая база включала опубликованные результаты тестирования больших языковых моделей при реферировании научных статей (в том числе на массиве из 200 публикаций ведущих журналов) и документированные кейсы: генерация фактически/исторически некорректных изображений, ложноположительное определение «AI-авторства» студенческих текстов, а также дискриминационные эффекты прокторинга в отношении обучающихся с ОВЗ.

Материалы отбирались при наличии проверяемого описания условий применения СИИ, связи с образовательным контекстом и возможности выделить тип ошибки и ее последствия.

Анализ проводился путем соотнесения каждого случая с соответствующей педагогической задачей, выделения характера сбоя (смысловая утрата при обобщении, фактические ошибки, ложноположительные срабатывания, ошибки из-за неучета индивидуальных особенностей) и оценки потенциального ущерба для обучения. Итогом стал синтез требований к верификации функциональной корректности СИИ и необходимости человеческой экспертной проверки результатов ИИ в критически значимых процедурах, особенно при оценивании и аттестации.

## Результаты и их обсуждение

Весной 2025 года в Канаде было проведено тестирование больших языковых моделей (ChatGPT-4o, ChatGPT-4.5, DeepSeek, LLaMA 3.3 70B, Claude 3.7 Sonnet и др.) для применения в контексте задачи реферирования научных статей. На вход систем последовательно подавались 200 статей из ведущих междисциплинарных (Science, Nature, Nature Human Behavior, and Psychological Science in the Public Interest) и медицинских (The New England Journal of Medicine, Lancet, JAMA, BMJ) научных журналов и вводилось текстовое описание задачи (создание реферата).

Результаты данного тестирования показали *чрезмерное обобщение научных выводов*, то есть утрату в процессе обобщения некоторых существенных деталей исследований, необходимых для правильного понимания содержания реферируемых источников. При этом было отмечено, что такое чрезмерное обобщение с частичной утратой значимой информации (фактически – потерей смысла) усиливается в новых языковых моделях и создает значительный риск неверного толкования результатов исследований [10].

Приведенное исследование доказательно иллюстрирует, что при решении задачи реферирования содержания из отобранных источников научной информации с применением систем искусственного интеллекта, основанных на больших языковых моделях (LLM), могут быть сформированы *некорректные научные выводы*, применение которых в образовательном процессе приводит к формированию ложных представлений об изучаемом явлении или предмете, что представляет собой *реальную опасность для общества*.

Другой документально подтвержденный случай применения СИИ, порождающих информацию, вводящую пользователя в заблуждение, описан в контексте создания сервисом Gemini AI от Google *исторически неточных изображений*. Например, когда Gemini попросили создать изображение американского сенатора XIX века, ИИ предложил картинки чернокожих женщин и женщин коренных американских наций. В действительности женщина впервые стала сенатором в США лишь в 1922 году, и она была белой.

Похожая история повторилась в запросе к СИИ создать изображение Папы Римского. Вместо того чтобы опубликовать фотографию одного из 266 понтификов за всю историю – все они были белыми мужчинами, – Gemini опубликовала фотографии женщины из Юго-Восточной Азии и чернокожего мужчины в церковных облачениях.

Пользователи социальных сетей развлекались тем, что создавали запросы, которые приводили к неожиданным результатам. «Новая игра: попробуйте заставить Google Gemini создать изображение мужчины европеоидной расы. Пока у меня ничего не получается», – написал пользователь Х Фрэнк Дж. Флеминг, журналист Babylon Bee, чья серия постов о Gemini в социальной сети быстро стала вирусной.

В другом примере Gemini попросили создать изображение викинга – скандинавского мореплавателя и разбойника, который когда-то наводил ужас на Европу. Среди странных изображений викингов, созданных чат-ботом Gemini, были чернокожий мужчина без рубашки с радужными перьями, прикрепленными к его меховому одеянию, чернокожая женщина-воин и азиат, стоящий посреди чего-то похожего на пустыню.

В результате развилась дискуссия между известными людьми в соцсетях, в процессе которой пользователи Gemini охарактеризовали данную СИИ как «абсурдно прогрессивную», а компания Google (разработчик системы) была вынуждена признать наличие проблемы и временно отключить функцию генерации изображений людей, до выполнения доработки системы [11].

Таким образом, применение СИИ с неподтвержденной функциональной корректностью для решения задачи повышения наглядности и интерактивности обучения также может приводить к *получению обучающимися ложных знаний*.

Примером некорректной работы систем ИИ при решении задачи проверки результатов самостоятельной работы обучающихся является ситуация, случившаяся в 2025 г. с Лей Бёррелл, студенткой факультета компьютерных наук Университета Хьюстона – Даунтаун.

Преподаватель поставил ей *ноль баллов за работу* на основании того, что онлайн сервис проверки текстовых работ Turnitin отметил ее, как созданную с помощью искусственного интеллекта. Помогла разрешить данную коллизию сохранившаяся *история создания и редактирования* работы в Google Docs, которая показала, что *Лей самостоятельно разрабатывала и редактировала работу*. После подачи студенткой апелляции, содержащей материалы, которые подтверждали факт самостоятельного написания работы (отметки времени, скриншоты и т.д.), оценка была выставлена без учета мнения нейросети относительно авторства текста [12].

По мнению автора данного исследования, преподаватели, которые не доверяют, равно как и те, кто безусловно доверяют выводам и решениям, созданным искусственным интеллектом, представляют опасность для развития общества. Первые – тормозят развитие технологии или будут преградой для эффективного применения ресурсов, вторые – могут применить ложные решения СИИ, не подвергнув их критической оценке с привлечением своего опыта и человеческого интеллекта [12], как описано в данном примере. В случае *применения СИИ с неподтвержденной функциональной корректностью* в оценивании результатов работы учащихся последствия могут быть критическими вследствие *негативного влияния на психологическое состояние* учащегося из-за незаслуженно поставленной низкой оценки.

В настоящее время уже известны случаи некорректной работы СИИ при их использовании в промежуточной и итоговой аттестации. Так, в статье «Lockdown Lessons: How Remote Proctoring Continues to Discriminate Against Disabled Students in a Post-COVID Era» поднимается важная проблема: при применении технологий прокторинга может возникать дискриминация учащихся с ограниченными возможностями.

Указанная ситуация возникает в связи с тем, что в целях предотвращения просмотра вспомогательных материалов за кадром могут использоваться алгоритмы локализации глаз и отслеживания движения глаз. Однако такие *алгоритмы не учитывают проявления*, связанные с нетипичным движением глаз и другими физическими реакциями, которые могут быть присущи людям с *ограниченными возможностями здоровья* (ОВЗ) [13, с. 86].

Например, студент колледжа Florida Gateway не сдал экзамен, поскольку система прокторинга HonorLock зафиксировала несколько случаев, когда лицо студента не попадало в поле зрения камеры, и он смотрел вниз или в сторону от экрана компьютера. Данный студент обратился к преподавателям и администрации колледжа, с разъяснениями о том, что нестандартные движения его глаз связаны с ограниченными возможностями здоровья, а HonorLock некорректно идентифицировал их как факт списывания, а затем подал жалобу в Управление по гражданским правам (Office for Civil Rights), была проведена проверка, в ходе которой было принято решение в пользу студента.

Также отмечается, что дискриминации могут подвергаться учащиеся с неконтролируемыми двигательными тиками, учащиеся, которые носят очки, учащиеся с ограничениями подвижности, нарушениями пищеварения и мочеиспускания, а также с нарушениями психического здоровья. Индивидуальные особенности обучающихся с ограниченными возможностями, как правило, не учитываются при разработке алгоритмов работы систем прокторинга.

Некорректная работа систем искусственного интеллекта при решении подобных задач также может негативно влиять на психоэмоциональное состояние учащегося. Эти вопросы рассматриваются в работе «Challenges and opportunities of AI in inclusive education: A case study of data-enhanced active reading in Japan» [14, с. 274].

## Выводы

Универсальность алгоритмов ИИ позволяет делегировать им задачи, ранее выполнявшиеся человеком. Вслед за этим происходит *перенос ответственности за некорректные действия по обработке данных с человека-оператора на других субъектов права*.

В случае, если алгоритмы ИИ предназначены для автоматизации процессов интеллектуальной обработки данных, ранее выполнявшихся человеком вручную, то *должна быть обеспечена возможность сравнения характеристик качества используемых алгоритмов ИИ и функциональных возможностей человека*, т.е. проведение так называемой верификации используемых алгоритмов ИИ (или самих систем искусственного интеллекта – при невозможности выделить алгоритмы в их классическом понимании) для применения в образовании.

## Библиография

1. Nja, C. O., Idiege, K. J., Uwe, U. E., Meremikwu, A. N., Ekon, E. E., Erim, C. M., ... & Cornelius-Ukpere, B. U. Adoption of artificial intelligence in science teaching: From the vantage point of the African science teachers // Smart Learning Environments, 10(1), 42.
2. Пшихачева А.А. Возможности использования искусственного интеллекта в практике высшего образования // Общество: социология, психология, педагогика. – 4(108). – С. 137-141.
3. Мерцалова Т.А., Косарецкий С.Г., Анчиков К.М. и др. (2022). Школьное образование в контексте национальных целей и приоритетных проектов: аналитический доклад. М.: НИУ ВШЭ.
4. Гарбук С.В. Об одном методе оценки репрезентативности тестирования алгоритмов искусственного интеллекта в средствах измерений/ Мягкие вычисления. –2024. – №10/2(71(2)). –С.6-12.
5. Мерецков О. В. Применение систем искусственного интеллекта для снижения трудоемкости сопровождения электронного обучения преподавателем / О. В. Мерецков // Человеческий капитал. – 2025. – № 5(197). – С. 113-126.
6. Мерецков, О. В. Системы искусственного интеллекта в информальном образовании взрослых: монография / О. В. Мерецков. – Москва : Директ-Медиа, 2025. – 268 с.
7. Мерецков О. В. Перспективные научные исследования в целях реализации новой стратегии научно-технологического развития Российской Федерации: Монография / А. А. Борисенко, А. А. Борисенко, Е. К. Васильева [и др.]. – Санкт-Петербург: Гуманитарный национальный исследовательский институт НАЦРАЗВИТИЕ, 2024. – 80 с.
8. Мерецков О. В. Применение искусственного интеллекта в мультилингвальной образовательной среде СНГ / А. В. Зажигалкин, Т. Т. Мансуров, О. В. Мерецков, Е. С. Червакова // Компетентность. – 2024. – № 2. – С. 3-11.
9. Мерецков О. В. Регулирование искусственного интеллекта в образовании / А. В. Зажигалкин, Т. Т. Мансуров, О. В. Мерецков // Компетентность. – 2024. – № 6. – С. 3-10.
10. Peters, U., Chin-Yee, B. Generalization bias in large language model summarization of scientific research // Royal Society open science. – 2025. – 12(4). – P. 241776.
11. Thomas Barrabi ‘Absurdly woke’: Google’s AI chatbot spits out ‘diverse’ images of Founding Fathers, popes, Vikings Published Feb. 21, 2024, 1:51 p.m. ET

12. Pelletier K. et al. EDUCAUSE Horizon Report Teaching and Learning Edition. – EDUC22, 2022. – С. 1-58

13. Johanson, A. Lockdown Lessons: How Remote Proctoring Continues to Discriminate Against Disabled Students in a Post-COVID Era // Journal of Law & Education. – 2025. – Vol. 54. – № 1.

14. Toyokawa, Y.; Horikoshi, I.; Majumdar, R.; Ogata, H. Challenges and opportunities of AI in inclusive education: A case study of data-enhanced active reading in Japan // Smart Learn. Environ. – 2023. – № 10(67). – P. 1-20.

Поступила в редакцию: 28.11.25

Принята к публикации: 20.01.26

## SOME NEGATIVE CONSEQUENCES OF USING ARTIFICIAL INTELLIGENCE SYSTEMS IN EDUCATION

Garbuk S.V.<sup>1</sup>, Meretskov O.V.<sup>2,3</sup>, Mlyakova M.R.<sup>1</sup>

<sup>1</sup>All-Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences

<sup>2</sup>Moscow Pedagogical State University

<sup>3</sup>Academy of Standardization, Metrology and Certification (educational)

### Abstract

*This article examines the potential and risks of implementing artificial intelligence systems in education in the context of e-learning. It demonstrates that AI can enhance student engagement, support individualized learning, and reduce costs. However, machine learning systems, whose performance is often poorly interpretable and can degrade unpredictably in real-world use, are predominantly used in education. Based on an analysis of documented case studies and the results of testing large language models during the review of scientific articles (200 publications from leading journals), a key threat is identified: "overgeneralization," which leads to the loss of essential details and the risk of misinterpretation of scientific findings. Additionally, examples of the generation of historically inaccurate visual content, false positive authorship of student papers using anti-plagiarism/AI detection tools, and the discriminatory effects of proctoring systems against students with disabilities are analyzed. The conclusion is drawn about the need to verify AI algorithms and/or artificial intelligence systems before their use in educational procedures, as well as the inadmissibility of either unconditional trust in AI decisions or complete ignoring them; responsibility for educational decisions should be accompanied by a critical expert assessment of the AI's results.*

### Keywords

*artificial intelligence (AI), artificial intelligence system, e-learning, digital educational content, Trusted Artificial Intelligence, Large Language Model (LLM)*