

Сергин Савелий Андреевич

магистрант

ФГБОУ ВО «Ульяновский государственный педагогический университет им. И.Н. Ульянова»

Ленгесова Наталья Анатольевна

канд. биол. наук, доцент, заведующая кафедрой,
старший научный сотрудник

Научно-исследовательский центр фундаментальных
и прикладных проблем биоэкологии и биотехнологии

ФГБОУ ВО «Ульяновский государственный педагогический университет им. И.Н. Ульянова»

г. Ульяновск, Ульяновская область

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОЦЕНИВАНИИ ЕДИНОГО ГОСУДАРСТВЕННОГО ЭКЗАМЕНА ПО БИОЛОГИИ

Аннотация: рассматривается применение GPT-4, усовершенствованной языковой модели искусственного интеллекта для оценивания заданий второй части Единого государственного экзамена (ЕГЭ) по биологии. Результаты показывают, что данный инструмент сопоставим с работой эксперта при выполнении четко определенных заданий, но демонстрирует вариативность субъективных оценок, поэтому может облегчить работу педагога, при этом контроль учителя остается крайне важным.

Ключевые слова: Единый государственный экзамен по биологии, оценка знаний, искусственный интеллект, GPT-4, формирующее оценивание, автоматизированная проверка заданий, образовательные технологии, критерии оценивания.

Единый государственный экзамен (ЕГЭ) в России является важнейшим инструментом оценки знаний старшеклассников. Традиционно оцениванием зани-

мались эксперты, но этот процесс занимает много времени. Появление искусственного интеллекта, в частности продвинутых моделей, таких как GPT-4, предлагает потенциальное решение этих проблем, обеспечивая последовательное и объективное оценивание.

Цель данного исследования – оценить эффективность GPT-4 при оценивании заданий части 2 экзамена ЕГЭ по биологии.

Интеграция искусственного интеллекта в образовательную среду актуальна, так как данная технология может повысить эффективность, последовательность и справедливость выставления оценок [1, 2023]. Формирующее оценивание – важный компонент эффективного преподавания, обеспечивающий постоянную обратную связь, которая может направлять процессы преподавания и обучения. Беннетт [2, 2011] подчеркивает важнейшую роль формирующего оценивания в улучшении результатов обучения. Она позволяет учителям выявлять пробелы в обучении и оперативно их устранять, способствуя более индивидуализированному обучению. Николь и Макфарлейн-Дик [3, 2006] более подробно рассматривают принципы эффективной обратной связи, подчеркивая важность своевременной и конкретной обратной связи для развития саморегулируемого обучения. Они утверждают, что эффективное формирующее оценивание не только измеряет процесс обучения, но и активно способствует ему, вовлекая студентов в рефлексивную практику.

Появление технологий изменило формирующее оценивание, сделав его более интерактивным и основанным на данных. Пеллегрино [4, 2010] говорит о том, как технологии могут поддерживать формирующее оценивание, предоставляя инструменты, обеспечивающие мгновенную обратную связь и позволяющие непрерывно отслеживать прогресс учащихся. Эти достижения позволяют создавать более динамичные и увлекательные формы оценки, такие как цифровые тесты и симуляции.

Скализ и Гиффорд [5, 2006] представили концепцию вопросов с «промежуточными ограничениями» в компьютерных оценках, которые уравнивают

структурированный характер традиционных вопросов с несколькими вариантами ответов и открытый характер вопросов с эссе. Этот подход позволяет использовать технологию для создания оценок, которые одновременно масштабируются и способны фиксировать сложные ответы учащихся.

Недавние разработки в области искусственного интеллекта еще больше расширили возможности образовательной оценки. Ванг [6, 2008] исследует использование формирующего оценивания на основе веб-викторин, похожих на игры, которые вовлекают студентов в интерактивный и игровой процесс обучения. Эти инструменты, управляемые искусственным интеллектом, могут адаптироваться к индивидуальным потребностям учащихся, обеспечивая персонализированные задачи, которые повышают эффективность обучения.

Применение искусственного интеллекта, в частности таких моделей, как GPT-4, при выставлении оценок открывает перспективные возможности для повышения эффективности и согласованности. Он может быстро обрабатывать большие объемы ответов учащихся и выставлять объективные оценки на основе заранее определенных критериев.

В исследовании приняли участие учащиеся 11 естественнонаучного класса Лицея ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», которые выполняли задания второй части ЕГЭ по биологии под наблюдением педагога, чтобы избежать списывания. Ответы оценивались как экспертом, так и GPT-4. Для оценки GPT-4 были созданы подсказки, включающие описание задания, элементы правильного ответа, критерии оценивания и решение учащегося.

Парные t-тесты использовались для сравнения оценок GPT-4 и экспертной оценки, определяя значимость любых различий.

Результаты оценивания с помощью GPT-4 по разным заданиям демонстрируют различный уровень согласия с экспертными оценками. Средние баллы, вариация и диапазон оценок представлены в таблицах 1 и 2.

Таблица 1

Результаты оценивания с помощью GPT4

Задание	Средняя оценка	вариация	минимум	максимум
---------	----------------	----------	---------	----------

22	2,33	0,25	2	3
23	1,00	1,50	0	3
24	0,33	0,25	0	1
25	1,44	2,03	0	3
26	0,89	1,36	0	3

Эти результаты иллюстрируют изменчивость согласованности оценок GPT-4 в разных типах заданий: в некоторых из них наблюдаются узкие отклонения, а в других – более значительные расхождения в выставленных баллах.

Таблица 2

Результаты оценивания эксперта

Задание	среднее	вариация	минимум	максимум
22	2,33	0,25	2	3
23	1,00	1,50	0	3
24	0,22	0,19	0	1
25	1,22	1,69	0	3
26	1,00	1,75	0	3

Оценка эксперта также демонстрирует разброс по заданиям, с особыми различиями в вариации и средних баллах по сравнению с GPT-4, особенно заметными в таких заданиях, как задание 24.

При боковом сравнении результатов оценивания, задания 22 и 23 показывают идентичные статистические показатели между GPT-4 и экспертом, что указывает на высокий уровень согласованности в этих заданиях. Однако расхождения возникают в таких заданиях, как задание 24, где средний балл, выставленный GPT-4 (0,33), немного выше, чем у эксперта (0,22), хотя оба показателя имеют низкую дисперсию, что говорит о некоторой согласованности в более низких баллах.

Задание 25 и Задание 26 демонстрируют более значительный разброс в оценках GPT-4 по сравнению с экспертом, что указывает на возможные различия в чувствительности к оценке или интерпретации критериев. Эти различия могут свидетельствовать о том, что, хотя GPT-4 в некоторых случаях может точно соответствовать экспертным оценкам, существуют нюансы в оценке субъективных

или сложных ответов, где искусственный интеллект может быть слишком мягким или строгим по сравнению с человеческим суждением.

Результаты показывают, что GPT-4 может надежно оценивать определенные типы экзаменационных заданий по биологии, обеспечивая последовательность и эффективность. Однако задания, требующие тонких суждений, показали некоторую вариативность, что подчеркивает необходимость человеческого контроля. Инструменты искусственного интеллекта, подобные GPT-4, могут дополнить традиционные методы оценивания, справляясь с рутинными оценками и освобождая преподавателей от необходимости сосредоточиться на сложных оценках. Такой гибридный подход может улучшить результаты обучения, сохраняя при этом точность оценок.

В дальнейшем предполагается изучить не только возможности применения искусственного интеллекта, но и его влияние на вовлеченность и обучение учащихся.

Список литературы

1. Лапина М.А. Особенности внедрения искусственного интеллекта в образовательный процесс / М.А. Лапина, М.Е. Токмакова, Д.А. Демин [и др.] // Auditorium. – 2023. – №3 (39) [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/n/osobennosti-vnedreniya-iskusstvennogo-intellekta-v-obrazovatelnyy-protsess> (дата обращения: 24.05.2024).
2. Bennett R.E. Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*. 2011. Vol. 18. No. 1. P. 5–25.
3. Nicol D.J. & Macfarlane-Dick D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*. 2006. Vol. 31. No. 2. P. 199–218.
4. Pellegrino J.W. Technology and formative assessment. *Review of Educational Psychology*. 2010. Vol. 21. P. 35–40.
5. Scalise K. & Gifford B. Computer-based assessment in e-learning: A framework for constructing «intermediate constraint» questions and tasks for technology

platforms. The Journal of Technology, Learning, and Assessment. 2006. Vol. 5. No. 6. P. 4–45.

6. Wang T.H. Web-based quiz-game-like formative assessment: Development and evaluation. Computers & Education. 2008. Vol. 51. No. 3, P. 1247–1263.