УДК 004.89

ОЦЕНКА ЭФФЕКТИВНОСТИ ГЕНЕРАТИВНЫХ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ В СОЗДАНИИ ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ВЫСШЕГО ОБРАЗОВАНИЯ

Гусятников В.Н. – д. ф.-м. н., профессор[1], Соколова Т. Н. – к.э.н., доцент[2]
[1] Поволжский институт управления имени П.А. Столыпина – филиал ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации»
[2] ФГБОУ ВО «Саратовская государственная юридическая академия», Россия, г. Саратов

Введение современных образовательных технологий актуализирует вопрос автоматизации процессов создания оценочных средств [1]. Генеративные нейросетевые модели представляют значительный интерес как инструмент, способный снизить временные затраты преподавателей на разработку тестовых заданий (ТЗ) [2, 3]. Однако их эффективность для создания заданий, соответствующих продвинутому уровню трудности в системе высшего образования, требует тщательного изучения [4-6].

В рамках данного исследования проведен сравнительный анализ четырех популярных нейросетевых моделей: DeepSeek, GigaChat, Qwen3-Max и Алиса. Анализ осуществлялся на основе единого промпта, требовавшего генерации не менее пяти ТЗ продвинутого уровня различных типов с последующим выводом в форматах, пригодных для загрузки в систему дистанционного обучения Moodle и для печати.

Методология и критерии оценки

В качестве исходного материала для генерации ТЗ использован текст из одного учебного пособия [7] на тему «Автоматизированные информационные системы».

Для оценки результатов работы моделей был применен комплекс критериев:

1. Спектр типологии ТЗ: способность генерировать разнообразные типы заданий (множественный выбор, соответствие, верно/неверно, короткий ответ, эссе).

2. Соблюдение количественных требований: выполнение условия о минимальном количестве заданий каждого типа.

3. Уровень трудности: соответствие сгенерированных ТЗ продвинутому, а не базовому уровню.

4. Наличие ключей ответов: предоставление правильных ответов для заданий закрытого типа.

5. Критерии оценки для открытых заданий: наличие четких параметров для оценки развернутых ответов и эссе.

6. Полнота обработки контента: способность модели анализировать весь предоставленный текстовый массив.

7. Функциональность экспорта: наличие готовых форматов вывода (таких как GIFT или XML для LMS Moodle) и инструкций по их использованию.

Результаты и обсуждение

Проведенный анализ выявил существенные различия в эффективности рассмотренных нейросетевых моделей.

Модель **DeepSeek** продемонстрировала наивысшую результативность. Ею были сгенерированы задания всех запрошенных типов в требуемом количестве. Модель корректно обработала весь исходный текст, предоставила ключи ответов, а для заданий типа «эссе» – критерии оценки. Важным практическим преимуществом является вывод результатов в формате GIFT с возможностью скачивания файла, что обеспечивает прямую интеграцию с Moodle. Экспертная оценка качества ТЗ составила 4 балла из 5.

Модель **GigaChat** показала низкие результаты. Несмотря на генерацию нескольких типов заданий, вопросы соответствовали базовому уровню трудности. Ключи ответов и критерии оценки для открытых вопросов отсутствовали. Наиболее критичным является тот факт, что модель обработала лишь около 37% исходного текста, что ставит под сомнение релевантность сгенерированных заданий. Формат вывода (XML) не был снабжен инструкциями по загрузке.

Модель **Qwen3-Max** показала хорошие результаты по полноте обработки текста и генерации релевантных заданий, включая такой сложный тип, как «верно/неверно с обоснованием». Однако ключи ответов были предоставлены только по дополнительному запросу, а для заданий, требующих обоснования, отсутствовали четкие критерии оценки его полноты. Наличие формата GIFT для Moodle является положительным, но отсутствие инструкций снижает удобство использования.

Модель «**Алиса**» не выполнила базовые требования промпта по количеству генерируемых заданий. Часть заданий не соответствовала содержанию исходного текста. Отсутствие какого-либо формата для загрузки в Moodle означает необходимость ручной обработки результатов преподавателем, что нивелирует потенциальный выигрыш во времени.

Результаты исследования свидетельствуют о том, что в настоящее время лишь отдельные генеративные модели (в данном случае DeepSeek) способны комплексно решать задачу создания качественных тестовых заданий продвинутого уровня с минимальными трудозатратами со стороны преподавателя. Ключевыми проблемами остаются неполная обработка контекста, несоответствие заданий заявленному уровню сложности, отсутствие прозрачных критериев оценки и недостаточная функциональность экспорта.

Таким образом, генеративные нейросети могут рассматриваться как перспективный вспомогательный инструмент в арсенале преподавателя, однако их применение требует обязательного последующего экспертного контроля и доработки генерируемых материалов. Дальнейшее развитие направлено на

создание специализированных образовательных ИИ-решений, изначально учитывающих педагогические и метрологические требования к оценочным средствам.

Список литературы:

1. Избранные вопросы цифровой трансформации образования: Монография. – Москва: ООО "Научно-издательский центр Инфра-М", 2025. – 242 с. – EDN AHFCBS.

2. Гусятников В. Н., Соколова Т. Н., Безруков А. И., Каюкова И. В. Использование методов искусственного интеллекта для оценки компетенций в ходе тестирования // Информатика и образование. – 2023. – Т. 38, № 6. – С. 75-85. – DOI 10.32517/0234-0453-2023-38-6-75-85. – EDN ABPXYT.

3. Гусятников В. Н., Соколова Т. Н., Безруков А. И., Каюкова И. В. Методика оценки набора компетенций на основе результатов тестирования // Фундаментальные исследования. – 2020. – № 12. – С. 209-215. – DOI 10.17513/fr.42935. – EDN IHYEON.

4. Искусственный интеллект в образовании: направления применения и ограничения / В. И. Абрамов, А. В. Гриншкун, А. В. Елисеев [и др.] // Современная {цифровая} дидактика. – Москва: ООО «А-Приор», 2023. – С. 89-98. – EDN DRHIEQ.

5. Паскова А. А. Возможности интеграции технологий генеративного искусственного интеллекта в процессы формирующего оценивания в высшем образовании // Вестник Майкопского государственного технологического университета. – 2024. – Т. 16, № 2. – С. 98-109. – DOI 10.47370/2078-1024-2024-16-2-98-109. – EDN PNMTVF.

6. Ортина Н. А., Гриншкун В.В. Использование нейронных сетей в образовании: вызовы и тренды // Фундаментальные проблемы обучения математике, информатике и информатизации образования: Сборник тезисов докладов XI международной научно-практической конференции в рамках Международного научно-образовательного форума «Математика. Информатика. Образование», Елец-Владикавказ, 29 июня – 05 2025 года. – Елец: Елецкий государственный университет им. И.А. Бунина, 2025. – С. 102-105. – EDN OUUDHB.

7. Прикладная информатика: справочник: учеб. пособие / Под ред. В.Н. Волковой и В.Н. Юрьева. – Эл. Изд. – 1 файл pdf: 767 с. М.: Финансы и статистика, 2021. – Текст: электронный.