

УДК 004.89

**ИСПОЛЬЗОВАНИЕ LLM для оценки ответов на открытые вопросы тестов**  
**Никифорова А.Д.** (Университет ИТМО), **Нестеренко Ю.А.** (Университет ИТМО),  
**Бармина Д.А.** (Университет ИТМО)  
**Научный руководитель – к.т.н. Хлопотов М.В.**  
(Университет ИТМО)

**Введение.** В текущую цифровую эпоху всё больше сфер деятельности человека подвергается цифровизации и автоматизации, и образование не является исключением [1]. Широко распространенным методом оценивания знаний студентов является тестирование, а в условиях повсеместного внедрения дистанционного формата обучения тестовый опрос зарекомендовал себя как основное средство контроля [2]. Среди инструментов проведения онлайн-тестирования особую популярность набирают интернет-формы, в частности Google Forms [3]. Такой формат поддерживает включение в тест открытых вопросов с развернутым ответом, что позволяет оценить способность студента самостоятельно формулировать мысли, в противовес закрытым вопросам с выбором ответа, где ответы часто «угадываются» [4]. Однако если оценивание закрытых вопросов легко автоматизировать, то открытые вопросы требуют ручной проверки, что значительно увеличивает нагрузку на преподавателя. Отсюда следует актуальность решения задачи автоматизации оценки ответов на открытые вопросы тестов.

**Основная часть.** Существующие системы оценки развернутых ответов на русском языке являются классификаторами (размечают ответы как правильные и неправильные), причем для пользователя процесс разметки проходит в «черном ящике» [5]. Так как конечное решение всегда остается за преподавателем, эффективность и объективность его работы могут быть повышены, если система будет предоставлять обоснование своего варианта оценки. Большие языковые модели (LLM) на данный момент являются самым эффективным и перспективным инструментом для работы с текстовыми данными [6]. Различные методы оценки производительности показывают, что LLM имеют достаточную гибкость для их применения в широком круге задач обработки естественного языка [7]. Предлагается консольное приложение на языке программирования Python, представляющее собой интерфейс для взаимодействия с LLM-моделью, оценивающей сходство двух текстов. В качестве входных данных инструмент принимает ответ студента и «эталонный» ответ преподавателя. Выходными данными являются:

- а) численная оценка сходства ответов от 0 до 1;
- б) мотивировка выставленной оценки и выявленных отличий ответа от «эталона» в текстовом виде.

Такой формат выходных данных позволяет преподавателю легко конвертировать его в оценку, а студенту получить развернутую обратную связь по данному им ответу.

**Выводы.** Разработана консольная версия приложения - помощника преподавателя в оценке открытых ответов на вопросы тестов. В перспективе приложение получит графический пользовательский интерфейс и будет интегрировано в систему кросс-тестирования, разрабатываемую в лаборатории LISA университета ИТМО.

### Список использованных источников:

1. M.Karimova. Digitalization in education // Science and innovation. 2022. Vol. 1 №8. pp. 1419-1422. URL: <https://zenodo.org/records/7433105> (дата обращения: 03.02.2024).
2. Секлетова Н.Н., Куваева Е.Н., Кондратьев А.И. Тестирование как инструмент оценки знаний // Экономика и социум : науч. журн. : электр. версия. 2020. №10 (77). С. 712. URL: <https://cyberleninka.ru/article/n/testirovanie-kak-instrument-otsenki-znaniy> (дата обращения: 03.02.2024).
3. F.Alimova, M.Yusupov. On the question of the use of internet testing in teaching chemistry // Science and innovation. 2022. Vol. 1 №8. pp. 1017-1022. URL: <https://zenodo.org/records/7407797> (дата обращения: 03.02.2024).
4. Тригуб Г.Я. Тестирование как метод обучения и контроля знаний в вузе // Концепт : электрон. науч.-метод. журн. 2017. №S3. С. 66-68. URL: <https://e-koncept.ru/2017/470051.htm> (дата обращения: 03.02.2024).
5. Кожевников В.А., Сабинин О.Ю. Система автоматической проверки ответов на открытые вопросы на русском языке // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11, №3. С. 57-72. DOI: 10.18721/JCSTCS.11306.
6. Wayne Xin Zhao, et al. A Survey of Large Language Models. 2023. URL: <https://arxiv.org/abs/2303.18223> (дата обращения: 04.02.2024).
7. Yupeng Chang, et al. A Survey on Evaluation of Large Language Models // ACM Transactions on Intelligent Systems and Technology. 2024. URL: <https://doi.org/10.1145/3641289> (дата обращения: 04.02.2024).

Никифорова А.Д. (автор)

Подпись

Нестеренко Ю.А. (автор)

Подпись

Бармина Д.А. (автор)

Подпись

Хлопотов М.В. (научный руководитель)

Подпись