УДК 004.83

ИССЛЕДОВАНИЕ САМООЦЕНКИ И ТОЧНОСТИ ОТВЕТОВ СУЩЕСТВУЮЩИХ УЧЕБНЫХ АССИСТЕНТОВ

Лапшина Ю.С. (Университет ИТМО), **Лактионова Е.А.** (Университет ИТМО)
Научный руководитель – к.т.н. Хлопотов М.В.
(Университет ИТМО)

Введение. Современные учебные ассистенты на основе искусственного интеллекта (ИИ) все чаще применяются в образовательном процессе [1], облегчая доступ к информации, помогая разъяснять сложные темы и автоматизируя выполнение рутинных задач. Однако одной из важных проблем является точность их ответов и способность адекватно оценивать свою уверенность в предоставляемой информации. Исследования показывают, что ИИ-модели могут демонстрировать высокий уровень уверенности даже в тех случаях, когда ответы являются некорректными [2], что может вводить обучающихся в заблуждение. Это особенно важно в образовательных системах, где студенты полагаются на полученные от ассистентов ответы при изучении учебного материала.

В связи с этим актуальной задачей является исследование самооценки учебных ассистентов, их способности к объективной оценке своей точности, а также факторов, влияющих на расхождение между заявленной уверенностью и реальной правильностью ответа.

Основная часть. Для изучения данной проблемы был проведен эксперимент, включающий подготовку датасета и анализ ответов ИИ-ассистентов на предмет их точности и соответствия уровню заявленной уверенности.

На первом этапе исследования был создан датасет, включающий вопросы различных тематик и уровней сложности. Датасет охватывал несколько ключевых областей: точные науки (математика, физика и химия), гуманитарные дисциплины (история, философия, лингвистика), программирование, а также вопросы на проверку общеобразовательных знаний. Такой выбор позволил оценить работу учебных ассистентов в разных предметных областях и выявить возможные закономерности в точности их ответов. Для каждого вопроса был сформирован эталонный ответ, обеспечивающий объективность эксперимента.

Было проведено экспериментальное тестирование учебных ассистентов, среди которых были рассмотрены такие популярные модели, как ChatGPT, Perplexity и DeepSeek. В ходе тестирования каждому из ассистентов были предложены вопросы из подготовленного датасета. После генерации ответа модель должна была самостоятельно оценить уровень своей уверенности по шкале от 1 (очень низкая уверенность) до 5 (полная уверенность). Это позволило зафиксировать не только фактическую правильность ответа, но и то, насколько уверенно ассистент его предоставляет.

После сбора данных был проведен анализ зависимости точности ответов от уровня уверенности моделей. Был вычислен общий процент правильных ответов в каждой категории вопросов и сопоставлен с уровнем уверенности, который указали учебные ассистенты. Результаты показали, что наибольшее совпадение между заявленной уверенностью и фактической точностью продемонстрировала модель DeepSeek, которая в большинстве случаев корректно оценивала степень достоверности своих ответов. Особое внимание было уделено случаям, когда модели демонстрировали завышенную уверенность в некорректных ответах, так как именно такие ошибки представляют наибольшую опасность для образовательных систем. Выяснилось, что хотя большинство ассистентов уверенно отвечали на вопросы, связанные с точными науками и фактологическими данными (даты, определения, формулы), наибольшие проблемы возникли в области философии. В этой дисциплине все тестируемые модели проявили высокую степень неуверенности даже в случаях, когда ответ был правильным. Это может быть связано с тем, что философские вопросы часто предполагают анализ разных точек зрения, что затрудняет формирование уверенности у языковых моделей.

Расхождение между заявленной уверенностью и фактической точностью ответов учебных ассистентов может объясняться несколькими факторами. Во-первых, механизмы генерации текста ориентированы на создание связного и логичного ответа, но не обязательно

достоверного. Во-вторых, не у всех моделей представлены эффективные способы обратной связи, которые позволяли бы им корректировать свою уверенность и точность на основе пользовательских оценок. Наконец, разные ИИ-ассистенты применяют различные подходы к оценке собственной уверенности, что усложняет их сравнение и выявление общих закономерностей в самооценке точности ответов.

Выводы. Проведенный эксперимент показывает, что современные учебные ассистенты демонстрируют значительное расхождение между заявленной уверенностью и фактической точностью, особенно при ответах на сложные вопросы. Это может создавать дополнительные трудности в образовательном процессе, так как пользователи, доверяя высокой уверенности модели, могут не подвергать ответы самостоятельному анализу. В связи с этим необходимо продолжить исследования в данном направлении и разработать более надежные алгоритмы самооценки, которые позволят ИИ корректировать свою уверенность на основе опыта и обратной связи. Улучшение этих аспектов позволит значительно повысить эффективность использования учебных ассистентов в образовании и снизить риск распространения некорректных данных [3].

Список использованных источников: (пример оформления)

1. Амиров Р.А., Билалова У.М. Перспективы внедрения технологий искусственного интеллекта в сфере высшего образования // Управленческое консультирование. – 2020. – № 3. – С. 80-88. DOI: 10.22394/1726-1139-2020-3-80-88
2. Zhang, Y., & Liu, J. Calibration of AI Self-Confidence in Answering Complex Questions. Proceedings of the 2022 International Conference on Artificial Intelligence and Education, 2022. DOI: 10.1109/AIE.2022.00012.
3. Давыдова Г.И., Шлыкова Н.В. Риски и вызовы при внедрении искусственного интеллекта в систему высшего образования [Электронный ресурс] // Вестник практической психологии образования. 2024. Том 21. № 3. С. 62–69. DOI: 10.17759/bppe.2024210308

Лапшина Ю.С. (автор) Подпись
Лактионова ЕА. (автор) Подпись

Хлопотов М.В. (научный руководитель) Подпись