

Разработка интеллектуальной системы с возможностью генерации и персонализации под пользователя вопросов на основе учебных материалов

Балцат К.И. (Университет ИТМО)

Научный руководитель - к. т. н. Федоров Д.А. (Университет ИТМО)

Введение. Современное образование требует инновационных подходов для улучшения процесса обучения, особенно в контексте персонализации учебных материалов. Создание индивидуально адаптированных учебных вопросов на основе набора учебных материалов представляет собой сложную задачу, для решения которой возможно использование Large Language Models (LLM). Однако, использование LLM чревато проблемой недостоверности генерируемых текстов и галлюцинаций [1]. Простейший подход повышения достоверности генерируемых текстов может заключаться в добавлении дополнительного контекста к промпту, однако, как показано в последних работах, при большой длине промпта LLM склонны упускать информацию из контекста [2].

Для решения этой проблемы в работе рассматривается применение Retrieval-Augmented Generation (RAG) – техники нахождения из всего корпуса текста релевантной информации и использовании её в контексте в LLM [3]. В дополнение к RAG использована техника дообучения (fine tuning) LLM на корпусе учебного материала, чтобы повысить качество генерируемых тестов, как показано в предыдущей работе [4]. Отдельно в работе рассматривается оценка качества разработанного подхода к задаче генерации педагогических средств оценки.

Основная часть. В качестве интеллектуальной системы к разработке предложен веб-сервис на streamlit, в который возможно загружать архив с учебными материалами в формате pdf. Система анализирует весь корпус учебного материала, выделяя ключевые понятия и темы, и генерирует на их основе вопросы. Техники RAG и fine tuning используется для интеграции информации из учебного курса, обеспечивая точность и глубину вопросов и вариантов ответов. Применение LLM делает разрабатываемую систему масштабируемой и применимой к любому учебному курсу.

На основе представленных материалов система генерирует вопросы и варианты ответа к ним, а также предоставляет пользователю функционал выбора ответа и его проверки, отображая то место в учебном материале, в котором содержится верный ответ. На основе ответов пользователя система персонализирует генерацию следующих вопросов, делая их проще или сложнее, а также повторно предлагает вопросы, на которые ранее был дан ошибочный ответ, в соответствии с моделью кривой забывания Эббингауза, что позволяет сделать процесс обучения наиболее эффективным [5].

Разработанный подход требует средств его оценки на рассматриваемой задаче, поэтому рассмотрены и опробованы на практике существующие на данный момент бенчмарки: оценка разнообразия вопросов по их длине, в скрытом пространстве эмбедингов, оценка релевантности вопросов документу, оценка совместной информации [6]. На основе анализа предложен свой подход к оценке составления моделями качественных пар вопрос-ответ с педагогической точки зрения. На основе бенчмарка эмпирически протестирована способность генерировать тесты разных LLM с открытым исходным кодом с применением RAG и без и с применением fine tuning или без. Результаты были сравнены с результатами, полученными на ограниченной выборке преподавателей НИУ ИТМО.

Выводы. Разработанная система повысит эффективность обучения, предоставляя студентам персонализированные и точные учебные вопросы, непосредственно связанные с содержанием их учебного курса, персонализированные под их уровень. Это улучшит понимание учебного материала и повысит мотивацию студентов, обеспечивая более глубокое и индивидуальное взаимодействие с учебным контентом.

Разработанный метод оценки позволил сравнить качество генерации оценочных средств профессионалами и LLM и доказал эффективность разрабатываемой системы. Так, в перспективе система может быть использована не только для персональной подготовки пользователями, но и заведениями высшего и среднего образования и платформами онлайн-образования для автоматизации контроля усвоения учебного материала.

Список использованных источников:

1. Huang L., et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. 2023. arXiv preprint arXiv:2311.05232.
2. Liu N.F., Lin K., Hewitt J., Paranjape A., Bevilacqua M., Petroni F., & Liang P. Lost in the Middle: How Language Models Use Long Contexts. 2023. arXiv preprint arXiv:2307.03172v3.
3. Chen J., Lin H., Han X., & Sun, L. Benchmarking Large Language Models in Retrieval-Augmented Generation. 2023. arXiv preprint arXiv:2309.01431.
4. “RAG vs Fine-tuning: Pipelines, Tradeoffs, and a Case Study on Agriculture”. Published on Jan 16, 2024. arXiv: 2401.08406
5. Murre J.M.J., & Dros J. Replication and Analysis of Ebbinghaus’ Forgetting Curve. PLOS ONE, 10(7), 2015, e0120644.
6. Lee D.B., Lee S., & Hwang S.J. Generating Diverse and Consistent QA pairs from Contexts with Information-Maximizing Hierarchical Conditional VAEs. ACL 2020.

Балцат К.И. (автор)



Федоров Д.А. (научный руководитель)
