IMPERIAL

Enhancing Higher Education: Personalized Learning with Large Language Models

Project Plan

Author:

Sarah Nesti
(CID: 02266658)
(GitHub:edsml-sjn22)

Supervised by:

Dr Marijan Beg
John Francois Van Schalkwyk

A Project Report submitted in fulfilment of requirements for the degree of Master of Science in Environmental Data Science and Machine Learning

Department of Earth Science and Engineering
Imperial College London
2024

Table of Contents

Abstract

This report details the development of an AI-based student support assistant in higher education to better align with human teaching methods, fostering deep understanding and critical thinking. Current AI solutions often provide inaccurate or generic responses. To address this, the project implements a Retrieval-Augmented Generation (RAG) model that improves relevance and accuracy by querying a database of course-specific materials. The project focuses on refining this methodology to create an AI tool that effectively supports learning, engages students, and enhances their educational experience. Key outcomes include a functional AI support assistant tailored to a specific course, a comprehensive evaluation framework, and strategies for promoting critical thinking. Effectiveness will be measured through surveys and evaluated based on measurable metrics.

1. Problem Description

Artificial Intelligence (AI) has become an integral part of our daily lives, as demonstrated by tools such as OpenAI's ChatGPT. The technology offers aid in writing and brainstorming, and assistance with research and analysis (Ka Yuk Chan & Hu, 2023). Despite its widespread use, AI has not yet been successfully applied to support learning in higher education. Current AI solutions, such as ChatGPT, provide generic answers that are often inaccurate, inadequate or incorrect. This is particularly problematic in university courses, where accuracy and relevance are critical to student understanding and success (Dempere, et al., 2023). Further concerns about the application of AI in higher education are that over-reliance on AI may hinder the development of critical thinking and creativity and also that it misaligns with human values (Ka Yuk Chan & Hu, 2023).

The issues of accuracy and transparency of AI-generated information can be addressed by using embedding models that query a vectorised database of relevant data, such as course materials (Choi, 2023). However, a basic implementation of a Retrieval-Augmented Generation (RAG) model that includes such an embedding model developed for a Master's programme at Imperial College has shown that while this approach addresses the issues of relevance and accuracy, it often results in AI answers that are too close to direct solutions. This undermines the learning process by eliminating the need for critical thinking.

Despite these challenges, addressing the integration of AI in higher education remains crucial. Enhancing AI literacy among students is essential for the responsible and effective use of Generative AI (GenAI) technologies (Ka Yuk Chan & Hu, 2023). Properly designed AI learning tools can empower students to tackle more challenging problems without bypassing the learning process. It's vital that these tools support understanding by guiding the students to the answer and foster critical thinking, rather than simply providing answers. By focusing on these goals, students are encouraged to engage deeply with the material and use AI tools to their fullest potential, ultimately enhancing their learning experience and better preparing them for their future academic and professional pursuits.

2. Methodology

2.1 Chosen Methodology: RAG Chain Architecture

The chosen approach for developing an effective AI learning support system involves implementing a Retrieval-Augmented Generation (RAG) pipeline. This methodology is preferred over alternatives like fine-tuning, since RAG combines the power of large language models with a database of specific course content, ensuring that the information provided is both accurate and relevant. Unlike fine-tuning, which requires extensive retraining and computational resources to adapt a model to specific content, RAG allows for the dynamic retrieval of information from a predefined set of documents, making it more efficient and flexible. (Maryamah, et al., 2024). The Architecture of an LLM application from a GitHub blog entry was utilized to develop the unrefined benchmark model alongside OpenAI's ChatGPT (Figure 1).

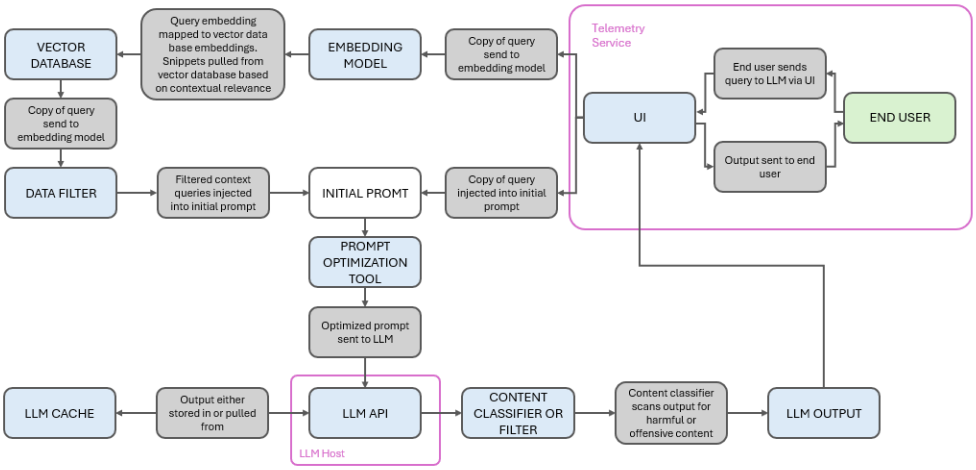

Figure 1: Flowchart of Benchmark RAG Chain Setup (Choi, 2023)

To improve the effectiveness of the RAG model, several refinements from a GitHub tutorial on RAG chains are being considered (Martin, 2024). The same tutorial was also aided in setting up the code for the initial benchmark model. Key aspects include selecting the right embedding model and chunking size, optimizing retriever and generator architectures, using multi-queries, and ranking and filtering prompts. Additional techniques like prompt optimization, multiple input formats, and hybrid search algorithms will be explored primarily for optimizing information retrieval, and also for supporting critical thinking and learning.

In addition to these refinement techniques, other strategies can be applied to further enhance the effectiveness of the RAG model in promoting critical thinking and supporting the learning process. The development of specialised agents within the RAG system for different types of tasks can be highly beneficial (Lehto, 2024). For example, some agents can be designed to provide detailed explanations and context, while others can challenge students with scenarios or questions that encourage critical thinking. Agents can be programmed to recognise the type of task, such as conceptual understanding, problem solving or application, and tailor their responses accordingly. This ensures that students receive the most appropriate support for their learning needs.

A prototype application will be developed that utilizes the RAG chain setup in the backend. This will involve creating a chat interface hosted on a web server, where students can interact with the AI. The chat interface will be user-friendly and designed to facilitate smooth communication between the students and the AI.

2.3 Challenges and Considerations

Supporting critical thinking with a RAG architecture is complex due to the need for relevant information retrieval and advanced natural language processing to understand context and handle interdisciplinary questions. Creating adaptive responses that challenge students while maintaining coherent and engaging interactions presents significant technical challenges.

3. Deliverables

The primary deliverable of this project is a functional AI-based learning support system designed and implemented specifically for the "Environmental Data Science and Machine Learning" Master's course at Imperial College. This system will utilize an enhanced Retrieval-Augmented Generation (RAG) chain to provide accurate, relevant, and explanatory responses to student queries, thereby enhancing the learning experience and emphasizing critical thinking.

3.1 Anticipated Outcomes

By leveraging AI technology, this project aims to achieve the following outcomes to improve education:

Baseline Model: Development of a foundational RAG chain setup to serve as a benchmark for evaluating enhancements, providing a reference point for measuring improvements.

Enhanced RAG Model: Creation of a refined RAG chain with enhanced database querying capabilities and precise information retrieval, improving accuracy and relevance of AI responses.

Effective AI Teaching Features: Identification of essential attributes and capabilities for an effective AI tutor, supporting various question types and promoting critical thinking and engagement.

Optimized Query Handling: Design of specialized agents for different types of student queries (e.g., conceptual, problem-solving, application-based) to ensure tailored responses that foster deeper understanding.

Evaluation Framework: Establishment of a comprehensive evaluation framework early to continuously assess the AI's performance and impact, validating the chosen architecture's effectiveness.

Functional Prototype: Development of a user-friendly frontend interface for students to interact with the AI, ensuring accessibility and practical application in real-world educational settings.

By focusing on these key outcomes, the project aims to deliver a robust and effective AI-based learning support system that significantly enhances the educational experience for students. This project collaborates with Nina Baumann on initial objectives, including creating the baseline and refined RAG models, as well as the final prototype integration. The main project phases

diverge, with my focus on enhancing teaching and learning experiences, while Nina focuses on integrating student backgrounds.

3.2 Evaluation and Limitations

The evaluation of the RAG chain set-up is challenging due to the complexity of measuring its impact on learning. During development, three surveys will be conducted with Master students from Imperial College to gather feedback on performance and usefulness. The final evaluation will compare the effectiveness of the system with OpenAI's ChatGPT and a standard RAG chain setup, focusing on accuracy, relevance and the ability to improve understanding without providing direct solutions. Evaluation frameworks will be developed to rigorously assess the AI's capabilities and ensure alignment with educational goals. Despite these measures, it remains difficult to accurately measure the true impact of the system without extensive A/B testing against human teaching assistants, which is beyond the scope of this project. In addition, the effectiveness of the RAG model depends on the quality of the course materials, and the optimisation of prompts for explanatory responses requires iterative refinement.

4. Future Plan

4.1 Estimated Timeline

Phase	Weeks	Tasks
Initial Setup	1-2	Set up baseline model, draft project plan, conduct initial research
Enhanced RAG Model	3-4	Develop enhanced RAG model, finalize project plan
Teaching Methodologies and Evaluation Structure	5-6	Develop evaluation framework, review literature on effective teaching methods
Subproject Implementation	7-10	Refine teaching techniques, apply evaluation methodologies, design agents for query types
Functional Prototype	11-14	Set up web server, deploy prototype, conduct survey
Final Adjustments and Project Report	15-16	Make final adjustments, write comprehensive project report

Table 1: Estimated Project Timeline

4.2 Progress to Date

The project has successfully completed the initial setup and preliminary research phase, establishing a basic RAG chain setup and drafting the project plan. Initial literature review and state-of-the-art research have been conducted to identify key methodologies and challenges.

References

Choi, N., 2023. *GitHub Blog*. [Online]
Available at: https://github.blog/2023-10-30-the-architecture-of-todays-llm-applications/
[Accessed 13 06 2024].

Dempere, J., Modugu, K., Hesham, . A. & Ramasamy, L. K., 2023. *The impact of ChatGPT on higher education*. [Online]
Available at: https://www.frontiersin.org/articles/10.3389/feduc.2023.1206936/full
[Accessed 12 6 2024].

Ka Yuk Chan, C. & Hu, . W., 2023. Students' voices on generative AI: perceptions, benefts, and challenges in higher education. *International Journal of Educational Technology in Higher Education,* p. 18.

Leddo, J. & Garg, K., 2021. COMPARING THE EFFECTIVENESS OF AI-POWERED EDUCATIONAL SOFTWARE TO HUMAN TEACHERS. *International Journal of Social Science and Economic Research,* p. 11.

Lehto, T., 2024. *Developing LLM-powered Applications Using Modern Frameworks*. [Online]
Available at: https://www.theseus.fi/handle/10024/862271
[Accessed 13 06 2024].

Martin, L., 2024. *GitHub*. [Online]
Available at: https://github.com/langchain-ai/rag-from-scratch/blob/main/rag_from_scratch_1_to_4.ipynb
[Accessed 13 06 2024].

Maryamah, M. et al., 2024. Chatbots in Academia: A Retrieval-Augmented Generation Approach for Improved Efficient Information Access. *IEEE,* Issue 16th International Conference on Knowledge and Smart Technology (KST), p. 6.

Neumann, M., Rauschenberge, M. & Schoen, . E.-M., 2023. "We Need To Talk About ChatGPT": The Future of AI and Higher Education. *IEEE/ACM,* Issue 5th International Workshop on Software Engineering Education for the Next Generation (SEENG), p. 4.

Steenbergen-Hu, S. & Cooper, H., 2014. A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning. *Journal of Educational Psychology,* p. 17.

Tajik, E. & Tajik, . F., 2023. *A comprehensive Examination of the potential application of Chat GPT in Higher Education Institutions*. [Online]
Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4699304
[Accessed 12 6 2024].