Course Syllabus Query Database Using RAG (Retrieval-Augmented Generation)

Cade Boiney, Ken Lam, Ognian Trajanov, Benjamin Zhao
Hamilton College, Clinton, NY, USA

***Abstract*—We present a RAG system for Hamilton College syllabi that uses LLaMA 3.3 70B to retrieve relevant passages and generate concise, context-aware answers, improving student and faculty access to course information.**

I. Introduction and Goals

This project aims to develop a RAG system capable of answering queries such as "When are office hours?" or "What percentage of the grade is the final exam?" By combining semantic retrieval with instruction-tuned large language model (LLM) generation, the system will provide accurate, context-aware responses that help students and professors quickly access course information.

This tool will create a centralized syllabus database, helping students plan for future courses efficiently. With a large number of syllabi available, it will save time, reduce manual search, and serve as a useful academic planning resource for faculty and students alike.

II. Data Collection

We will create a Google Form for Computer Science, Math, and Economics department chairs to distribute to professors, allowing them to upload syllabi in PDF format. Syllabi available on departmental websites will be collected directly. Text extraction will be performed using `PyPDF2`, and data will be structured into JSON or CSV with fields for course title, instructor, schedule, and grading policy.

III. Data Preprocessing

Text will be cleaned, standardized, and split into 300–500 token passages. Embeddings will be computed using `all-mpnet-base-v2` to support semantic retrieval. Optionally, cloud-based embeddings such as OpenAI's `text-embedding-3-large` may be used for efficiency or improved quality.

IV. Methods

We use **LLaMA 3.3 70B** (2) as the primary LLM for its large context window, open weights, and strong instruction-following ability. Optionally, we compare it with **GPT-4 Turbo** (1) via API to assess accuracy and response quality.

Syllabus passages are vectorized and stored in FAISS or Chroma for fast similarity search. Given a query, the retriever selects the most relevant passages, which are passed to the generator to produce context-aware answers. Prompts include retrieved text and instructions to answer only from it. Few-shot examples may be added to improve accuracy. Prompt design emphasizes clarity, conciseness, and adherence to academic policy. Generation parameters (e.g., temperature, max tokens) are tuned to reduce hallucinations.

Evaluation combines objective metrics (Exact Match, F1, recall) with subjective Likert-scale ratings for clarity, completeness, and helpfulness.

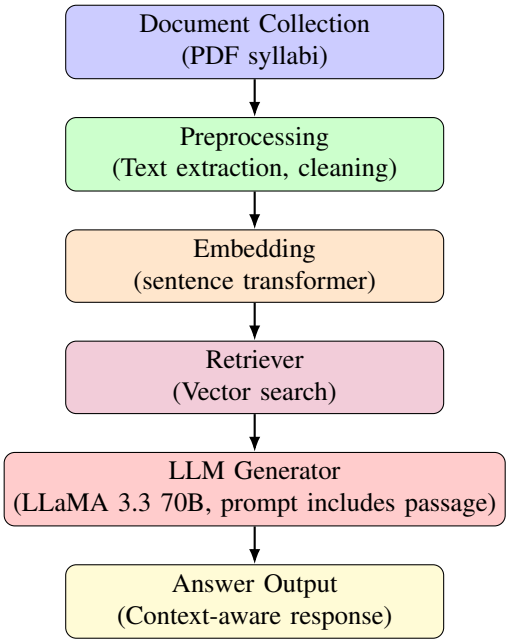

Fig. 1: RAG pipeline: retrieved passages are passed to LLaMA 3.3 70B to generate context-aware answers.

V. Conclusion

The proposed system provides an efficient, accurate method for querying course syllabi, improving access to information for students and faculty. Our tool would complement Student Planning with a conversational interface for course questions and could scale to additional departments or campus-wide syllabi.

References

[1] OpenAI, "GPT-4 Turbo," OpenAI API, 2023. [Online]. Available: https://platform.openai.com/docs/models/gpt-4-turbo

[2] Meta AI, "LLaMA 3.3: 70B," Meta AI Research, 2025. [Online]. Available: https://ai.meta.com/llama