Capstone Project Plan Template

Last updated: November 14, 2024

Instructions: Each capstone team can use this template to capture and summarize information about the project. This can be shared with the sponsor and mentor. When submitting the plan during the course, a PDF file is preferable.

Stakeholder Names and Roles

In the table below, enter information for each team member, mentor, and sponsor. On the sponsor side, there may be several stakeholders including project manager, data contact, etc.

Stakeholder	Role
Vishwanath Guruvayur	*Team Member (Student)*
Doruk Ozar	*Team Member (Student)*
Bereket Tafesse	*Team Member (Student)*
Luke Napolitano	*Team Member (Student)*
Brian Wright	*Mentor*
Ali Rivera	*Mentor*
Lucas McCabe	*Sponsor*

Project Title: Multimodal LLM RAG in Education

Abstract

The rapid advancement of Large Language Models (LLMs) is increasingly being considered for multimodal data processing. However, the effective retrieval of information from diverse data sources, for example, in educational contexts, still has many challenges. This capstone project explores the application of Retrieval-Augmented Generation (RAG) methods for

multimodal data retrieval, focusing on a classroom setting where audio, video, and text data are made available. Our primary aim is to investigate methods that allow a foundational LLM(s), which have not been pre-trained on such multimodal data, to effectively search and retrieve relevant information across these different data types.

In this study, we will collaboratively design and execute experiments to determine the most effective approaches for multimodal retrieval, taking into account the unique modal characteristics of the data. For instance, we might explore the capability of these models to accurately infer the content of a lecture by analyzing and integrating inputs from audio recordings, video footage, and textual transcripts. Additionally, we will examine the utility of various retrieval methods in delivering contextually appropriate and accurate information in response to user queries.

Throughout the discovery phase, we will collaboratively define a set of research questions that guide your experiments, such as assessing whether RAG methods can optimize retrieval accuracy or how these models perform under different multimodal input conditions. By the end of this project, we aim to demonstrate practical applications of multimodal retrieval in education, providing insights into the strengths and limitations of current approaches and paving the way for future advancements in this area, and other domains.

Outline of the Project

- *The Business purpose of the project*
    - *Improve independent individual learning outcomes (students using AI chatbot to learn when teacher isn't available)*
    - *Analyze use cases of Multimodal RAG*
- *Why the project is important?*
    - *Expand capabilities of LLMs using multimodal data*
    - *Developments in AI, is it really a helpful tool or is it a gimmick (learning outcomes)?*
- *Who are the stakeholders?*
    - *Students*
    - *Professors*
    - *AI researchers, LLM/NLP research*
- *Important Assumptions*
    - *Assumption that vectorization for all multimodal data will work just as we need it to*
    - *Retriever will route from the right document stores*
    - *Generation won't hallucinate too much and we can actually evaluate performance in the classroom as an educational tool*

- *What is in scope / out of scope (as needed)*
    - *What is in scope?*
        - *How to **evaluate** these models, what do we consider successful implementation*
        - *What embedding models/vector representations are best for multimodal data?*
            - *Using backup plan, active RAG flows*
        - *Only using content from DS 3001, pdfs, lecture slides, hopefully moving into videos and audio to exactly pinpoint where relevant info is*

    - *What is NOT in scope?*
        - *Website/package/chatbot front-end that we can put the RAG-bot into? Assuming that the UI is already built in the Text based RAG phase*
        - *Using the bot for multiple courses (Data Systems instead of Foundations of ML)*

Success Criteria

Success criteria define the needs for a successful project. These should be identified with help from the sponsor. It is important to revisit these criteria with the sponsor as the project progresses to ensure alignment.

SC1	*Effective embedding of all multimodal data*
SC2	*FAISS scores achieve parity, embeddings are close*
SC3	*Retrieval is from the correct sources (e.g. pdfs find pages, video transcripts find min:sec interval)*
SC4	*Minimal hallucinations and corrective measures in instance of hallucination*
SC5	*Generalizable framework for other courses, can just apply RAG flow*

Assumptions and Limitations

For any project, there may be assumptions [A] *and limitations* [L] *on the data and the modeling approach. These can be documented here. Example:* [L] *Ideally, the dataset would include variable X, but we did not have access to this data, which was a limitation.*

Identifier	**Description**
A	*RAG is going to be better than newer chatbots with larger context windows and improved reasoning abilities*
A	*Vectorization will work the way it is supposed to with Multimodal Data*
A	*Retriever will grab from the right databases/vector stores*
A	*Minimal hallucination so we can evaluate classroom outcomes*
A	*The UI and Architecture for the Chatbot is already built.*
L	*Training RAG-bot for undergraduate machine learning class, nothing more or less advanced*

Potential Background Literature and Resources

Resources provided from Active Learning Lab Onboarding Doc:

- Literature Review
- Rag resources:
    - LangChain videos
        - A good overview to get an idea of the big picture & different parts of RAG
    - RAG paper
        - The original publication – this is more jargon-y so don't worry if you don't understand everything, just need to read through and understand where it started
    - HuggingFace
        - Another good walkthrough of a RAG architecture with examples and plainly written commentary
    - RAG vs Fine tuning
        - A nice comparison between RAG and fine-tuning, helpful to understand why RAG may be a better solution than fine-tuning for us.
- Papers:
    - Artificial Intelligence in Education: A Review
    - What is AI Literacy? Competencies and Design Considerations
    - Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support
    - The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education
    - Factors Influencing Learner Attitudes Towards ChatGPT-Assisted Language Learning in Higher Education

Brief Outline of the Data

https://github.com/ali-rivera/LMI-Capstone/tree/main

- PDFs of textbooks
- Txt files representing lecture transcripts
- Lecture slides

Brief Plan of How the Data will be Modeled and Processed

- A Vector Database of all media types (Audio Transcripts, PDF Texts, Lecture Slides) will be created
This will be done using an embedding model (TBD) that transforms all chunks of data into vectors

- Retriever Choice (TBD)

- LLM choice for specifically multimodal machine learning data retrieval and outcome (tbd)

- Shouldn't need to supplement with additional data

- Thoughts on processing raw data in refined data useful for modeling

- Use python to read text files, chunking for larger pdfs, same embedding model for everything

Brief Plan of Modeling Approaches

1. *Pre-process data in python, create our objects for our data*
2. *Pick a vector store (ChromaDB? LangGraph? Etc.)*
3. *FAISS to evaluate embeddings*
4. *Pick a retriever, use comparative internet search to find the best retriever for our data specifically*
5. *Pick an LLM for optimal generation and referencing for our sources*
6. *Evaluate flow of model (start to finish from a student's question -> processing the question in line with chatbot prompt -> retrieving the right docs -> generating the right answer)*

Potential Concerns [C] and Blockers [B]

Concerns are things to keep in mind but might not yet impede progress.
*Blockers i*ndicate that you are stuck and need help.

Identifier	**Description**
C	*Finding the right embedder and retriever for our data so that we don't get gummed up when trying to generate*
C	*Size of the database, how much is adding multimodal going to do to our DB*
C	*Scaling of the project, what happens when we train for the whole course instead of just a few weeks*