Virtual Academic Assistant Using Large Language Models

Fundaments & Application Of LLM Project Proposal Presentation - Spring 2025

Team: BabyCoda 

- Aashi Sharma
- Harsh Gupta
- Pranay Bandaru
- Rakshit Rao

What We Will Cover

- Introduction & Motivation
- Background & Related Work
- Research Questions & Goals
- Data & Methods
- Discussion & Feasibility
- References

Introduction & Motivation

What problem are you addressing, and why is it important?

Students hesitate to ask questions because nobody wants to be "that" person, and professors face high workloads due to repetitive queries requiring in-detail doubt solving. This creates a gap in doubt resolution, impacting learning efficiency. This is a great opportunity for an AI-powered assistant to step in, answering doubts instantly while giving professors and TAs some much-needed breathing room.

Provide a compelling real-world connection or societal relevance.

Imagine a world where no student’s question goes unanswered, and professors don’t feel like email helplines. An AI-powered course assistant makes learning more accessible, reduces academic burnout, and keeps education interactive, even in massive classes.

Background & Related Work

- LLM applications like ChatGPT [1] have been making waves in the learning space, knocking out giants like Chegg.
- Instant answers are great but this spoonfeeds students and when students just copy-paste, is there any real learning happening?
- The challenge? Helping students think on their own, instead of just Guided-Googling.
- Meanwhile, professors are drowning in emails all in the exam week - AI could be **The Ultimate TA** they never had.
- Universities are actively exploring solutions - Coginiti.ai [3] (University of Sydney) is one such example.
- “Using ChatGPT in Education” [4] Highlights the ability of ChatGPT to produce human-like responses and assist with personalized learning.
- “Large language models for education: A survey and outlook” [5] outlines the various possible areas of LLM in education and also highlights the limitations.

Research Questions & Goals

Research questions we look to answer

- Identify the ideal architecture to best answer the queries
  - Agents and tools required
  - Fine-tuned or general LLM
  - Best suited models
- Identify which type of prompts and context result in the best, relevant answer and reduce hallucination.

Expected outcomes and contributions of our project

- Build a foundational application of an LLM in the field of education which can be further extended to solve more problem in this domain.
- Allow professors to prompt the LLM model on how to generate response.
- Promote use of LLMs in education in a creative and constructive way while keeping out the ability to cheat and bypass.
- Provide a healthy, thought-stimulating learning environment with accurate and swift query resolution.

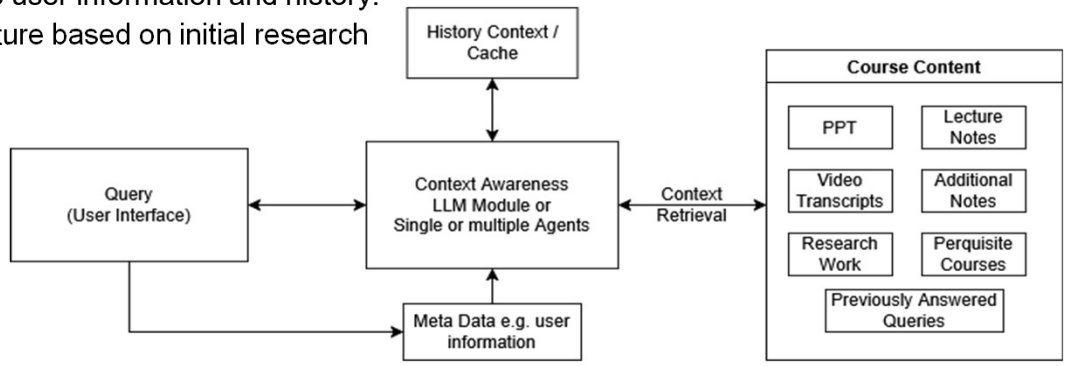
Data & Methods

Dataset(s) that will be used, and why are they appropriate

- Course material like lecture slides, recommended learning like research papers/reference articles/additional lectures.
- Transcripts of video lecture recordings.
- Perquisite course material
- Meta data regarding the users, course and prompting the LLM model on how the answers should be generated

Briefly describe your planned approach and methodology.

- Compare LLM models – fine tuned models, general models or an ensemble of both.
- Compare different prompting strategies and evaluate which gives the most relevant answers for the scenario.
- Tailor the answer based on the user information and history.
- Preliminary high-level architecture based on initial research



Discussion & Feasibility

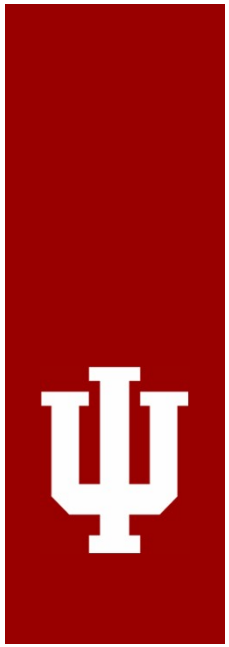
- How will your data and methods help answer your research questions?
- Accuracy of multiple LLM models and prompting strategies can be evaluated
  - Use of external resources to match and validate the answers.
  - Initial round of verification by us using our understanding of the course materials.
  - Take feedback from TAs and Professors wherever possible.
- Potential challenges and how you plan to address them?
  - We aim is to reduce hallucinations and ensure accurate responses, it may not always be possible to achieve his.
  - Ensuring sensitive data is not exposed and user identifiers are masked.

    Possible Sol: Filtering/post processing answers, masking identity when storing history.

References

Cite any relevant papers, datasets, or sources used

1. https://chatgpt.com/
2. https://learning.google.com/
3. https://cogniti.ai/
4. Loos, E.; Gröpler, J.; Goudeau, M.-L.S. Using ChatGPT in Education: Human Reflection on ChatGPT's Self-Reflection. *Societies* **2023**, *13*, 196. https://doi.org/10.3390/soc13080196
5. Wang, Shen, et al. "Large language models for education: A survey and outlook." *arXiv preprint arXiv:2403.18105* (2024).
6. Course materials of subject [Subject to approval]




Thank You !!