**Capstone Project Proposal**

**Project Title:** STARS Tutoring Chatbot Web Application

**Project Overview:** This capstone project aims to enhance a Streamlit-based chatbot application that integrates OpenAI and LangChain technologies. The primary focus is on extending the existing system to allow real tutors to maintain and improve the chatbots in an intuitive and user-friendly manner through the interface. Students will also work on enabling multi-chat memory, where chatbot selection is determined for each individual chat at the start, rather than switching between two static bots. Additionally, extending the embeddings will be a foundational task, ensuring that students understand the system deeply before developing tools for tutors to use effectively.

**Technologies and Tools:**

* **Programming Language:** Python
* **Frontend Framework:** Streamlit
* **AI Integration:** OpenAI and LangChain
* **Database:** MongoDB
* **Analytics:** Integration of analytics frameworks (specific tools to be determined)

**Project Objectives:**

1. **Integrated Chat Functionality and Embedding Enhancements:**
   * Enable users to maintain separate chat histories for each chatbot, saved in a database and displayed in a sidebar for easy access upon login.
   * Extend the existing embeddings to support topic-specific restrictions and class-specific chatbots.
   * Add functionality to ensure responses are relevant to the selected topic area, leveraging embeddings to implement these restrictions.
   * Develop a system where tutors can review chat logs, create their own embedding texts on the backend, and dynamically use these embeddings during chats or for model tuning.
2. **Input and Analytics Features:**
   * Implement document and image input functionality, treating this as a stretch goal to enhance usability.
   * Design and set up a MongoDB database to collect and manage application data while ensuring compliance with relevant regulations.
   * Develop and integrate an analytics framework into the application, focusing solely on the development and integration aspects.

**Timeline and Milestones:** The project will span 14 weeks, with tasks designed to flow collectively rather than sequentially:

* **Weeks 1-2:**
  + Orientation and project setup.
  + Familiarization with Streamlit, OpenAI, LangChain, and MongoDB.
  + Finalize database schema with project owner.
  + Begin basic database setup for chat history.
  + Initiate design and prototyping for the tutor review and embedding training system.
* **Weeks 3-6:**
  + Develop and test the multi-chat history feature.
  + Continue integration of extended embeddings to support topic-specific restrictions and chat functionality.
  + Refine multi-chat memory to align with class-specific chatbot requirements.
  + Simultaneously iterate on the design and early development of the tutor review and embedding training interface.
* **Weeks 7-9:**
  + Finalize embedding-based restrictions to ensure topic relevance.
  + Advance the tutor review and embedding training interface based on prior feedback and integration efforts.
  + Conduct testing of tutor-focused features and refine for usability.
* **Weeks 10-11:**
  + Implement and polish the tutor review and model training framework.
  + Conduct end-to-end testing and collect user feedback to ensure robustness.
* **Weeks 12-14:**
  + Integrate analytics frameworks into the application.
  + Work on stretch goals (document and image input support) if time permits.
  + Perform final testing and debugging of all features.
  + Prepare a final project presentation and technical documentation.

**Deliverables:**

1. Fully functional chatbot application with multi-chat history feature.
2. Extended embeddings supporting topic-specific restrictions and class-specific chatbots.
3. Tutor review and model training framework.
4. Document and image input functionality (if feasible).
5. Compliance-ready MongoDB database for data collection.
6. Analytics framework integrated into the application.
7. Technical documentation and final presentation.

**Learning Outcomes:**

* Practical experience in integrating AI technologies (OpenAI and LangChain) into applications.
* Hands-on knowledge of database design and regulatory compliance.
* Understanding and implementation of embeddings and their extensions to support application functionality.
* Experience in integrating analytics frameworks into applications.
* Development and debugging of scalable, user-friendly applications.