Project Proposal: Education Support and Learning Assistant (VA)

Project Overview:
This project proposes developing an Education Support and Learning Assistant, a virtual assistant designed to enhance educational experiences by providing personalized support and learning assistance. Leveraging advanced NLP models and integrating with educational databases, this assistant will help students access relevant course materials, manage assignments, and find educational resources. Additionally, it will utilize web search capabilities to provide broader educational insights, ensuring students have comprehensive support throughout their learning journey.

Use Case & Functionality:
The VA will be designed to assist students by:

1. Providing information such as office hours of professors, assignment deadlines, and class schedules, ensuring they stay organized and on track with their academic responsibilities.
2. Providing relevant learning resources based on a student's learning style and needs such as textbooks, online courses, or educational apps.
3. The VA will leverage Retrieval-Augmented Generation (RAG) to provide context from recent educational research and guidelines, offering students the most up-to-date information and insights relevant to their studies.

Data and Tools:

a. **Structured Data:** A structured dataset will be created by crawling the SJSU university website for information on classes, instructors, and course materials, then stored in a relational database like MySQL or MongoDB.
b. **Database & Retrieval Tools:** For storing vector embeddings of textual contents or resources, tools like Chromadb or Pinecone will be used. This will enable efficient querying and retrieval of relevant educational materials.
c. **Web Search Component:** An integrated tool, that will search the SJSU university website and educational resources to assist students with queries and assignments.

Large Language Models & Retrieval Components:

To Build the Education Support and Learning Assistant Virtual Agent, I will use the following two models and retrieval modules:

- **Base Models (LLMs):**
  - **Large Model:** Llama-3.1-8B-Instruct (Number of Parameters: 8 Billion)
  - **Smaller Model:** flan-t5-small (Number of Parameters: 40 million)
- **Retrieval Module:**
  - Sentence-Transformer framework for creating embeddings.
  - Pinecone or Chromadb for indexing and querying education/academic resources.

Planned User Queries:

1. “What is the application process for an international student to apply for the MS Software Engineering Program?”
2. “What are the program requirements for the MS AI program?”
3. “How do I enroll in the summer course for Introduction to Computer Science?”
4. “What are the prerequisites for taking Advanced Linear Algebra?”
5. “What are the requirements for completing a thesis in the Master's program?”
6. “How do I apply for a research grant for my undergraduate project?”
7. “Can you provide a study guide for the final exam in Psychology 101?”
8. “What is the deadline for the upcoming assignment in my calculus class?”
9. “How do I access the online textbook for my Biology class?”
10. “What are the office hours for Professor Johnson?”
11. “How do I submit my homework online through the learning management system?”
12. “What are the dates for the final exams this semester?”
13. “How do I request an extension on a project due to illness?”
14. “Can you explain the concept of derivatives in calculus?”
15. “How do I access the library's online databases for research purposes?”
16. “What is the format for citing sources in APA style?”
17. “Can you summarize the main points from last week's lecture in Economics?”
18. “Can you recommend some online resources for learning Python programming?”
19. “Can you help me understand the difference between a hypothesis and a theory?”
20. “Can you provide tips for improving my writing skills for academic papers?”