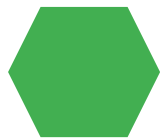


Praveen Kumar S

Final Project



PROJECT TITLE

LLM Question-Answering Application

AGENDA

- Define project objectives, scope, and requirements.
- Select appropriate technologies and frameworks for development.
- Develop wireframes and prototypes for the user interface.
- Implement backend functionality for document processing, embedding, and storage.
- Conduct testing, including unit testing, integration testing, and system testing.
- Deploy the application to a production environment and provide documentation and user training for usage and maintenance.



PROBLEM STATEMENT

The existing methods of extracting insights from documents often lack efficiency and accessibility, requiring users to navigate complex interfaces or incur high costs for processing and analysis. Additionally, users face challenges in securely storing and retrieving document embeddings for future reference. This creates a need for a user-friendly and cost-effective solution that seamlessly processes documents, extracts relevant information, and provides efficient access to insights without compromising security or usability.



PROJECT OVERVIEW

The LLM Question-Answering Application offers a user-friendly interface for seamlessly extracting insights from documents. Users kickstart the process by providing their OpenAI API keys. Following this, they can upload documents in PDF, DOCX, or TXT formats. The application then begins processing, chunking, and embedding the content employing the all-MiniLM-L6-v2 model from HuggingFace. This innovative approach ensures users incur no charges for generating embeddings, with processing times averaging between 1 to 2 minutes, contingent on file size and computational resources.

Once the document is prepped, the embeddings are securely stored in a vector store through FAISS, a powerful open-source library adept at efficient similarity searches and clustering of dense vectors. With the document processing complete, users can then submit their queries. After roughly a minute, the application serves up the desired response, effectively providing an efficient and hassle-free experience for accessing pertinent information from uploaded documents.



WHO ARE THE END USERS?

- Researchers
- Professionals in law, finance, consulting, journalism, etc.
- Students
- Businesses across industries
- Knowledge workers like librarians, data analysts, knowledge managers, etc.

YOUR SOLUTION AND ITS VALUE PROPOSITION



Solution:

Streamlines document analysis process using advanced AI technologies.
Supports document upload in PDF, DOCX, and TXT formats.
Processes documents using all-MiniLM-L6-v2 model from HuggingFace.
Embeds content for efficient analysis and retrieval of insights.
Securely stores embeddings in vector store using FAISS.
Enables users to submit queries and receive precise responses within minutes.

Value Proposition:

Efficiency: Automates document analysis tasks, saving time and effort.
Cost-effectiveness: No charges for generating embeddings.
User-friendly Interface: Intuitive interface for easy navigation.
Secure Storage: Ensures data privacy and confidentiality.
Accuracy: Provides precise responses to user queries.
Versatility: Supports diverse user needs across industries and domains.

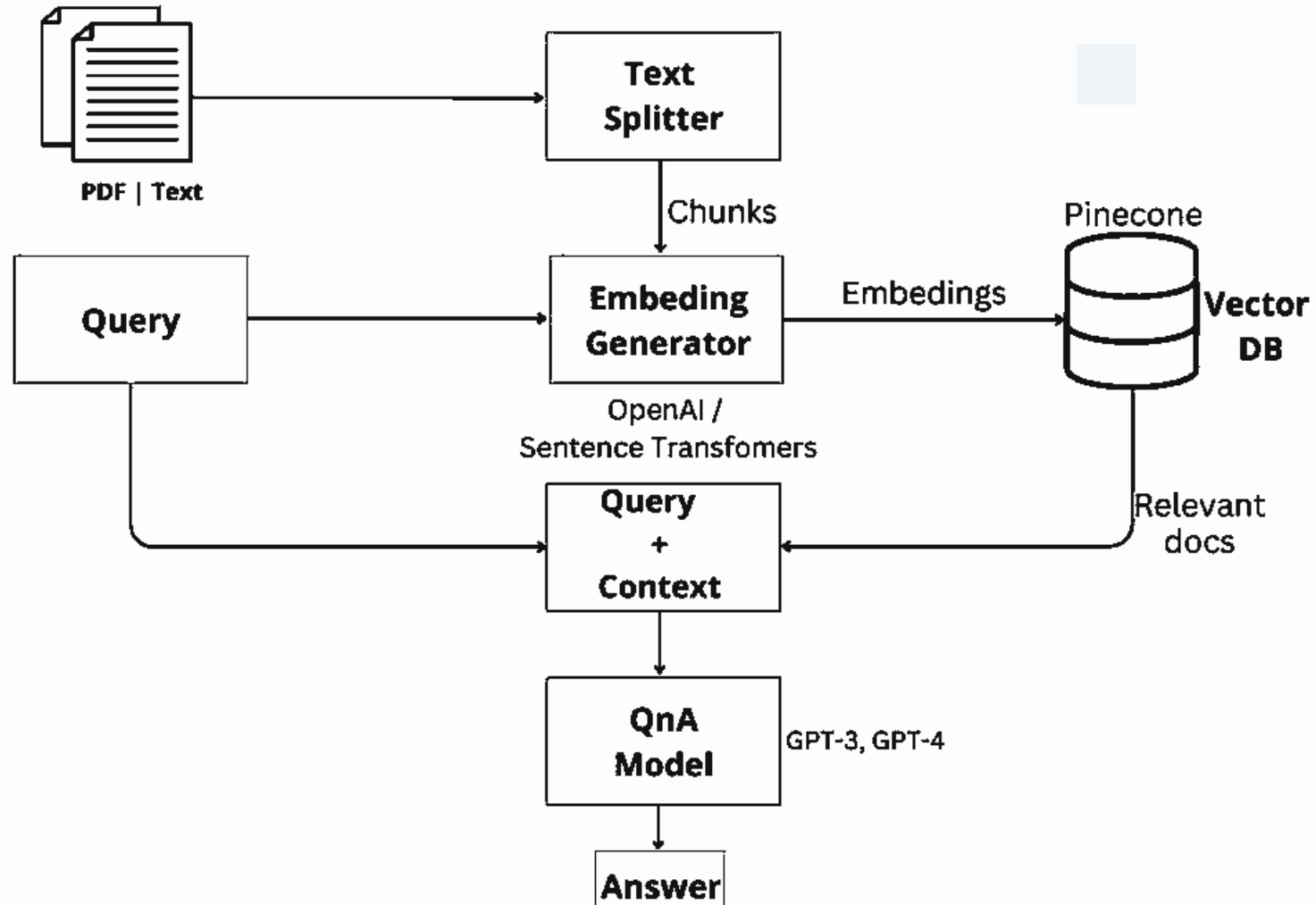
THE WOW IN YOUR SOLUTION



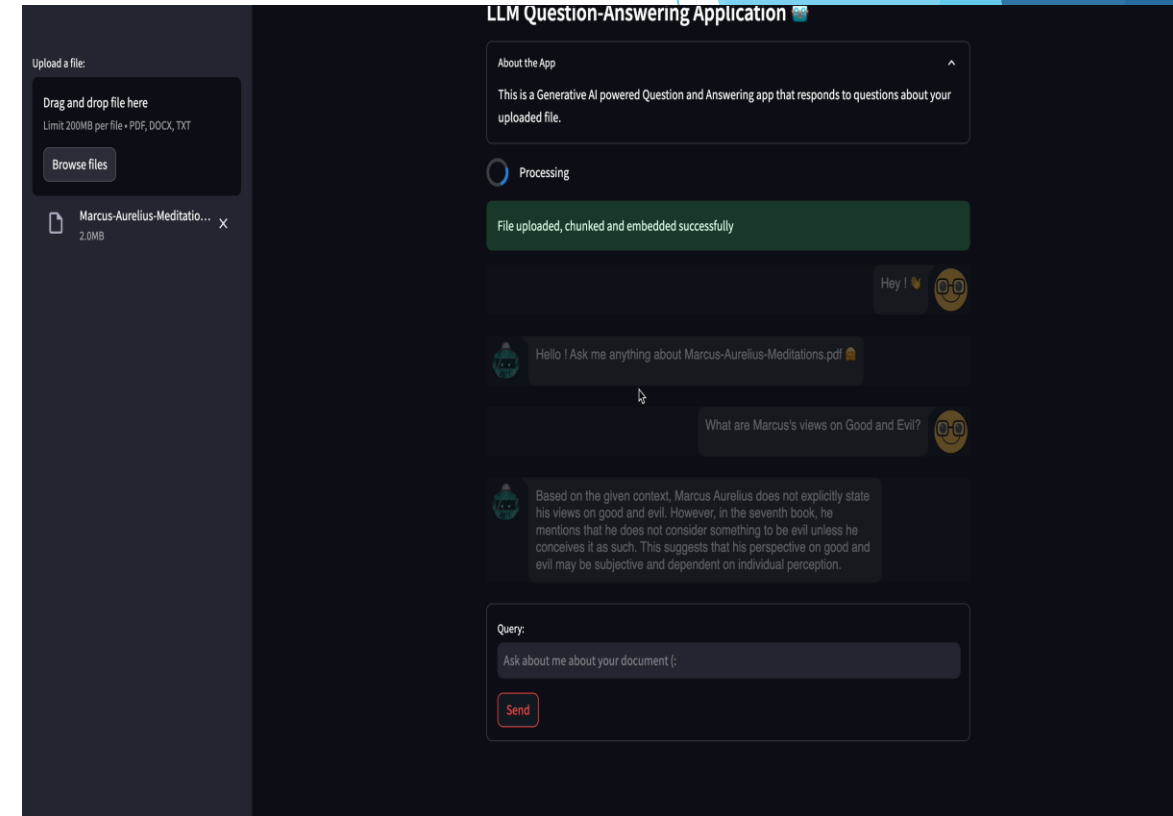
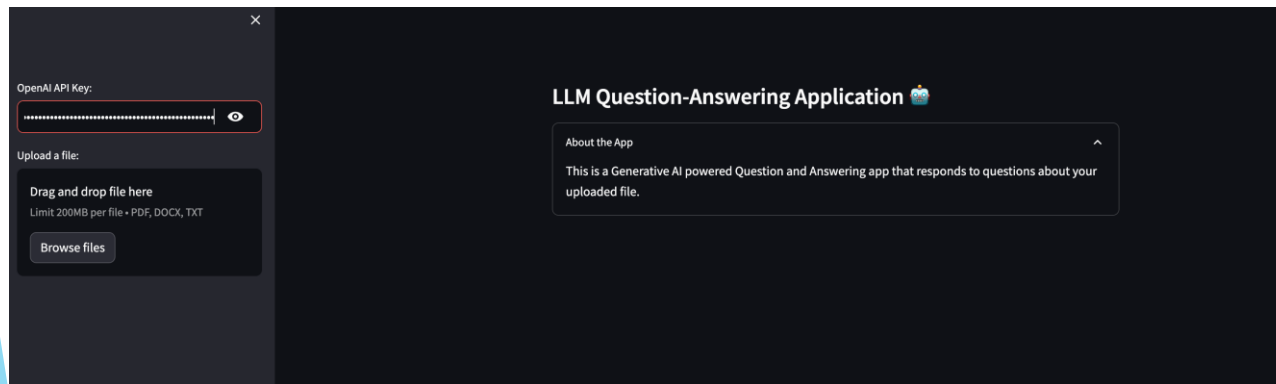
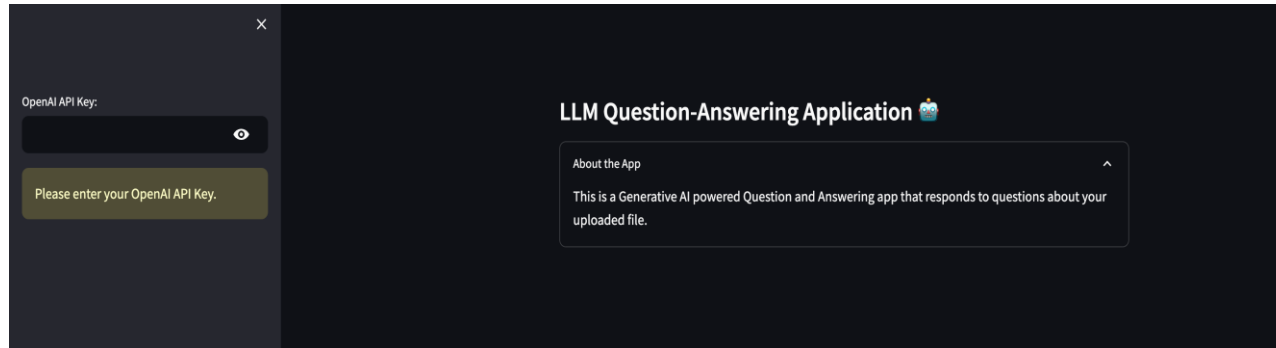
Prepare to be wowed by our groundbreaking solution, where cutting-edge AI technology meets seamless user experience! Imagine effortlessly uploading your documents in any format—PDFs, DOCX, or TXT—and within moments, our advanced system harnesses the power of the all-MiniLM-L6-v2 model to analyze and extract key insights with unparalleled precision. But it doesn't stop there! Your data's security is our top priority, as we securely store embeddings using FAISS vector storage, ensuring complete confidentiality. And the speed? It's simply breathtaking—get ready for lightning-fast query responses in just minutes! With our solution, efficiency, affordability, and accuracy converge in an intuitive platform that caters to your every need, delivering a truly wow-worthy experience.



MODELLING



RESULTS



<https://github.com/praveen0908/Q-A-using-LLM>

3/21/2024 Annual Review