Roshan Kumar A

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

Existing tools for document analysis frequently face issues such as inefficiency and limited accessibility. Users often struggle with complex interfaces or high costs for analysis. Additionally, securely storing and retrieving document embeddings presents considerable obstacles. This highlights the necessity for a solution that is both easy to use and cost-effective. Therefore, there is a critical requirement for a simplified method that efficiently handles documents, extracts relevant information, and provides convenient access to insights while maintaining security and user satisfaction.

PROJECT OVERVIEW

The LLM Question-Answering Application offers a user-friendly interface for effortlessly extracting insights from documents. Users start by entering their OpenAI API keys and uploading documents in PDF, DOCX, or TXT formats. The application then utilizes the all-MiniLM-L6-v2 model from HuggingFace to process and embed the content. This approach ensures that users incur no charges for generating embeddings, with processing times typically ranging from 1 to 2 minutes, depending on file size and available computational resources.

After preparing the documents, the embeddings are securely stored in a vector store using FAISS, a robust open-source library known for its efficiency in similarity searches and clustering of dense vectors. Once the document processing phase is complete, users can submit their queries. Within approximately a minute, the application provides the desired response, delivering an efficient and seamless experience for accessing relevant 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.

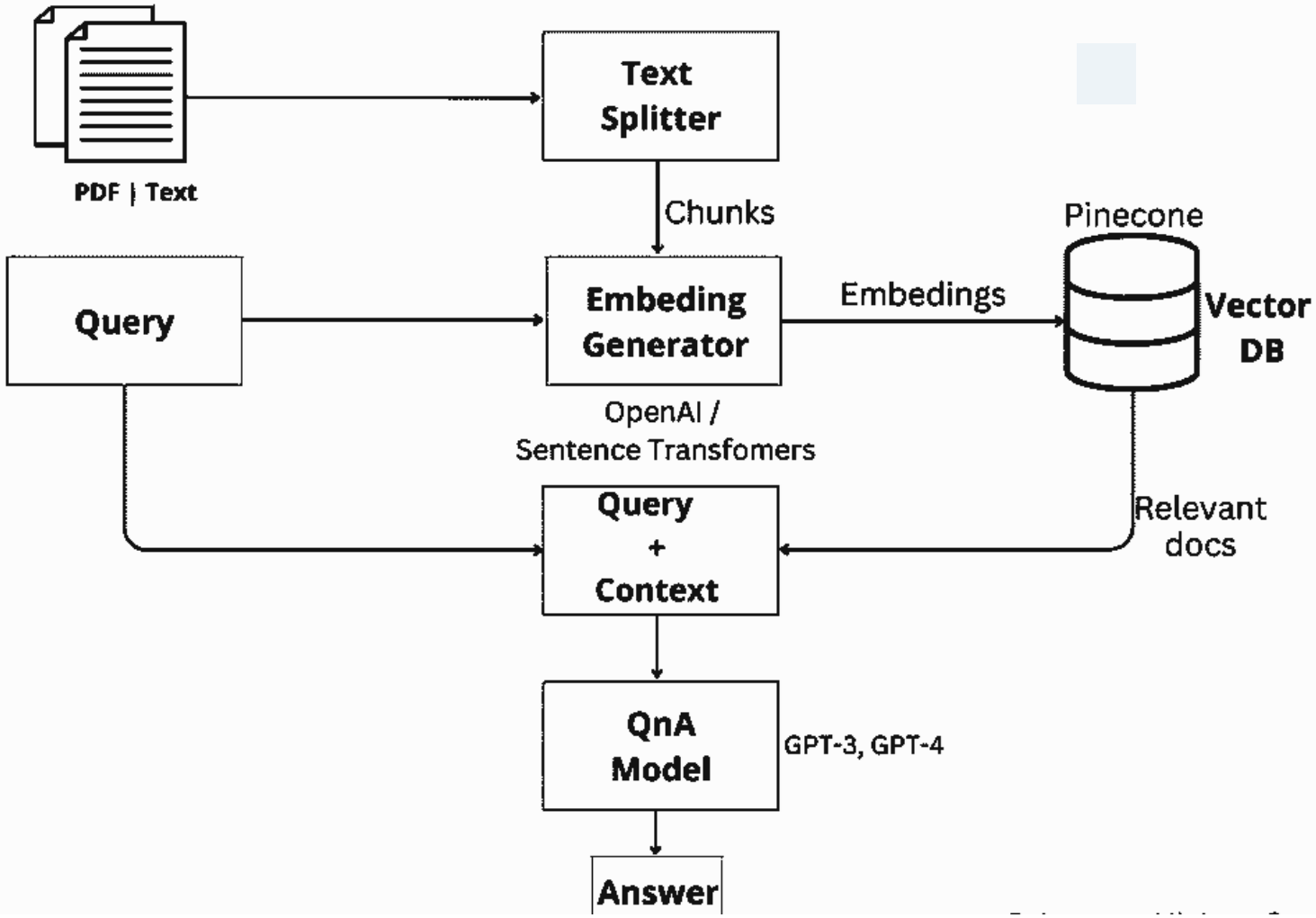
Value Proposition:

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.

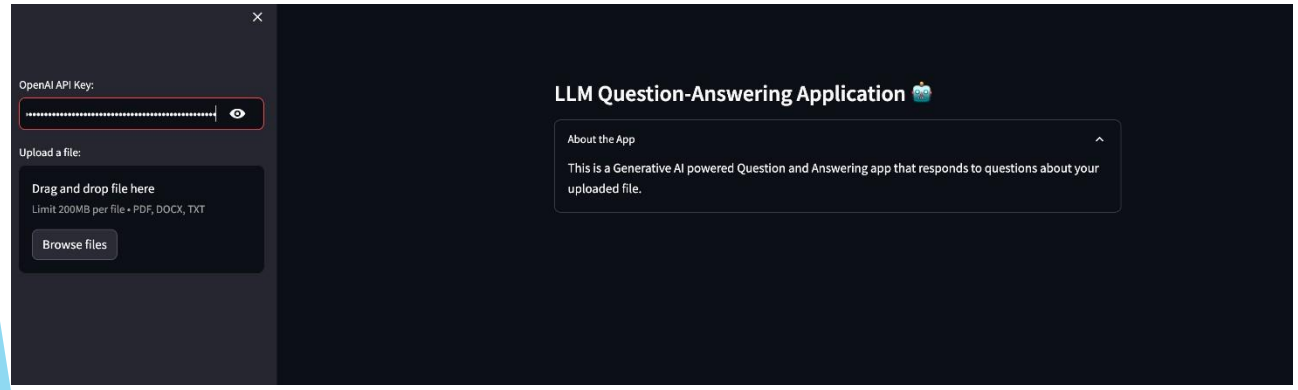
THE WOW IN YOUR SOLUTION

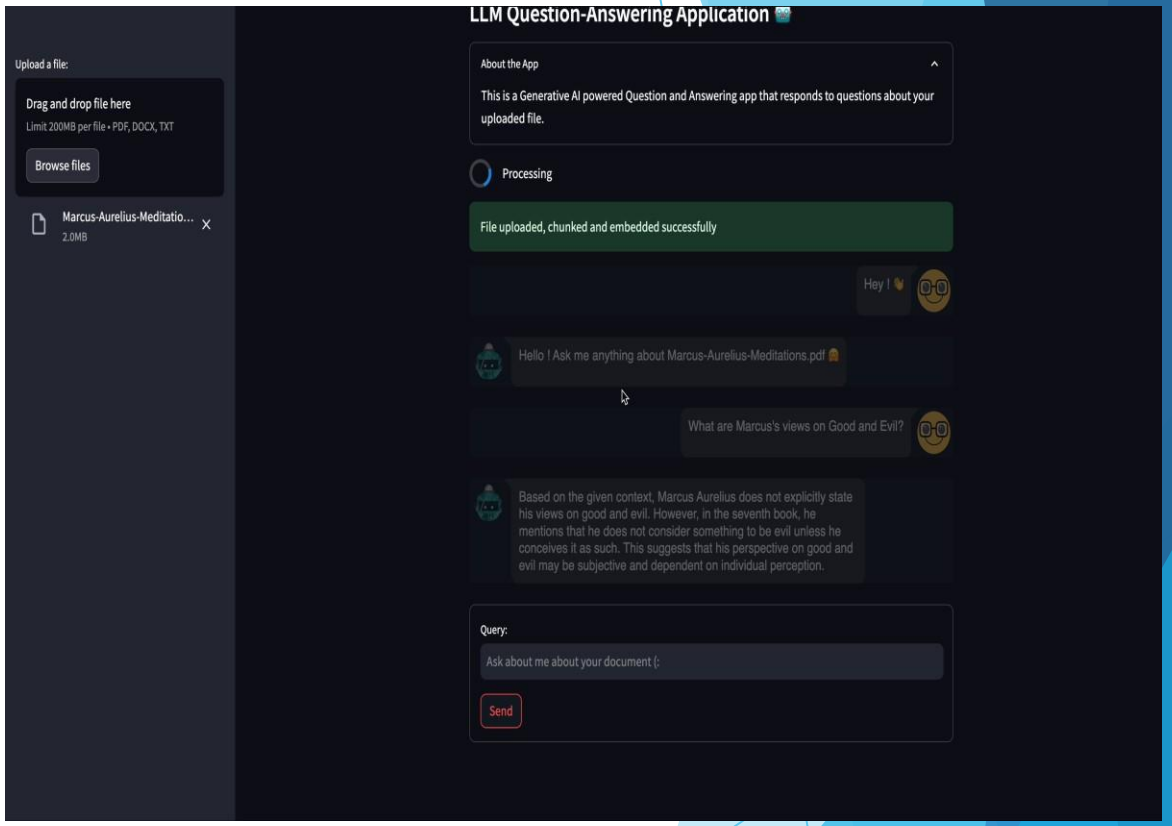
Prepare to be amazed by our groundbreaking solution, where cutting-edge AI seamlessly combines with an intuitive interface! Imagine effortlessly uploading documents in any format—PDFs, DOCX, or TXT files—and within moments, our advanced system utilizes the all-MiniLM-L6-v2 model to meticulously analyze and extract vital insights with unmatched precision. And that's not all! Rest assured, your data's security is our top priority, as we employ robust measures to securely store embeddings using FAISS vector storage, guaranteeing complete confidentiality. And when it comes to speed? Get ready for lightning-fast query responses delivered in just minutes! With our solution, efficiency, affordability, and accuracy merge seamlessly within an intuitive platform tailored to meet all your needs, delivering an experience that truly astounds.

MODELLING


PDF | Text
Text Splitter
Chunks
Pinecone
Query
Embeding Generator
Embedings
Vector DB
OpenAI / Sentence Transfomers
Query + Context
Relevant docs
QnA Model
GPT-3, GPT-4
Answer

RESULTS





https://github.com/roshankumar2003/TNSDC_GENAI