TAMILARASU N

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

Current document analysis tools often encounter challenges such as inefficiency and restricted accessibility. Users frequently encounter difficulties navigating through complex interfaces or bearing significant expenses for analysis. Furthermore, securely storing and retrieving document embeddings poses notable hurdles. This underscores the need for a solution that offers both simplicity and cost-efficiency. Consequently, there is a crucial demand for a streamlined approach that effectively manages documents, extracts pertinent information, and grants easy access to insights, all while ensuring security and user contentment are upheld.

PROJECT OVERVIEW

The LLM Question-Answering Application presents an intuitive interface tailored for effortlessly extracting insights from documents. Users commence by inputting their OpenAI API keys and uploading documents in PDF, DOCX, or TXT formats. Subsequently, the application employs the all-MiniLM-L6-v2 model from HuggingFace to process and embed the content. This ensures that users do not encounter any costs for generating embeddings, with processing times typically averaging between 1 to 2 minutes, contingent upon file size and computational resources available.

Following document preparation, the embeddings are securely stored in a vector store utilizing FAISS, a robust open-source library esteemed for its efficacy in conducting similarity searches and clustering of dense vectors. Upon conclusion of the document processing phase, users can proceed to submit their queries. In approximately a minute, the application furnishes the desired response, furnishing an efficient and seamless encounter 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.

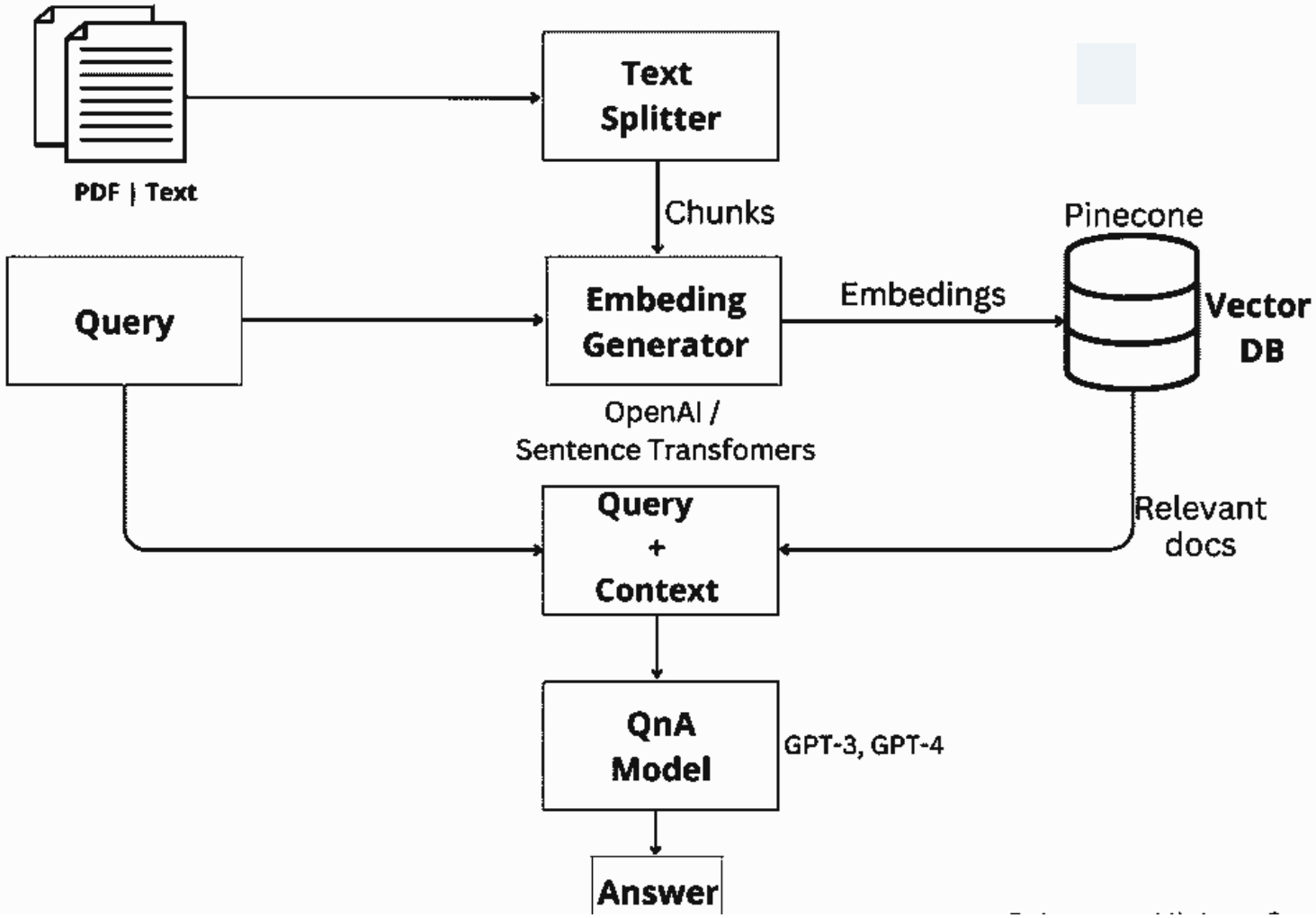
Value Proposition:

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.

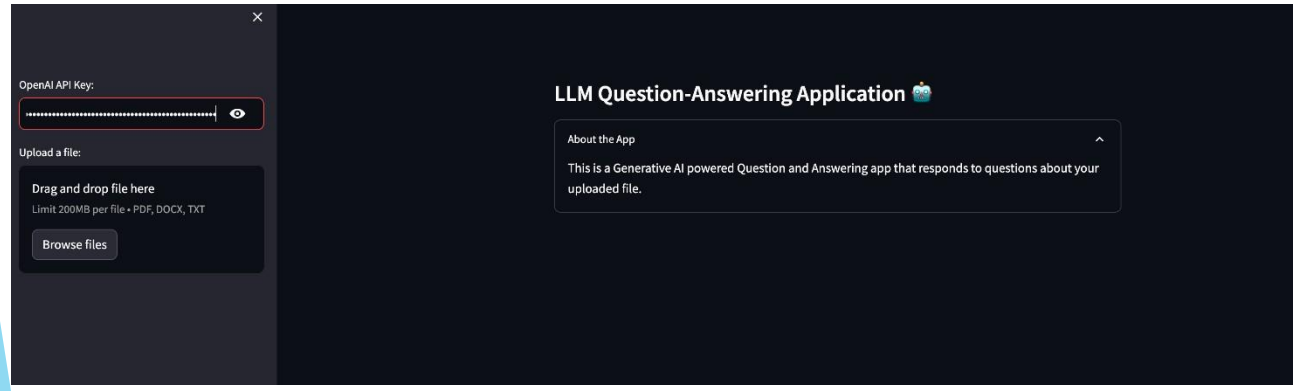
THE WOW IN YOUR SOLUTION

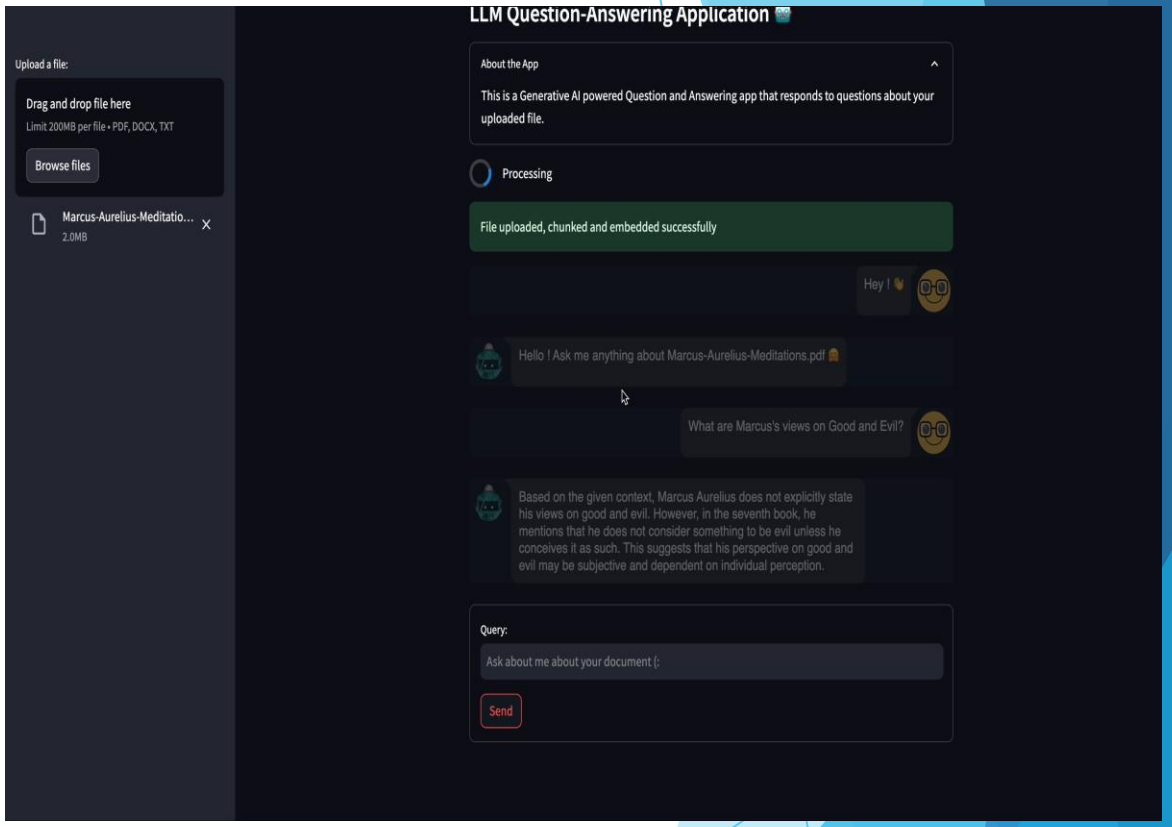
Get ready to be impressed by our revolutionary solution, where state-of-the-art AI seamlessly integrates with an intuitive interface! Picture effortlessly uploading documents in any format—be it PDFs, DOCX, or TXT files—and in mere moments, our cutting-edge system utilizes the all-MiniLM-L6-v2 model to meticulously analyze and extract crucial insights with unparalleled precision. But wait, there's more! Your data's security is our utmost concern, as we implement robust measures to securely store embeddings using FAISS vector storage, ensuring complete confidentiality. And as for speed? Brace yourself for lightning-fast query responses delivered in just minutes! With our solution, efficiency, affordability, and accuracy converge seamlessly within an intuitive platform tailored to cater to all your needs, providing an experience that truly dazzles.

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/Tamilarasu-2323/TNSDC_GENAI