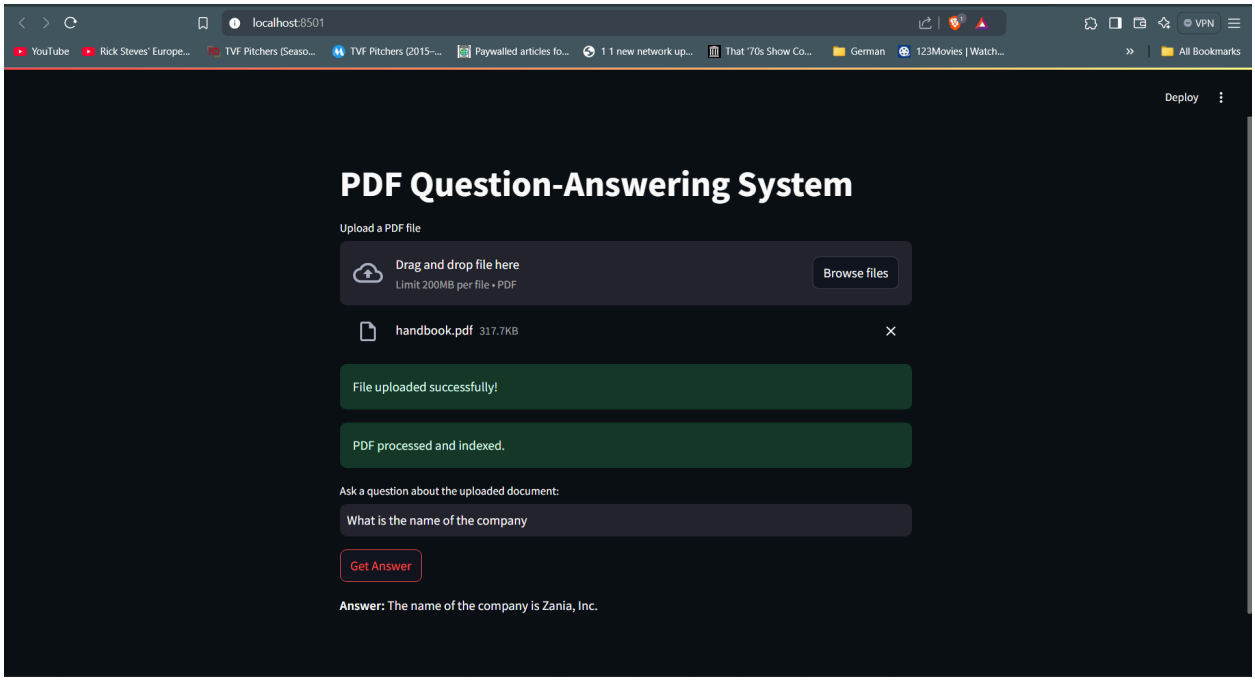
Assignment

- List all the ways you can make the solution more accurate?
  In addition to the solution there are other ways to make this solution more accurate,few of them are -
  Advanced Embedding Models: Upgrade from text-embedding-ada-002 to more advanced models for higher-quality embeddings.
  Context-Aware Retrieval: Incorporate additional context into embeddings or query expansion to better match queries with relevant chunks.
  Metadata-Enriched Search: Use more detailed metadata (e.g., chunk type, source) to improve search result relevance.
  Dynamic Parameter Tuning: Optimize parameters like top_k dynamically based on query complexity or user feedback.
  Error Handling: Add robust error handling for incomplete or low-quality embeddings and queries.
  Normalization and Preprocessing: Standardize input text for uniformity (e.g., case folding, punctuation removal) before embedding generation.
  Session level context: Maintaining sessions and providing context of previous conversation will help answer the follow up questions in a more human-like way.
  Vector Database and similarity search : We can use some different databases like cassandra, pinecone for vector database and even write our own similarity search mechanism for retrieval. Also, we can add the source of information in the webapp to show the source of information.

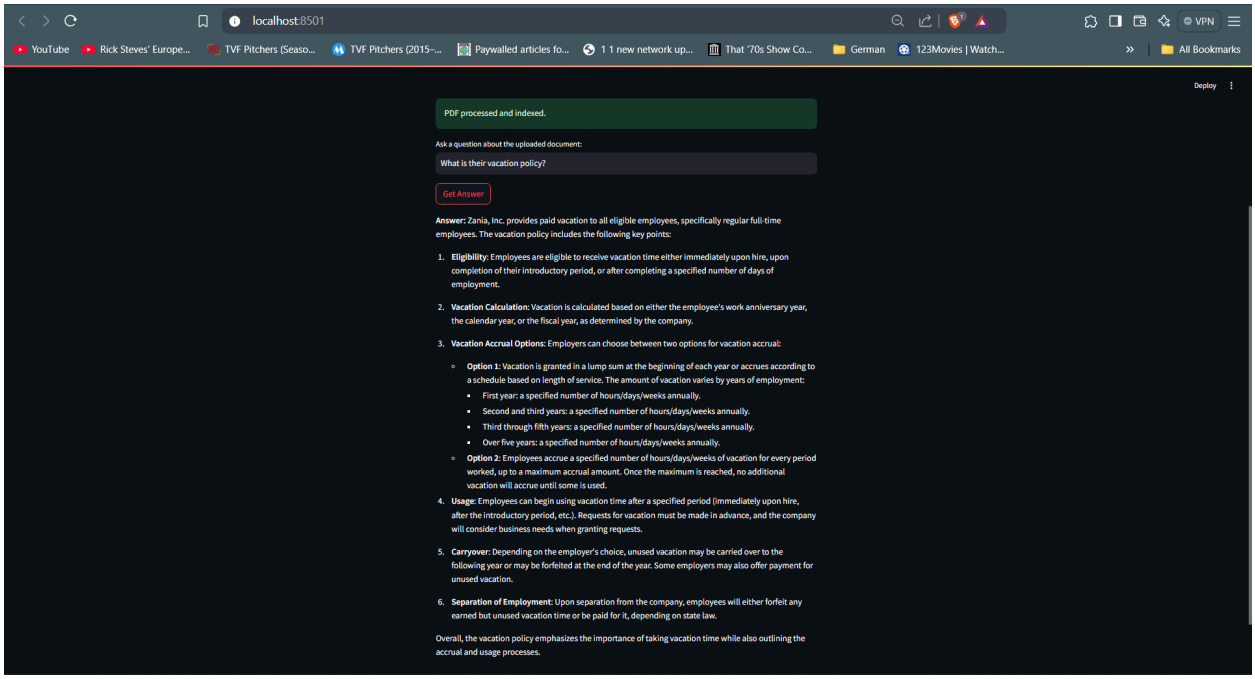
- How would you have made your code more modular, scalable and production grade?
  Below points could be considered to make code more modular,scalable and production grade -
  Configuration Management:
  Store API keys and other sensitive configurations securely using environment variables or secrets management tools.
  Code Organization:
  Structure utility functions into separate modules (e.g., `embedding_utils`, `search_utils`) for clear responsibilities.
  Dependency Injection:
  Pass dependencies (e.g., OpenAI clients, FAISS index) as arguments to functions or constructors for better testability.
  Asynchronous Execution:
  Use asynchronous calls for embedding generation and database queries to improve performance.
  Logging and Monitoring:
  Integrate logging for debugging and monitoring tools for performance tracking.
  Version Control and Documentation:
  Document API usage and maintain versioning for compatibility and easier updates.

DEMO of the webApp in action :-)

★ What is the name of the company?



★ What is their vacation policy?



Authored by- Amay Dubey