Final Report

Problem Statement:

The task involves implementing a simple Retrieval Augmented Generation (RAG) system using LangChain. The system should allow a user to upload a PDF file, process it by splitting it into manageable text chunks, embed those chunks, and then answer user questions based on the retrieved chunks using a language model.

Objective:

- To create a system that enables users to upload a PDF file and interact with its content by asking questions.
- To split the PDF content into manageable chunks for efficient embedding and retrieval.
- To generate answers based on the retrieved text chunks using a pre-trained language model.

Methodology:

1. **Setup and Library Installation:**
    - Install the necessary libraries such as LangChain, PyMuPDF, and other dependencies for PDF processing, embedding, and language modeling.

2. **PDF Upload and Processing:**
    - Implement a function to allow users to upload a PDF file.
    - Use PyMuPDF to open the PDF and extract its text content.
    - The extracted text is split into manageable chunks using a RecursiveCharacterTextSplitter with specified chunk size and overlap parameters to maintain context.

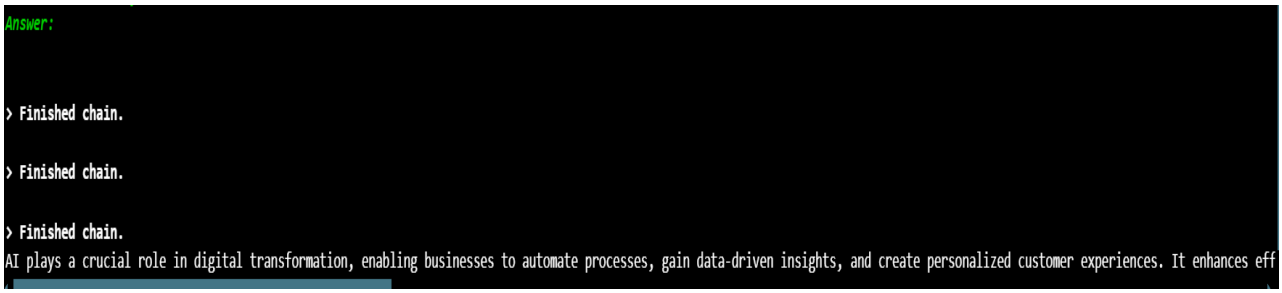
3. **Text Embedding:**
    - Embed the text chunks using the Google Generative AI Embeddings model. The embedding model is initialized using an API key stored securely in environment variables.
    - Store these embeddings in a FAISS (Facebook AI Similarity Search) index for efficient retrieval.

4. **Question Answering System:**
    - Set up a retriever to search through the embedded chunks and retrieve the most relevant ones based on the user's question.
    - Use a pre-trained language model (ChatGoogleGenerativeAI) to generate an answer using the retrieved chunks and the question provided.
    - A prompt template is defined to structure the interaction between the retriever, language model, and the user.

5. **Integration and Testing:**
    - Integrate all components to form a cohesive RAG system.
    - Test the system using different PDF files and questions to ensure the functionality meets the objectives.

Result:

The PDF "AI and Its Importance in Digital Transformation" was successfully processed by the RAG system. The question asked was "Give me the summary." and the output as screenshot below



Conclusion:

The implemented RAG system successfully allows users to interact with content from uploaded PDF files by asking questions. The system efficiently processes the PDF, splits the content into manageable chunks, embeds them, and uses a retrieval-based approach to generate accurate and contextually relevant answers. The solution leverages LangChain's capabilities to manage text embeddings and integrate with a powerful language model, providing an effective platform for question-answering tasks based on document content.