

**Project:** PDF QA ChatBot

**Project Link:** https://huggingface.co/spaces/AliXaidi/PDF\_ChatBot\_Ali\_nd\_Arooj

**Participants:** S M Ali Zaidi

Arooj Imtiaz

**Section:** Cohort DS4 / Violet

**Instructor:** Muhammad Umair

**PDF Q&A Chatbot Documentation**

# **Overview**

This application is a PDF Q&A Chatbot developed using Streamlit, a popular framework for building web applications in Python. It enables users to upload PDF documents and ask questions based on the content of these documents. The application utilizes LangChain for natural language processing and Google's Generative AI models to extract answers from the PDF content.

## **System Requirements**

* Python 3.8 or newer
* Streamlit
* PyPDF2 for PDF processing
* LangChain and its community packages for question-answering
* Google's generativeai library for embeddings and chat models
* dotenv for environment variable management

## **Setup**

**Environment Variables:** The application uses a .env file to store sensitive information like the Google API key. Ensure this file contains the GOOGLE\_API\_KEY variable.

**Dependencies:** Installed all required Python packages via pip

**Application Components**

Title and Header

The application sets a centred page configuration and displays a custom title using HTML styles.

**API Key Configuration**

The application retrieves the Google API key from the environment variables. If not found, it displays a warning and stops execution.

**File Upload**

Users can upload PDF files, which are then read into the application.

**Prompt Template**

A detailed prompt template is constructed, guiding the model on how to formulate answers based on the context extracted from the PDF.

**PDF Processing**

The application uses PyPDF2 to read the uploaded PDF file and extracts text from each page. This text serves as the context for answering questions.

**Text Splitting and Embedding**

Text is split into manageable chunks and processed through Google Generative AI embeddings. A vector index is created for the retrieval of relevant documents.

**Question Input and Processing**

Users enter their questions, which are then processed by the application to retrieve and display answers.

**Conclusion**

This document outlines the core functionalities and flow of the PDF Q&A Chatbot. By walking through each segment, one can understand how the application operates from the moment a PDF is uploaded to the generation of answers to user queries.