

AI bot that interacts with multiple pdf's

Pratham V¹, Likhith HJ²

¹Dept of Artificial Intelligence and Data Science, East West institute of technology

²Dept of Artificial Intelligence and Data Science, East West institute of technology

Abstract - This project involves the development of an intelligent AI bot capable of interacting with and extracting information from multiple PDF documents. Leveraging advanced natural language processing (NLP) techniques, the AI bot can read, comprehend, and respond to user queries based on the content of multiple PDF files. The bot ensures efficient data retrieval, providing concise and accurate answers, summaries, or insights. This tool aims to streamline the process of document management and information extraction, making it an invaluable asset for research, business, and educational purposes.

Key Words: AI Bot, NLP, PDF Interaction, Information Extraction, Document Management, Data Retrieval, Multi-PDF Analysis, User Queries, Automated Summarization, Research Tool, Business Intelligence, Educational Tool.

1. INTRODUCTION

In today's digital age, information retrieval from extensive documents can be tedious and time-consuming. "Chat with PDFs" is an innovative solution designed to enhance the user experience by enabling interaction with PDF documents via a conversational interface. This application leverages advanced natural language processing (NLP) techniques, powered by Google Generative AI and LangChain, to enable users to ask questions and receive relevant answers directly from PDF content.

2. Body of Paper

Efficiently extracting relevant information from extensive PDFs is challenging due to the lack of interactive and context-aware tools. Basic PDF viewers provide search functionality but are limited to exact keyword matches. Time-consuming for users to sift through large amounts of irrelevant data. A Python-based solution using **Streamlit** for a user-friendly interface Leverages **Google Generative AI** for contextual understanding of PDF content. Implements **FAISS** for efficient semantic search using embeddings generated by LangChain. Enable conversational interaction with PDF documents. Facilitate multi-document semantic search. Improve user productivity by reducing the time spent searching.

3. CONCLUSIONS

The "Chat with PDFs" project demonstrates a significant advancement in document analysis by combining AI-driven conversational interfaces with robust semantic search mechanisms. It simplifies information retrieval, reduces manual effort, and showcases the potential of emerging AI technologies.

ACKNOWLEDGEMENT

This project was a collaborative effort made possible by the hard work and dedication of its authors, Pratham V and Likhith HJ. We would like to extend our deepest gratitude to Akshatha K for their invaluable mentorship and support throughout the development of this project. Their guidance and insights have been instrumental in shaping our research and outcomes.

We would also like to thank our colleagues and friends for their constant support, constructive feedback, and encouragement. Their contributions have been essential in refining our work and achieving our goals.

Our heartfelt thanks go to our families for their unwavering support and understanding during this journey. Their belief in us has been a constant source of motivation.

Lastly, we acknowledge the use of various resources and tools, including advanced AI and NLP technologies, which have empowered the development of this project.

Pratham V & Likhith HJ

REFERENCES

- LangChainDocumentation: https://langchain.com
- FAISSDocumentation: https://github.com/facebookresearch/faiss
- PyPDF2Library: https://pypdf2.readthedocs.io
- Streamlit Framework: https://streamlit.io
- Google Generative AI: https://cloud.google.com/generative-ai