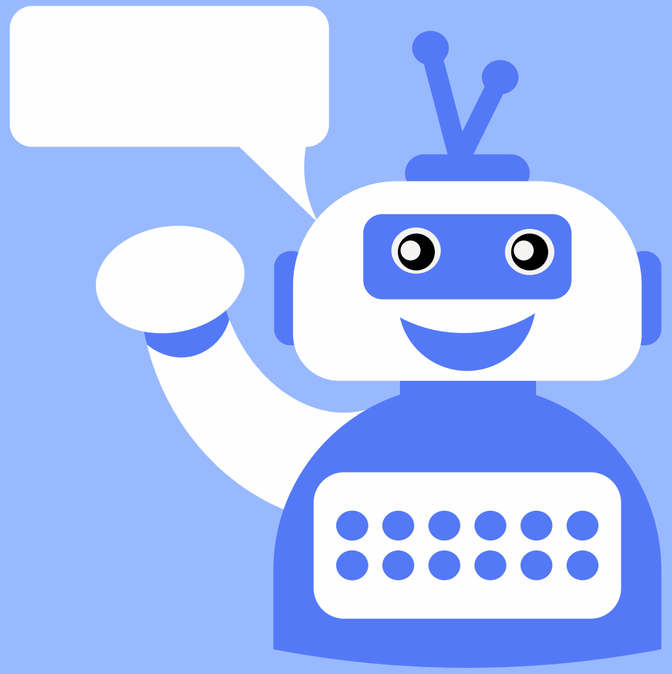
PDF-KNOWS

A Chatbot for Comprehensive Understanding of Complex Documents

CONTENTS



1 About the Project

2 Why PDF-KNOWS

3 Dataset Description

4 Structure of Chatbot

5 Usecases of Chatbot

6 Technologies Tired Before

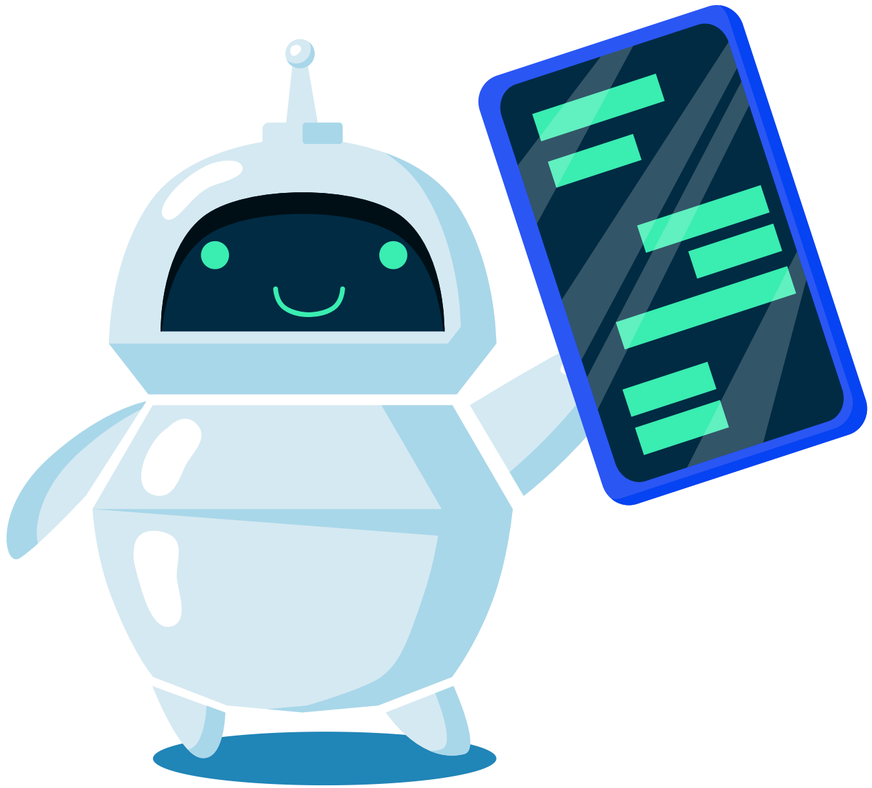
7 Unique Value Proposition

8 Results

9 Conclusion

ABOUT THE PROJECT

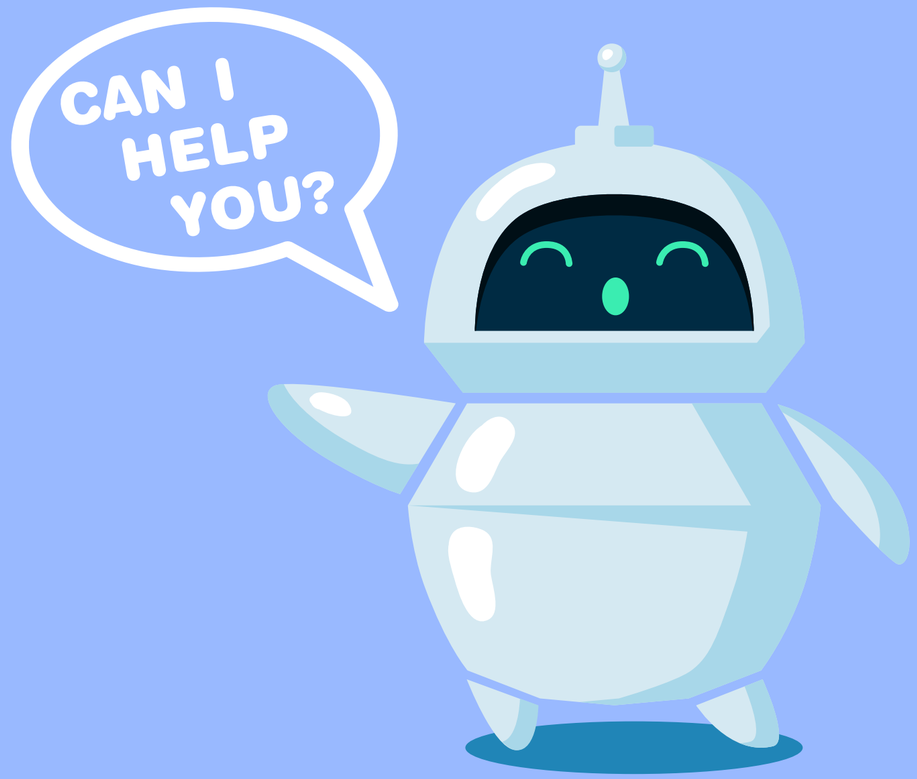
The focus of our college project lies in developing an advanced document navigation system aimed at simplifying the process of traversing intricate documents. Leveraging cutting-edge Artificial Intelligence and Machine Learning technologies, we aim to revolutionize the way individuals interact with and extract information from complex textual content.



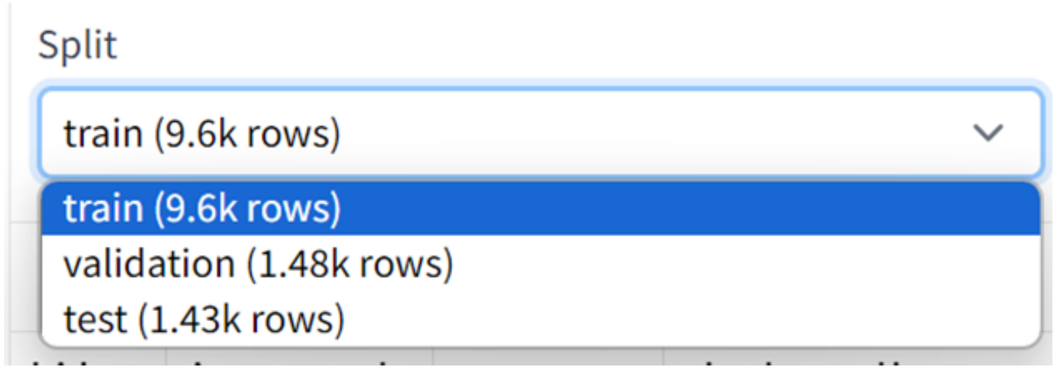
WHY PDF-KNOWS

In today's information-driven world, individuals and businesses frequently encounter the challenge of efficiently extracting key insights and information from lengthy documents such as PDFs, Word documents, and PowerPoint presentations. The process of manually sifting through extensive content to identify essential details can be time-consuming, labor-intensive, and error-prone.

To address this issue, we propose the development of an AI-powered document summarization and chatbot integration system.



DATASET DESCRIPTION



Size of downloaded dataset files:
379 MB

Size of the auto-converted Parquet files:
157 MB

Number of rows:
12,515

"kmfoda/booksum"

dataset is specifically designed for training and evaluating models in the domain of book summarization

Hugging Face Model Hub

The dataset likely contains pairs of books and their corresponding summaries. .

LM = Language Model

STRUCTURE OF CHATBOT

User Interface

Chatbot Frontend

Natural Language Processing (NLP) Module

PDF Parsing Module

Document Preprocessing

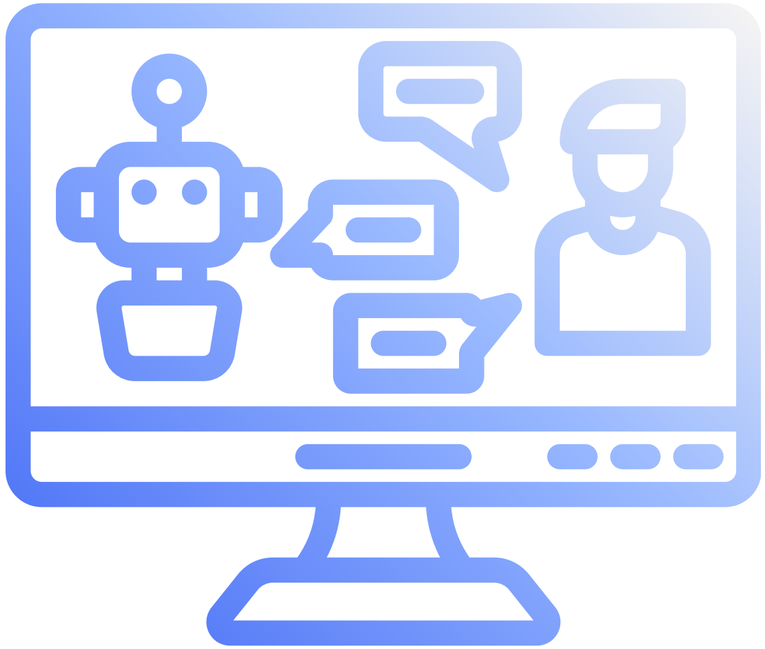
Knowledge Graph or Indexing System

Intent Matching and Query Resolution

Response Generation

USECASES OF CHATBOT

- Legal Document Analysis
- Financial Document Understanding
- Insurance Policy Understanding



- Academic Research Assistance
- Technical Document Support
- Education and E-Learning

Technologies Tried Before

- Machine Learning Frameworks
- Chatbot Frameworks
- Cloud Services for Scalability
- Knowledge Graphs

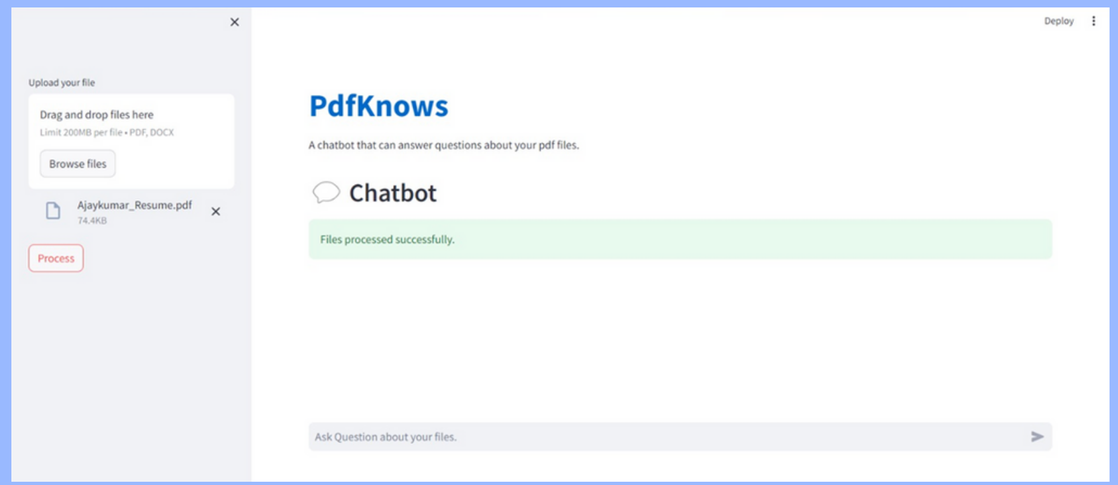
Unique Value Proposition

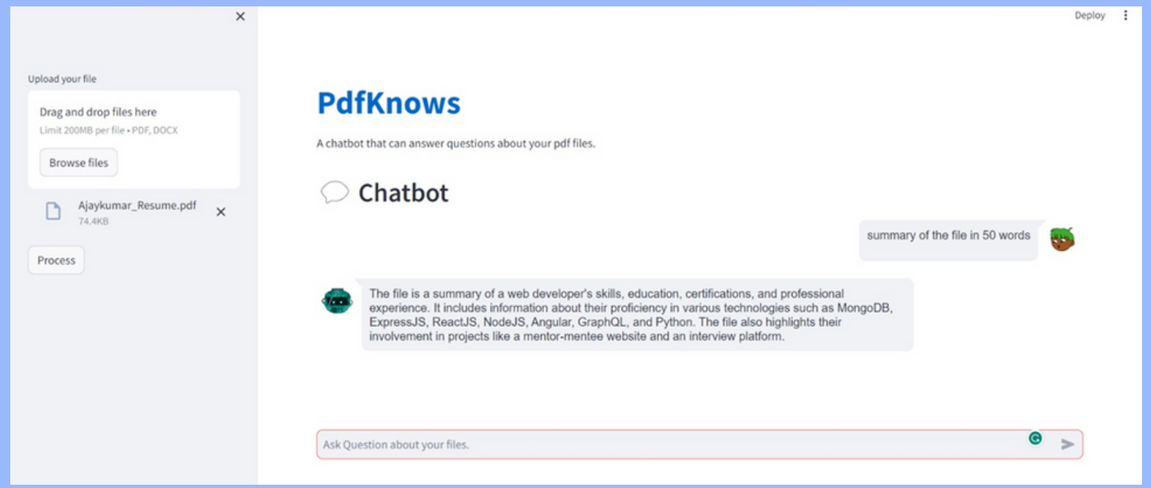
The proposed project addresses the challenge of extracting vital information from lengthy documents. It combines document summarization and chatbot integration, offering a comprehensive solution. Document summarization condenses content for easier consumption, while the chatbot interface facilitates interactive communication. This integrated system aims to enhance efficiency and accuracy in information retrieval from documents like PDFs and presentations.

RESULTS


Deploy
Upload your file
Drag and drop files here
Limit 200MB per file • PDF, DOCX
Browse files
Process
PdfKnows
A chatbot that can answer questions about your pdf files.
Chatbot
Made with Streamlit


Deploy
Upload your file
Drag and drop files here
Limit 200MB per file • PDF, DOCX
Browse files
Ajaykumar_Resume.pdf
74.4KB
Process
PdfKnows
A chatbot that can answer questions about your pdf files.
Chatbot
Made with Streamlit


Deploy
Upload your file
Drag and drop files here
Limit 200MB per file • PDF, DOCX
Browse files
Ajaykumar_Resume.pdf
74.4KB
Process
PdfKnows
A chatbot that can answer questions about your pdf files.
Chatbot
Files processed successfully.
Ask Question about your files.


Deploy
Upload your file
Drag and drop files here
Limit 200MB per file • PDF, DOCX
Browse files
Ajaykumar_Resume.pdf
74.4KB
Process
PdfKnows
A chatbot that can answer questions about your pdf files.
Chatbot
summary of the file in 50 words
The file is a summary of a web developer's skills, education, certifications, and professional experience. It includes information about their proficiency in various technologies such as MongoDB, ExpressJS, ReactJS, NodeJS, Angular, GraphQL, and Python. The file also highlights their involvement in projects like a mentor-mentee website and an interview platform.
Ask Question about your files.

CONCLUSION

The PDF-KNOWS chatbot is a significant breakthrough in extracting vital information from intricate documents. By merging advanced natural language processing with PDF parsing techniques, it enables conversational interaction for streamlined access to critical insights. Integration of technologies like Langchain and OpenAI API enhances the chatbot's language understanding, resulting in a robust system for precise document navigation. This innovation marks a substantial leap forward in improving efficiency and productivity in document management.

THANK YOU!