INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

DocBot: Intelligent Document Chatbot Using Artificial Intelligence and Natural Language Processing

[1]P.Durga, [2]V.Rishi, [3]V.Kavyasree, [4]A.Ganesh, [5]Dr,P.Srinivas Kumar,
[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
[1] Computer Science and Engineering,
[1]SRK Institute of Technology, Enikepadu, Vijayawada.

Abstract : Understanding large PDF documents manually is time-consuming and inefficient. Existing document search systems mainly rely on keyword-based retrieval, which lacks contextual understanding and fails to answer user-specific queries accurately. This paper proposes DocBot, an intelligent document chatbot that enables users to interact with PDF documents using natural language queries. The system automatically extracts text from documents, converts it into semantic vector representations, and retrieves relevant information using similarity search. Context-aware question answering is achieved using Natural Language Processing techniques integrated with FAISS and transformer-based embeddings. In addition, DocBot supports multilingual translation and image or diagram explanation present in documents. Experimental evaluation shows that the proposed system significantly reduces information retrieval time while providing accurate and meaningful responses. The system is user-friendly, scalable, and suitable for educational and technical document analysis.

IndexTerms **- Document Chatbot, Natural Language Processing, Retrieval-Augmented Generation (RAG), LangChain, FAISS, BERT Embeddings, Multimodal AI, Image Understanding, Machine Learning, Artificial Intelligence, PDF Analysis, Semantic Search, Conversational AI.**

I. INTRODUCTION

The rapid growth of digital information has resulted in an overwhelming increase in the number of documents available across various domains such as education, healthcare, research, and business. Most of these documents are stored in formats such as PDFs, which makes extracting relevant information a challenging and time-consuming task. Traditional document retrieval systems primarily rely on keyword-based search techniques, which often fail to understand the context and intent of user queries, leading to inefficient and inaccurate results.

With the advancement of Artificial Intelligence (AI) and Natural Language Processing (NLP), there is a growing need for intelligent systems that can interact with documents in a human-like manner. In recent years, conversational AI systems have gained significant attention due to their ability to understand natural language queries and provide meaningful responses. However, many existing systems lack the capability to handle complex document structures, including images and diagrams, and often do not support multilingual interactions.

To address these limitations, this paper proposes DocBot, an intelligent document chatbot that enables users to interact with PDF documents using natural language queries. The system leverages Retrieval-Augmented Generation (RAG) techniques, integrating semantic search using FAISS with transformer-based embeddings to provide context-aware answers. Additionally, DocBot incorporates multimodal capabilities by analyzing both textual and visual content, allowing it to explain images and diagrams within documents.

The system also supports multilingual translation, making it accessible to a broader range of users. By combining NLP, machine learning, and deep learning techniques, DocBot provides an efficient and user-friendly solution for document understanding and information retrieval. This approach enhances user experience by reducing manual effort and improving the accuracy and relevance of retrieved information.

II. NEED OF THE STUDY.

The rapid growth of digital documents has led to an increasing need for efficient information retrieval systems. Traditional methods of searching information within documents rely on keyword-based approaches, which often fail to understand the contextual meaning of user queries. This results in time-consuming manual searching and reduced productivity.
With advancements in Artificial Intelligence and Natural Language Processing, there is a growing demand for intelligent systems that can understand user queries in natural language and provide accurate, context-aware responses. The proposed DocBot system addresses this challenge by enabling users to interact with documents through conversational interfaces.

Furthermore, the inclusion of multimodal capabilities such as image and diagram explanation, along with multilingual support, enhances accessibility and usability. This system is particularly beneficial in domains such as education, research, healthcare, and enterprise environments, where quick and precise document understanding is essential.

III. RELATED WORKS

Recent advancements in Artificial Intelligence and Natural Language Processing have significantly improved document understanding and conversational systems. Researchers have focused on developing intelligent systems capable of extracting meaningful information from large volumes of unstructured data such as PDF documents.

3.1Literature Review

Lewis et al. (2020) introduced Retrieval-Augmented Generation (RAG), which combines document retrieval with generative models to produce context-aware responses. This approach improved answer accuracy by grounding responses in external knowledge sources.

Johnson et al. (2017) developed FAISS, an efficient similarity search library that enables fast retrieval of high-dimensional vector embeddings. It is widely used in large-scale semantic search applications.

Reimers and Gurevych (2019) proposed Sentence-BERT, which enhances semantic understanding by generating meaningful sentence embeddings, improving performance in tasks such as document similarity and question answering.

Chen et al. (2022) introduced DocQA systems that extract answers directly from documents. While these systems provide accurate results, they lack support for image-based understanding and multimodal interaction.

Recent tools such as ChatPDF (2023) and LangChain-based frameworks (2024) enable conversational interaction with PDF documents. However, these systems often have limitations such as lack of deep contextual understanding, limited multilingual support, and inability to explain visual content like images and diagrams.

Despite these advancements, existing approaches primarily focus on text-based processing and fail to integrate multimodal capabilities and efficient retrieval mechanisms into a unified system. The proposed DocBot system addresses these limitations by combining RAG, semantic search, image analysis, and multilingual support into a single framework.

Table.1. Comparison Table

System Type	Limitations	Advantages
Keyword-Based Search	No contextual understanding	Simple and fast
DocQA Systems	No support for images or diagrams	Accurate text extraction
RAG-Based Systems	High computational cost	Context-aware responses
Proposed DocBot	Slightly complex architecture	Multimodal, scalable, accurate

3.2 Comparison with Previous Methodology

Traditional document retrieval systems rely on keyword-based matching, where users must search manually for relevant information. These systems do not understand the semantic meaning of queries and often return irrelevant results. Additionally, earlier systems lack the ability to process images or diagrams present in documents, limiting their effectiveness.

Recent systems such as DocQA and ChatPDF have improved document interaction by enabling question-answering capabilities. However, these systems are primarily text-focused and do not fully utilize multimodal data. They also lack efficient retrieval mechanisms in large datasets and provide limited support for multilingual users.

In contrast, the proposed DocBot system utilizes a hybrid approach combining Retrieval-Augmented Generation (RAG), FAISS-based vector search, and transformer-based embeddings for semantic understanding. It enhances traditional methods by incorporating image analysis using deep learning models and multilingual translation capabilities. This results in improved accuracy, faster retrieval, and a more interactive user experience.

3.3 Proposed framework

The proposed DocBot system is designed using a hybrid architecture that integrates Natural Language Processing (NLP), Machine Learning, and Retrieval-Augmented Generation (RAG) techniques to enable intelligent document interaction. The system provides a seamless pipeline that transforms static PDF documents into an interactive conversational interface.

Initially, the user uploads one or more PDF documents through the application interface. These documents are processed to extract both textual and visual content using document parsing techniques. The extracted text is then preprocessed and divided into smaller segments to improve retrieval efficiency and contextual understanding.

In the next stage, each text segment is converted into vector embeddings using transformer-based models such as BERT or Sentence-BERT. These embeddings capture the semantic meaning of the document content and are stored in a FAISS vector database, enabling efficient similarity-based search.

When a user submits a query, the system converts the query into an embedding and performs semantic similarity matching with the stored document embeddings. The most relevant text segments are retrieved and passed to a language model through LangChain to generate accurate and context-aware responses.

Additionally, the system incorporates multimodal capabilities by analyzing images and diagrams present in the document using deep learning models such as Convolutional Neural Networks (CNN) and Vision Transformers. This allows the system to provide explanations not only for textual content but also for visual elements.

Furthermore, the system supports multilingual interaction by integrating translation models such as MarianMT, enabling users to query and receive responses in different languages.

The entire framework operates as an integrated pipeline combining document processing, semantic retrieval, response generation, and multimodal analysis, resulting in an efficient, scalable, and user-friendly document chatbot system.

3.4 Main Methodology

The proposed DocBot system follows a structured pipeline that converts user-uploaded documents into an interactive conversational system. The methodology integrates document processing, semantic retrieval, and intelligent response generation to provide accurate and context-aware answers.

Document Upload:
Users upload PDF documents through the user interface. The system accepts multiple documents and temporarily stores them for processing.

Text and Image Extraction:
The uploaded PDF is processed using document parsing libraries such as pdfplumber or PyMuPDF. Text content is extracted, and images or diagrams present in the document are also identified for further analysis.

Text Preprocessing:
The extracted text is cleaned and segmented into smaller chunks using text-splitting techniques. This improves processing efficiency and ensures better retrieval performance.

Embedding Generation:
Each text chunk is converted into vector embeddings using transformer-based models such as BERT or Sentence-BERT. These embeddings capture the semantic meaning of the text.

Vector Storage:
The generated embeddings are stored in a FAISS vector database, enabling fast and efficient similarity-based retrieval.

User Query Processing:
Users interact with the system by asking questions in natural language. The query is processed using Natural Language Processing techniques to understand user intent.

Semantic Retrieval:
The system retrieves the most relevant document chunks from the FAISS database by comparing the query embedding with stored embeddings using similarity search.

Response Generation:
The retrieved content is passed to a language model through LangChain, which generates a context-aware and meaningful response based on the document.

Image and Diagram Analysis:
If the query relates to visual content, the system analyzes images using deep learning models such as Convolutional Neural Networks (CNN) and Vision Transformers to provide explanations.

Multilingual Translation:
The generated response can be translated into different languages using models like MarianMT, allowing users to interact in their preferred language.

Display of Results:
Finally, the system displays the generated response to the user through the interface, providing a seamless conversational experience.

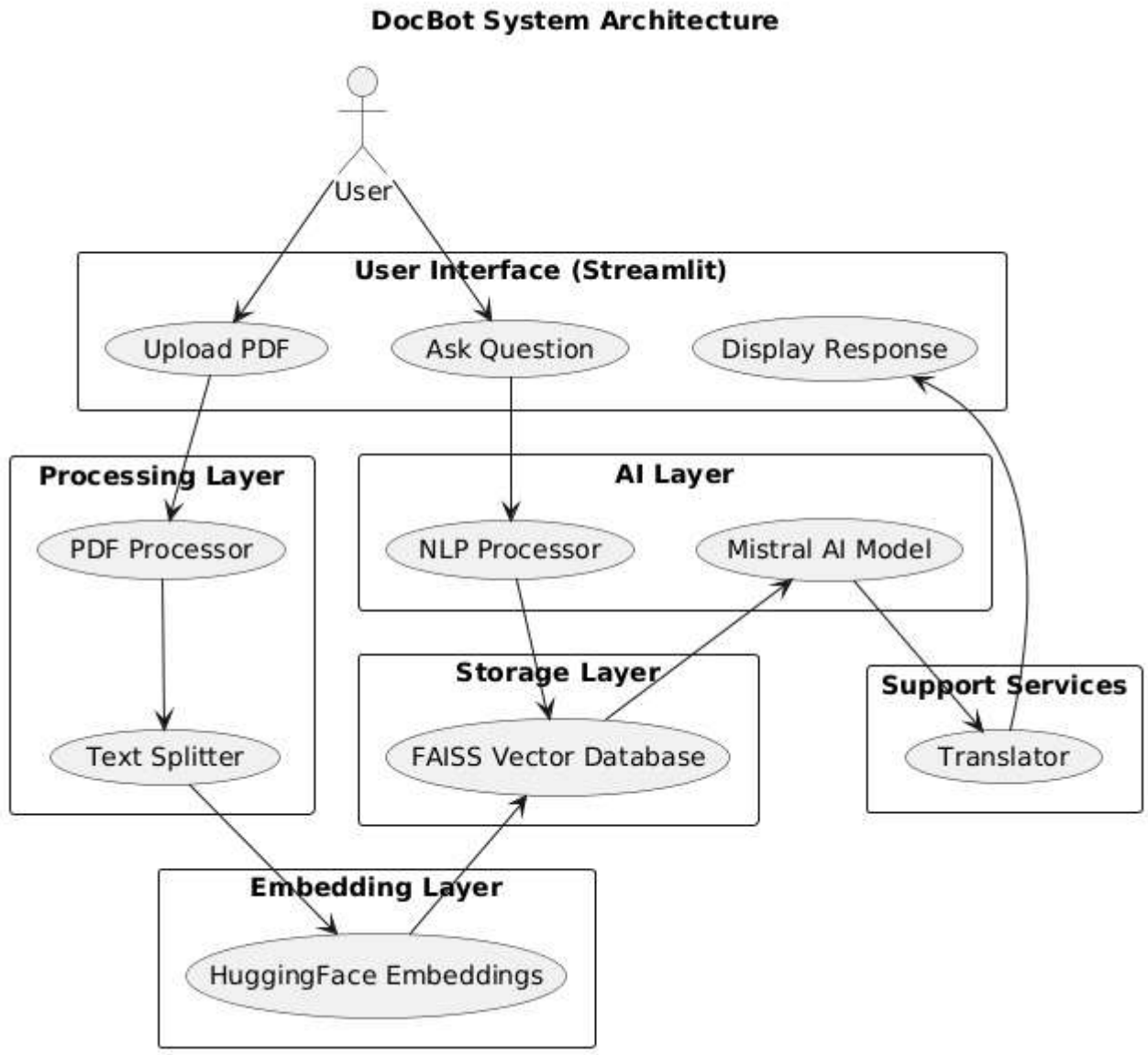

Fig.1. Architecture of the Proposed DocBot System

3.4.1 Implementation

1. Development Environment Setup

The system is developed using Python as the primary programming language. Frameworks such as Streamlit are used for building the user interface, while libraries like LangChain, FAISS, and Hugging Face transformers are used for backend processing. Additional libraries such as pdfplumber are used for document extraction.

2. Document Input

Users can upload PDF documents through the web interface. The system supports multiple document uploads, enabling users to analyze large collections of files.

3. Document Processing

Once uploaded, the documents are processed to extract textual and visual content. The extracted text is divided into smaller chunks using recursive text splitting techniques to improve efficiency.

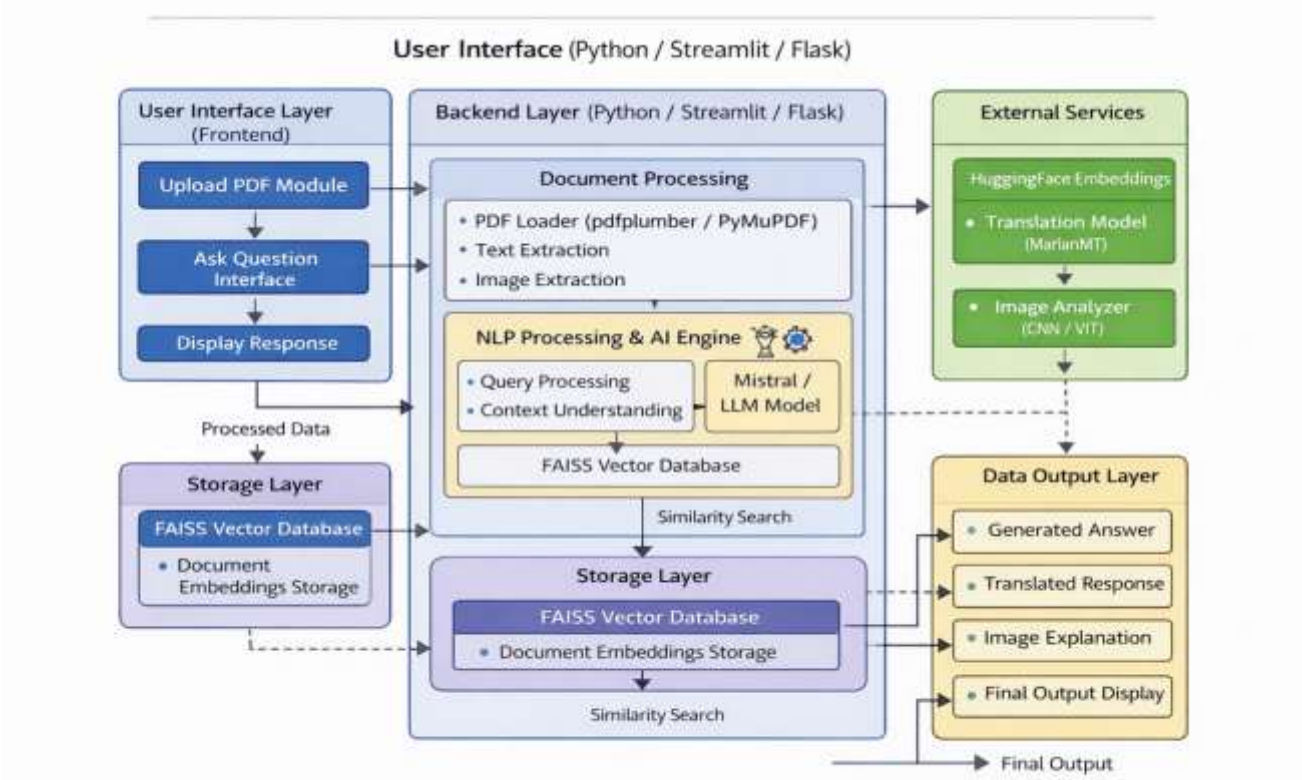

Fig.2. Implementation of DocBot System

4. Embedding and Storage
The processed text is converted into embeddings using pre-trained transformer models. These embeddings are stored in a FAISS vector database for efficient similarity search and retrieval.

5. Query Handling
When a user submits a query, it is converted into an embedding and compared with stored embeddings in the vector database. The system retrieves the most relevant document segments based on semantic similarity.

6. Response Generation
The retrieved content is passed to a language model through LangChain, which generates a coherent and context-aware response. The system ensures that responses are accurate and relevant to the user's query.

7. Image Explanation
For queries related to images or diagrams, the system uses deep learning models to analyze visual content and generate descriptive explanations.

8. Multilingual Support
The system integrates translation models such as MarianMT to provide responses in multiple languages, enhancing accessibility for diverse users.

9. User Interface Integration
The final output is displayed through an interactive interface, allowing users to view answers, ask follow-up questions, and explore document content efficiently.

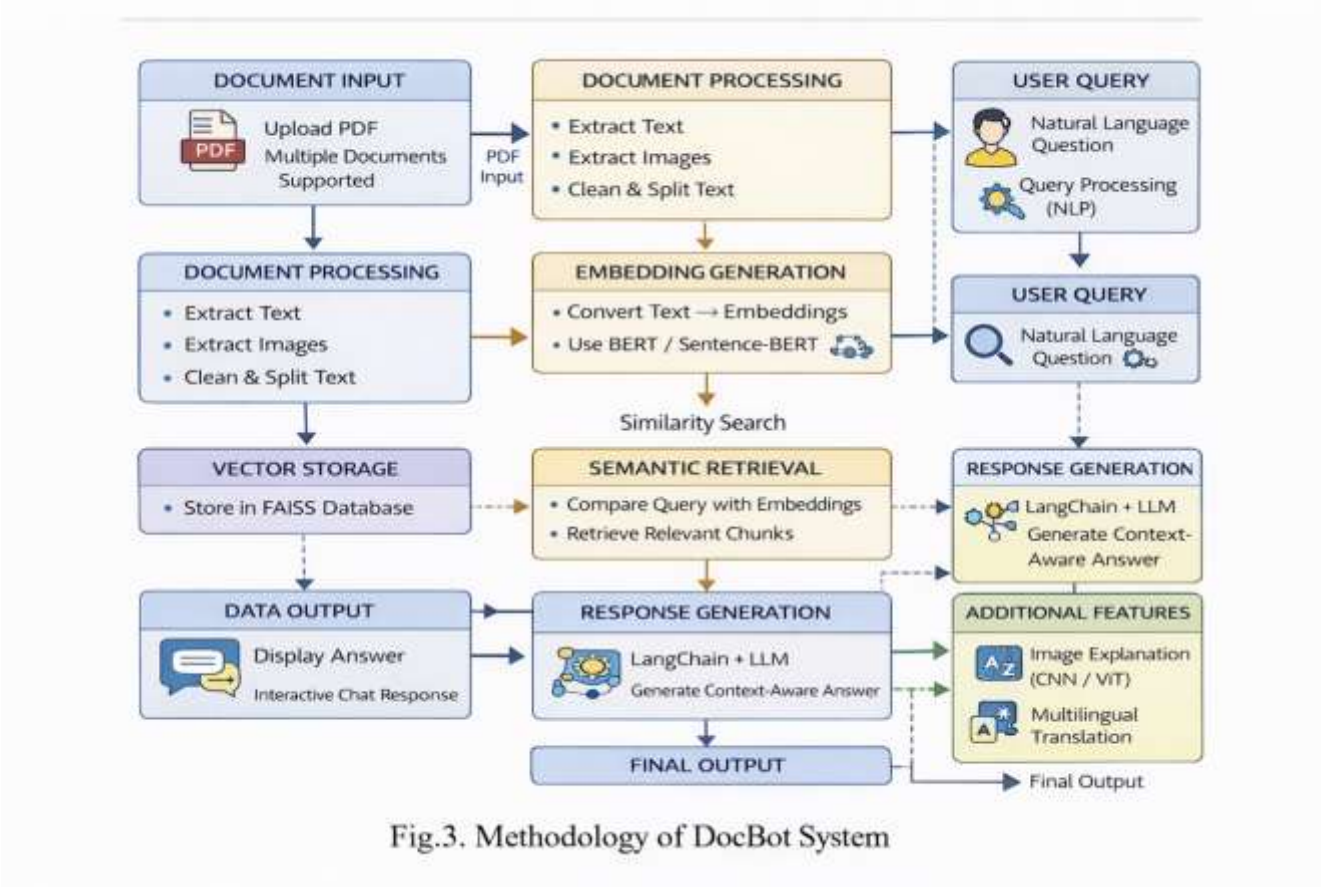

Fig.3. Methodology of DocBot System

IV. RESULTS AND DISCUSSION

4.1 System output screenshots and eplanation

The proposed DocBot system was tested using multiple PDF documents to evaluate its performance in document understanding, question answering, image explanation, and multilingual translation. The system demonstrated efficient processing and accurate response generation in real-time.

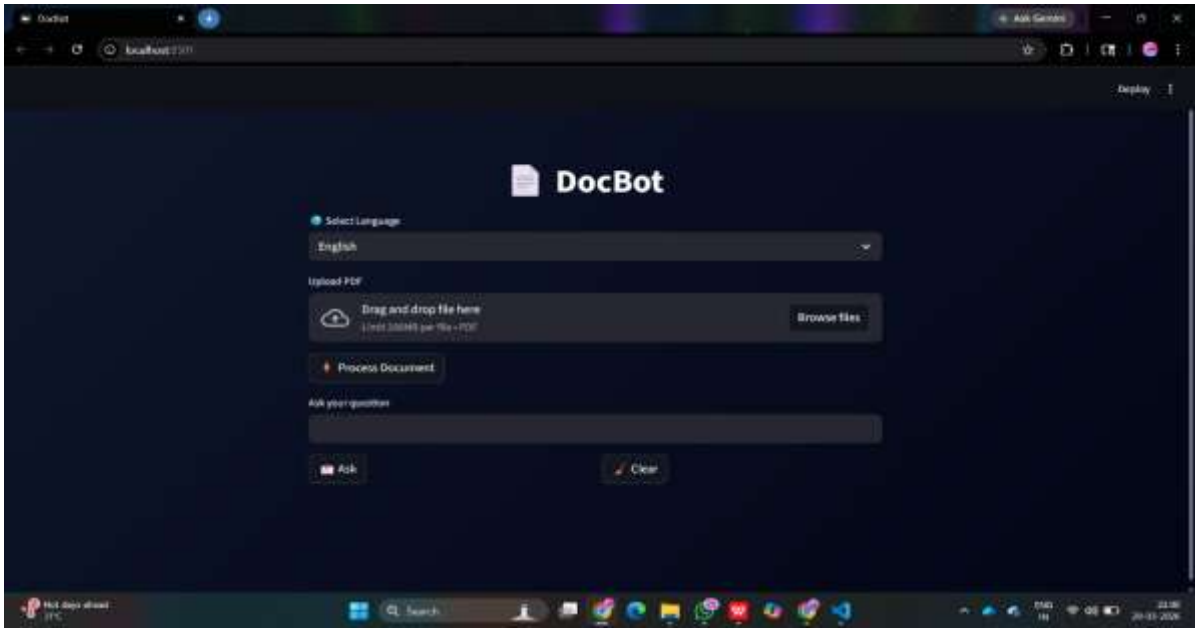

Fig.4.Main Interface

After uploading a document, users can ask questions such as summarization. The system successfully generates concise summaries by analyzing document content using NLP techniques. This demonstrates the system's ability to understand and condense large amounts of information effectively

.

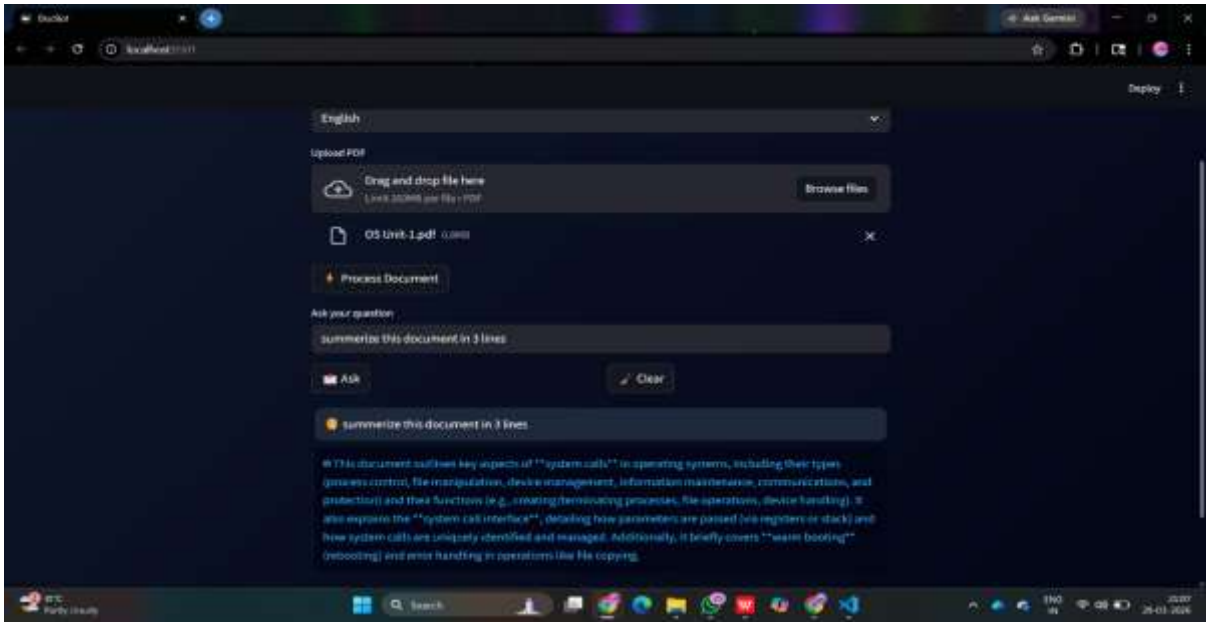

Fig.5. Document Processing and Summarization

The system supports multiple languages, allowing users to receive responses in their preferred language. As shown in the figure, the system successfully translates responses into Telugu, demonstrating its ability to enhance accessibility and usability for diverse users.

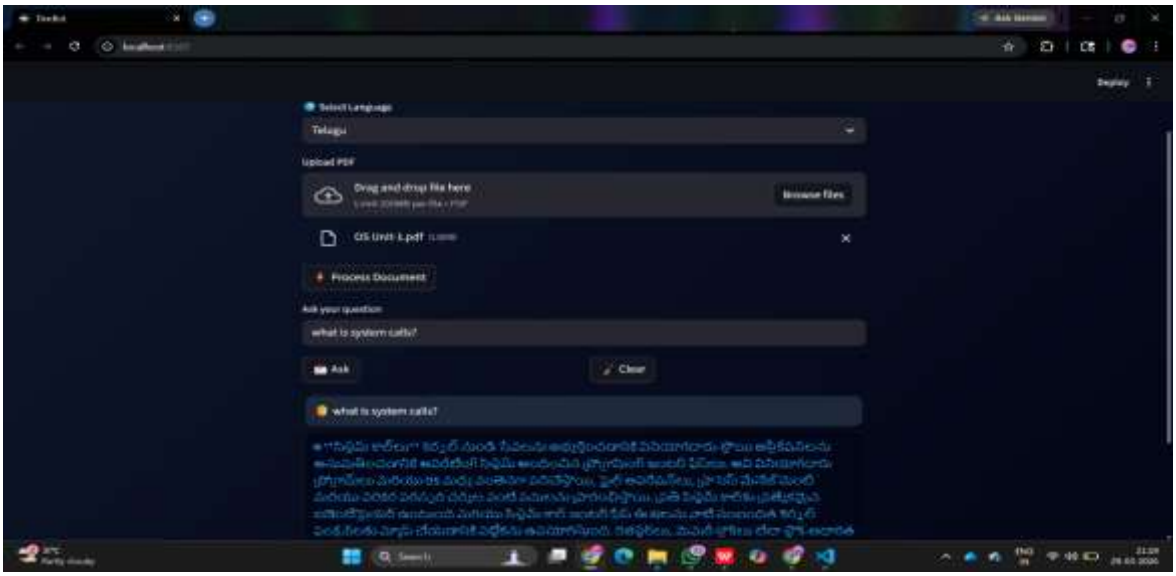

Fig.6.Multilingual Response Generation

The overall system performance indicates that the integration of Retrieval-Augmented Generation (RAG), FAISS-based semantic search, and transformer-based models provides accurate and context-aware responses. The system also maintains low latency during document processing and query handling, ensuring a smooth user experience.
The combination of text understanding, image analysis, and multilingual capabilities makes DocBot a comprehensive solution for intelligent document interaction.

4.2 Conclusion

The proposed DocBot system demonstrates an effective approach for intelligent document understanding by integrating Artificial Intelligence, Natural Language Processing, and Retrieval-Augmented Generation techniques. The system successfully enables users to interact with PDF documents using natural language queries, providing accurate, context-aware, and meaningful responses.
By utilizing FAISS-based semantic search and transformer-based embeddings, the system improves information retrieval efficiency compared to traditional keyword-based methods. The inclusion of multimodal capabilities, such as image and diagram analysis, further enhances the system's ability to interpret complex document content. Additionally, multilingual support allows users to access information in different languages, increasing the accessibility and usability of the system.
The experimental results show that DocBot reduces the time required for manual document analysis while maintaining high accuracy and low latency in response generation. The integration of various components into a unified pipeline ensures a seamless and interactive user experience.

Overall, the DocBot system provides a scalable and efficient solution for intelligent document interaction, making it suitable for applications in education, research, enterprise systems, and digital knowledge management.

4.3 Future Scope

- Although the proposed DocBot system provides an efficient solution for intelligent document understanding, there are several areas for further improvement and enhancement. One of the key future directions is the integration of advanced deep learning models to improve accuracy in complex queries and large-scale document processing.
- The system can be extended to support additional document formats such as Word documents, images, and scanned PDFs using Optical Character Recognition (OCR) techniques. This would increase the versatility of the system in handling diverse data sources.
- Another important enhancement is the incorporation of voice-based interaction, enabling users to ask questions through speech and receive spoken responses, thereby improving accessibility and user experience.
- The multilingual capabilities can be further expanded by supporting more languages and improving translation quality using advanced transformer-based models. Additionally, real-time collaboration features can be introduced, allowing multiple users to interact with the same document simultaneously.
- Performance optimization techniques can be applied to reduce latency and improve scalability for deployment in cloud-based environments. Security features such as user authentication and document privacy protection can also be integrated for enterprise-level applications.
- Furthermore, the system can be enhanced with personalized recommendations and learning capabilities, enabling it to adapt to user preferences over time. These improvements will make DocBot a more robust, scalable, and intelligent system for future applications in education, research, and industry.

V. Acknowledgement

The authors would like to express their sincere gratitude to dr. srinivas kumar for his valuable guidance, support, and encouragement throughout the development of this project.

The authors also acknowledge the support provided by the department of computer science and engineering for facilitating the resources required to carry out this work.

The authors extend their appreciation to all faculty members and peers who contributed directly or indirectly to the successful completion of this project

REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” IEEE Transactions on Big Data, 2017.

[3] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” in Proceedings of EMNLP, 2019.

[4] J. Chen, H. Fang, and W. Xu, “Document Question Answering with Deep Learning,” in Proceedings of ACL, 2022.

[5] LangChain Documentation, “Building Applications with LLMs,” 2024.

[6] Mistral AI Documentation, “Large Language Models and APIs,” 2024.

[7] Hugging Face, “Transformers: State-of-the-art Machine Learning Models,” 2024.

[8] Streamlit Documentation, “Interactive Web App Framework for Machine Learning,” 2024.

[9] pdfplumber Documentation, “PDF Text Extraction in Python,” 2024.

[10] K. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proceedings of NAACL, 2019.