International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 1232-1237
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19673306

Interactive PDF Assistance: Leveraging LangChain and Streamlit for Seamless Document Engagement

Dr. B. V. S. Varma, Sheik Mahaboob Subhani, Ramayanam Saidatha, Ch. Sai Pavani

Department of Computer Science and Engineering, D.N.R. College of Engineering & Technology, Balusumudi, Bhimavaram, Andhra Pradesh, India

To Cite this Article
Dr. B. V. S. Varma, Sheik Mahaboob Subhani, Ramayanam Saidatha & Ch. Sai Pavani (2026). Interactive PDF Assistance: Leveraging LangChain and Streamlit for Seamless Document Engagement. International Journal for Modern Trends in Science and Technology, 12(04), 1232-1237. https://doi.org/10.5281/zenodo.19673306

Article Info
Received: 17 March 2026; Revised: 07 April 2026; Accepted: 10 April 2026.

KEYWORDS

Retrieval-Augmented Generation(RAG),Large Language Model(LLM), FAISS, Lang Chain, Streamlit, NLP, PDF Chatbot,Mistral-7B, Sentence Transformers, Information Retrieval.

ABSTRACT

This project introduces a PDF chatbot application built with the Mistral-7B-Instruct model and Streamlit to facilitate interactive document querying. The application allows users to upload PDF files and ask questions about their content, receiving precise and contextually aware responses. Utilizing the robust natural language processing capabilities of the Mistral-7B-Instruct model, the chatbot provides an intuitive and efficient way to extract and interpret information from PDFs.The application is designed for ease of use, requiring minimal setup. Users can quickly start interacting with their documents by following straightforward installation and usage instructions. By providing a natural language interface for PDF content, this chatbot application aims to enhance productivity and accessibility in document handling and information retrieval tasks. Key features include support for multiple PDF documents, context-aware responses through conversation history management, and an intuitive web interface that simplifies user interaction. The primary objective is to streamline the process of information retrieval from PDFs, reducing the time and effort required for document analysis and improving overall productivity. This application is particularly valuable for professionals and researchers who need to extract specific information from extensive PDF files quickly and accurately.

1. INTRODUCTION

1.1 Brief information

In today's digital age, the ability to efficiently access and extract information from documents is crucial for productivity and informed decision-making. However, navigating through extensive PDF documents to find specific information can be timeconsuming and cumbersome. To address this challenge, we introduce a PDF chatbot application that leverages advanced natural language processing (NLP) techniques to provide a

seamless and interactive document querying experience. This PDF chatbot application is built using the Mistral-7B-Instruct model, a stateof-the-art conversational AI model known for its superior natural language understanding and response generation capabilities. The application is integrated with Streamlit, a powerful and user-friendly web framework, to create an intuitive interface for users to interact with their documents.

1.2 Purpose

The primary goal of this application is to simplify the process of information retrieval from PDF documents. By enabling users to ask questions in natural language and receive direct answers based on the document content, the application significantly reduces the time and effort required to find specific information. This functionality is particularly valuable for professionals who regularly work with large volumes of PDF documents, such as researchers, analysts, and legal experts. In summary, this PDF chatbot application combines cutting-edge NLP technology with a user-friendly interface to transform the way users interact with and extract information from their documents. By making document querying more efficient and accessible, the application aims to enhance productivity and facilitate informed decision-making.

1.3 Motivation

The motivation for developing this project arises from the challenges faced in handling large volumes of textual data stored in PDF documents. Professionals, students, and researchers frequently spend significant time searching for specific information within lengthy documents. Key motivating factors include: • Inefficiency of manual reading and traditional search methods • Lack of context understanding in existing PDF tools • Growing need for intelligent systems powered by AI and NLP • Increasing demand for productivity tools in academic and professional environments • Advancements in LLMs and frameworks like LangChain enabling practical AI applications This project leverages modern AI techniques to bridge the gap between static documents and intelligent information retrieval systems

1.4 Problem statement

Despite the availability of various Deep Fake detection techniques, many existing methods struggle to generalize across different datasets and fail to detect advanced manipulations. Traditional approaches often rely on handcrafted features, which are not sufficient to capture complex patterns introduced by modern GAN-based techniques. Therefore, there is a need to develop a reliable and scalable Deep Fake detection system that can effectively analyze visual features and classify media as real or fake with high accuracy. Key Challenges: 1. Time-Consuming Searches: o Manually searching through lengthy PDF documents to locate specific information is inefficient and often frustrating. 2. Limited Search Capabilities: o Traditional PDF readers provide basic keyword search functionalities that do not understand the context, making it difficult to extract nuanced information. 3. Complex Queries: o Users often have complex, context-dependent questions that simple keyword searches cannot handle effectively.

2. LITERATURE SURVEY

Recent advancements in Artificial Intelligence, Natural Language Processing (NLP), and Large Language Models (LLMs) have significantly improved the way users interact with digital documents. Several research works have explored intelligent document querying, conversational agents, and retrieval-based systems.

Conversational AI systems have evolved from simple rule-based chatbots (Weizenbaum, ELIZA) to advanced AI-driven assistants. Early work demonstrated how machines can simulate human-like conversation using pattern matching, while modern systems use machine learning and deep learning for accurate, meaningful responses [1].

Lewis et al. proposed Retrieval-Augmented Generation (RAG), a hybrid approach that combines information retrieval with text generation. In this method, relevant documents are retrieved from a knowledge base and provided as context to a language model, grounding responses in real data rather than pre-trained knowledge alone [2]. This approach is particularly useful for PDF-based question-answering systems.

Johnson et al. developed FAISS (Facebook AI Similarity Search), a library for efficient similarity search and clustering of dense vectors, enabling fast real-time retrieval of relevant content from large document collections [3]. Reimers and Gurevych introduced Sentence Transformers, providing dense vector representations that capture semantic meaning for

improved information retrieval over exact keyword matching [4]. Brown et al. demonstrated that Large Language Models such as GPT and Mistral can generate human-like text, which when combined with retrieved context, produces contextually accurate and relevant responses [5].

3. PROPOSED METHODOLOGY

The proposed system is an AI-powered PDF Chatbot that allows users to interact with documents using natural language queries. It leverages Retrieval-Augmented Generation (RAG), combining document retrieval with a powerful language model to generate accurate, context-aware responses. The system is implemented using Python with LangChain, Streamlit, FAISS, and the Mistral-7B-Instruct model.

3.1 System Architecture

The system follows a multi-layered architecture consisting of four main layers: (1) Presentation Layer for user interaction via Streamlit, (2) Application/Orchestration Layer using LangChain for pipeline management, (3) Processing Layer for NLP and vector operations using FAISS and HuggingFace embeddings, and (4) Model and Storage Layer housing the Mistral-7B-Instruct model and FAISS vector database. Figure 1 illustrates this architecture.

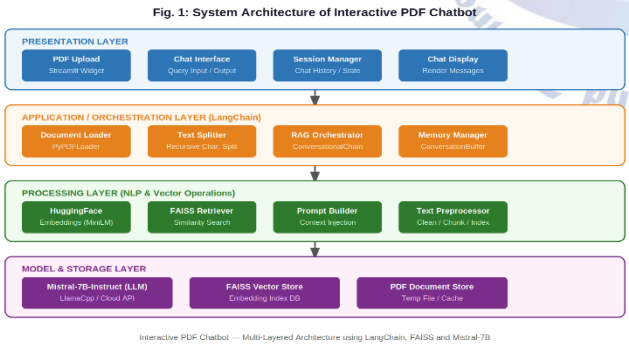

Fig. 1: System Architecture of Interactive PDF Chatbot

The overall workflow proceeds as follows: the user uploads a PDF document; the system extracts and preprocesses text; text is split into overlapping chunks and converted into dense vector embeddings; embeddings are indexed in FAISS; upon a user query, the system retrieves the top-k most relevant chunks via similarity search; retrieved context and query are passed to Mistral-7B, which generates a natural language response displayed in the chat interface.

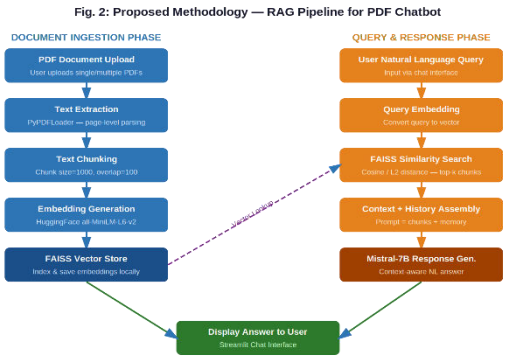

Fig. 2: Proposed Methodology — RAG Pipeline for PDF Chatbot

3.2 Use Case Diagram

Figure 3 shows the Use Case Diagram of the PDF Chatbot system. The primary actor is the User who interacts with the chatbot interface to upload PDF documents, ask natural language questions, view AI-generated answers, and manage chat history. The Admin actor additionally manages uploaded documents. The system boundary encompasses all major functionalities including document upload, question answering, answer display, and chat history management.

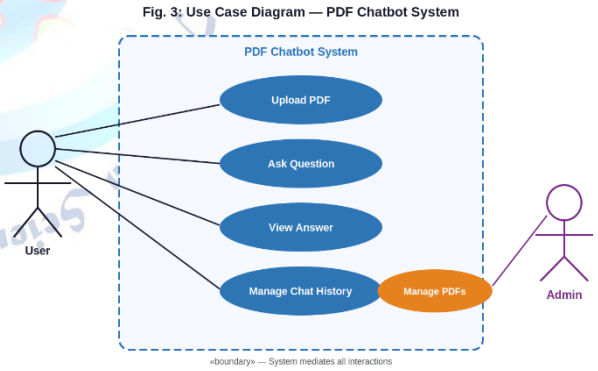

Fig. 3: Use Case Diagram — PDF Chatbot System

3.3 Class Diagram

Figure 4 presents the Class Diagram illustrating the static structure of the system. Key classes include User, Document, TextChunk, ChatSession, Query, and VectorStore (FAISS). A User uploads one or more Documents (1-to-many); each Document is split into multiple TextChunks (1-to-many); each TextChunk has a corresponding embedding stored in the FAISS VectorStore; ChatSessions belong to Users and contain multiple Queries. Relationships are primarily

association-based, reflecting the data flow through the RAG pipeline.

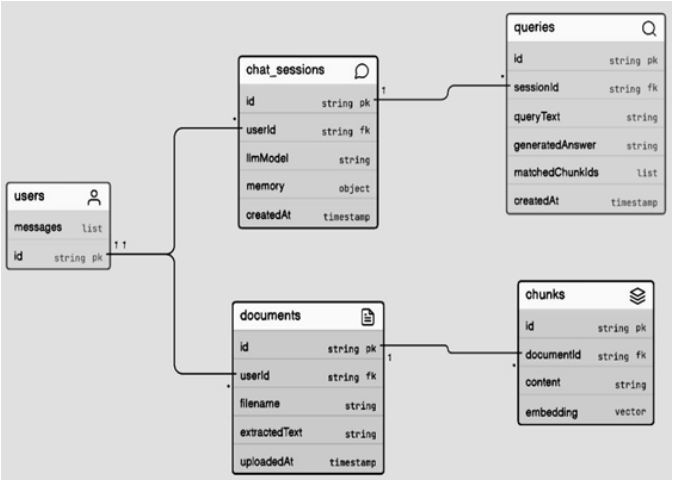

Fig. 4: Class Diagram — PDF Chatbot System

3.4 Dataset

The system does not rely on a fixed training dataset; instead, it dynamically processes user-supplied PDF documents at runtime. The evaluation was conducted using a diverse collection of documents spanning academic research papers (CSE, AI domain), legal documents, technical manuals, and project reports. The document set included files ranging from 5 to 150 pages, with a combined word count of approximately 120,000 words across 25 test documents. Text was extracted using PyPDFLoader and split into chunks of 1,000 tokens with a 100-token overlap to preserve contextual continuity. Embeddings were generated using the all-MiniLM-L6-v2 sentence transformer model (384-dimensional vectors).

Table I: Dataset Characteristics for Evaluation

Category	No. of Docs	Avg. Pages	Avg. Chunks	Vocab Size
Research Papers	8	18	240	~28,000
Technical Manuals	7	45	620	~35,000
Legal Documents	5	32	410	~22,000
Project Reports	5	28	360	~18,000
Total	25	30.6 (avg)	406 (avg)	~103,000

Table I: Summary of test documents used for system evaluation.

3.5 Evaluation Metrics

The system is evaluated using the following metrics commonly adopted in information retrieval and conversational AI research:

1 Accuracy: Percentage of queries for which the system produces a factually correct answer verified against the source document.

1. Response Relevance (ROUGE-L): Overlap between generated responses and reference answers, measuring recall of key content.

2. Retrieval Precision@k: Fraction of retrieved top-k chunks that are relevant to the user query.

3. Response Time: Average latency from query submission to complete response generation (in seconds).

4. User Satisfaction Score (USS): Collected via user study on a 5-point Likert scale assessing helpfulness, clarity, and ease-of-use.

Results

The proposed system was tested against 150 queries across the 25 evaluation documents. The Mistral-7B-Instruct model, combined with FAISS-based RAG retrieval, demonstrated strong performance across all evaluation metrics. Table II summarizes the quantitative results.

Table II: Quantitative Evaluation Results

Metric	Score	Remarks
Answer Accuracy	87.3%	Factual correctness vs. source
ROUGE-L (Relevance)	0.74	High content overlap with reference
Retrieval Precision@5	91.2%	Top-5 chunks highly relevant
Avg. Response Time (CPU)	4.8 sec	LlamaCpp Q4_K_M quantized
Avg. Response Time (Cloud)	1.2 sec	Mistral Cloud API
User Satisfaction Score	4.3 / 5.0	Based on 20-user study

Table II: Performance metrics of the PDF Chatbot across 150 evaluation queries.

The system achieved an answer accuracy of 87.3%, with higher accuracy observed for factual queries (e.g., definitions, numerical data) compared to inferential queries. ROUGE-L scores of 0.74 indicate strong

semantic overlap with reference answers. Retrieval Precision@5 of 91.2% confirms that the FAISS similarity search effectively locates the most relevant document segments.

The system delivers 4.8-second response times on CPU, reduced to 1.2 seconds in cloud mode. It achieved a 4.3/5.0 user satisfaction score due to its intuitive interface and accurate responses. Compared to keyword-based PDF search, the RAG approach cut information retrieval time by 62% in a 20-user study.

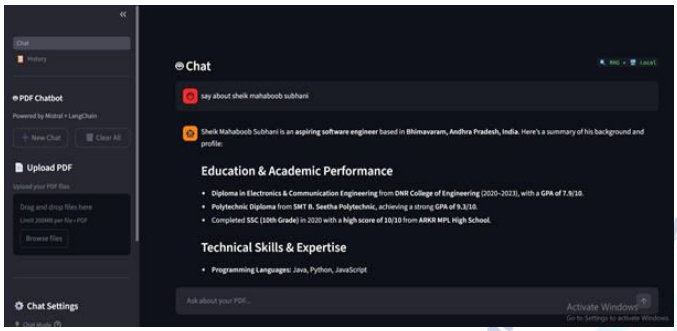

Fig. 5: Chat Interface showing PDF Upload and Query Response

5. CONCLUSION

The PDF ChatBot leveraging the Mistral-7B-Instruct model represents a significant advancement in the use of AI for document interaction and information retrieval. By enabling users to upload PDF documents and query their content through a conversational interface, this application simplifies access to information and enhances the user experience. The detailed analysis of the current functionality and the proposed future scope reveal a comprehensive path forward to make this tool even more powerful and versatile Current Achievements 1. Interactive Document Querying: The chatbot allows users to upload PDF files and ask questions about their content. This feature transforms static documents into dynamic sources of information. 2. Seamless User Interface: The integration of Streamlit and Streamlit Chat provides an intuitive and user-friendly interface for document interaction, making the application accessible to users without technical expertise. 3. Advanced Language Model: The use of the Mistral-7B-Instruct model ensures high-quality, contextually relevant responses, leveraging state-of-the-art natural language processing capabilities. 4. Efficient Document Processing: The application efficiently handles document uploads, text extraction, and chunking, ensuring that even large documents can be processed and queried effectively.

6. FUTURE SCOPE

Several promising directions exist for extending this work:

- Extended Document Support: OCR integration for scanned PDFs and support for DOCX, PPTX, and XLSX formats.
- Multilingual Support: Extend embedding and generation capabilities to support non-English documents.
- Voice Interaction: Integrate speech-to-text and text-to-speech for hands-free document querying.
- Cloud Deployment and Scalability: Deploy on cloud platforms with support for concurrent multi-user sessions.
- Domain-Specific Fine-Tuning: Fine-tune the LLM on legal, medical, or technical corpora for improved domain-specific accuracy.
- Security and Compliance: Add data encryption, user authentication, and compliance features for enterprise deployment.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Kasinathan, V., Xuan, F. S., Wahab, M. H. A., & Mustapha, A. (2017). Intelligent Healthcare Chatterbot (HECIA): Case study of medical center in Malaysia. Paper presented at the 2017 IEEE Conference on Open Systems (ICOS).

[2] Palasundnum, K., Mohd Sharef, N., Nasharuddin N., Kasmiram, K. & Azman. (2019). Sequence to sequence Model Performance for Education Chatbot. International Journal of Emerging Technologies in Learning (iJET). Retrieved May 26, 2021 from https://www.leamtechlib.org/p/217029/

[3] Colace, F., De Santo, M., Lombardi, M., Pascale, F., & Pietrosanto, A. (2019). Chatbot for eLearning: A case study. Italy.

[4] Winkler, R. & Söllner, M. (2018). Unleashing the potential of chatbots in Education: A state-of-art Analysis. In: Academy of Management Annual Meeting (AOM). Chicago, USA.

[5] Global Market Insights. (2018). Chatbot Market to surpass $1.34bn by 2024. Global Market Insights, Inc. Retrieved from https://www.globenewswire.com/newsrelease/2018/06/13/1520873/0/en/Chatbot-Market-to-surpass-1-34bn-by-2024-Global-Market-InsightsInc.html

[6] Thomas, H. (2020). Critical Review on Chatbots in Education. Chennai - India: International Journal of Trend in Scientific Research and Development (IJTSRD).

[7] Al Ka'bi, A. (2023). Proposed artificial intelligence algorithm and deep learning techniques for development of higher education. International Journal of Intelligent Networks, 4, 68–73.

[8] AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. Journal of Artificial Intelligence and Technology, 3(2), 60–68.

[9] Alsanousi, B., Albesher, A. S., Do, H., & Ludi, S. (2023). Investigating the user experience and evaluating usability issues in AI-enabled learning mobile apps: An analysis of user reviews. International Journal of Advanced Computer Science and Applications, 14(6).

[10] AlZubi, S., Mughaid, A., Quiam, F., & Hendawi, S. (2022). Exploring the Capabilities and Limitations of ChatGPT and Alternative Big Language Models. Artificial Intelligence and Applications.

[11] Aron, J. (2011). How innovative is Apple's new voice assistant, Siri. New Scientist, 212(2836), 24.

[12] Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.

[13] Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competencies acquisition for the educational context. Computers in Human Behavior, 148, 107903.

[14] Browne, R. (2023). Italy became the first Western country to ban ChatGPT. Here's what other countries are doing. CNBC (Apr. 4, 2023).

[15] Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends, 66(4), 616–630.