

Department of Computer Science and Engineering (2025-26)

Synopsis for B.E Final Year (Project Phase -1)

Batch No: 24	Guide Name: S Brunda
Project Title: LlamaChain: An Interactive AI Knowledge Assistant for Multi-Document Analysis and Comprehension	

USN	Student Name
02JST22UCS088	Rhiya Giridhara Bhat
01JST21CS043	Ghanashyam D
01JST22UCS032	Chirag N
01JST22UCS159	Tarun GP

Signature of Guide

(Prof. S Brunda)

Signature of HOD

(Dr. Anil Kumar K.M)

Department of Computer Science and Engineering (2025-26)

Synopsis Evaluation Certificate

Batch No: 24	Guide Name: S Brunda
Project Title: LlamaChain: An Interactive AI Knowledge Assistant for Multi-Document Analysis and Comprehension	

USN	Student Name
02JST22UCS088	Rhiya Giridhara Bhat
01JST21CS043	Ghanashyam D
01JST22UCS032	Chirag N
01JST22UCS159	Tarun GP

Can the project be considered as final year project (yes/no)	
Is the problem statement and objectives of the project clear? (yes/no)	
Can the project be completed in time? (yes/no)	
Does this project have the potential to be converted into a product/publication (yes/no)	
Does this project have social relevance? (yes/no)	

Note: if 'No' please write your Remarks/ Suggestions on the backside of this page

Evaluation Committee

Name

Signature

Table of Contents

Sl. No	Section	Page No.
1	Title of the project	1
2	Introduction	1
3	Literature review	1
4	Limitations of the existing work	2
5	Objectives	3
6	Methodology	3
7	Work plan	5
8	Conclusion	5
9	References	6

1. Title of the Project

LlamaChain: An Interactive AI Knowledge Assistant for Multi-Document Analysis and Comprehension

2. Introduction

In today's information-rich world, vast amounts of critical knowledge are stored across multiple unstructured documents such as PDFs and presentations. Extracting relevant insights from these sources manually is both time-consuming and cognitively demanding. Traditional search tools are limited to keyword-based retrieval, lacking the ability to comprehend semantics, context, or relationships between information spread across different files. Moreover, most existing systems focus on processing a single document and fail to interpret non-textual elements like tables and images. As a result, users face significant challenges when trying to compare documents, obtain summarized content, or interact dynamically with multiple files in a meaningful, conversational way.

The proposed system will be an Interactive AI-Powered Multi-Document Knowledge Assistant that will allow users to chat with, summarize, and compare multiple PDFs and PPTs. It will utilize Llama and LangChain within a Retrieval-Augmented Generation (RAG) framework to generate accurate, context-aware, and explainable responses across text, tables, and images.

3. Literature Review

1. MuDoC: An Interactive Multimodal Document-grounded Conversational AI System

Authors: Xiaohan Zhang, Zihan Liu, Yiyang Zhou, Ashish Vaswani, and Diyi Yang
Publication Date: May 2024

This research paper introduces MuDoC, a conversational AI system designed to interact with multimodal documents that include text, tables, and images. The system employs a Retrieval-Augmented Generation (RAG) framework that retrieves document-grounded content before generating responses, ensuring accuracy and factual consistency. MuDoC integrates Dense Passage Retrieval (DPR) for textual information and CLIP embeddings for visual content, enabling the model to reason across modalities effectively. The paper highlights features such as document-grounded response generation, context traceability, and an interactive interface that allows users to view the specific document regions related to each answer. By combining multimodal retrieval and large language models, MuDoC demonstrates a scalable and explainable approach to document-based conversational AI, significantly improving user interactivity and the interpretability of responses.

2. Multi-Purpose NLP Chatbot: Design, Methodology & Conclusion

Authors: Shivom Agarwal, Shourya Mehra, and Pritha Mitra

Publication Date: October 2023

This paper presents a comprehensive study on the development of a multi-purpose chatbot leveraging Natural Language Processing (NLP) and reinforcement learning to enhance conversational intelligence and user experience. The system integrates emotion recognition, multilingual capabilities, voice-to-voice interaction, and offline functionality, making it adaptable across industries such as healthcare, finance, and customer service. The authors employ models like DistilBERT for emotion detection, Tacotron2 and Transformer-based architectures for speech processing, and MarianMT for multilingual translation. Additionally, the chatbot includes a financial advisory module using ChatGPT and BERT to provide personalized guidance. The study emphasizes the importance of ethical design, addressing privacy, bias, and transparency in AI-driven conversations. Through this hybrid framework, the chatbot demonstrates a high level of emotional intelligence, contextual understanding, and accessibility — advancing the state of conversational AI beyond traditional rule-based systems.

3. TOMDS (Topic-Oriented Multi-Document Summarization): Enabling Personalized Customization of Multi-Document Summaries

Authors: Xin Zhang, Qiyi Wei, Qing Song, Pengzhou Zhang

Publication Date: February 2024

This paper presents a novel approach to multi-document summarisation where the user can decide the target topic for the summary, rather than receiving a generic overview. The process is divided into two stages: first, it extracts paragraphs relevant to the chosen topic (extractive phase), then it applies an abstractive summarisation model enhanced with a discourse-aware attention mechanism (abstractive phase). The model uses topic-aware attention and a discourse parsing module to capture intra- and inter-paragraph relationships, resulting in summaries that are more aligned with user goals, semantically richer, and more focused on the selected topic.

4. Limitations of the Existing Work

- Most existing systems handle only text-based PDFs, ignoring images, graphs, and tables.
- They cannot process multiple files simultaneously or generate combined summaries.
- No context tracing, that is, users don't see where answers originate from.
- Depend heavily on cloud APIs, leading to data privacy and cost issues.
- Limited or no multilingual and speech input support.
- Lack features like document comparison, topic clustering, or confidence scoring

5. Objectives

- To design and develop a multimodal conversational system capable of interacting with multiple PDFs and PPTs, extracting and processing text, tables, and images for comprehensive information retrieval.
- To implement Llama-based document understanding for combined, document-wise, and topic-wise summarization with context highlighting of relevant paragraphs, pages, or slides
- To ensure secure and efficient local execution using Llama with persistent vector storage (ChromaDB/FAISS).
- To enhance user interaction through document comparison, multilingual and speech-based inputs, chat export, and confidence scoring features.

6. Methodology

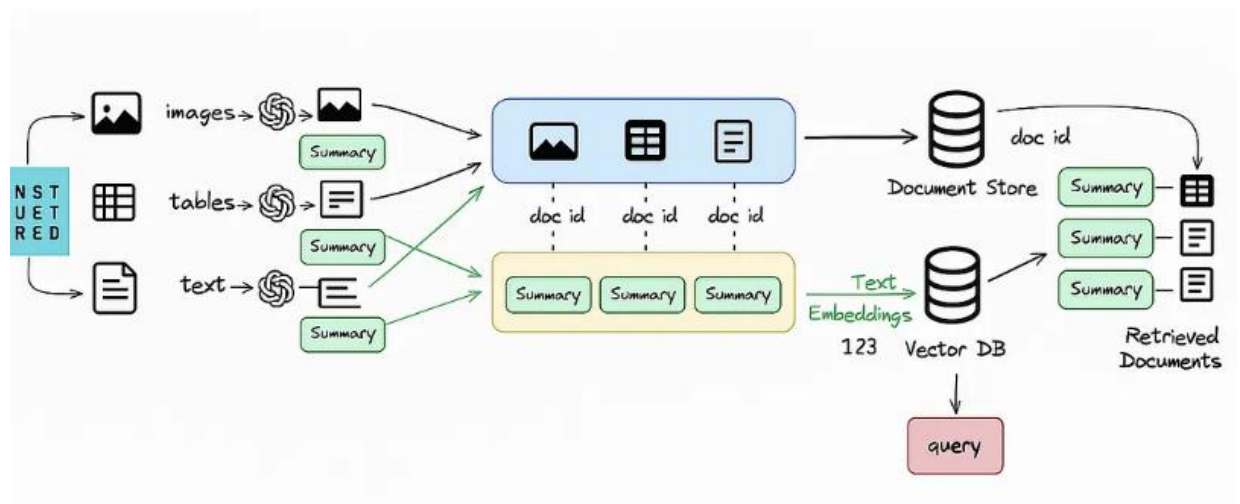


Figure 1. Multimodal retrieval using unstructured for extraction

1. Data Acquisition

- Users upload one or more PDF or PPT files through a Streamlit-based web interface.
- Text, tables, and images are extracted using libraries like PyMuPDF, python-pptx, and Unstructured.

2. Pre-processing and Chunking

- The extracted data is segmented into structured chunks — textual content, table content, and image captions.
- Each chunk is stored with metadata: file name, type (text/table/image), and page or slide number.

3. Embedding Generation

- Each chunk is transformed into high-dimensional embeddings using Hugging Face Instructor or Sentence-Transformer models (e.g., all-MiniLM-L6-v2).

4. Vector Storage

- Embeddings are stored in ChromaDB (persistent vector database) along with metadata.
- Enables quick retrieval across multiple files and sessions.

5. Retrieval-Augmented Generation (RAG)

- When a user enters or speaks a question:
 - The system embeds the query and retrieves top relevant chunks.
 - These chunks are provided as context to Llama via LangChain's ConversationalRetrievalChain.
 - Llama generates a precise answer with source citations and confidence scores.

6. Summarization

- Combined Summary: Generates a unified overview of all uploaded documents.
- Document Summary: Individual summaries for selected files.
- Topic-wise Summary: Clusters content by topic and summarizes each cluster.
- Summaries are generated using map-reduce summarization chains.

7. Document Comparison

- Allows selection of two documents for difference analysis.
- Llama compares semantic embeddings and generates a contrastive summary of changes or similarities.

8. Speech & Multilingual Input

- Speech-to-text conversion via Whisper API.
- Language detection and translation using langdetect and MarianMT before querying the LLM.

9. Context Highlighting & Export

- Displays the exact page/slide and paragraph from which the answer was derived.
- Chat can be exported in PDF format, including time stamps and confidence levels.

10. Evaluation

- Retrieval accuracy (Precision and Recall).
- Summarization quality (ROUGE/BLEU metrics).
- Response latency and usability surveys.

7. Work Plan

Table 1. Project Timeline – Gantt Chart for Multiple pdf chatbot

	Week 0-2	Week 2-4	Week 4-6	Week 6-8	Week 8-10	Week 10-12	Week 12-14	Week 14-16	Week 16-18
Literature Review and Architechtural Design									
Data extraction (PDF, PPT) and preprocessing									
Embedding and vector storage setup									
Integration of Llama and RAG pipeline									
Summarization, comparison, multilingual & speech modules									
UI development, context highlighting, export options									
Testing, evaluation, and optimization									
Documentation and final presentation									

8. Conclusion

This project presents a comprehensive multimodal AI assistant that revolutionizes document understanding by enabling users to chat, compare, and summarize multiple PDFs or PPTs seamlessly. It integrates Llama, LangChain, and vector databases to deliver contextually aware, explainable, and multilingual interactions entirely offline. Equipped with advanced capabilities such as table and image interpretation, speech input, source highlighting, and confidence scoring, the system combines technical sophistication with real-world practicality. Overall, it represents a significant advancement toward intelligent, privacy-preserving document comprehension systems with broad applications in education, research, law, and corporate knowledge management.

9. References

- [1] X. Zhang, Z. Liu, Y. Zhou, A. Vaswani, and D. Yang, “MuDoC: An Interactive Multimodal Document-grounded Conversational AI System,” *arXiv preprint* arXiv:2502.09843, May 2024. [Online]. Available: <https://arxiv.org/abs/2502.09843>
- [2] S. Agarwal, S. Mehra, and P. Mitra, “Multi-Purpose NLP Chatbot: Design, Methodology & Conclusion,” *arXiv preprint* arXiv:2310.08977, Oct. 2023. [Online]. Available: <https://arxiv.org/abs/2310.08977>
- [3] X. Zhang, Q. Wei, Q. Song, and P. Zhang, “TOMDS (Topic-Oriented Multi-Document Summarization): Enabling Personalized Customization of Multi-Document Summaries,” *Applied Sciences*, vol. 14, no. 5, p. 1880, Feb. 2024. [Online]. Available: <https://doi.org/10.3390/app14051880>
- [4] LangChain Documentation, *LangChain: Building applications with large language models*, <https://python.langchain.com/docs/>
- [5] Unstructured Documentation, *Unstructured: A library for preprocessing and parsing unstructured documents*, <https://unstructured.io/>