Journal of Science Engineering Technology and Management Science
Volume 02, Issue 08, August 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i08.388-394*

CHAT WITH PDFs: INTERACT WITH PDFs USING GOOGLE GEMINI AI

[1]R. VAISHNAVI,[2]Y. PRANITHA,[3]G. CHANDRASHEKAR AZAD,[4]Mrs. N. PRIYANKA
[123] Students, [4] Assistant Professor
Department Of Information Technology
Teegala Krishna Reddy Engineering College, Meerpet, Balapur, Hyderabad-500097

To Cite this Article
R. Vaishnavi, Y. Pranitha, G. Chandrashekar Azad, Mrs. N. Priyanka, "Chat With Pdfs: Interact With Pdfs Using Google Gemini AI", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 08, August 2025,pp: 388-394, DOI: http://doi.org/10.63590/jsetms.2025.v02.i08.pp388-394
Submitted: 12-07-2025 *Accepted: 18-08-2025* *Published: 25-08-2025*

ABSTRACT
Chat-With-PDF is an innovative AI-driven tool that enables users to interact effortlessly with PDF documents. Powered by Google Gemini AI, this solution allows users to extract insights, summarize content, and ask context-aware questions while receiving instant responses. Unlike traditional search methods, this AI-enhanced approach enhances productivity by providing an intuitive and fluid document interaction experience. With advanced natural language processing (NLP), semantic search, and intelligent document understanding, Chat-With-PDF ensures information retrieval is handled coherently and smoothly. It is particularly beneficial for students, professionals, and researchers, offering an efficient way to analyze academic papers, legal documents, and business reports. By leveraging Google Gemini AI's powerful capabilities, Chat-With-PDF transforms document handling, making it more harmonious and flawless. This smart innovation redefines the way users access information, allowing them to work perfectly and naturally within their digital documents

I. INTRODUCTION:

With the rapid growth of digital information, PDF (Portable Document Format) files have become a standard format for sharing academic papers, business reports, legal documents, and technical manuals. While PDFs are convenient for preserving document layout and integrity across devices, they often pose challenges when users need to search for specific information, extract insights, or analyze content. Traditional PDF readers typically support basic functionalities like viewing, scrolling, and simple keyword searches. However, they fall short when it comes to deeper interactions such as summarizing content, retrieving context-aware information, or handling scanned documents that require Optical Character Recognition (OCR).

To address these limitations, **Chat-With-PDF** introduces a novel and AI-enhanced solution that leverages the capabilities of **Google Gemini AI** to offer a conversational interface for interacting with PDFs. Instead of manually navigating through pages or relying on limited search functionalities, users can now ask natural language questions, receive instant responses, and access summarized content in real time. This makes the process of working with large documents more fluid, efficient, and intuitive.

The core innovation lies in the integration of **Natural Language Processing (NLP)**, **semantic search**, and **machine learning algorithms**, which together enable the system to understand user intent and retrieve relevant information with high accuracy. Moreover, the use of **OCR technology** allows the system to process and analyze scanned PDFs by converting image-based text into machine-readable format, greatly enhancing accessibility.

Chat-With-PDF is particularlydesigned to benefit a wide range of users:

- **Students** can use it to quickly understand academic material, summarizeresearch papers, and find answers without reading through entire documents.
- **Researchers** can analyze technical content, extract relevant findings, and gather supporting information efficiently.
- **Legalprofessionals** can searchfor clauses, summarize contracts, and retrieve case-specific data from lengthy legal texts.
- **Business users** can explore reports, generate quick overviews, and access key performance indicators without manual effort.

By providing a user-friendly chat interface, **Chat-With-PDF** not only improves productivity but also transforms the document analysis experience. The system's support for both structured and unstructured data types ensures versatility, while the real-time response capability makes it an ideal tool for anyone who works with large volumes of text-based content.

In conclusion, *Chat-With-PDF* represents a significant advancement in document interaction technology. Through the intelligent use of AI and natural language interfaces, it empowers users to interact with their documents in a smarter, faster, and more meaningful way—setting a new standard for how we access and use information in the digital age.

EXISTING SYSTEM

Existing PDF readers allow users to view and search text but lack interactive features for extracting insights, summarizing content, or answering queries. Manual navigation and keyword searching in large PDFs can be time-consuming and inefficient, especially for academic and professional users. Some AI-powered tools offer basic text extraction and summarization, but they often lack real- time, context-aware conversational. Most existing systems do not support scanned PDSs effectively, as they lack OCR technology to convert images into searchable and editable text.

A "Chat with PDF" system is a powerful application of natural language processing that enables users to interact with the content of PDF documents in a conversational manner, similar to chatting with a knowledgeable assistant. This technology typically works by first ingesting the PDF file and extracting its textual content using libraries such as PyMuPDF, PDF Miner, or pdf plumber. Once the raw text is extracted, it is then segmented into smaller, manageable chunks— either by logical sections, paragraphs, or fixed-size text blocks—to make it easier for the system to process and retrieve relevant information later. These embeddings capture the semantic meaning of the text and are stored in a specialized vector database such as FAISS, Pinecone, or Chroma, which allows for fast similarity searches based on user queries. When a user asks a question, the system converts the query into an embedding and retrieves the most relevant chunks from the vector database. These selected chunks are then passed to a large language model, such as GPT-4, in a process known as retrieval-augmented generation (RAG). The model uses the context from the PDF to formulate a coherent and accurate response to the user's question. This setup allows users to gain insights from dense or technical documents quickly, without having to manually search through pages of content. Existing implementations of this concept include commercial platforms like Chat PDF, Humate, and Ask Your PDF, as well as open-source projects built using frameworks like

Lang Chain, Haystack, and Private.

PROPOSED SYSTEM

Chat-With-PDF, powered by Google Gemini AI, allowsusers to interact with PDFs using simple chat commands. It enables users to extract key information, generate summaries, searchfor specific content, and get precise answers in real time. The system supports both text-based and scanned PDFs using OCR technology, making documents more accessible. With a user-friendly chat interface, it helps students, researchers, legal professionals, and business users quickly analyze large documents, saving time and effort.

The proposed system for a "Chat with PDF" project using Google Gemini AI is designed to allow users to interact with PDF documents conversationally through a web-based interface. When a user uploads a PDF, the system first extracts the textual content using libraries such as PyMuPDF or pdf plumber, ensuring that even complex layouts with tables and images are handled effectively. The extracted text is then segmented into smaller, semantically meaningful chunks to enable more accurate information retrieval. Each of these chunks is transformed into high- dimensional embeddings using Google's Universal Sentence Encoder or Gemini embedding models via Vertex AI. These embeddings are stored in a vector database, such as FAISS, Pinecone, or Google's Vertex AI Matching Engine, to support fast semantic search. When the user submits a query, it is also converted into an embedding and compared against the stored vectors to retrieve the most relevant content from the document. This context, along with the user's question, is then sent to the Gemini large language model using the Vertex AI API, which generates a natural-language answer grounded in the document content. The response is displayed in a chat-like interface, creating a seamless and intuitive experience for the user. Additional features like multilingual support, voice input, citation tracking, and document summarization can be integrated to enhance usability and functionality. By leveraging Google Gemini's advanced language understanding capabilities and scalable cloud infrastructure, this system provides a robust and intelligent solution for exploring and extracting knowledge from complex PDF documents.

PROBLEM STATEMENT

Many users struggle with extracting, summarizing, and searching for information in PDF documents, making manual navigation time-consuming and inefficient. Chat-With-PDF, powered by Google Gemini AI, provides an interactive way to engage with PDFs using natural language, enabling users to extract key details, generate summaries, and retrieve specific information. The s o l u t i o n supports various PDF formats, including scanned d o c u m e n t s with OCR technology, ensuring accessibility and ease of use. By offering a user-friendly chat-based interface, it helps students, researchers, legal professionals, and business users efficiently analyze large documents, saving time and effort.

Many users across academic, corporate, and legal sectors often need to extract specific information from large and complex PDF documents. This task can be tedious and time- consuming, as traditional PDF readers do not offer intelligent tools for quickly finding relevant content. Users must manually search through pages, which is inefficient and prone to errors. The lack of an intuitive and interactive way to engage with the content of PDFs limits productivity and accessibility. To address this challenge, the "Chat with PDF" project proposes the development of a conversational interface that allows users to upload PDF documents and interact with them using natural language queries. By leveraging natural language processing (NLP) and machine learning techniques, the system will understand the context of the user's questions and retrieve accurate, relevant information from the documents. This approach aims to make document exploration more efficient, user-friendly, and intelligent.

II. LITERATURE SURVEY

The need for intelligent document interaction tools has surged with the exponential growth of digital documents. Traditional PDF readers provide basic functionalities like viewing and keyword searching but fail to offer intuitive, context-aware interaction, especially with complex and lengthy documents. Here are key developments from the literature:

1. Traditional PDF Readers:

- Provide only static viewing and keyword search features.
- Lack intelligence in summarization, querying, and context understanding.

2. AI-Powered Tools:

- Some tools such as Adobe's Sensei and other commercial tools allow text extraction and summarization.
- However, most lack real-time interactivity and support for scanned PDFs without OCR (Optical Character Recognition) integration.

3. OCR-BasedDocument Handling:

- Researchshows that OCR tools like Tesseract can convert scanned documents into searchable text, improving accessibility.
- Limitations lie in context understanding and intelligent querying.

4. Chat-BasedAI Systems:

- Natural Language Processing(NLP) modelssuch as GPT, BERT, and now Google Gemini are capable ofunderstanding human queries contextually.
- Integration of such models with document handling allows deeper interaction, transforming static documents into dynamic, queryable data sources.

5. Multimodal AI and Transformer Models:

- Transformer-based architectures support diverse input types—text, images, audio.
- Used in advanced tools to unify multiple data representations for contextual understanding and output generation.

METHODOLOGY

The proposed methodology for Chat-With-PDF is structured around AI-enhanced interaction, specificallyleveraging Google Gemini AI capabilities. Here's how it works step-by-step:

1. User Input Handling
   - User uploads a PDF file or provides a URL.
   - The system detects whether it's text-based or scanned and applies OCR if needed.
2. Text Extraction & Preprocessing
   - For text PDFs: Content is extracted and segmented.
   - For scanned PDFs: OCR converts image-based text to machine-readable format.
   - Tokenization and vectorization are performed to structure the content.
3. Contextual Query Processing
   - Users submit natural language queries through a chat interface.
   - The query is parsed and understood by the NLP model (Gemini AI).
4. Vector-Based Retrieval
   - The system uses vector databases (e.g., FAISS) to retrieve relevant document sections.
   - Context from previous chat history is also considered for coherence.
5. AI Response Generation
   - Retrieved information is passed to the AI model to generate a human-like response.

- For database queries, it can convert text into SQL and retrieve data accordingly.

6. Display and Feedback

- Final response is shown via a user-friendly interface built using Streamlit.

7. The chat allows continuous interaction, enabling follow-up questions and deeper exploration

III. MODULES

1. User Interface Module

- Developed using **Streamlit** or any frontend framework.
- Allows users to:
  - Upload PDF files or enter a URL.
  - Enter natural language queries.
  - View AI-generated responses.

2. PDF Processing Module

- Handles both **text-based** and **scanned PDFs**.
- Responsibilities:
  - Extract text from PDFs.
  - Use **OCR** (like Tesseract) for scanned PDFs.
  - Clean and preprocess extracted content.

3. Text Chunking and Vectorization Module

- Splits large documents into manageable chunks.
- Converts chunks into **vectorrepresentations** using embedding models.
- Stores vectors in a **vectordatabase** like **FAISS**.

4. Natural Language Query Module

- Processes the user's query using NLP techniques.
- Handles:
  - Query understanding and tokenization.
  - Semantic matching withthe PDF content.
  - Context-aware conversation tracking (like maintaining chat history).

5. Context Retrieval and Ranking Module

- Fetches relevant document chunks using similarity search.
- Ranks them based on relevance to the query.
- Integrates previous user interactions for better context.

6. AI Response Generation Module

- Uses **Google Gemini AI (or any LLM)** to:
  - Generate coherent and relevant responses.
  - Summarize, explain, or extract insights based on user queries.
- Optional: Converts text-to-SQL for database queries.

7. Multimodal Integration Module (Optional but Powerful)

- Supports different input types: PDFs, images, audio, etc.
- Converts all input to a unified format for the transformer model.
- Allows visual, tabular, or audio content analysis along with text.

8. Authentication & Access Control Module (If multi-user setup)

- Handles login, registration, and user access to their uploaded PDFs.
- Ensures session-based document interaction.

9. Logging and Analytics Module

- Tracks:
  - User queries.
  - Response quality.
  - Performance metrics (like latency, relevance score).
- Useful for improving the system using feedback loops.

10. Database and File Storage Module

- Stores:
  - User-uploaded PDFs.
  - Extracted data, metadata, and query history.
  - Vector indexes for fast search.

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

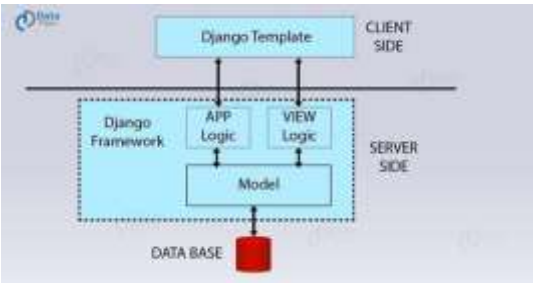
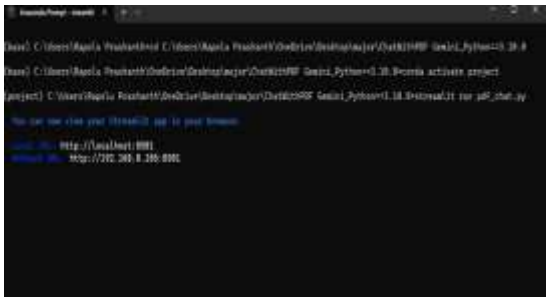
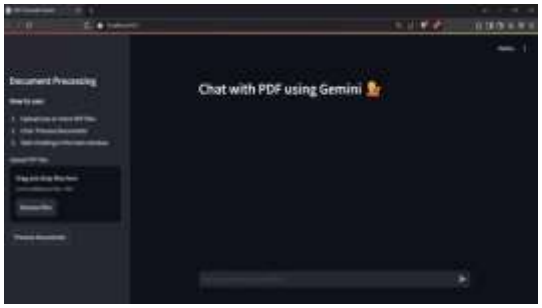


FIG: SYSTE ARCHITECTRE

V. OUTPUT SCREENSHOTS :





VI. CONCLUSION

Chat-With-PDF, powered by Google Gemini AI, provides an interactive way to extract, summarize, and search for information in PDFs using simple chat commands. •It enhances efficiency by reducing the time and effort needed to navigate large documents, making it ideal for students, researchers, legal professionals, and business users. •With support for both text - based and scanned PDFs through OCR technology, it ensures accessibility and usability across different document types. •By offering a user-friendly chat interface and real-time responses, this AI-powered tool improves document analysis, making information retrieval faster and more convenient.

REFERENCES:

1. Google AI Blog and Research Publications: Keep a close eye on the official blog (ai.googleblog.com) and Google's researchpublication Google AI IEEE Xplore Digita lLibrary. Availableonline: ‘https https://dl.acm.org/ ://ieeexplore.ieee.org/Xplore/home.jsp arXiv.org (Cornell University e-Print Archive)
2. Preprintsof researchpapers in AI, computer science, physics, and more: https://arxiv.org/
3. MIT Technology Review – AI Section Latest news and analysis on AI: https://www.technologyreview.com/ai/
4. Nature Machine Intelligence High-quality research articles in AI and machine learning: https://www.nature.com/natmachintell/
5. OpenAI Research Blog Researchinsights and updates from OpenAI: https://openai.com/research
6. Stanford AI Lab (SAIL) Publications Researchpapers from Stanford’s AILab: https://ai.stanford.edu/research/
7. DeepMind Research Blog Researchpapers and blog posts from DeepMind: https://www.deepmind.com/research
8. ResearchGate Networkwhere researchers publish and share their work: https://www.researchgate.net