PDF GENIE: INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION

A PROJECT REPORT

Submitted by

RESHMA R   **2116210701212**

SANGEETHA G   **2116210701229**

in partial fulfilment of the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

AUTONOMOUS, CHENNAI 602 105

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **"PDF GENIE: INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION"** is the bona-fide work of **"RESHMA R (210701212), SANGEETHA G (210701229)"** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. Rakesh Kumar M, M.E., Ph.D.,

SUPERVISOR

Assistant Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602105

Submitted to Project Viva-Voce Examination held on _____.

Internal Examiner   External Examiner

TABLE OF CONTENTS

ABSTRACT

The "PDF GENIE: ADVANCED AI INTEGRATION" project is a unique ecosystem for the organization that aims to boost the interaction of users with PDF documents in real time. This project relies on cutting edge technologies such as Large Language Models (LLMs), LangChain, Pinecone DB and Vercel AI SDK to help a user seeking information on the content of a pdf and search for answers in an efficient manner. Here, customers may upload PDF files and ask questions through the chatbot. The chatbot analyses these questions through LLMs that are specifically trained to be able to recognize and identify the information that is pertinent when analysing certain information from the PDFs. LangChain is used to orchestrate the language model and makes sure that the model responses the language models in an understandable and meaningful way. Pinecone DB provides vector storage and query services for the system to retrieve exact terms from massive documents associated with documents generated in real-time. The Vercel AI SDK ensures the efficient streamlining and scaling of the platforms, making the usage easy and convenient for the users. The goal of this project is to provide users with an easier way to work with documents by arranging them in conversational discussions and eliminating PDF as a stagnant document, which forces the user to read it to understand it fast. This system is user friendly and has broad application area from education and learning to generation of business intelligence. This awesome application is powered by the use of advanced AI technologies to enable understanding and extraction of data contained in complex documents and thus realized the dynamic PDF format from static one.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.,** our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr. Rakesh Kumar M, M.E., Ph.D.,** Assistant Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

RESHMA R: 210701212
SANGEETHA G: 210701229

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

LLM	Large Language Model
DB	Database
SDK	Software Development Kit
NLU	Natural Language Understanding
UI	User Interface
GPT	Generative Pre-trained Transformer
AI	Artificial Intelligence
PDF	Portable Document Format
OCR	Optical Character Recognition
NLP	Natural Language Processing

CHAPTER - 1

INTRODUCTION

1.1 GENERAL

The "PDF GENIE: An Interactive Question and Answer System Using Advanced Intelligent Integration Technology: An Application for PDF Documents" aims to revolutionize the way we read and engage with PDFs. This project implemented with the help of LLMs, LangChain, Pinecone DB database, and Vercel AI SDK provides a conversational chatbot for interaction with PDF uploaded by the users. The chatbot takes users' queries, extracts the desired information from the PDF content and delivers correct answers to the users which means the chatbot helps the users to make the PDF content and the information more accessible and easier to retrieve. It is an innovative approach to utilize the potential link transforming static PDFs into dynamic and web-enabled resources to suit various applications such as research and business activity.

1.2 OBJECTIVE

The aim and objective of this project is to present the development and implementation of "PDF Genie: Q &A With AI on Advanced Integration Level. The objectives are to describe implementing LLMs, LangChain, Pinecone DB, and the Vercel AI SDK to build an intelligent bot that can result query and retrieve PDFs and deliver accurate responses from given PDFs. Furthermore, the report is designed to demonstrate the user-friendliness of the system and its further uses in a number of areas, focusing on the potential of the system in increasing the possibilities of work with documents in terms of their availability, the optimization of the work on the search and retrieval of necessary information, and the transformation of static PDF-files into active sources of information.

1.3 EXISTING SYSTEM

One of the areas in which the existing system fails is that interaction with PDF documents is generally by manual search and read which can be very cumbersome especially on large documents. Users of such systems have to go through a lot of content in huge volumes without assistance and locating the desired information becomes time-consuming. Another issue is that traditional search functions do not seem to recognize the context necessary to understand advanced requests. Keyword-based search engines are simple but often lack relational or linguistic capabilities and are weak at returning contextually sound results for even simple searches for structured data. In general, there is no proper solution to deal with the extraction task from PDF documents and hence the ability to obtain useful information from the documents is limited as it affects the overall performance regarding time in doing the task.

1.4 PROPOSED SYSTEM

The proposed system, has distinctive features of interacting with PDF documents through chat with Artificial Intelligence. This system has a extremely complex backend known as the Knowledge Store for LLMs where the knowledge base including numerous PDFs and documents is kept and analysed. This backend uses Pinecone DB, a vector database designed to handle a significant amount of high-dimensional data. Pinecone DB is used to store the PDFs and any other documents so that they can be accessed easily by various systems when needed. To implement vector indexing Pinecone DB enhances the operations of the system in a way that allows for fast and effective searching of documents contained in large volumes. The user interface is designed such that the users upload the PDF documents they want to convert and ask questions to an intelligent chatbot. These queries are then processed by the chatbot using Large Language Models (LLMs) like those from Open AI as well as the uploaded PDFs.

Langchain TypeScript can aid the development of applications using LLM, especially in chaining language models for better and more coherent answers. This is done in Langchain, which controls the sequence of these actions to ensure the correct understanding of the user's query and the subsequent generation of an appropriate and contextually correct response. In order to extend the functionality of the application, Vercel has developed an AI SDK that is based on machine learning and natural language processing, which allows the AI component to be easily integrated into the web application. The chatbot SDK for our application is developed in this way so that it is easily scalable, deployable, and capable of handling multiple users' requests in parallel. The design of the layout is exceptional and easy to use since it is designed using the Shadcn UI component library. Instead, users can either pose questions or use questions in the form of a chat to get answers on the PDF documents, which makes the interaction closer to PDF documents friendly and appealing to the end user.

In conclusion, the PDF GENIE system is an enhancement in document management since it changes the PDF documents to be more interactive and search friendly. Thus, using the state-of-the-art AI technologies and tools such as Pinecone DB, LangChain, and Vercel AI SDK, the proposed system offers a strong approach for accessing and utilizing the information encoded in PDFs. Such a tool is expected to transform industries including educational, research, and business intelligence where accessing relevant information from a large pool of documents is critical. Introducing intelligent augmentation and interactive elements that engage machine intelligence changes the conventional methodology of reading and searching for content in documents and makes PDF documents more available and useful in different disciplines.

CHAPTER - 2

LITERATURE SURVEY

[1] Alaswad, S., Kalganova, T., & Awad, W.S. (2023). Using ChatGPT and other LLMs in Professional Environments. Large language models like ChatGPT, Google's Bard, and Microsoft's new Bing, to name a few, are developing rapidly in recent years, becoming very popular in different environments, and supporting a wide range of tasks. A deep look into their outcomes reveals several limitations and challenges that can be further improved. The main challenge of these models is the possibility of generating biased or inaccurate results, since these models rely on large amounts of data with no access to unpublic information. Moreover, these language models need to be properly monitored and trained to prevent generating inappropriate or offensive content and to ensure that they are used ethically and safely. This study investigates the use of ChatGPT and other large language models such as Blender, and BERT in professional environments. It has been found that none of the large language models, including ChatGPT, have been used in unstructured dialogues. Moreover, involving the models in professional environments requires extensive training and monitoring by domain professionals or fine-tuning through API.

[2] Ayub, I., Hamann, D., Hamann, C.R., & Davis, M.J. (2023). Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: A qualitative analysis. Cureus, 15(8). This article investigates the limitations of Chat Generative Pre-trained Transformer (ChatGPT), a language model developed by OpenAI, as a study tool in dermatology. The study utilized ChatPDF, an application that integrates PDF files with ChatGPT, to generate American Board of Dermatology Applied Exam (ABD-AE)- style questions from continuing medical education articles from the Journal of the American Board of

Dermatology. A qualitative analysis of the questions was conducted by two board-certified dermatologists, assessing accuracy, complexity, and clarity. Out of 40 questions generated, only 16 (40%) were deemed accurate and appropriate for ABD-AE study preparation. The remaining questions exhibited limitations, including low complexity, lack of clarity, and inaccuracies. The findings highlight the challenges faced by ChatGPT in understanding the domain-specific knowledge required in dermatology. The study emphasizes that while ChatGPT may aid in generating simple questions, it cannot replace the expertise of dermatologists and medical educators in developing high-quality, board-style questions that effectively evaluate candidates' knowledge and reasoning abilities.

[3] Cheng, S.L., Tsai, S.J., Bai, Y.M., Ko, C.H., Hsu, C.W., Yang, F.C., Tsai, C.K., Tu, Y.K., Yang, S.N., Tseng, P.T., & Hsu, T.W. (2023). Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study. Journal of Medical Internet Research, 25, e51229. ChatGPT may act as a research assistant to help organize the direction of thinking and summarize research findings. However, few studies have examined the quality, similarity (abstracts being similar to the original one), and accuracy of the abstracts generated by ChatGPT when researchers provide full-text basic research papers. We aimed to assess the applicability of an artificial intelligence (AI) model in generating abstracts for basic preclinical research. Using ChatGPT to generate a scientific abstract may not lead to issues of similarity when using real full texts written by humans. However, the quality of the ChatGPT-generated abstracts was suboptimal, and their accuracy was not 100%.

[4] Chien, C.-C., Chan, H.-Y., & Hou, H.-T. (2024). Learning by playing with generative AI: Design and evaluation of a role-playing educational game with generative AI as scaffolding for instant feedback interaction. Journal of Research on Technology in Education, pp. 1–20. doi: 10.1080/15391523.2024.2338085. In this study, an online contextualized educational game was designed to provide interactive simulated dialogues using generative AI scaffolding (using Chat PDF) in a contextualized game as scaffolding for immediate feedback, where learners can access guides and explore knowledge. This study analyzed learners' behaviors while performing AI prompting in the interactive scaffolding, as well as learners' psychological responses. A total of 59 students participated in this study. The results of the study showed that learners had significantly high flow and low activity anxiety in the game tasks, while game feedback and scaffolding usefulness had significant effects on learning aids. The generative AI instant feedback interactive scaffolding had a certain high percentage of direct answers or indirect suggestions, which is suitable for interactive scaffolding.

[5] Chubb, L.A. (2023). Me and the machines: Possibilities and pitfalls of using artificial intelligence for qualitative data analysis. International Journal of Qualitative Methods, 22, 16094069231193593. This paper describes how one researcher learned to overcome artificial intelligence (AI) paralysis and embrace Chat PDF. This freely available AI application uses natural language processing (NLP) to respond to user queries about an uploaded PDF. Researcher insights from experimenting with the AI tool Chat PDF for qualitative data analysis are presented, highlighting the advantages, pitfalls, and application-related considerations. As a two-phase curiosity experiment, the researcher engaged in a theory-building exercise to explore key concepts for understanding when using Chat PDF to assist researchers in qualitative data analysis. The experiment generated insights about the purposeful use of AI tools that

incorporate NLP for analysis and the risks of inaccuracy when researchers are not familiar with the data or skilled in prompt engineering. Insights raise questions about whether Chat PDF is a viable research assistant for qualitative researchers, ethical issues with specific forms of qualitative data, and the potential of AI tools for community and student researchers.

[6] Chubb, L.A. (2024, April). Leveraging AI for Social Work Research: Possibilities for AI-Assisted Qualitative Data Analysis. In European Conference for Social Work Research (ECSWR). Can social workers harness the power of generative artificial intelligence (Al) as a tool for qualitative research? Might prompt engineering be a valuable skill of future social work researchers? This poster showcases how one researcher learned to overcome Al paralysis and embrace Chat PDF - a freely available Al application that uses natural language processing {NLP) to respond to user queries about an uploaded PDF. Examples experimenting with the Al tool Chat PDF for qualitative data analysis are presented, highlighting the advantages, pitfalls, and application-related considerations. As a two-phase curiosity experiment, the researcher engaged in a theory-building exercise to explore key concepts for understanding when using Chat PDF to assist researchers in qualitative data analysis, specifically in the organization of data into vignettes. The experiment generated insights about the purposeful use of Al tools for analysis and the risks of inaccuracy when researchers have not reached data familiarization or lack skills in prompting. The poster shares an assessment on whether Chat PDF and other NLPs are viable research assistants, the ethical issues with specific forms of qualitative data, and the potential of Al tools for social work students and community social work researchers.

[7] Fabiano, N., Gupta, A., Bhambra, N., Luu, B., Wong, S., Maaz, M., Fiedorowicz, J.G., Smith, A.L., & Solmi, M. (2024). How to optimize the systematic review process using AI tools. JCPP Advances. Systematic reviews are a cornerstone for synthesizing the available evidence on a given topic. They simultaneously allow for gaps in the literature to be identified and provide direction for future research. However, due to the ever-increasing volume and complexity of the available literature, traditional methods for conducting systematic reviews are less efficient and more time-consuming. Numerous artificial intelligence (AI) tools are being released with the potential to optimize efficiency in academic writing and assist with various stages of the systematic review process including developing and refining search strategies, screening titles and abstracts for inclusion or exclusion criteria, extracting essential data from studies and summarizing findings. Therefore, in this article we provide an overview of the currently available tools and how they can be incorporated into the systematic review process to improve efficiency and quality of research synthesis. We emphasize that authors must report all AI tools that have been used at each stage to ensure replicability as part of reporting in methods.

[8] Fu, J., Han, H., Su, X., & Fan, C. (2024). Towards human-ai collaborative urban science research enabled by pre-trained large language models. Urban Informatics, 3(1), 8. Pre-trained large language models (PLMs) have the potential to support urban science research through content creation, information extraction, assisted programming, text classification, and other technical advances. In this research, we explored the opportunities, challenges, and prospects of PLMs in urban science research. Specifically, we discussed potential applications of PLMs to urban institution, urban space, urban information, and citizen behaviors research through seven examples using ChatGPT. We also examined the challenges of PLMs in urban science research from both technical and social perspectives. The prospects of the application of PLMs in urban

science research were then proposed. We found that PLMs can effectively aid in understanding complex concepts in urban science, facilitate urban spatial form identification, assist in disaster monitoring, sense public sentiment and so on. They have expanded the breadth of urban research in terms of content, increased the depth and efficiency of the application of multi-source big data in urban research, and enhanced the interaction between urban research and other disciplines. At the same time, however, the applications of PLMs in urban science research face evident threats, such as technical limitations, security, privacy, and social bias. The development of fundamental models based on domain knowledge and human-AI collaboration may help improve PLMs to support urban science research in future.

[9] Gustafson-Sundell, N., & McCullough, M. (2023). The Library & Generative AI. A demonstration of several AI tools, including ChatGPT, ChatPDF, Consensus, and more. The focus of the session is on potential student uses of the tools and related library initiatives, so we address the limits of ChatGPT as an information source. Librarians can help students learn how to use these tools responsibly and provide leadership on campus as AI is integrated into assignments.

[10] Inamdar, S. (2023). Impact of artificial intelligence text generators (AITGs) on libraries. Library Hi Tech News, 40(8), 9-13. The purpose of studying the impact of artificial intelligence text generators (AITGs) on libraries is to examine the effect of AITGs on the library landscape, including the services offered, the resources provided and the roles of library staff. The current study examined how AITGs impact libraries. The researcher was able to comprehend the problem by critically analyzing and reviewing the pertinent published works, such as books, journals and articles. This study concludes AITGs can assist libraries in streamlining operations, enhancing services and making collections

more accessible. It is vital to highlight that AITGs are not intended to dissuade its users from visiting physical libraries or to replace them with virtual ones. Instead, they are a tool that can improve and supplement the services and resources provided by virtual libraries. The study's observations add to the corpus of information on AITGs in libraries and help users comprehend their technological foundations. Further empirical research is recommended on the effects of AITGs and their impact on libraries.

[11] Nguyen-Trung, K., Saeri, A.K., & Kaufman, S. (2023). Applying ChatGPT and AI-powered tools to accelerate evidence reviews. Artificial Intelligence (AI) tools have been used to improve the productivity of evidence review and synthesis since at least 2016, with EPPI-Reviewer and Abstrackr being two prominent examples. However, since the release of ChatGPT by OpenAI in late 2022, a large language model with an intuitive chatbot interface, the use of AI-powered tools for research – especially those that deal with text-based data – has exploded. In this working paper, we describe how we used the AI-powered tools such as ChatGPT, ChatGPT for Sheets and Docs, Casper AI, and ChatPDF to assist several stages of an evidence review. Our goal is to demonstrate how AI-powered tools can boost research productivity, identify their current weaknesses, and provide recommendations for researchers looking to utilize them.

[12] Özcan, A., & Polat, S. (2023). Artificial Intelligence and Chat Bots in Academic Research. Journal of Research in Social Sciences and Language, 3(2), 81-90. The aim of our study is to discuss the use of artificial intelligence-supported platforms, which have become increasingly popular in recent months, in the context of ethics, opportunities, challenges, and the role of the researcher. In this context, we analysed platforms such as ChatGPT, ChatPDF, Consensus, SciSpace, and Scite Assistant. Within the scope of our analyses, we concluded

that various regulations regarding the use of AI-supported platforms in scientific research should be enacted as soon as possible. Although such platforms offer opportunities for researchers, they also bring challenges such as referencing and reproducibility of scientific work. Besides, the use of AI-supported platforms in scientific research also puts the role of researchers into question.

[13] Panda, S. (2023). Enhancing PDF interaction for a more engaging user experience in library: Introducing ChatPDF. IP Indian Journal of Library Science and Information Technology, 8(1), 20-25. The paper introduces Chat PDF as a solution to the limited interaction capabilities of traditional PDF readers in library systems. It highlights the benefits of Chat PDF for enhancing user engagement and satisfaction with digital resources in libraries, and identifies potential areas for implementation and opportunities for further research. The paper addresses the issue of limited PDF interaction capabilities of traditional PDF readers in library systems and their potential to cause frustration and disengagement among users. It proposes Chat PDF, an online software platform that leverages ChatGPT API to offer a more intuitive and natural way of engaging with PDF documents. The paper also outlines the various features and advantages of Chat PDF and its potential impact on library systems. The paper concludes that Chat PDF offers a more advanced and user-friendly approach to interacting with PDF files than traditional PDF readers. It recommends libraries to consider implementing Chat PDF to enhance user engagement and satisfaction with their digital resources. The paper also identifies limitations and opportunities for future research in evaluating the impact of Chat PDF on user experience and library systems. This paper introduces Chat PDF as a solution to the limited PDF interaction capabilities in library systems, highlighting its unique features and benefits, such as summarization, recommendations, multi-lingual support, and AI assistance. It also discusses the potential impact of Chat PDF implementation on

library systems, making it an original and valuable contribution to the field of library science.

[14] Ruiz-Rojas, L.I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. Sustainability, 15(15), 11524.** This study focuses on the potential of generative artificial intelligence tools in education, particularly through the practical application of the 4PADAFE instructional design matrix. The objective was to evaluate how these tools, in combination with the matrix, can enhance education and improve the teaching–learning process. Through surveys conducted with teachers from the University of ESPE Armed Forces who participated in the MOOC course "Generative Artificial Intelligence Tools for Education: GPT Chat Techniques", the study explores the impact of these tools on education. The findings reveal that generative artificial intelligence tools are crucial in developing massive MOOC virtual classrooms when integrated with an instructional design matrix. The results demonstrate the potential of generative artificial intelligence tools in university education. By utilizing these tools in conjunction with an instructional design matrix, educators can design and deliver personalized and enriching educational experiences. The devices offer opportunities to enhance the teaching–learning process and tailor educational materials to individual needs, ultimately preparing students for the demands of the 21st century. The study concludes that generative artificial intelligence tools have significant potential in education. They provide innovative ways to engage students, adapt content, and promote personalized learning. Implementing the 4PADAFE instructional design matrix further enhances the effectiveness and coherence of educational activities. By embracing these technological advancements, education can stay relevant and effectively meet the digital world's challenges.

[15] Souifi, L., Khabou, N., Rodriguez, I.B., & Kacem, A.H. (2024). Towards the Use of AI-Based Tools for Systematic Literature Review. The constant growth in the number of published research studies and their rapid rate of publication creates a significant challenge in identifying relevant studies for unbiased systematic reviews. To address this challenge, artificial intelligence (AI) methods have been used since 2016 to improve the efficiency of scientific review and synthesis. Nevertheless, the growth in the number of AI-powered tools dedicated to processing text-based data has been remarkable since the introduction of generative pre-trained transformers by OpenAI in late 2022. Moreover, alongside this development, ChatGPT, a language model that provides a user-friendly chatbot interface, was introduced. The incorporation of this interactive feature has greatly enhanced the capability of developers and end-users alike to effectively utilize and access ChatGPT. This study aims to investigate the effectiveness of six AI-based tools namely Chatpdf, Pdf2gpt, Hipdf, SciSpace, Easy-peasy AI, and DocAnalyzer AI, developed utilizing ChatGPT technology. These tools will be evaluated in a specific scenario where they are automated to carry out a particular step within a Systematic Literature Review. Furthermore, the limitations associated with each tool will be analyzed, and strategies will be proposed to overcome them. Additionally, this study aims to provide recommendations for researchers who intend to incorporate these tools into their research processes.

CHAPTER - 3
SYSTEM DESIGN

3.1 GENERAL

System design involves the formulation and creation of systems that meet the specific needs of users. Fundamentally, the essence of studying system design lies in comprehending the individual elements and how they interact with each other.

3.2 DEVELOPMENT ENVIRONMENT

3.2.1 HARDWARE SPECIFICATIONS

This document offers a comprehensive overview of the hardware and its implementation, detailing the key components, their interactions, and the necessary requirements for seamless connectivity to utilities and installation.

Table 3.2.1 Hardware Specifications

| PROCESSOR	Intel Core i5
RAM	4GB or above (DDR4 RAM)
GPU	Intel Integrated Graphics
HARD DISK	6GB
PROCESSOR FREQUENCY	1.5 GHz or above

3.2.2 SOFTWARE SPECIFICATIONS

The below table constitutes a thorough evaluation of requirements that precedes the more detailed phases of system design, aiming to minimize the need for subsequent revisions. Furthermore, it should offer a practical foundation for estimating product expenses, potential risks, and project timelines.

Table 3.2.2 Software Specifications

FRONT END	Next JS, Shadcn UI
BACK END	Next JS
CODE EDITOR	Visual Studio Code

3.3 ARCHITECTURE DIAGRAM
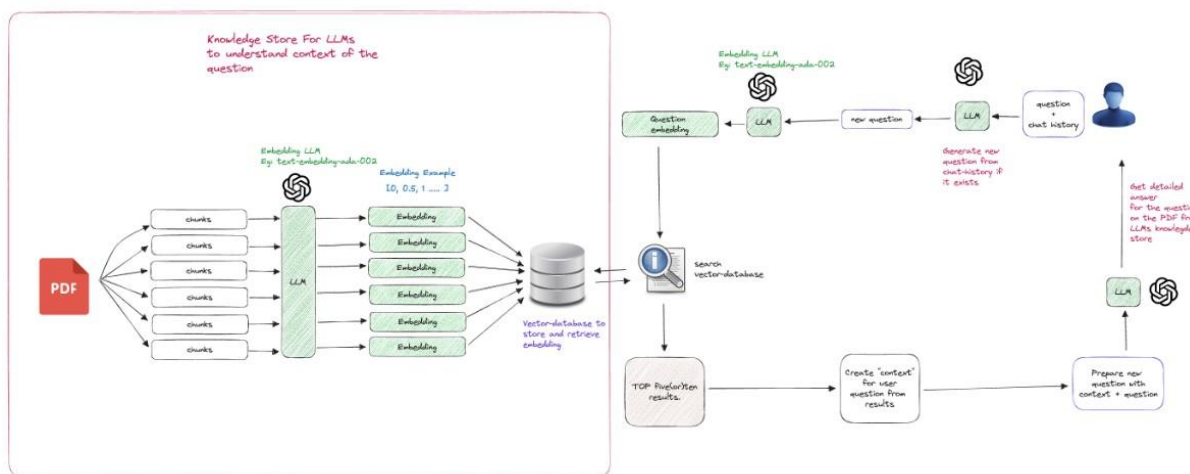
Fig 3.3.1 Architecture Diagram

A user submits a PDF and asks questions through large language models (LLMs) such as OpenAI and LangChain. Pinecone DB is a vector database that stores document embeddings to enhance information search. The LLMs then process the uploaded PDF to respond to the queries of the users. Another advantage of LangChain is that it chains LLMs together enabling it to offer more extensive responses. It also uses a knowledge store to put into consideration the context of the questions received from the users. Together with the LLMs, the vector database, and the knowledge store, it makes it possible for users to pose interactive questions on the PDFs, hence shifting the paradigm of how to access and use information across different fields such as education, research, and business intelligence.

CHAPTER - 4
PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

4.1.1 USER INTERFACE MODULE

The User Interface module is the component that actually deals with the way the PDF GENIE system looks and functions. It guarantees an easy and beautiful interface for the users who upload PDF files and deal with the chatbot. It contains page for chat window, and a place to display the responses. Using the Shadcn UI component library, the interface is silky smooth and thoroughly optimized, creating a more engaging and efficient use of the system.

4.1.2 PDF PROCESSING MODULE

In the PDF Processing module, files that are uploaded in PDF formats are processed and ingested. It reads text and metadata from the PDFs, and transforms these into indexed data storage forms. This module also aids in preparing the extracted information for further analysis to enhance its accuracy and integrity. By doing so, the efficiency of parsing and conversion is addressed and makes the integration of PDF documents into the knowledge base much easier and possible for the module to retrieve knowledge from it.

4.1.3 KNOWLEDGE STORE MODULE

Knowledge Store is the backend module that maintains and indexes the PDF documents which have been identified as important by the search engine. It takes advantage of Pinecone DB to optimally store datasets with high dimensionality to enable the quick and efficient search for documents. This module is chiefly responsible for indexing and storage of documents so that the latter can be retrieved by other components in the system. Using Pinecone DB, the Knowledge Store module improves the system's performance and data processing speed, which helps when working with extensive documents in PDF format.

4.1.4 CHATBOT INTERACTION MODULE

The Chatbot Interaction module works in terms of interaction between a user and the intelligent chatbot. It takes user inputs, searches the knowledge base for related information and constructs meaningful responses utilizing sophisticated language models. This module is responsible for understanding the user inputs and queries, as well as providing meaningful responses pertinent to the query. Through adopting LLMs such as the ones from OpenAI, the Chatbot Interaction module further strengthens the system performance of interpreting and replying relevant user queries.

4.1.5 LANGCHAIN INTERACTION MODULE

The LangChain Integration is a module that coordinates the series of operations that occur between identifying the user's query and providing a response. It combines multiple language models to provide syntactically and semantically correct answers to the posed questions. This module, therefore, improves on the system's performance by only allowing information flow from one language model to another after it has been analyzed and refined. In this way, using LangChain's possibilities, the function improves the generation of responses for each input, which contributes to the improvement of the general experience for users.

CHAPTER - 5

IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

PDF GENIE's workflow starts with the user interface whereby the users, can load PDF documents and also engage with the intelligent chatbot. When the documents are uploaded, the PDF Processing module extracts text and its metadata in the PDFs, handling the parsing phase carefully. This information is then structuring and processed in the Knowledge Store module fit for indexing and storage.

Once indexed, the PDF documents are stored in the Knowledge Store with appropriate use of Pinecone DB for efficient storage and retrieval. It means that this backend repository contains a large number of PDFs uploaded by the users to access various contents. When the users ask questions to the chatbot, the Chatbot Interaction module understands the meaning of the questions, searches the Knowledge Store for the appropriate information and uses the semiotic function to produce natural language questions as answers. Incorporation of LLM helps in providing exceptional and contextually sound responses to the queries posed by the user.

During this process, the Lai module manages the series of steps concerning handling user requests and constructing answers. It links several language models to maintain relevance and context of the responses generated, increasing the usability factor of the bot. This chain of operation starting from the uploading of the documents to the query processing to the generation of the correct response is a key function in the PDF GENIE to enable and facilitate an efficient search of information locked in PDF documents.

5.2 SAMPLE CODING

//pdf-loader.ts

```
import { PDFLoader } from "langchain/document_loaders/fs/pdf";
import { RecursiveCharacterTextSplitter } from "langchain/text_splitter";
import { env } from "./config";
export async function getChunkedDocsFromPDF() {
 try {
   const loader = new PDFLoader(env.PDF_PATH);
   const docs = await loader.load();
   const textSplitter = new RecursiveCharacterTextSplitter({
     chunkSize: 1000,
     chunkOverlap: 200,
   });
   const chunkedDocs = await textSplitter.splitDocuments(docs);
   return chunkedDocs;
 } catch (e) {
   console.error(e);
   throw new Error("PDF docs chunking failed !");
 }
}
```

//vector-store.ts

```
import { env } from './config';
import { OpenAIEmbeddings } from 'langchain/embeddings/openai';
import { PineconeStore } from 'langchain/vectorstores/pinecone';
import { Pinecone } from '@pinecone-database/pinecone';
export async function embedAndStoreDocs(
 client: Pinecone,  // @ts-ignore docs type error
 docs: Document<Record<string, any>>[]
```

```
) {  /* create and store the embeddings in the vectorStore */
  try {
    const embeddings = new OpenAIEmbeddings();
    const index = client.Index(env.PINECONE_INDEX_NAME);
    //embed the PDF documents
    await PineconeStore.fromDocuments(docs, embeddings, {
      pineconeIndex: index,
      textKey: 'text',
    });
  } catch (error) {
    console.log('error ', error);
    throw new Error('Failed to load your docs !');
  }
}
// Returns vector-store handle to be used a retrievers on langchains
export async function getVectorStore(client: Pinecone) {
  try {
    const embeddings = new OpenAIEmbeddings();
    const index = client.Index(env.PINECONE_INDEX_NAME);
    const vectorStore = await PineconeStore.fromExistingIndex(embeddings, {
      pineconeIndex: index,
      textKey: 'text',
    });
    return vectorStore;
  } catch (error) {
    console.log('error ', error);
    throw new Error('Something went wrong while getting vector store !');
  }
}
```

```typescript
//langchain.ts
import { ConversationalRetrievalQAChain } from "langchain/chains";
import { getVectorStore } from "./vector-store";
import { getPineconeClient } from "./pinecone-client";
import {
  StreamingTextResponse,
  experimental_StreamData,
  LangChainStream,
} from "ai-stream-experimental";
import { streamingModel, nonStreamingModel } from "./llm";
import { STANDALONE_QUESTION_TEMPLATE, QA_TEMPLATE } from
"./prompt-templates";
type callChainArgs = {
  question: string;
  chatHistory: string;
};
export async function callChain({ question, chatHistory }: callChainArgs) {
  try {
    // Open AI recommendation
    const sanitizedQuestion = question.trim().replaceAll("\n", " ");
    const pineconeClient = await getPineconeClient();
    const vectorStore = await getVectorStore(pineconeClient);
    const { stream, handlers } = LangChainStream({
      experimental_streamData: true,
    });
    const data = new experimental_StreamData();
    const chain = ConversationalRetrievalQAChain.fromLLM(
      streamingModel,
      vectorStore.asRetriever(),
```

```
  {
   qaTemplate: QA_TEMPLATE,
   questionGeneratorTemplate: STANDALONE_QUESTION_TEMPLATE,
   returnSourceDocuments: true, //default 4
   questionGeneratorChainOptions: {
    llm: nonStreamingModel,
   },
  }
);
chain
  .call(
   {
     question: sanitizedQuestion,
     chat_history: chatHistory,
   },
   [handlers]
  )
  .then(async (res) => {
   const sourceDocuments = res?.sourceDocuments;
   const firstTwoDocuments = sourceDocuments.slice(0, 2);
   const pageContents = firstTwoDocuments.map(
    ({ pageContent }: { pageContent: string }) => pageContent
   );
   console.log("already appended ", data);
   data.append({
    sources: pageContents,
   });
   data.close();
  }); // Return the readable stream
```

```
    return new StreamingTextResponse(stream, {}, data);
  } catch (e) {
    console.error(e);
    throw new Error("Call chain method failed to execute successfully!!");
  }
}
```
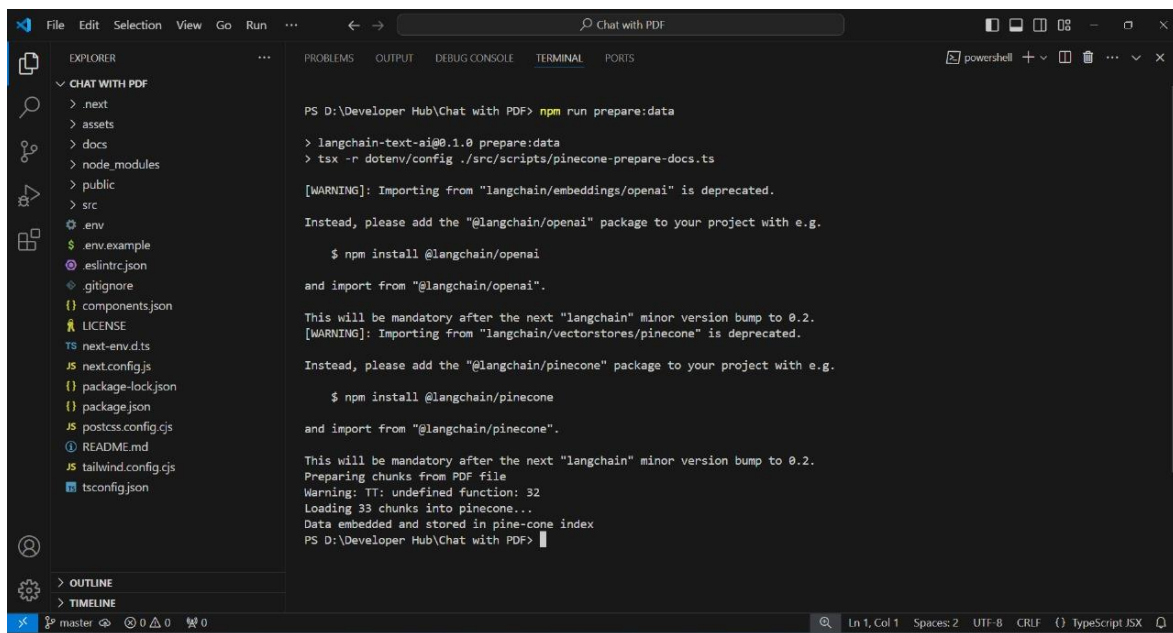
5.3 OUTPUT SCREENSHOTS
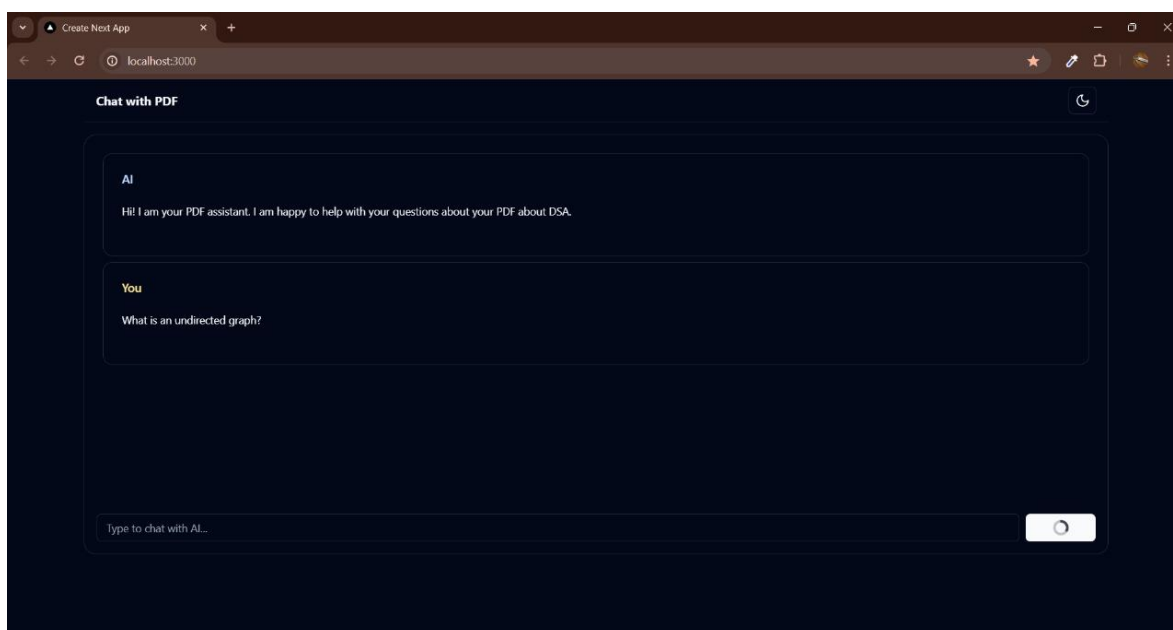

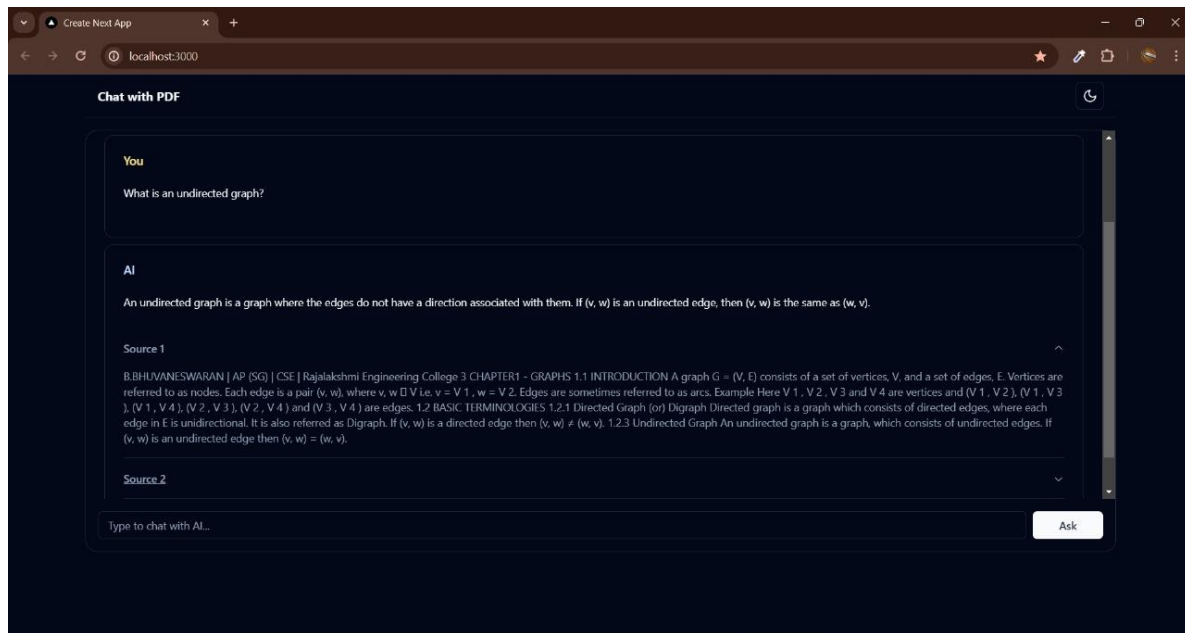**Fig 5.3.1** PDF Upload into Pinecone DB

Fig 5.3.2 User Prompt

Fig 5.3.3 Chatbot Response with source context

CHAPTER – 6
CONCLUSION AND FUTURE ENHACEMENT

6.1 CONCLUSION

The "PDF GENIE: Thus, the "INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION" project proves that it is possible to significantly improve the possibility of the users' engagement as well as the efficiency of using PDF documents by integrating the latest AI solutions. Utilizing Large Language Models (LLMs), LangChain, Pinecone DB, and the Vercel AI SDK; the project presents a comprehensive method for intelligent data retrieval on both the web and in its application. The usual static document is turned into an interactive one and users are allowed to query themselves in an efficient and effective manner and get appropriate answers pertaining to their PDFs. This helps not only in enhancing efficiency of solving search and retrieval tasks but opens a wide range of possibilities for document application in various spheres of human activity, such as education, research, and business. The application of these advanced technologies also guarantees scalability, speed, and precision to enhance the user experience that hard and fast has become a requirement in the current society's interaction with documents. PDF GENIE remains another exhibit of how AI is set to revolutionize how we interact and utilize pdf documents online.

6.2 FUTURE ENHANCEMENT

Future enhancements for the "PDF GENIE: INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION" project will involve enhanced capabilities and adaptation to meet user needs. One future enhancement is translating support and more comprehensiveness of the chatbot in approaching PDFs in different languages and therefore expand the reach and usability in global setup. Moreover, further expansion of the NLU can improve the system's capacity to recognize and properly respond to slightly more complicated inquiries. Implementing the OCR can diversify the application of the chatbot to embrace the scanned documents or

images within the PDF in that it can analyse them. Furthermore, using machine learning approaches to be able to formulate messages that fit the behaviour and characteristics of users will be of great value in improving the user experience. Enhancing interoperability with cloud services will then help achieve scale and performance optimization for dealing with bigger volumes of data and accommodating more users at a time. Last but not the least, it will ensure that the security with which the PDFs handle sensitive data is enhanced and thus retain the user confidence and meet the compliance with the data protection laws. These enhancements will jointly put PDF GENIE on the stronger and smarter position in terms of being more interactive and richer in providing information and also in meeting user's requirement of doing document querying and information extract in much easier way.

REFERENCES

[1] Alaswad, S., Kalganova, T., & Awad, W.S. (2023). Using ChatGPT and other LLMs in Professional Environments.

[2] Ayub, I., Hamann, D., Hamann, C.R., & Davis, M.J. (2023). Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: A qualitative analysis. Cureus, 15(8).

[3] Cheng, S.L., Tsai, S.J., Bai, Y.M., Ko, C.H., Hsu, C.W., Yang, F.C., Tsai, C.K., Tu, Y.K., Yang, S.N., Tseng, P.T., & Hsu, T.W. (2023). Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study. Journal of Medical Internet Research, 25, e51229.

[4] Chien, C.-C., Chan, H.-Y., & Hou, H.-T. (2024). Learning by playing with generative AI: Design and evaluation of a role-playing educational game with generative AI as scaffolding for instant feedback interaction. Journal of Research on Technology in Education, pp. 1–20. doi: 10.1080/15391523.2024.2338085.

[5] Chubb, L.A. (2023). Me and the machines: Possibilities and pitfalls of using artificial intelligence for qualitative data analysis. International Journal of Qualitative Methods, 22, 16094069231193593.

[6] Chubb, L.A. (2024, April). Leveraging AI for Social Work Research: Possibilities for AI-Assisted Qualitative Data Analysis. In European Conference for Social Work Research (ECSWR).

[7] Fabiano, N., Gupta, A., Bhambra, N., Luu, B., Wong, S., Maaz, M., Fiedorowicz, J.G., Smith, A.L., & Solmi, M. (2024). How to optimize the systematic review process using AI tools. JCPP Advances.

[8] Fu, J., Han, H., Su, X., & Fan, C. (2024). Towards human-ai collaborative urban science research enabled by pre-trained large language models. Urban Informatics, 3(1), 8.

[9] Gustafson-Sundell, N., & McCullough, M. (2023). The Library & Generative AI.

[10] Inamdar, S. (2023). Impact of artificial intelligence text generators (AITGs) on libraries. Library Hi Tech News, 40(8), 9-13.

[11] Nguyen-Trung, K., Saeri, A.K., & Kaufman, S. (2023). Applying ChatGPT and AI-powered tools to accelerate evidence reviews.

[12] Özcan, A., & Polat, S. (2023). Artificial Intelligence and Chat Bots in Academic Research. Journal of Research in Social Sciences and Language, 3(2), 81-90.

[13] Panda, S. (2023). Enhancing PDF interaction for a more engaging user experience in library: Introducing ChatPDF. IP Indian Journal of Library Science and Information Technology, 8(1), 20-25.

[14] Ruiz-Rojas, L.I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. Sustainability, 15(15), 11524.

[15] Souifi, L., Khabou, N., Rodriguez, I.B., & Kacem, A.H. (2024). Towards the Use of AI-Based Tools for Systematic Literature Review.