



MINI PROJECT REPORT

ON

“AI PDF Reader”

SUBMITTED IN

COMPUTER SCIENCE & ENGINEERING

BACHELOR OF TECHNOLOGY

BY

AMAN KANNAUJIYA

(ROLL NO: [22150010004])

SEMESTER VI

SUBMITTED

TO

MR. GAURAV DWIVEDI

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

FACULTY OF ENGINEERING AND TECHNOLOGY

UNITED UNIVERSITY,

RAWAPUR, PRAYAGRAJ, UTTAR PRADESH-211012

ABSTRACT

This project presents the development of an AI-powered PDF Reader designed to enhance document interaction through natural language processing. Unlike traditional PDF viewers that require manual navigation and keyword search, this AI-based system enables users to query and understand content conversationally. By leveraging machine learning models such as transformer-based language models (e.g., GPT or BERT), the system extracts, indexes, and interprets PDF content efficiently. The application supports semantic understanding, context-aware answers, and can summarize large documents or highlight relevant sections based on user queries. This innovation significantly improves user experience, especially in research, legal, academic, and corporate environments where large documents are prevalent. The project demonstrates how artificial intelligence can streamline information retrieval and make document analysis more intuitive and accessible.

INTRODUCTION:

PDF files are commonly used for sharing documents, but extracting specific information from them can be difficult and time-consuming. Traditional PDF readers require manual searching and lack intelligent features. This project introduces an AI-powered PDF Reader that uses natural language processing (NLP) to allow users to ask questions and receive meaningful answers from the document. It makes reading and analyzing PDFs faster, easier, and more interactive by understanding the context and content of the text.

MOTIVATION:

Reading and searching through long PDF documents can be slow and frustrating. Traditional PDF readers do not understand the content, making it hard to find specific information quickly. This project is motivated by the need to make PDF interaction smarter and faster using AI, allowing users to ask questions and get instant, accurate answers from the document.

WHY THIS PROJECT?

This project was chosen to solve a common problem faced by students, researchers, and professionals—difficulty in quickly finding relevant information in large PDF documents. By using AI and natural language processing, the project offers a smarter way to interact with PDFs, making reading, searching, and understanding content much easier. It combines practical usefulness with modern technology, making it both innovative and impactful.

OBJECTIVES AND GOALS:

- ☐ To develop an AI-powered PDF Reader that can understand and respond to user queries in natural language.
- ☐ To automate the process of extracting key information from large PDF documents.
- ☐ To enhance the user experience by enabling semantic search and intelligent navigation.
- ☐ To implement summarization and question-answering features for faster content understanding.
- ☐ To create a user-friendly interface that makes document interaction simple and efficient.

BACKGROUND:

PDFs are commonly used for sharing documents, but finding specific information in large files is often difficult with traditional readers. These tools lack understanding of the content and only support basic search functions. With the rise of AI and natural language processing (NLP), it's now possible to build systems that understand text and respond intelligently. This project uses these technologies to create a smarter PDF reader that makes document navigation and information retrieval easier.

Tools and Platform:

The project uses a modern full-stack architecture:

- **Frontend:** Next.js, React, TypeScript, Tailwind CSS for UI, and Lucide React for icons.
- **Backend:** Node.js, Express.js, Multer for file handling, BullMQ & Redis for job queuing.
- **AI & Document Processing:** LangChain, OpenAI GPT-4.1, and Qdrant for semantic search and AI-powered report generation.
- **API Endpoints:** /upload/pdf, /chat, and /report for document interaction.
- **Development Tools:** PNPM for package management, ESLint for code quality.

This stack ensures efficient document processing, semantic search, and AI integration.

Hardware requirements:

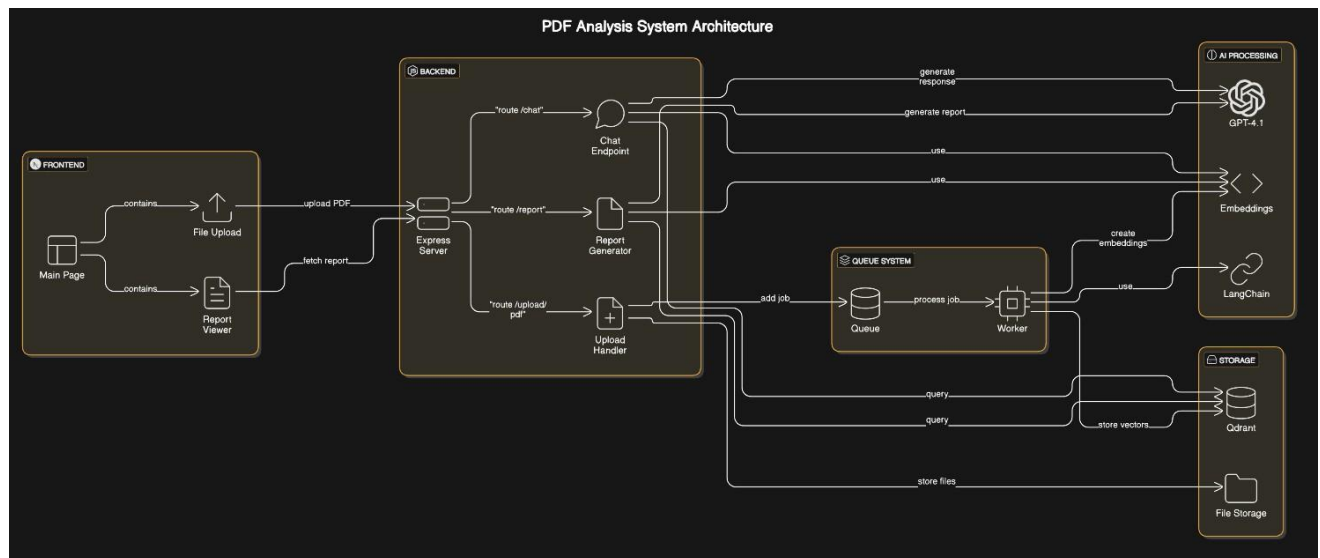
- ☐ **CPU:** A multi-core processor (4+ cores) is recommended for handling AI processing and asynchronous tasks efficiently.
- ☐ **RAM:** At least 8GB of RAM for smooth operation, especially when processing large PDF files and running AI models.
- ☐ **Storage:** SSD with at least 100GB of free space to store documents, database files, and models.
- ☐ **Network:** A stable internet connection is required for AI model inference (if hosted remotely) and data transfer.

Software platforms:

- Frontend : Next js , Tailwind-CSS, Lucide react
- Backend : Express , node , typescript
- Database server : VectorDB(Qdrant)
- AI_Engine : Langflow
- OpenAI

SRS DIAGRAMS:

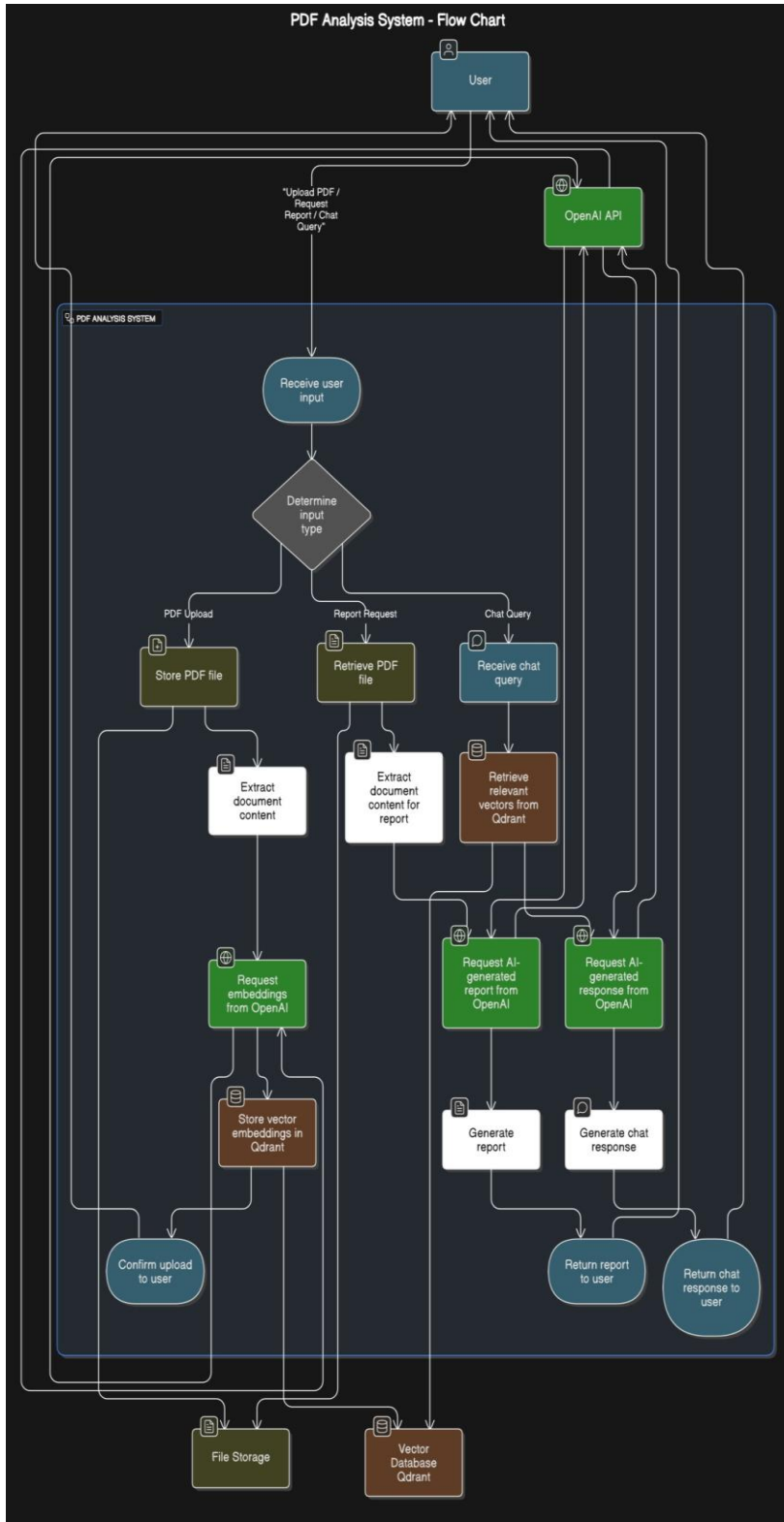
- **UML DIAGRAM:**



DFD:

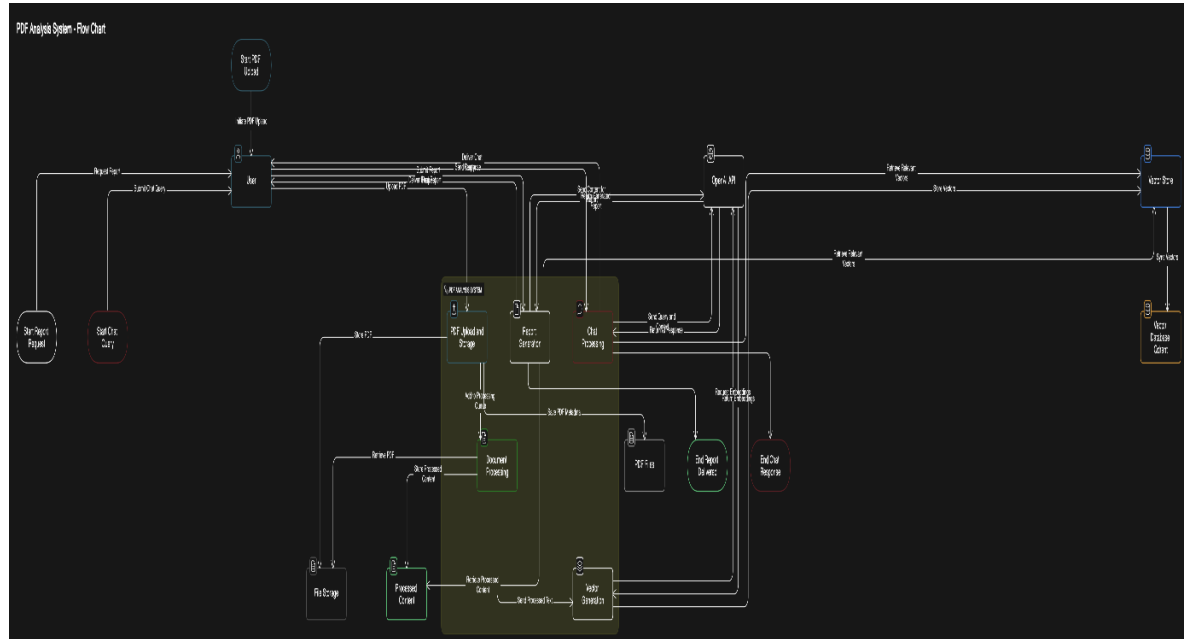
○

0 LEVEL

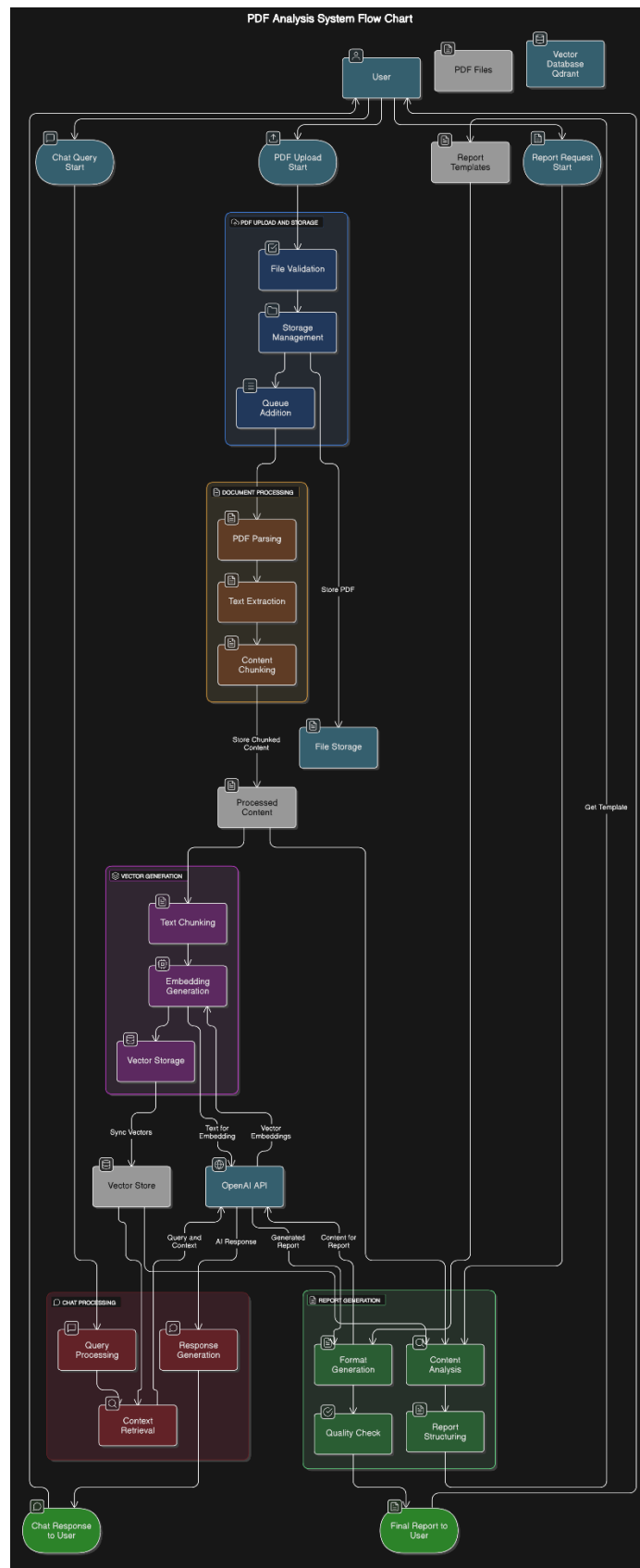




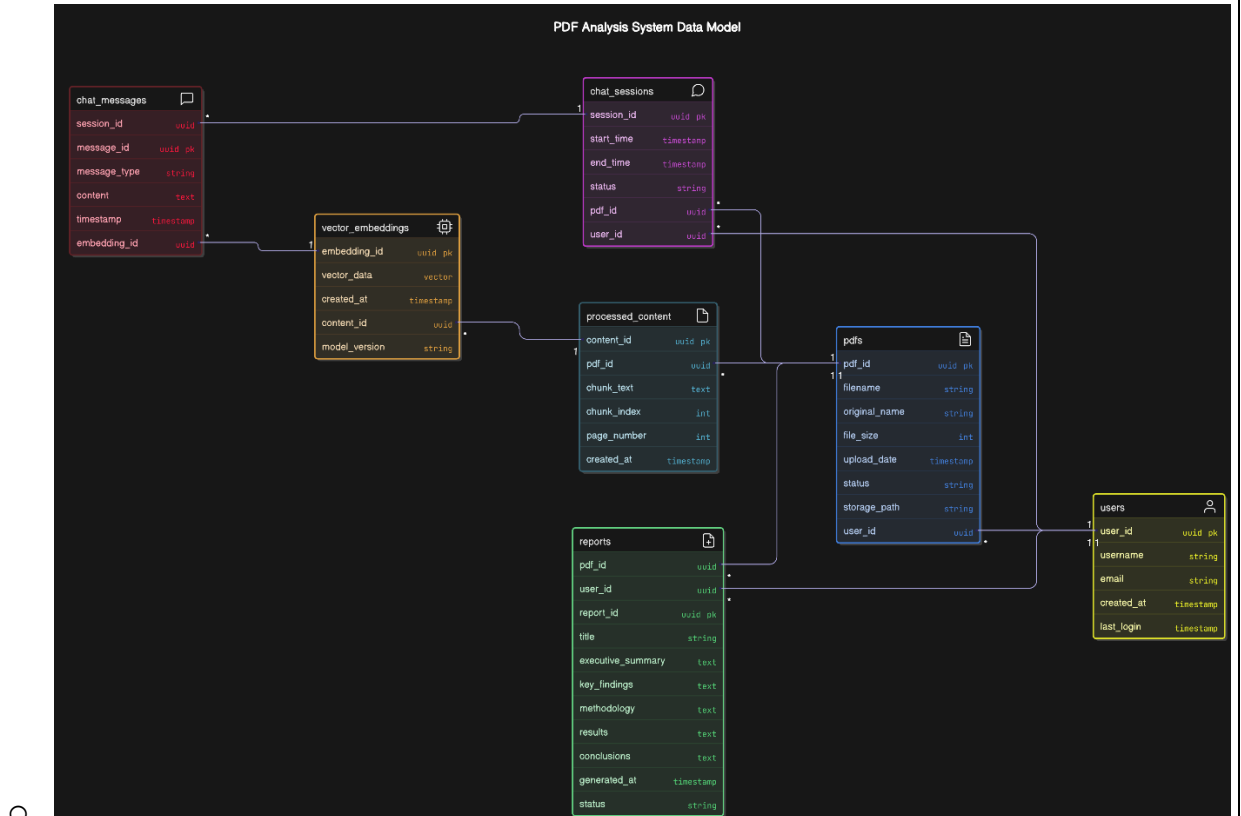
O



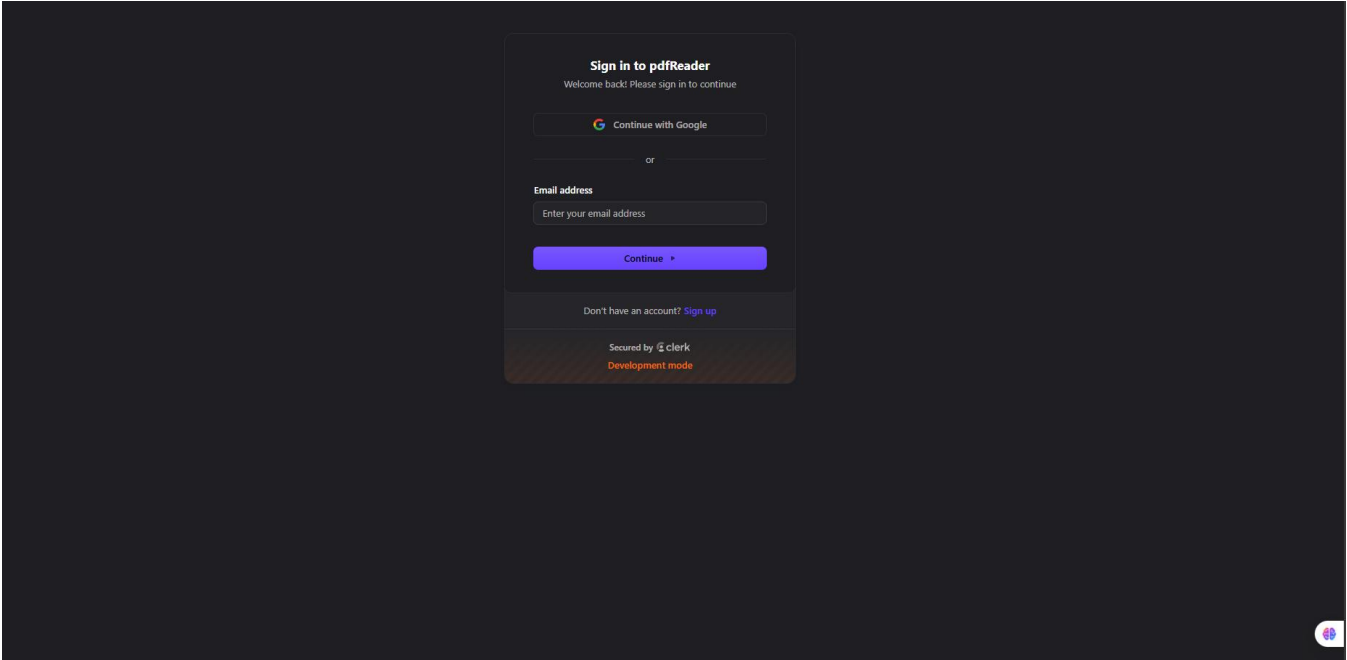
2 LEVEL



ER DIAGRAM



SCREENSHOT OF PROJECT



```

document.write(`
pageContent: 'Ricky Vishwakarma \n' +
'IT NIT Kurukshetra 2025 \n' +
' abhisheksahu1908@gmail.com UP, India electrophileee +917753962141 abhishek-sahu-698718224 \n' +
'\n' +
'EDUCATION \n' +
'B.Tech in Information Technology \n' +
'National Institute Of Technology,Kurukshetra \n' +
' CGPA: 8.66 till 4th sem \n' +
'\n' +
'\n' +
'Senior Secondary Examination CBSE - Class XII \n' +
'Allahabad Public School \n' +
' Percentage: 90.4% CBSE-2020 \n' +
'PROJECTS \n' +
'AccoNest | GitHub \n' +
'• Tools and technologies used :Reactjs ,React Router , Tail- \n' +
'wind, Nodejs ,Express ,MongoDb \n' +
'• Implemented an efficient booking system AccoNest, enabling \n' +
'users to reserve accommodations by selecting check-in and \n' +
'check-out dates. \n' +
'• Provided comprehensive access to accommodation details, \n' +
'amenities, and pricing \n' +
'• Enabled registered users to both view and manage book- \n' +
'ings, as well as add and manage accommodations within the \n' +
'platform. \n' +
'Higher Secondary Examination CBSE - Class X \n' +
'Little Blossom's Convent School \n' +
' Percentage: 85.54% CBSE-2018 \n' +
'\n' +
'EXPERIENCE \n' +
'\n' +
'Subject Matter Expert(SME) \n' +
'Physics Wallah \n' +
' July 2022 - December 2022 Work From Home \n' +
'• Through my freelancing experience as a Subject Expert in Mathematics, I \n' +
'had solved more than 800 questions. \n' +
'• This extensive practice significantly enhanced my problem-solving abil- \n' +
'ities. As I tackled a diverse range of mathematical problems, I became \n' +
'more adept at analyzing and breaking down complex tasks into manage- \n' +
'able steps. \n' +
'\n' +
'\n' +
'\n' +
`);

```

APPLICATIONS:

1. PDF Document Processing
2. Semantic Search and analysis
3. AI-powered questioning
4. Asynchronous process

Future of the project:

- 1. Enhanced Document Processing**
- 2. Improve AI capabilities**

REFERENCES:-

VectorDB Docs

Docker docs

Next js documentation

Mozilla dev networks

Redis docs