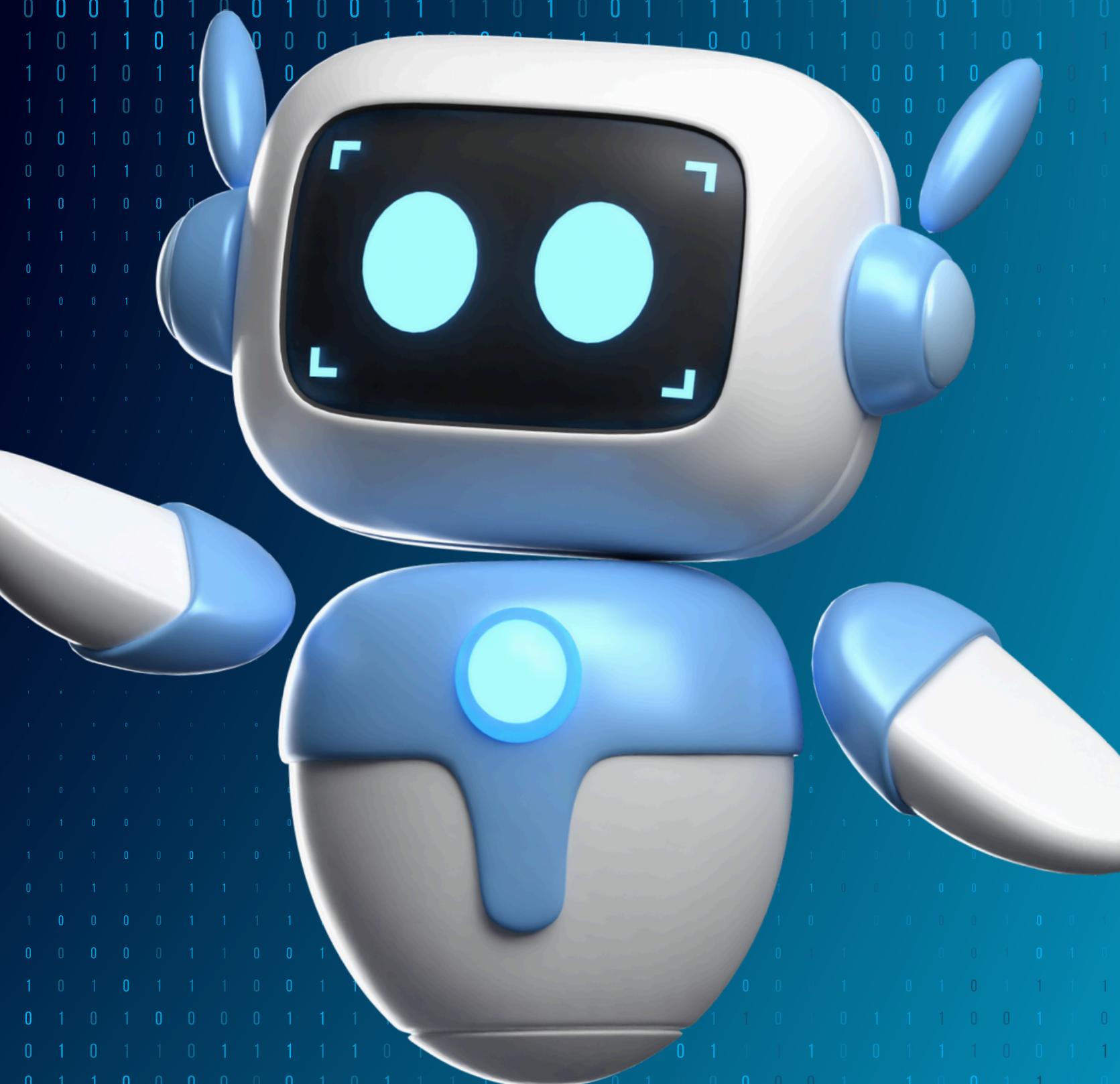


PDF GENIE: INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION

Team Members:

- Reshma R - 210701212
- Sangeetha G - 210701229



CONTENTS



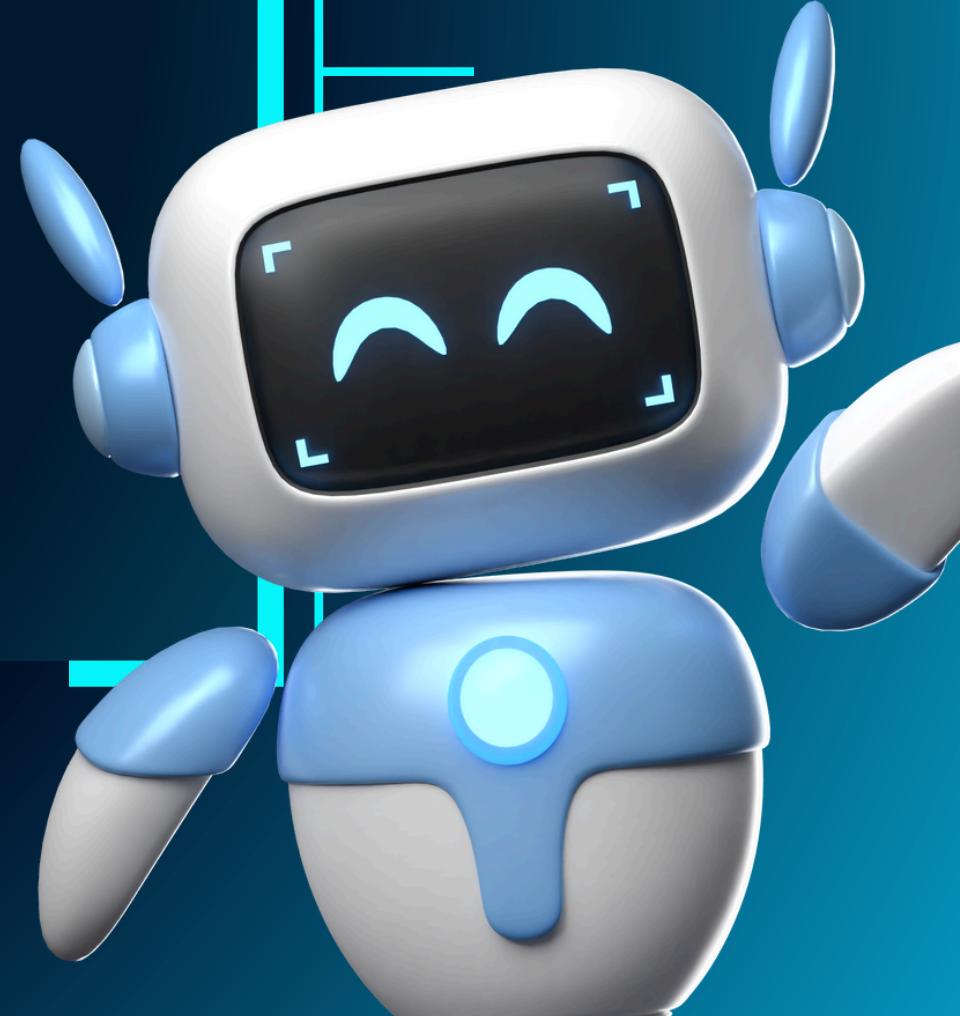
- ① Introduction
- ② Existing System
- ③ Technologies Used
- ④ Proposed System
- ⑤ Architecture of the system
- ⑥ Implementation
- ⑦ Output
- ⑧ Conclusion and Future Enhancement



INTRODUCTION

The "PDF GENIE: INTERACTIVE Q&A WITH ADVANCED AI INTEGRATION" project introduces a transformative approach to interacting with PDF documents.

The chatbot processes user queries, retrieves relevant information from the PDF content, and provides accurate responses, thereby enhancing document accessibility and efficiency in information retrieval.

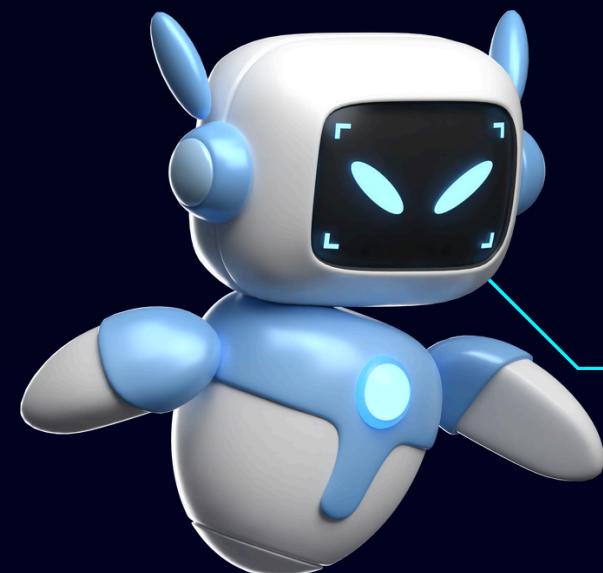


②

EXISTING SYSTEM



MANUAL SEARCH AND READ WHICH CAN BE VERY CUMBERSOME ESPECIALLY ON LARGE DOCUMENTS

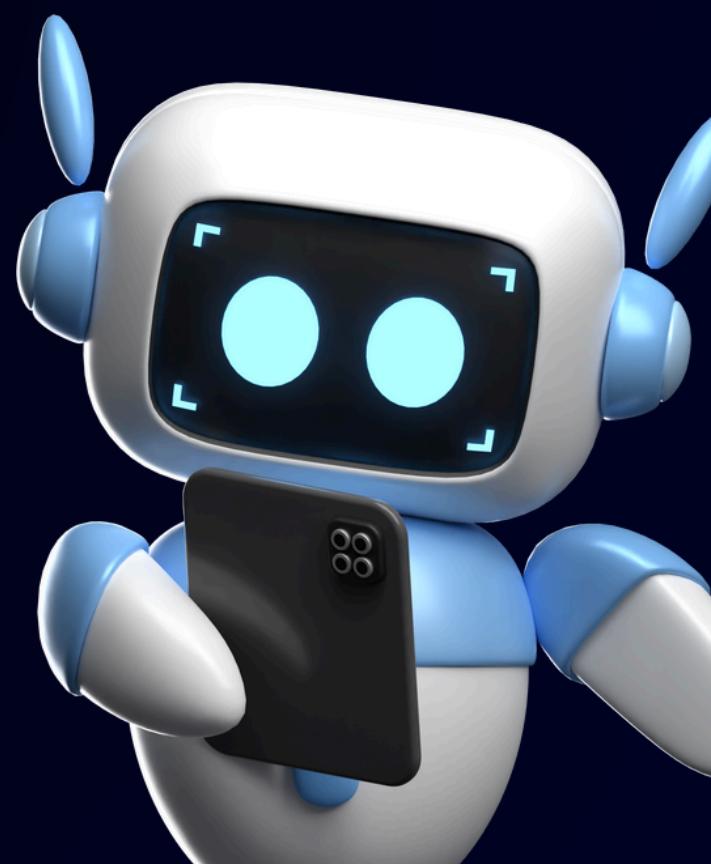


TRADITIONAL SEARCH FUNCTIONS DO NOT SEEM TO RECOGNIZE THE CONTEXT NECESSARY TO UNDERSTAND ADVANCED REQUESTS

3

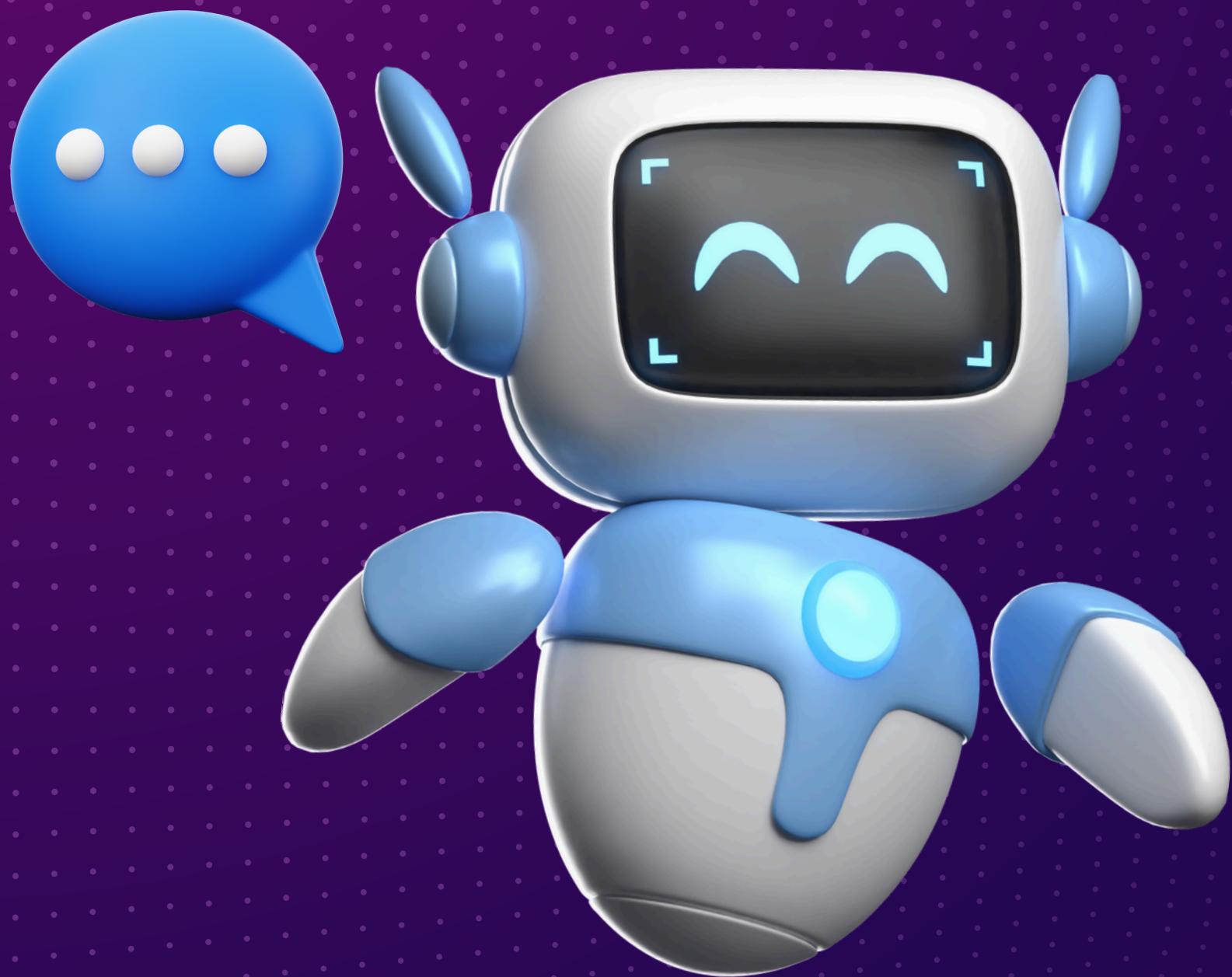
TECHNOLOGIES USED

- LARGE LANGUAGE MODEL**
 - Summarize lengthy documents
- LANGCHAIN**
 - Language Model Framework
- PINECONE DB**
 - Vector Database Service
- VERCEL AI SDK**
 - Rapid Deployment Toolkit
- NEXT JS**
 - React JS Framework



4

PROPOSED SYSTEM



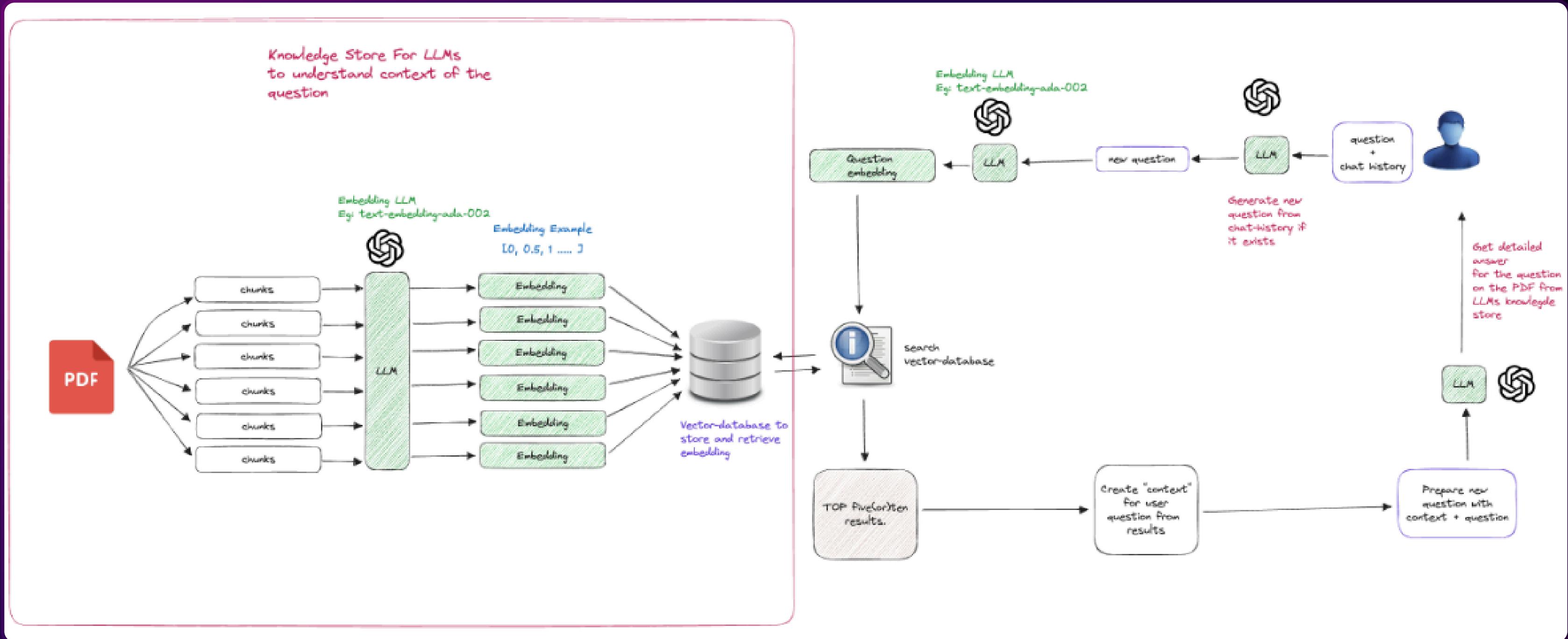
USER INTERFACE AND PDF
PROCESSING

KNOWLEDGE STORE AND
RETRIEVAL

INTELLIGENT INTERACTION AND
RESPONSE GENERATION

5

ARCHITECTURE OF THE SYSTEM



6

IMPLEMENTATION

USER INTERFACE MODULE

It guarantees an easy and beautiful interface for the users who upload PDF files and deal with the chatbot. Using the Shadcn UI component library, the interface is silky smooth and thoroughly optimized, creating a more engaging and efficient use of the system.

PDF PROCESSING MODULE

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. The efficiency of parsing and conversion is addressed.

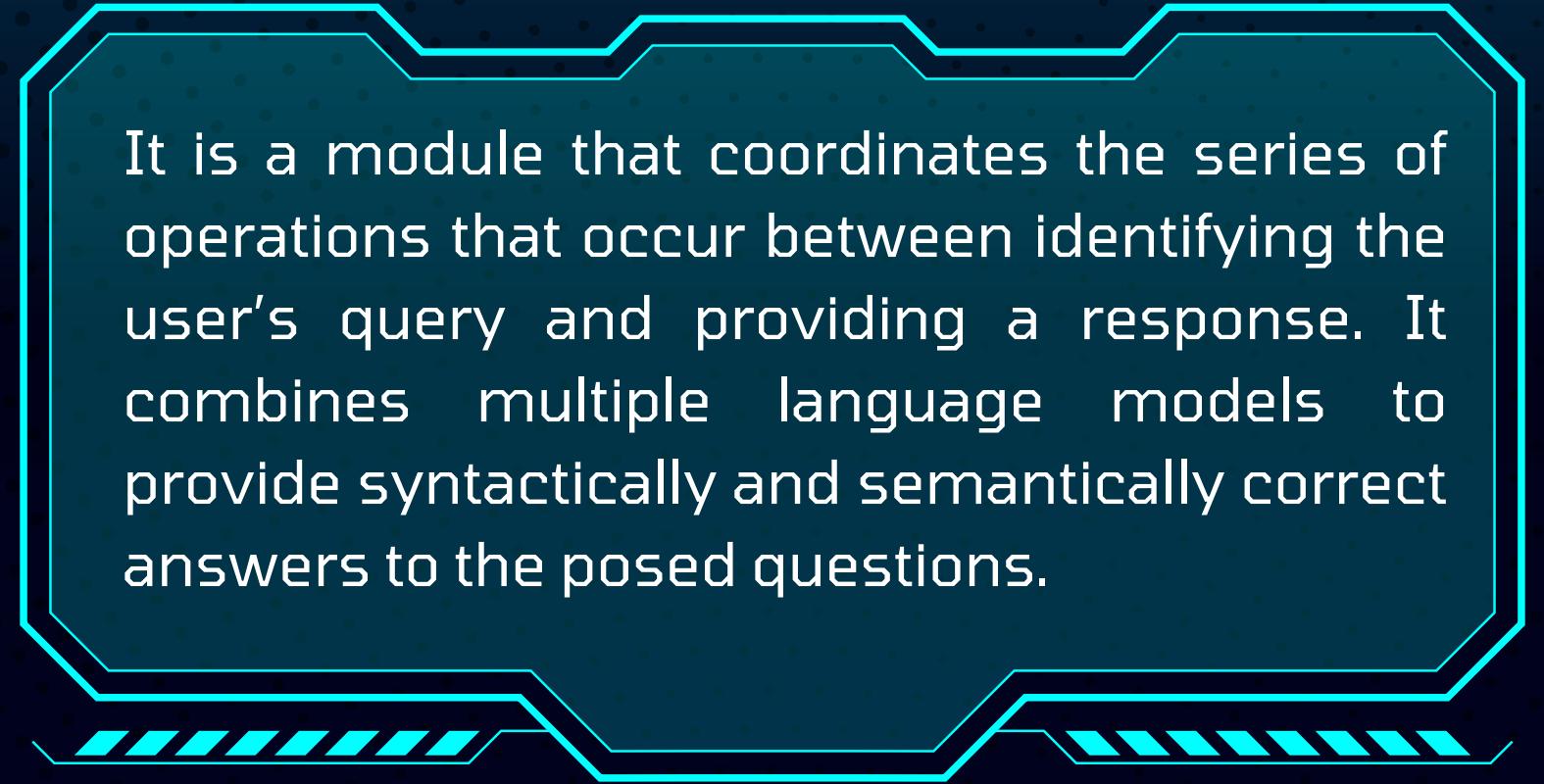
KNOWLEDGE STORE MODULE

It 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.

CHATBOT INTERACTION MODULE

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.

LANGCHAIN INTEGRATION MODULE



It 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.

7

OUTPUT

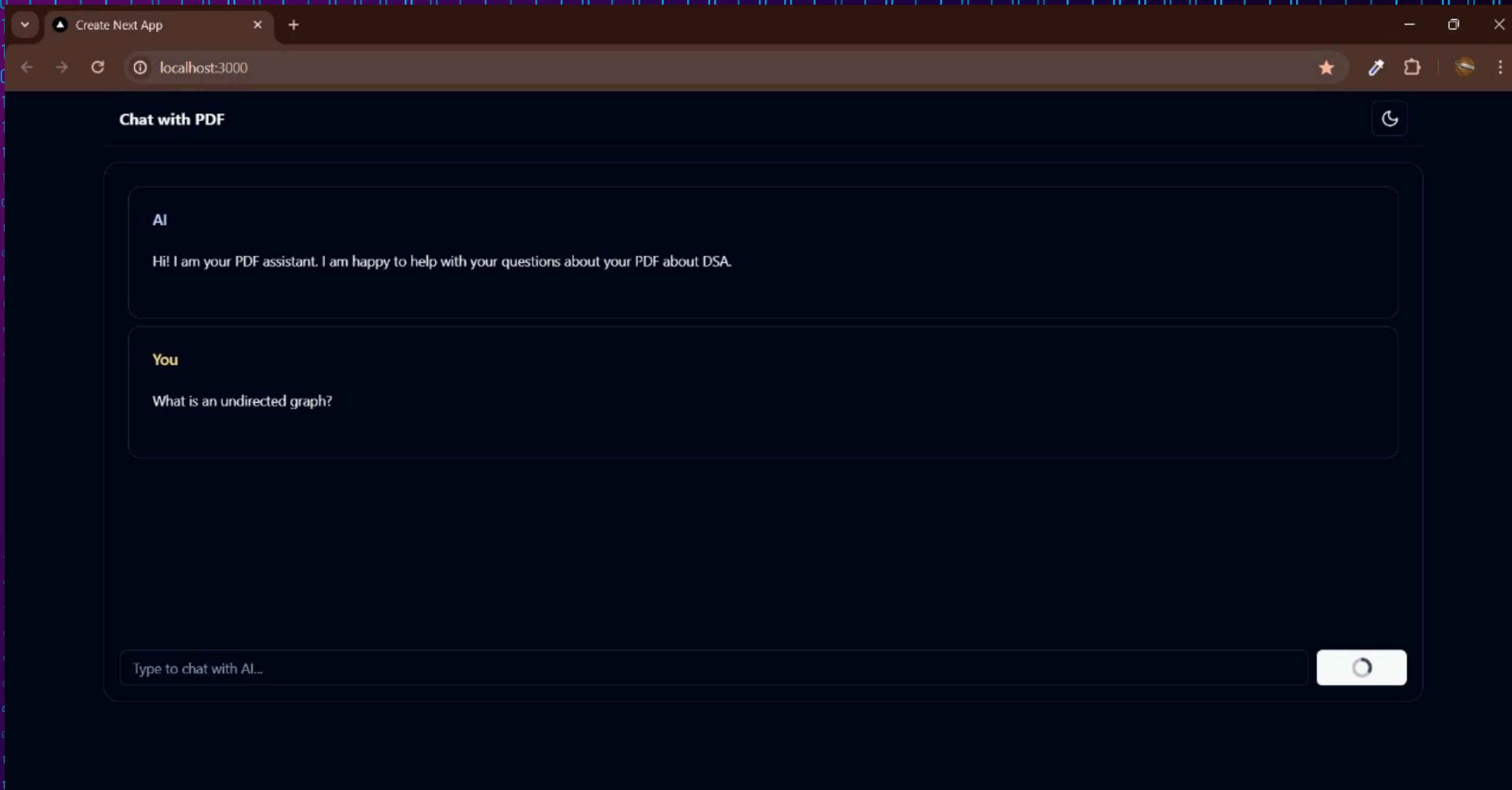
```

File Edit Selection View Go Run ... ← → Chat with PDF powershell + ...
EXPLORER PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
CHAT WITH PDF
.next
.assets
.docs
.node_modules
.public
.src
.env
.env.example
.eslintrc.json
.gitignore
.components.json
.LICENSE
.next-env.d.ts
.next.config.js
.package-lock.json
.packagejson
.postcss.config.cjs
.READMD
.tailwind.config.cjs
.tsconfig.json

PS D:\Developer Hub\Chat with PDF> npm run prepare:data
> langchain-text-ai@0.1.0 prepare:data
> tsx -r dotenv/config ./src/scripts/pinecone-prepare-docs.ts
[WARNING]: Importing from "langchain/embeddings/openai" is deprecated.
Instead, please add the "@langchain/openai" package to your project with e.g.
$ npm install @langchain/openai
and import from "@langchain/openai".
This will be mandatory after the next "langchain" minor version bump to 0.2.
[WARNING]: Importing from "langchain/vectorstores/pinecone" is deprecated.
Instead, please add the "@langchain/pinecone" package to your project with e.g.
$ npm install @langchain/pinecone
and import from "@langchain/pinecone".
This will be mandatory after the next "langchain" minor version bump to 0.2.
Preparing chunks from PDF file
Warning: TT: undefined function: 32
Loading 33 chunks into pinecone...
Data embedded and stored in pine-cone index
PS D:\Developer Hub\Chat with PDF>

```

> OUTLINE
> TIMELINE



User Prompt and Query Processing

Chat with PDF

You

What is an undirected graph?

AI

An undirected graph is a graph where the edges do not have a direction associated with them. If (v, w) is an undirected edge, then (v, w) is the same as (w, v) .

Source 1

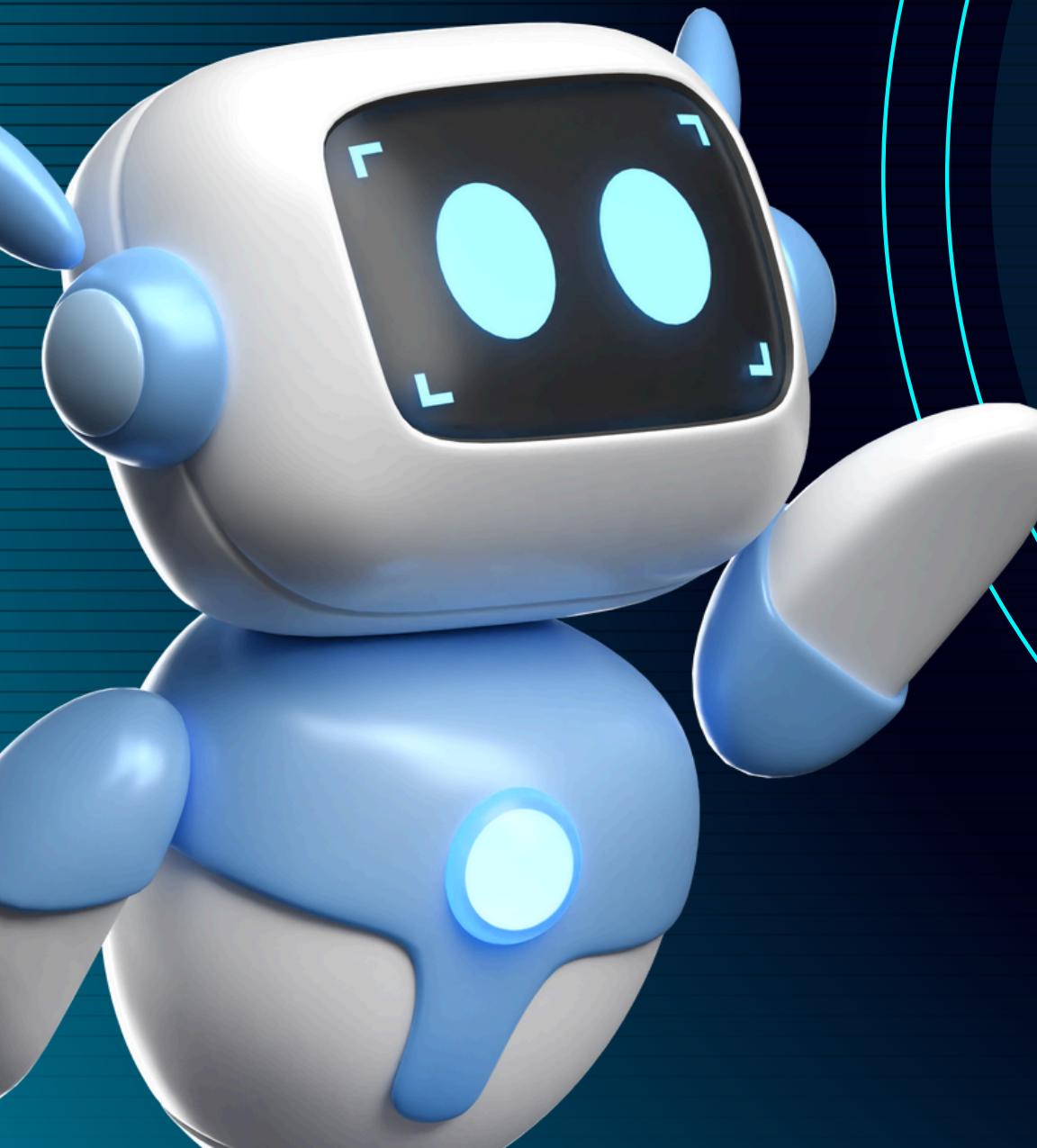
B.BHUVANESWARAN | AP (SG) | CSE | Rajalakshmi Engineering College 3 CHAPTER1 - GRAPHS 1.1 INTRODUCTION A graph $G = (V, E)$ consists of a set of vertices, V , and a set of edges, E . Vertices are referred to as nodes. Each edge is a pair (v, w) , where $v, w \in V$ i.e. $v = V_1, w = V_2$. Edges are sometimes referred to as arcs. Example Here V_1, V_2, V_3 and V_4 are vertices and $(V_1, V_2), (V_1, V_3), (V_1, V_4), (V_2, V_3), (V_2, V_4)$ and (V_3, V_4) are edges. 1.2 BASIC TERMINOLOGIES 1.2.1 Directed Graph (or) Digraph Directed graph is a graph which consists of directed edges, where each edge in E is unidirectional. It is also referred as Digraph. If (v, w) is a directed edge then $(v, w) \neq (w, v)$. 1.2.3 Undirected Graph An undirected graph is a graph, which consists of undirected edges. If (v, w) is an undirected edge then $(v, w) = (w, v)$.

Source 2

Type to chat with AI...

Ask

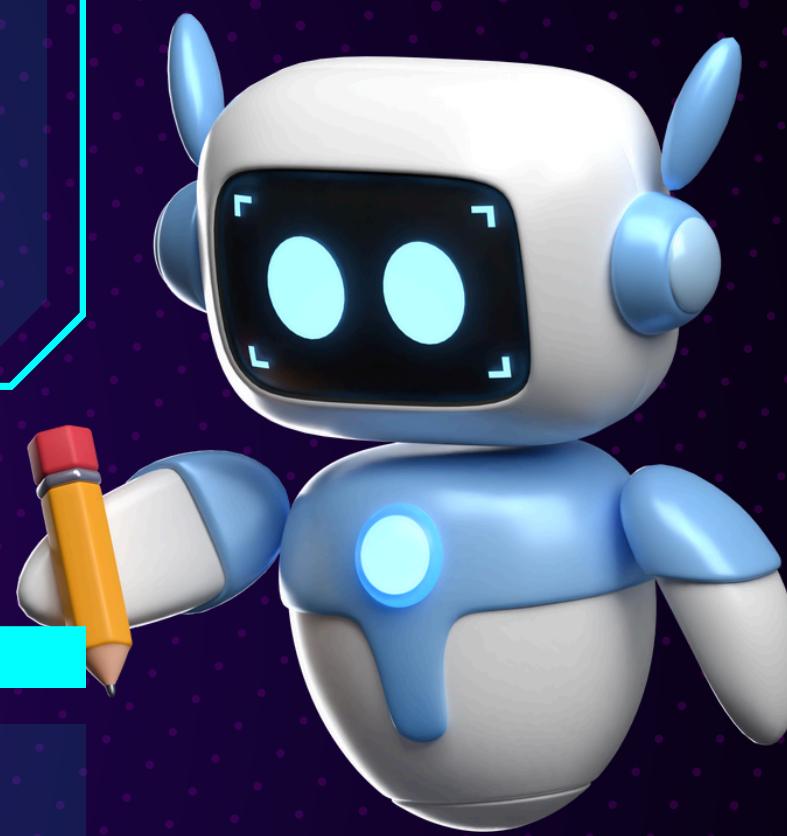
Intelligent Interaction and Response Generation



CONCLUSION & FUTURE ENHANCEMENT

The "PDF GENIE" project leverages AI, including LLMs, LangChain, PineconeDB, and Vercel AI SDK, to transform static PDFs into interactive tools, enhancing information retrieval and usability across various fields.

Future enhancements for "PDF GENIE" include multi-language support, advanced NLU capabilities, OCR integration for scanned documents, personalized responses, improved cloud integration, and enhanced security measures for user data protection.





THANK YOU