AI-Driven Predictive Diagnosis Chatbot for Healthcare

By:

Sonam Bhul MITU21BTIT0058
Ayushi Tiwari MITU21BTIT0018
Tanay Kumar MITU21BTIT0060

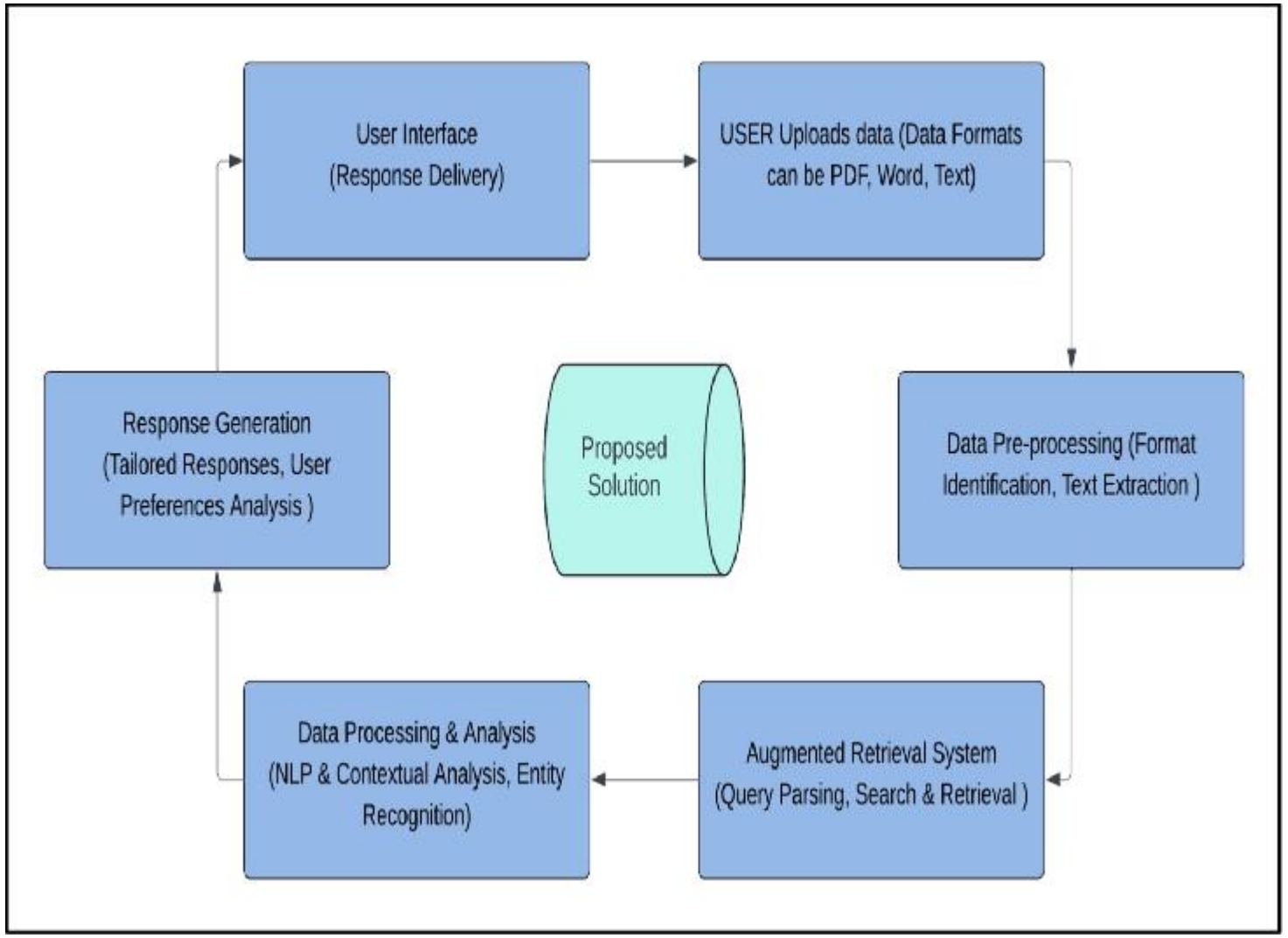
Guided By:
Prof. Dr. Mohit Kumar

AI-Driven Predictive Diagnosis Chatbot for Healthcare

Proposed Solution:

- Users upload medical documents in various formats.
- The system processes and extracts relevant medical information from these documents.
- Extracted data is indexed and organized into a searchable knowledge base.
- When a user queries the system, it retrieves pertinent information from the indexed data.
- A lightweight Large Language Model (LLM) generates a clear, concise response based on the retrieved information.
- The system delivers personalized and accurate medical guidance efficiently.

TECHNICAL APPROACH



Literature survey

YEAR	TITLE	AUTHOR	RESULT	LIMITATION
2024	Leveraging LLM: Implementing an Advanced AI Chatbot for Healthcare"	Ajinkya Mhatre, Sandeep R. Warhade, Sayali Kokate, Omkar Pawar, Samyak Jain	- Result: LLM-based chatbots showed 61% accuracy in general illness queries.	- Issues with accuracy, biases, and ethical concerns remain.
2024	A Medical Chatbot: Your Healthcare Assistance	Harsh Jain	- Llama 2-based ChatBot delivers accurate medical info and improves access.	- Needs more refinement for better functionality.
2024	Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, RAG, and Performance-Optimized LLM Fusion	S. Vidivelli, Manikandan Ramachandran, A. Dharunbalaj	- The chatbot efficiently delivers healthcare information using advanced technologies.	- Requires improvements in accuracy and user interaction.
2024	Integrating RAG with LLMs in Nephrology: Advancing Practical Applications	Jing Miao, Charat Thongprayoon, Supawadee Suppadungsuk,Oscar A. Garcia Valencia, Wisit Cheungpasitporn	- LLMs with RAG improve nephrology care and education.	- Accuracy and reliability of information are still challenging.

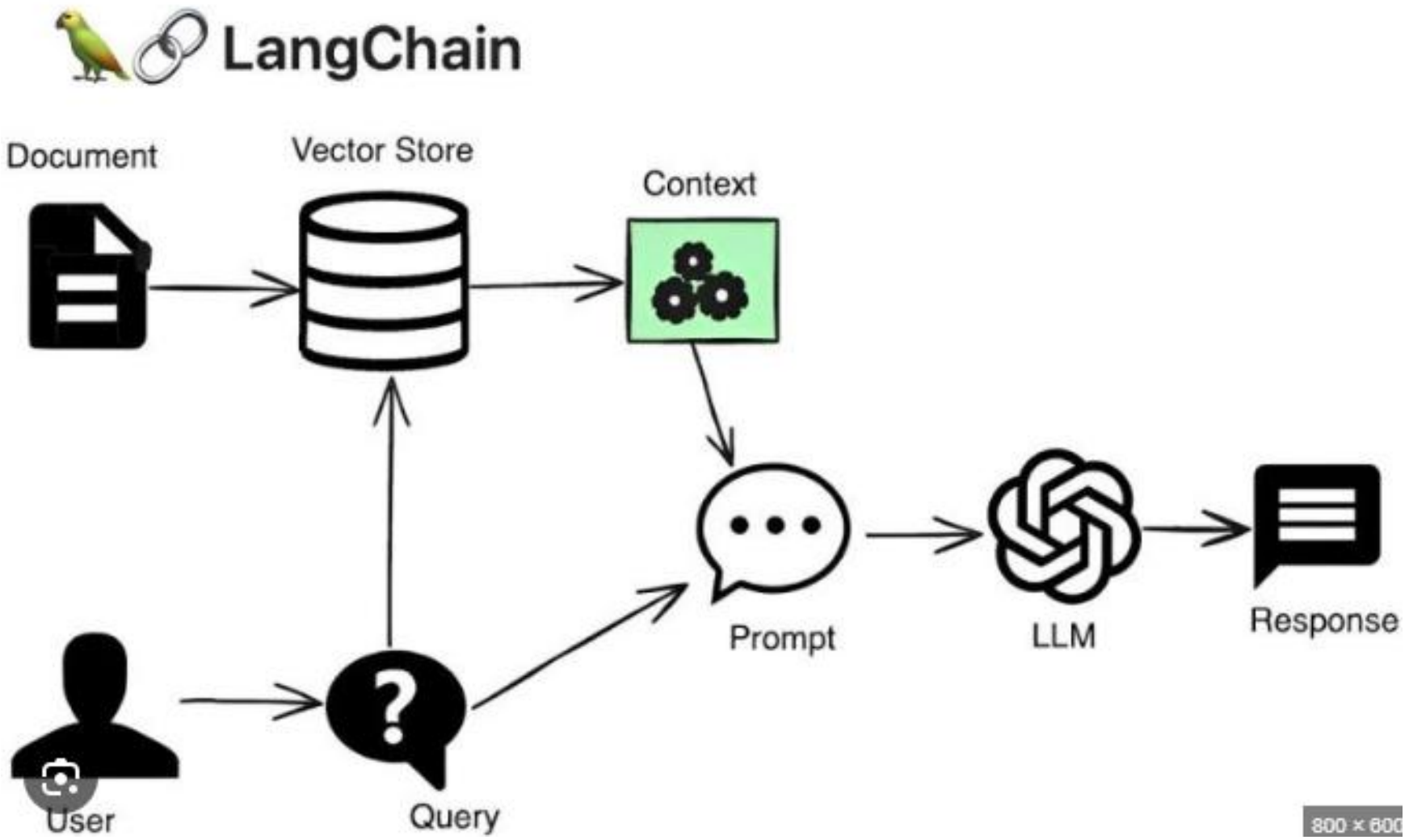
FEASIBILITY AND VIABILITY

- **Data Availability and Quality**: The effectiveness of AI systems heavily depends on the quality and quantity of data. For predictive diagnosis, having access to comprehensive and high-quality patient data (including medical history, lab results, and imaging) is crucial. Ensuring data privacy and compliance with regulations like HIPAA (in the U.S.) is also essential.
- **Integration with Existing Systems**: The AI system must be compatible with current healthcare IT infrastructures, such as Electronic Health Records (EHR) systems. Seamless integration is necessary to ensure that predictions are easily accessible to healthcare professionals and can be incorporated into clinical workflows without disrupting existing practices.

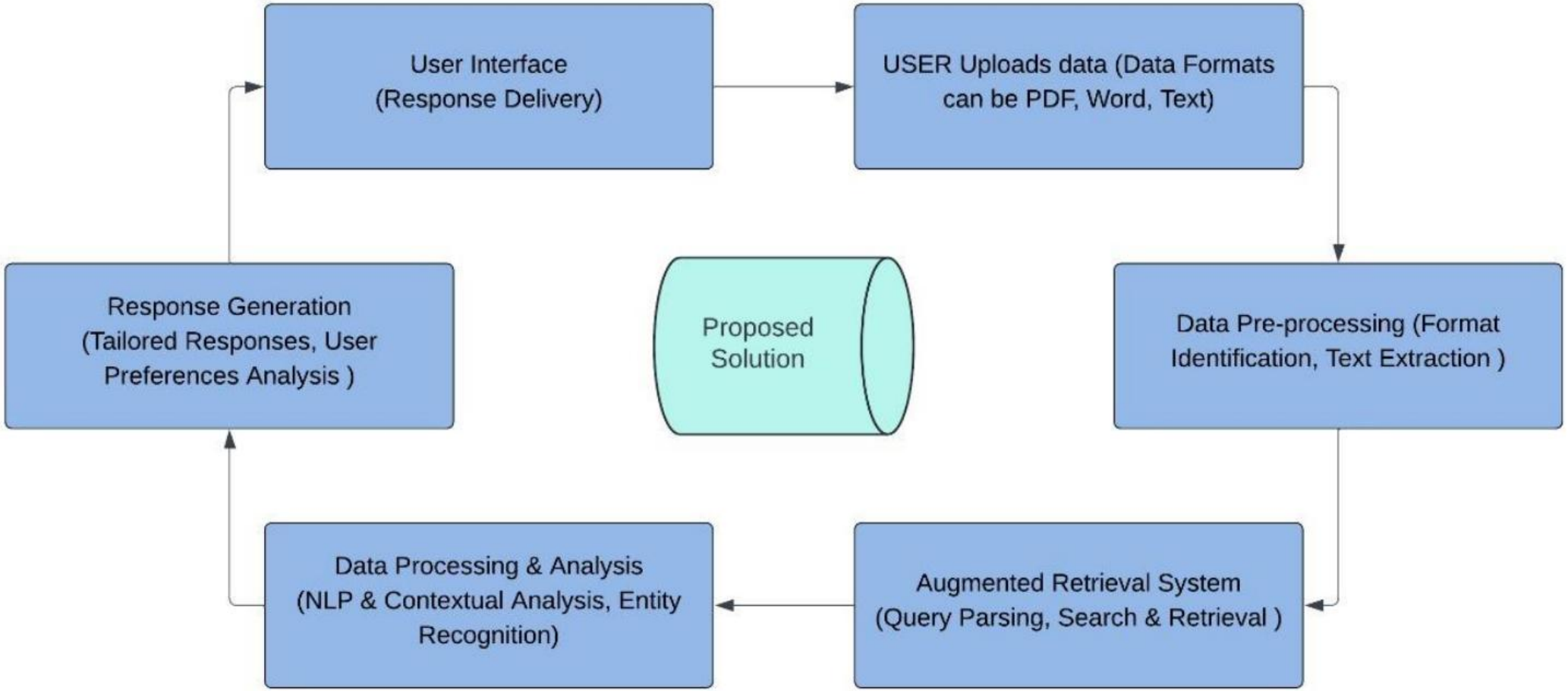
Potential Challenges and Risks:

- Data Privacy and Security
- Data Quality and Bias
- Symptom analysis, medication recommendations
- Reduces misinformation risk.
- Provides prompt healthcare guidance for better outcomes.
- Targeting multiple individuals with easy of access to professional medical advice.

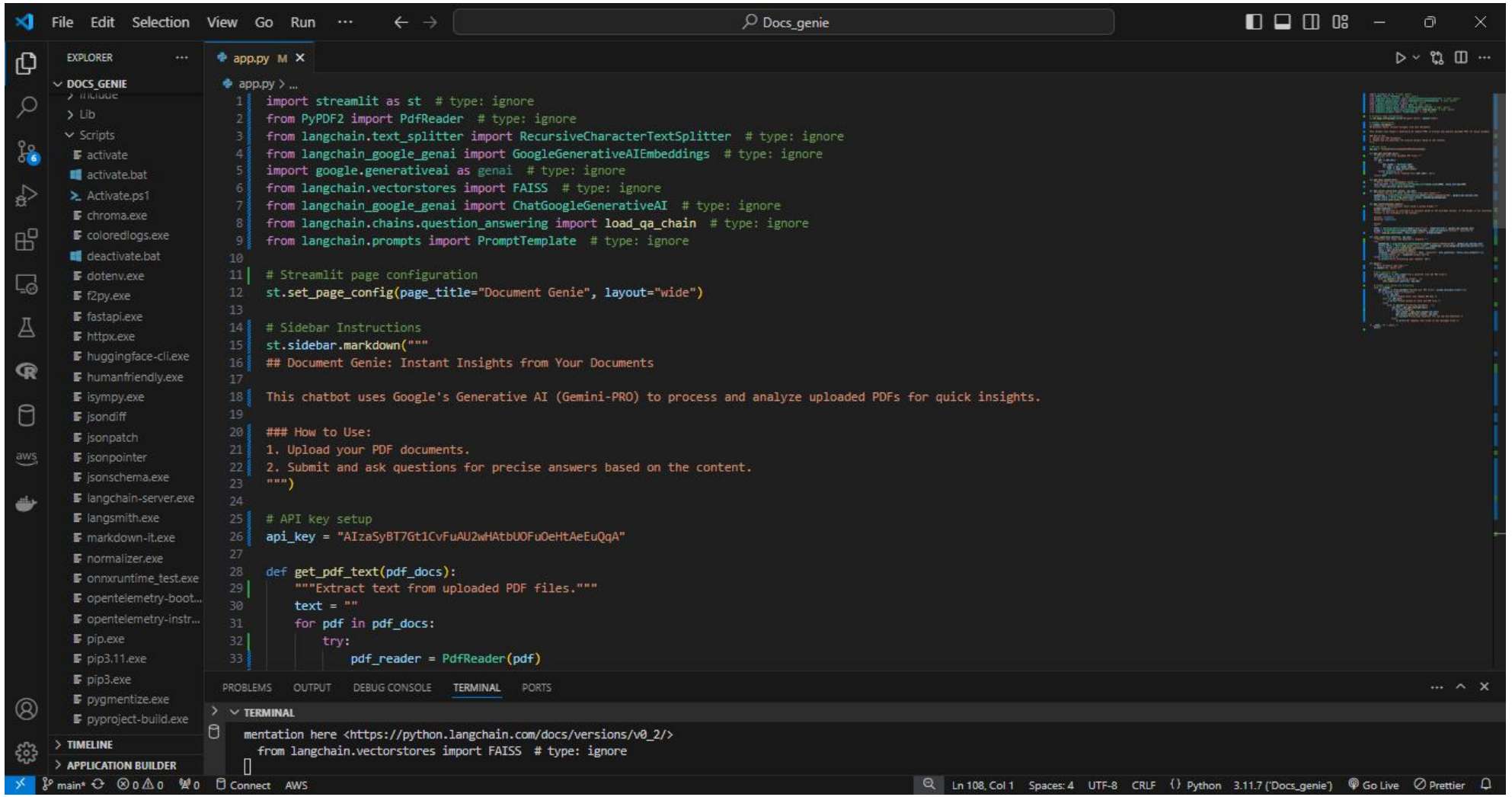
langchain Architecture



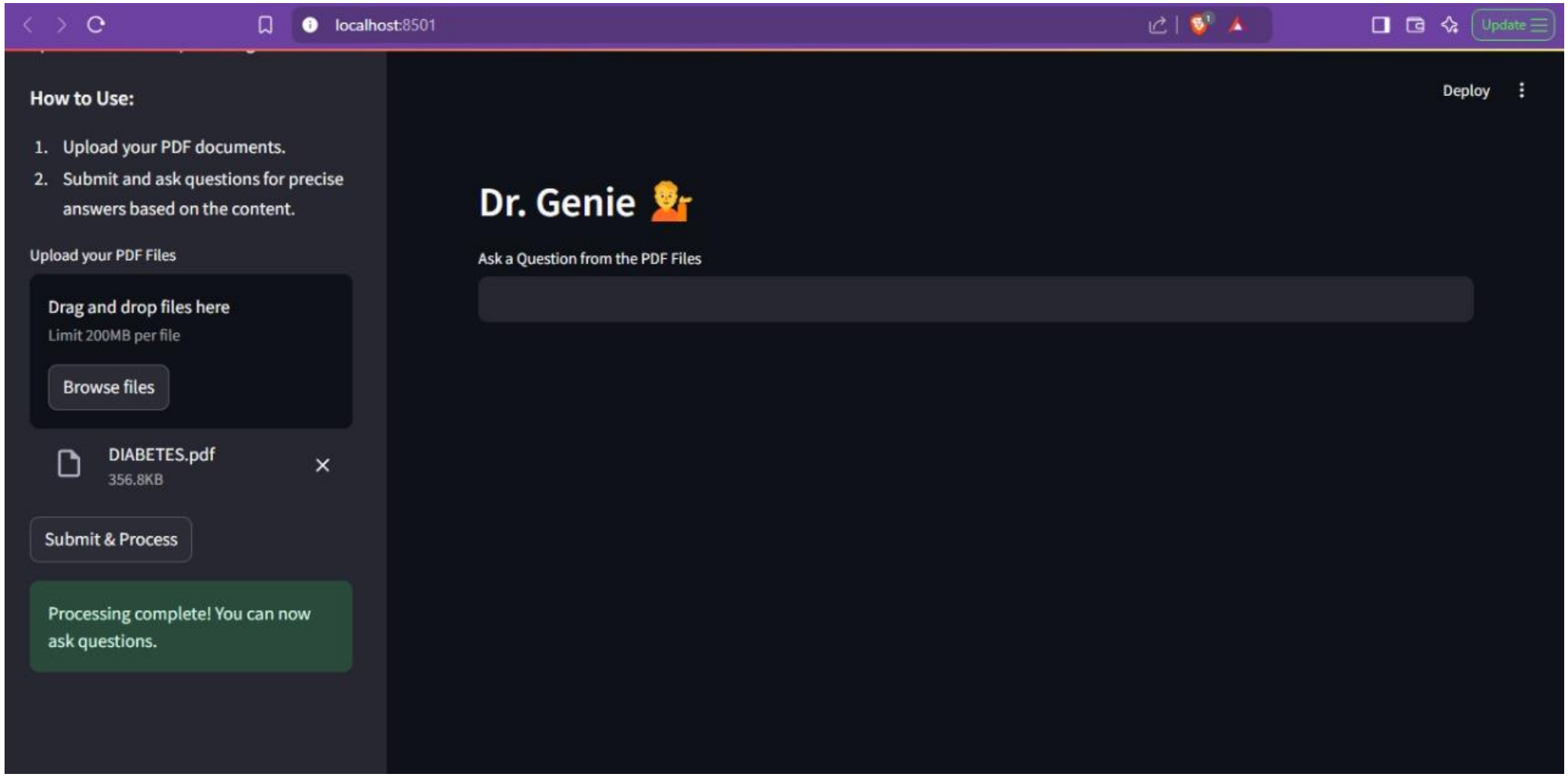
PROPOSED SOLUTION



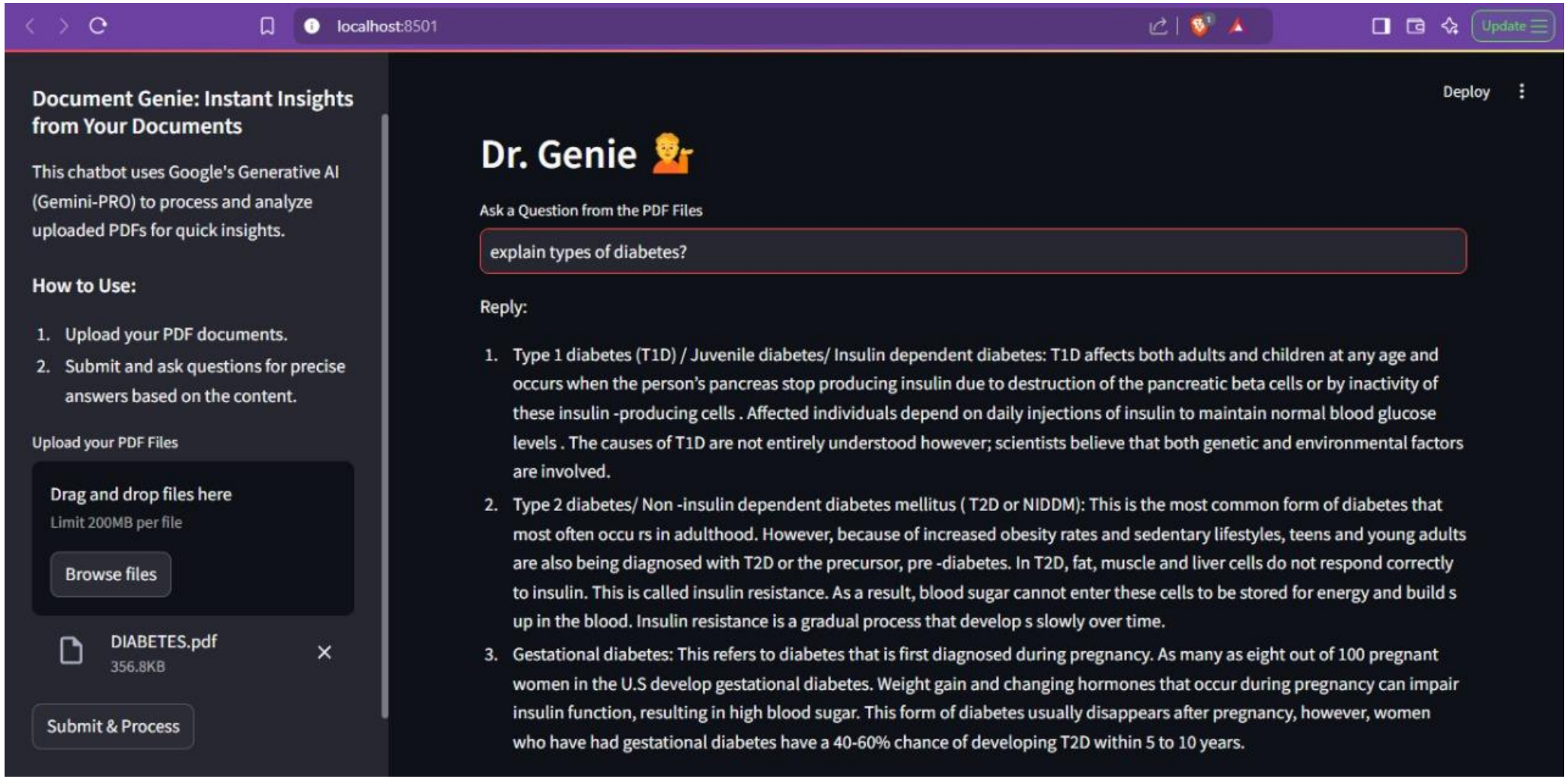
CODE


File Edit Selection View Go Run
Docs_genie
EXPLORER
DOCS_GENIE
app.py M
app.py > ...
import streamlit as st # type: ignore
from PyPDF2 import PdfReader # type: ignore
from langchain.text_splitter import RecursiveCharacterTextSplitter # type: ignore
from langchain_google_genai import GoogleGenerativeAIEmbeddings # type: ignore
import google.generativeai as genai # type: ignore
from langchain.vectorstores import FAISS # type: ignore
from langchain_google_genai import ChatGoogleGenerativeAI # type: ignore
from langchain.chains.question_answering import load_qa_chain # type: ignore
from langchain.prompts import PromptTemplate # type: ignore
# Streamlit page configuration
st.set_page_config(page_title="Document Genie", layout="wide")
# Sidebar Instructions
st.sidebar.markdown("""
Document Genie: Instant Insights from Your Documents
This chatbot uses Google's Generative AI (Gemini-PRO) to process and analyze uploaded PDFs for quick insights.
How to Use:
1. Upload your PDF documents.
2. Submit and ask questions for precise answers based on the content.
""")
# API key setup
api_key = "AIzaSyBT7Gt1CvFuAU2wHAtbUOFuOeHtAeEuQqA"
def get_pdf_text(pdf_docs):
"""Extract text from uploaded PDF files."""
text = ""
for pdf in pdf_docs:
try:
pdf_reader = PdfReader(pdf)
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
TERMINAL
mentation here <https://python.langchain.com/docs/versions/v0_2/>
from langchain.vectorstores import FAISS # type: ignore
main* Connect AWS
Ln 108, Col 1 Spaces: 4 UTF-8 CRLF Python 3.11.7 ('Docs_genie') Go Live Prettier

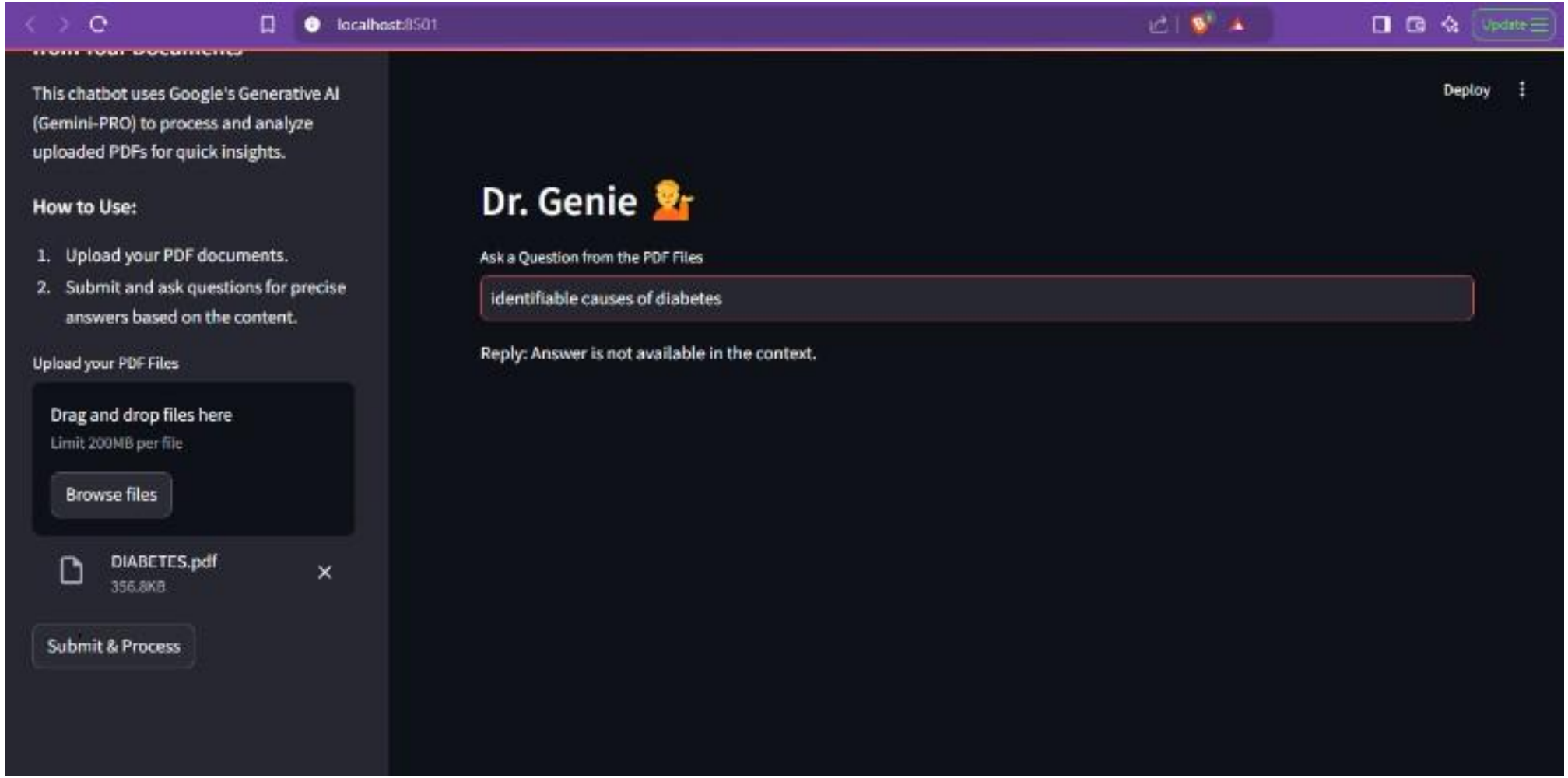
OUTPUT


localhost:8501
Update
How to Use:
1. Upload your PDF documents.
2. Submit and ask questions for precise answers based on the content.
Upload your PDF Files
Drag and drop files here
Limit 200MB per file
Browse files
DIABETES.pdf
356.8KB
Submit & Process
Processing complete! You can now ask questions.
Deploy
Dr. Genie
Ask a Question from the PDF Files

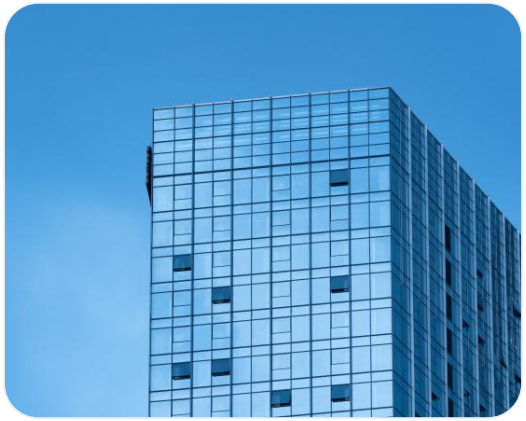
OUTPUT


localhost:8501
Update
Deploy
Document Genie: Instant Insights from Your Documents
This chatbot uses Google's Generative AI (Gemini-PRO) to process and analyze uploaded PDFs for quick insights.
How to Use:
1. Upload your PDF documents.
2. Submit and ask questions for precise answers based on the content.
Upload your PDF Files
Drag and drop files here
Limit 200MB per file
Browse files
DIABETES.pdf
356.8KB
Submit & Process
Dr. Genie
Ask a Question from the PDF Files
explain types of diabetes?
Reply:
1. Type 1 diabetes (T1D) / Juvenile diabetes/ Insulin dependent diabetes: T1D affects both adults and children at any age and occurs when the person's pancreas stop producing insulin due to destruction of the pancreatic beta cells or by inactivity of these insulin -producing cells . Affected individuals depend on daily injections of insulin to maintain normal blood glucose levels . The causes of T1D are not entirely understood however; scientists believe that both genetic and environmental factors are involved.
2. Type 2 diabetes/ Non -insulin dependent diabetes mellitus (T2D or NIDDM): This is the most common form of diabetes that most often occu rs in adulthood. However, because of increased obesity rates and sedentary lifestyles, teens and young adults are also being diagnosed with T2D or the precursor, pre -diabetes. In T2D, fat, muscle and liver cells do not respond correctly to insulin. This is called insulin resistance. As a result, blood sugar cannot enter these cells to be stored for energy and build s up in the blood. Insulin resistance is a gradual process that develop s slowly over time.
3. Gestational diabetes: This refers to diabetes that is first diagnosed during pregnancy. As many as eight out of 100 pregnant women in the U.S develop gestational diabetes. Weight gain and changing hormones that occur during pregnancy can impair insulin function, resulting in high blood sugar. This form of diabetes usually disappears after pregnancy, however, women who have had gestational diabetes have a 40-60% chance of developing T2D within 5 to 10 years.

OUTPUT


localhost:8501
Update
This chatbot uses Google's Generative AI (Gemini-PRO) to process and analyze uploaded PDFs for quick insights.
How to Use:
1. Upload your PDF documents.
2. Submit and ask questions for precise answers based on the content.
Upload your PDF Files
Drag and drop files here
Limit 200MB per file
Browse files
DIABETES.pdf
356.8KB
Submit & Process
Deploy
Dr. Genie
Ask a Question from the PDF Files
identifiable causes of diabetes
Reply: Answer is not available in the context.

FEASIBILITY AND VIABILITY



Potential Challenges and Risks:

Presentations are communication tools that can be used as demonstrations.

Data Privacy and Security

Presentations are communication tools that can be used as demonstrations.

Data Quality and Bias

Presentations are communication tools that can be used as demonstrations.

Your title here

Symptom analysis, medication recommendations

IMPACT AND BENEFITS

Early Detection of Health Issues:
Benefit: Identifies potential health problems before they become serious.
Outcome: Enables earlier treatment, potentially preventing severe conditions.
Personalized Treatment:
Benefit: Moves away from a one-size-fits-all approach.
Outcome: Tailors treatment plans based on individual health data, improving patient care.
Economic Savings:
Benefit: Reduces the need for expensive treatments and hospital stays.
Outcome: Lowers healthcare costs, reduces missed workdays, and boosts productivity.
Environmental Benefits:
Benefit: Minimizes unnecessary tests and treatments.
Outcome: Reduces medical waste and, if remote monitoring is used, decreases travel to healthcare facilities.
Improved Quality of Life:
Benefit: Healthier individuals with fewer severe health issues.
Outcome: Enhances overall well-being and quality of life.

FUTURE SCOPE

1. Enhanced Diagnostic Accuracy: Achieve greater precision by integrating real-time data streams and training the system in specialized medical domains.

2. Integration of Multiple LLMs: Leverage multiple lightweight Large Language Models for improved context understanding and diverse query handling.

3. Chat History Functionality: Provide users with access to past queries and responses for better continuity and reference.

4. Support for Multiple Input Formats: Enable compatibility with various file types to streamline data ingestion and processing.

RESEARCH AND REFERENCES

1. Leveraging LLM: Implementing an Advanced AI Chatbot for Healthcare, Ajinkya Mhatre, Sandeep R. Warhade, Sayali Kokate, Omkar awar, Samyak Jain, May 2024
https://ijisrt.com/assets/upload/files/IJISRT24MAY1964.pdf

2. A Medical Chatbot: Your Healthcare Assistance, Harsh Jain, June 2024
https://www.irjmets.com/uploadedfiles/paper//issue_6_june_2024/58850/final/fin_irjmets1717770049.pdf

3. Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, S. Vidivelli, Manikandan Ramachandran, A. Dharunbalaji, August 2024
https://file.techscience.com/files/cmc/2024/TSP_CMC-80-2/TSP_CMC_54360/TSP_CMC_54360.pdf

4. Integrating RAG with LLMs in Nephrology: Advancing Practical Applications, Jing Miao, Charat Thongprayoon, Supawadee Suppadungsuk, Oscar A. Garcia Valencia, Wisit Cheungpasitporn, March 2024
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10972059/

5. Health-LLM: Personalized Retrieval-Augmented Disease Prediction Model, Mingyu Jin, Qinkai Yu, Chong Zhang, Dong Shu, February 2024
https://www.researchgate.net/publication/377968117_Health_x0002_LLM_Personalized_Retrieval-Augmented_Disease_Prediction_Model

Thank You !!!