Project Report:

Project Title:

Medical Diagnosis and Treatment Assistance System by RAG Model

Introduction:

The Medical Diagnosis and Treatment Assistance System is an AI-driven solution designed to assist medical professionals and patients by providing accurate diagnoses and treatment recommendations. This system utilizes LangChain, advanced natural language processing models like Llama 3.2, and document processing technologies to analyze medical texts and deliver evidence-based insights.

Objectives

1. Streamline Medical Analysis: To enable automated extraction and summarization of medical data from documents like research papers and PDFs.
2. Provide Accurate Responses: Leverage AI to diagnose illnesses and suggest treatments based on medical contexts.
3. Efficient Data Retrieval: Build a retriever system for quick access to specific medical information.
4. Improve Accessibility: Offer an easy-to-use interface for patients and healthcare professionals.

Tools and Technologies

- Programming Language: Python
- Libraries and Frameworks:
    - LangChain: For chaining prompts, models, and document loaders.
    - PyPDF2 and PyMuPDF: For PDF processing.

- o DocArray: For embedding and vector storage.
- o RapidOCR: For extracting text from images in PDFs.
- Model: Llama 3.2 via langchain_ollama.
- Vector Store: DocArrayInMemorySearch for semantic search and similarity-based document retrieval.
-

Implementation Details

1. Document Loading and Processing

- Used PyPDFLoader and PyMuPDFLoader to load medical research papers in PDF format.
- Extracted and split text into manageable chunks (500 characters with 50-character overlap) using RecursiveCharacterTextSplitter.

2. Prompt Engineering

- Designed a professional PromptTemplate to simulate an experienced medical specialist's response.
- Ensured the AI answers with accuracy and compassion, avoiding disclaimers like "I'm not a medical professional."

3. Embedding and Vector Storage

- Created embeddings using OllamaEmbeddings for semantic representation.
- Stored document chunks in DocArrayInMemorySearch to facilitate similarity-based retrieval.

4. Retrieval-Augmented Generation (RAG) System

- Integrated a retriever mechanism to fetch relevant information dynamically based on user queries.
- Combined context with questions for enhanced AI response accuracy.

5. AI Model Integration

- Utilized Llama 3.2 for generating human-like responses to medical queries.
- Chain architecture (PromptTemplate | Llama Model | StrOutputParser) ensured smooth processing and output.

Workflow Diagram

1. Input: Medical PDF → Extract Context → Chunk Text
2. Query: User Question → AI Prompt
3. Processing: Context + Question → AI Model → Generate Response
4. Output: Accurate and professional answer with evidence-based recommendations.

Results

- Successfully processed a comprehensive medical research paper titled *"Medical Diagnosis and Treatment Methods in Basic Medical Sciences"*.
- Retrieved contextually relevant answers to questions like:
    - "What are the common symptoms of cancer?"
    - "Histopathological findings in specific diseases."
- Demonstrated the system's ability to provide detailed, evidence-backed responses in a conversational manner.

Key Features

1. Document Summarization: Extracts key insights from large volumes of text.
2. Intelligent Query Resolution: Combines context with AI reasoning to answer user queries effectively.
3. Customizable Prompts: Adapts to different medical scenarios with tailored prompt templates.
4. Semantic Search: Retrieves the most relevant document sections using vector embeddings.

Challenges

1. Data Quality: Ensuring accurate text extraction from poorly scanned documents.
2. Scalability: Handling large datasets with high computational efficiency.

3. Model Limitations: Ensuring the AI avoids errors in sensitive medical contexts.

4. Conclusion

5. The **Medical Diagnosis and Treatment Assistance System** is a robust tool for medical data analysis and query resolution. It leverages cutting-edge AI technologies to simplify complex medical information and deliver professional, accurate, and user-friendly solutions. With further improvements, this system has the potential to become an indispensable resource for healthcare professionals and researchers.