**Cognitive Multimodal RAG for Medical Diagnosis**

**ABSTRACT**

Cognitive Multimodal RAG for Medical Diagnosis is an AI-powered system that integrates computer vision and natural language processing to analyze medical images and textual data, such as prescriptions and reports. Using Retrieval-Augmented Generation (RAG), the system retrieves relevant medical knowledge from a database and generates context-aware diagnoses. This approach enhances diagnostic accuracy, supports healthcare professionals with real-time insights, and improves patient outcomes. Additionally, features like interactive diagnosis validation, explainability, and dynamic knowledge base updates ensure adaptability and reliability in medical decision-making.

The system leverages multimodal learning to combine image-based and text-based medical data, enabling comprehensive analysis beyond traditional AI models. By incorporating clinician feedback into the diagnostic loop, it continuously improves accuracy and relevance. With applications in telemedicine, automated report generation, and medical assistance, this framework enhances efficiency in healthcare workflows. Its scalable and interpretable design makes it a promising solution for AI-assisted medical diagnosis and decision support.

**Project Domain**

The project falls under the domain of **Artificial Intelligence in Healthcare**, specifically focusing on **Medical Diagnosis and Decision Support Systems**. It combines **Computer Vision (CV), Natural Language Processing (NLP), and Retrieval-Augmented Generation (RAG)** to enhance diagnostic accuracy. The system integrates multimodal data (images and text) to provide AI-driven insights, supporting clinicians in medical decision-making.

**Problem Statement**

Traditional medical diagnosis often relies on separate analyses of images (X-rays, MRIs) and textual data (prescriptions, reports), leading to potential gaps in interpretation. Manual diagnosis is time-consuming, prone to human error, and limited by individual expertise. Additionally, existing AI models lack real-time knowledge retrieval, explainability, and adaptability to evolving medical information. This project aims to develop a **Cognitive Multimodal RAG-based Medical Diagnosis System** that efficiently analyzes both visual and textual medical data, retrieves relevant knowledge, and generates context-aware diagnoses. The system will improve diagnostic accuracy, reduce workload for healthcare professionals, and enhance patient outcomes through an AI-assisted, feedback-driven approach.

**Input, Output, and Metrics for Evaluation**

**Input:**

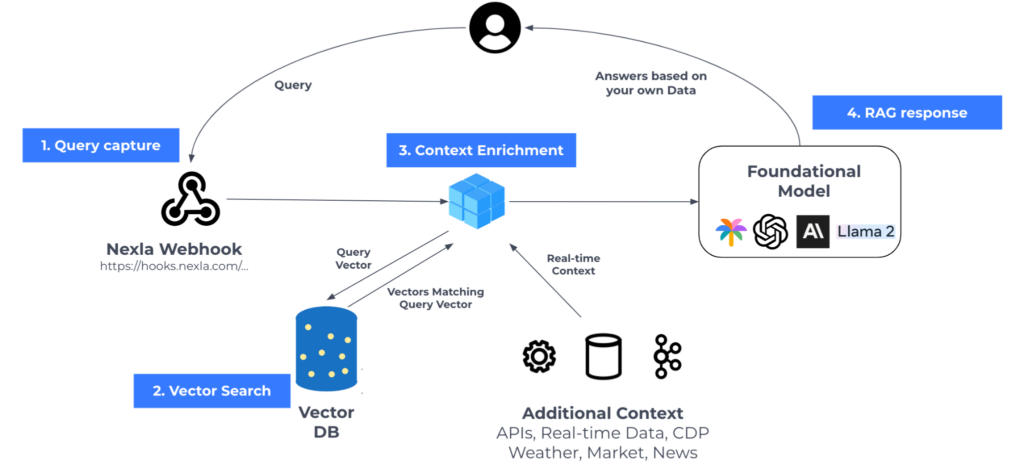
1. **Medical Images** (X-rays, MRIs, CT scans) – Processed using Computer Vision.
2. **Textual Data** (Prescriptions, Lab Reports, Doctor’s Notes) – Analyzed using NLP.
3. **Patient Information** (Age, Symptoms, Medical History) – Used for contextual understanding.
4. **Medical Knowledge Base** (Clinical Guidelines, PubMed Articles) – Retrieved for reference.

**Output:**

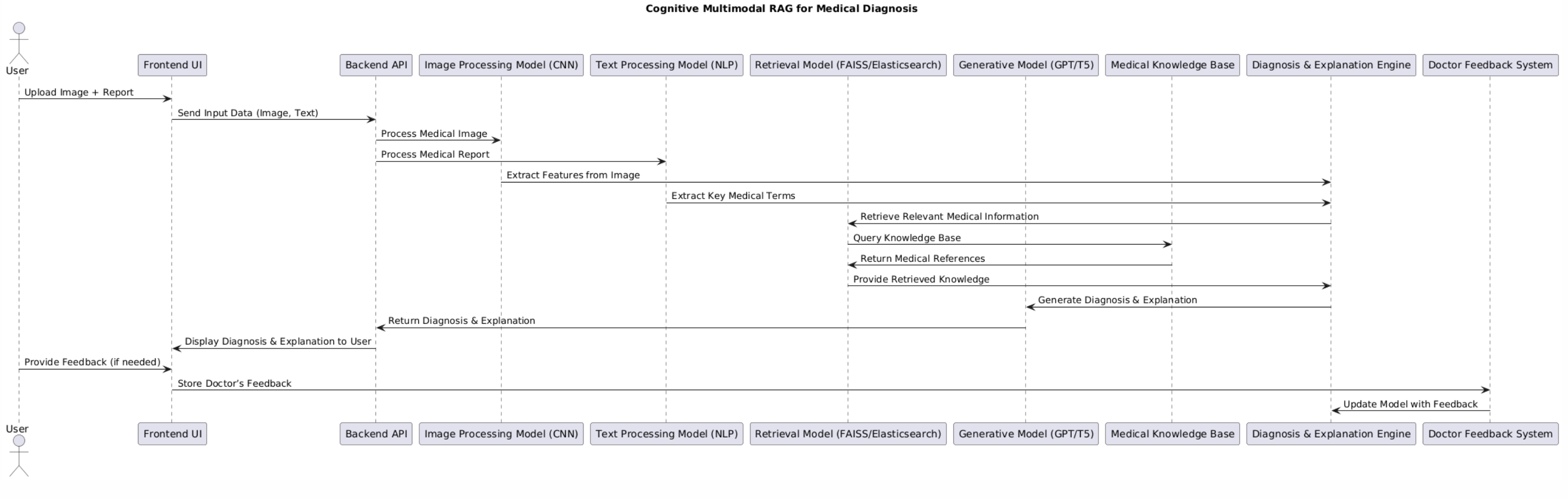
1. **Diagnosis Prediction** – AI-generated medical condition based on input data.
2. **Treatment Recommendations** – Suggested treatment plans based on retrieved medical knowledge.
3. **Explainability Features** – Highlighted key areas in images (via heatmaps) and relevant text references.
4. **Confidence Score** – Model-generated probability of the predicted diagnosis.
5. **Real-time Knowledge Retrieval** – Relevant medical guidelines retrieved dynamically.
6. **Feedback Mechanism** – Option for doctors to validate and refine model outputs.

**Evaluation Metrics:**

1. **Accuracy** – Percentage of correct diagnoses compared to actual medical reports.
2. **Precision & Recall** – Measures the correctness and completeness of the AI predictions.
3. **F1 Score** – Harmonic mean of precision and recall for balanced performance assessment.
4. **BLEU Score** – Evaluates text-generated explanations against expert-written medical reports.
5. **Mean Reciprocal Rank (MRR)** – Measures relevance ranking of retrieved knowledge sources.
6. **Heatmap Interpretability Score** – Evaluates how well the system highlights relevant areas in images.
7. **Latency** – Measures the response time of the AI system for real-time diagnosis.
8. **User Feedback Score** – Assesses acceptance and usability based on clinician ratings



**ARCHITECTURE**

****