# Gen AI Symptom Checker

## 1. Background

Medical misdiagnosis is a critical global issue with deadly consequences. The 2018 case of Elvina Naa Densua Mould, who died from undiagnosed malaria despite seeking hospital care, highlights how missed symptoms can be fatal. Studies show that about 12 million Americans suffer diagnostic errors annually, and 40,000–80,000 deaths in U.S. hospitals may be attributed to misdiagnosis. Emergency departments are often overcrowded, with up to 50% of visits deemed non-urgent, and physician burnout is increasing. As digital health gains traction, there is an urgent need for an AI-first-contact tool that aids triage, reduces misdiagnosis, and improves patient navigation.

Our solution—a GenAI-powered symptom checker—combines advanced language understanding with a structured knowledge base to analyze free-form symptom descriptions, deliver relevant suggestions, and escalate critical cases. Unlike rigid rule-based systems, our approach is designed to interpret complexity and uncertainty while ensuring safer and more flexible interactions.

## 2. Prototype Details

The prototype uses a Retrieval-Augmented Generation (RAG) system integrating OpenAI's language model, Pinecone for semantic search, and LangChain for prompt orchestration. Users describe symptoms through a Streamlit frontend, which passes the query to a backend pipeline. Symptoms are cleaned and embedded, relevant documents are retrieved, and a tailored, grounded response is generated by the LLM. A SQLite database supports structured relations among symptoms, diagnoses, and urgency levels.

The backend is modular, using NLTK and SciSpaCy for preprocessing, Scikit-learn for urgency classification, and Pickle for model persistence. The design allows session continuity, urgent-case escalation, and retraining. It supports scalability, real-time triage, and future extensions like voice/image inputs or HIPAA-compliant deployment.

## 3. Evaluation & Results

We evaluated the prototype with 20 realistic patient scenarios. Triage accuracy ranged between 75–80%, with 100% recall for true emergencies. Only one non-urgent case was incorrectly flagged as urgent. Diagnostic suggestions were appropriate in 75% of cases, often identifying the correct condition or category. The system returned relevant results even when exact answers were missing, demonstrating robust performance.

Overall, the prototype proved responsive and dependable. Response times were suitable for user interaction, and the retriever consistently returned context-aware documents. The system's balanced performance shows strong potential as a safe, AI-assisted triage tool.

## 4. Reflection on Feedback

The Elvina Mould case was pivotal in framing our business problem. Our system directly addresses the need for accurate, accessible triage and early detection, potentially preventing fatal misdiagnoses. Feedback from our class emphasized the system's relevance and its human-in-the-loop safeguards.

To simulate a hospital knowledge base, we used Faker to create over 5,000 synthetic patient records and 15,000+ symptom reports stored in SQLite. These contained ambiguous, overlapping, or filler data to mimic real-world inputs.

Key lessons included the importance of grounding AI with structured data to reduce hallucination, and the value of empathy in healthcare responses. Future improvements will include expanding the medical knowledge base, upgrading the urgency classifier, and integrating appointment booking features. These will push our prototype toward real-world deployment and improved healthcare access.