



PubMed Agentic RAG

A RAG-based AI for Biomedical Research

Team: Abhay, Abhishek, Dhawal, Junaid
Guide : Dr Greg Chism, Dr Enrique Noriega



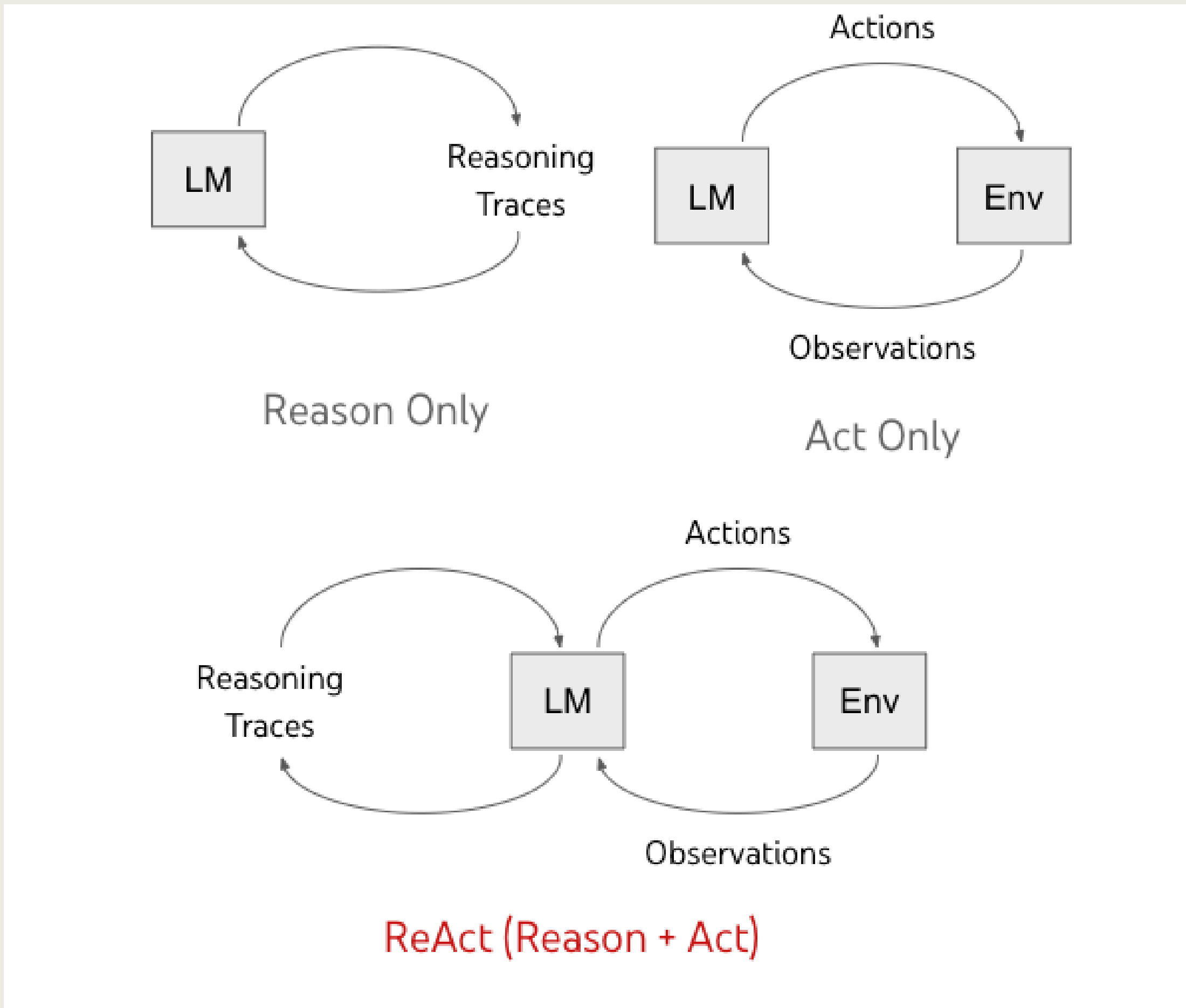
The Motivation

With over 36 million articles and thousands added daily, PubMed is a goldmine of medical knowledge—yet navigating it is overwhelming. Our PubMed RAG Assistant leverages Retrieval-Augmented Generation (RAG) to empower researchers and clinicians with instant, context-rich answers grounded in biomedical literature. By combining semantic search and LLMs, it bridges the gap between information overload and actionable insight.

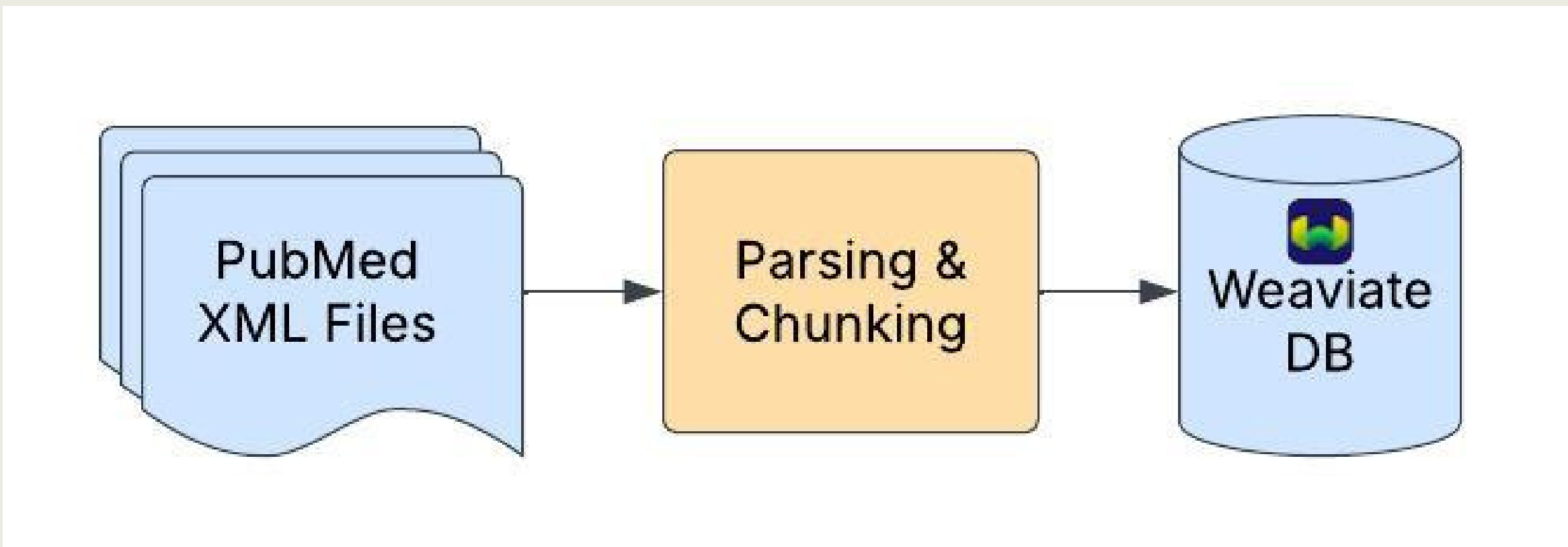
Dataset

Metric	Value
Number of papers	18,015 NXML files from PubMed Central
Average Length in Tokens	Approximately 4,571 tokens per paper
Number of Chunks	235,161 text segments
Average Chunks per Document	13.1 chunks per paper

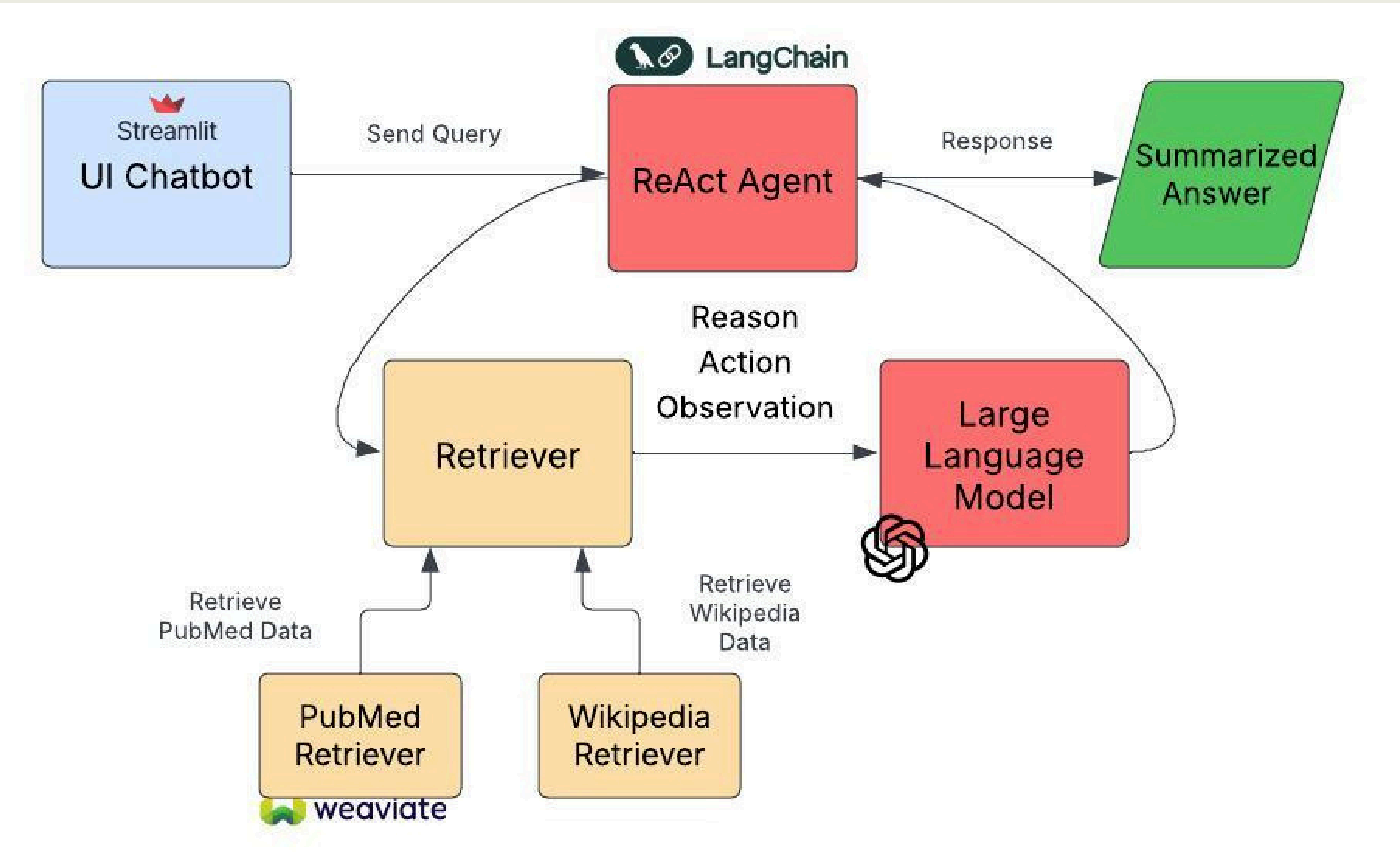
ReAct Agentic Framework



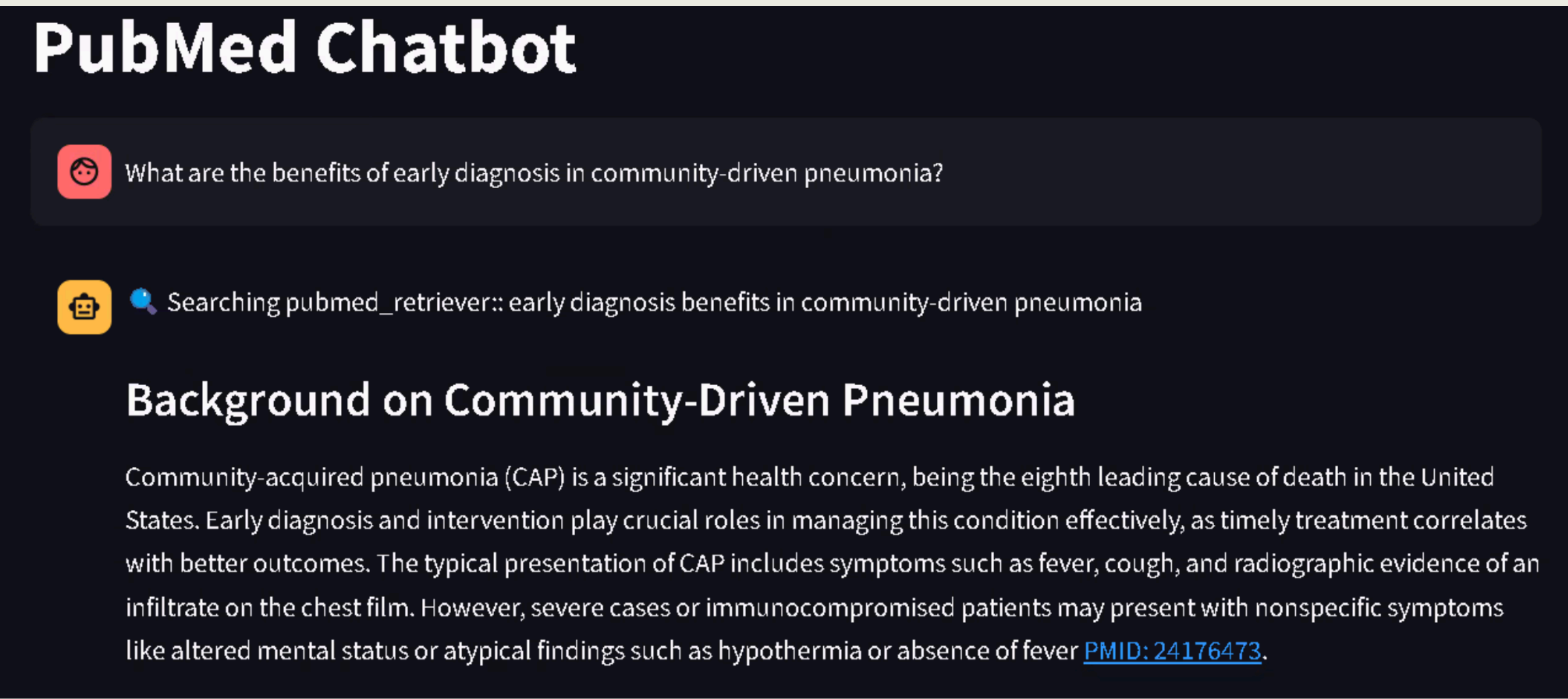
Data Preprocessing



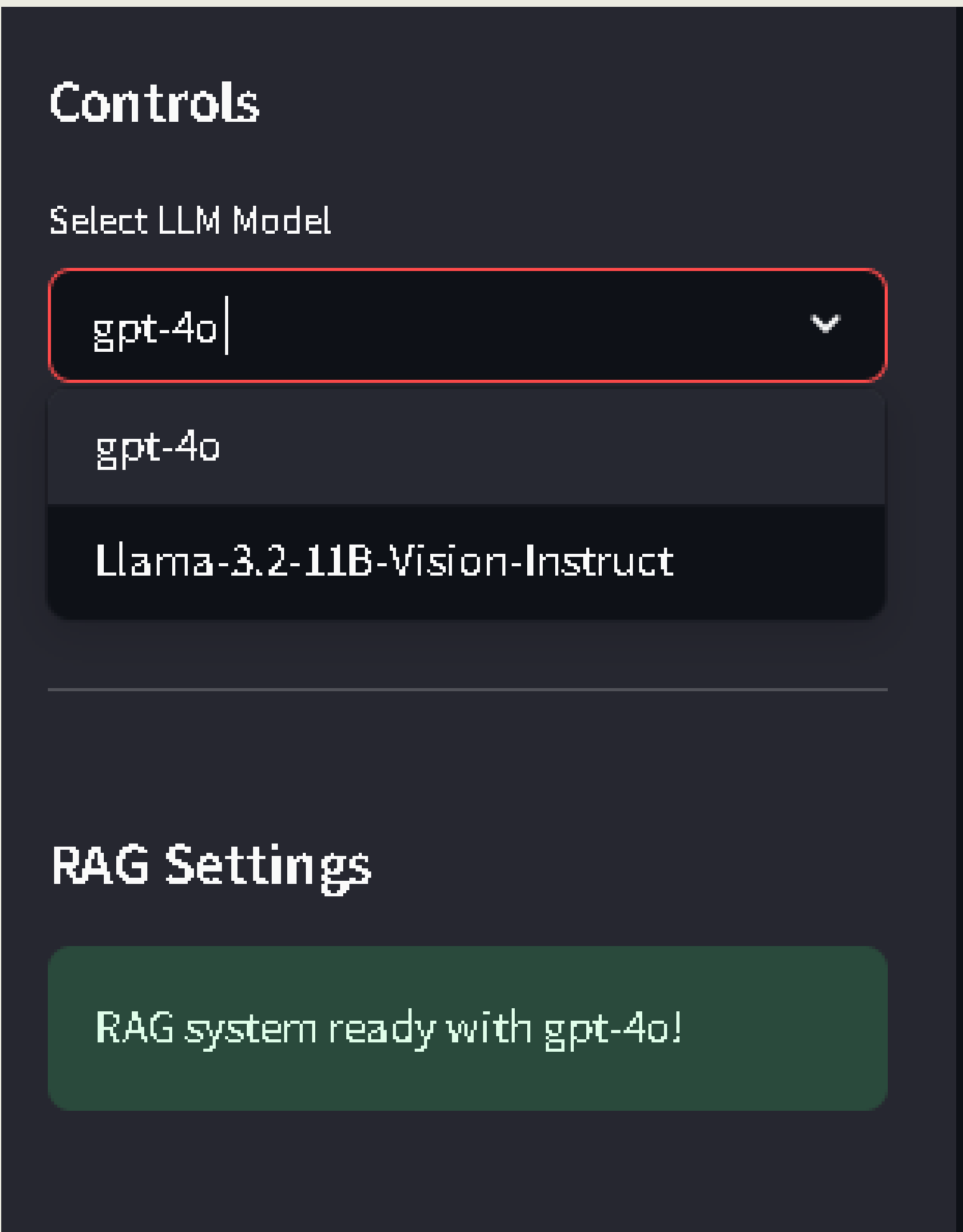
System Architecture



Screenshot



Ability to choose model



Try the App



Tech Stack

