MSOE Policy Chatbot


Nathan Cernik (CS)


Kevin Paganini (CS)


Tyler Cernik (CS)


Jackson Rolando (CS)

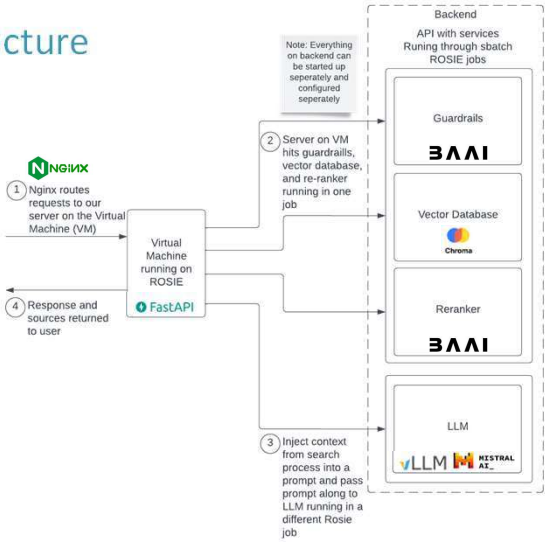

Jennifer Madigan (CS)


Dr. Derek Rlley (Project Advisor)

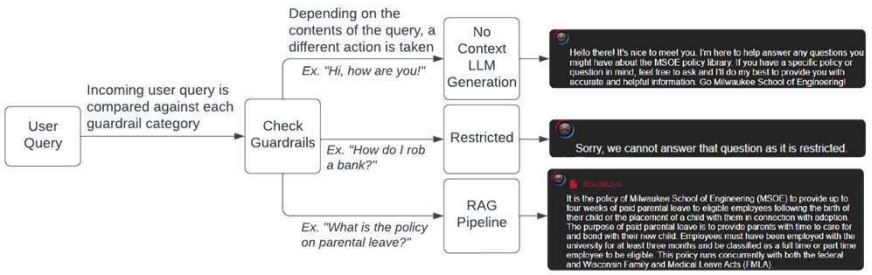
Project Overview

The goal of our Senior Design project was to create a state-of-the-art chatbot using Retrieval Augmented Generation (RAG) to answer questions regarding MSOE policies. RAG allows us to enhance the knowledge base of a Large Language Model (LLM), as LLMs do not have domain specific knowledge, especially about MSOE policies. By creating a chatbot to help answer policy questions, we can increase the access to information.

System Architecture

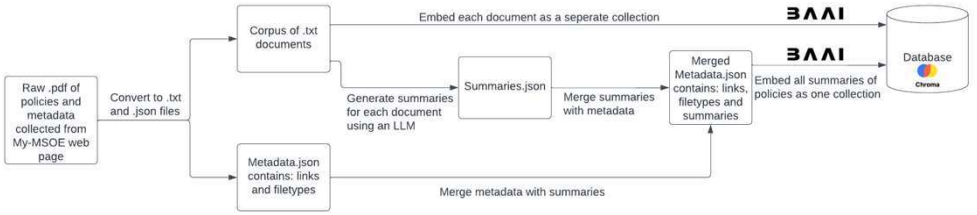


Data Pipeline



Chroma HUGGING FACE React FastAPI vLLM MISTRAL AI_ BAAI MSOE ROSIE

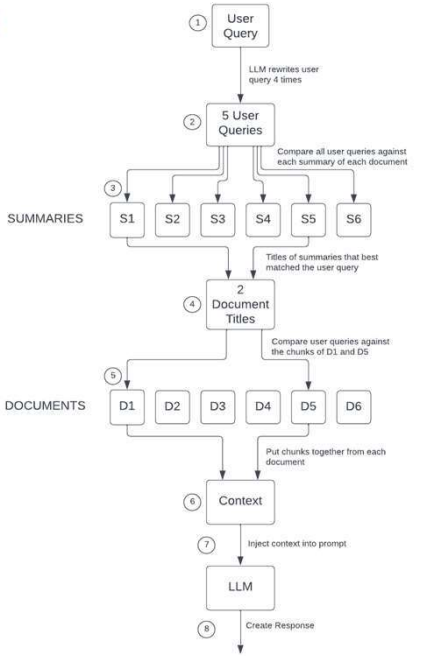
Data Processing



LLM Generation with Context

1. User query comes to VM and VM calls search process
2. User query is rewritten in four different ways
3. Five user queries are compared against the summaries of all 6 documents
4. The two summaries that are most similar to the five user queries titles are returned
5. We look in each of the returned files and grab the relevant chunks from there
6. The context for the LLM is created putting the chunks of the two documents together
7. The context is injected into the prompt and passed to the LLM
8. The LLM generates a response

Right: This is a toy example using only six documents in the corpus and two documents as search results. A re-ranker model is used, however is omitted from the diagram for clarity and ease of understanding.



Guardrails