

Geo-RAG Chatbot: Al-Powered Document Question Answering

Project Summary & Achievements

Haocheng Lin

Project Motivation & Objective



Objectives

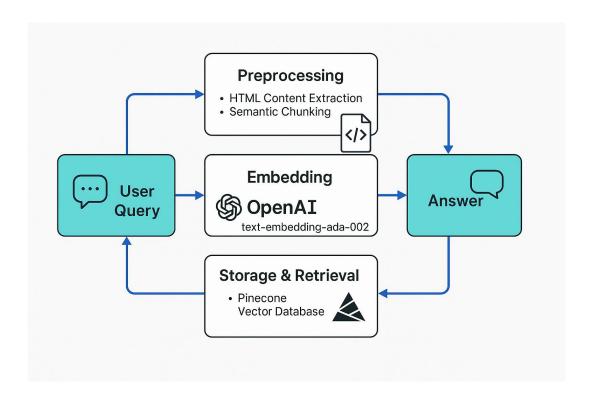
- Challenge: Difficult and time-consuming to perform manual search through the technical documents.
- **Requirement**: Users need accurate and prompt responses from Q & A sessions on large structured data.
- Objective: Build an AI chatbot for answering questions about domain-specific documents.

System Architecture



An End-to-End System Design

- **Input**: User Query
- **Preprocessing**: HTML content extraction and chunking
- **Embedding**: text-embedding-ada-002
- **Storage & Retrieval**: Pinecone vector database
- **Response**: GPT-4o-mini generates an answer.



Technical Stack



Technologies Used

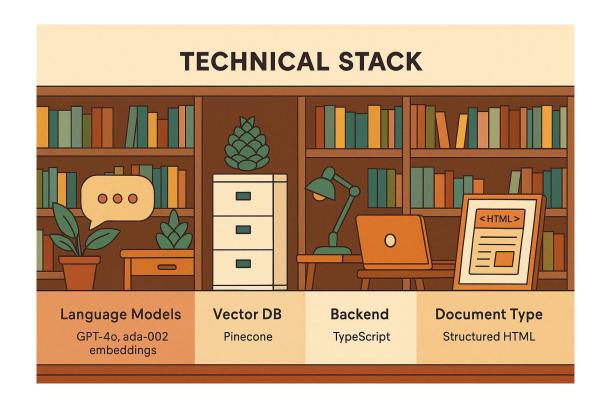
• Language Models: GPT-4o-mini

• **Vector DB**: Pinecone

• **Backend**: Typescript

Frontend: HTML & CSS

Document Type: Structured HTML (e.g., GEO_Limits)

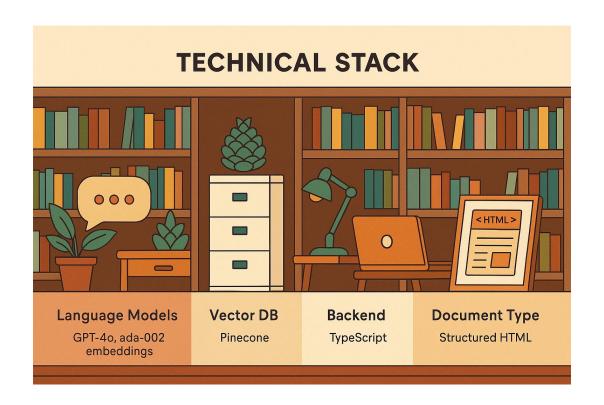


Achievements



What I have achieved:

- Functioning prototype with full document-querying capabilities.
- Performed semantic search to retrieve relevant extracts.
- Modularized architecture for ensuring future scalability.
- Fine-tuned and optimized RAG model for accuracy and performance.

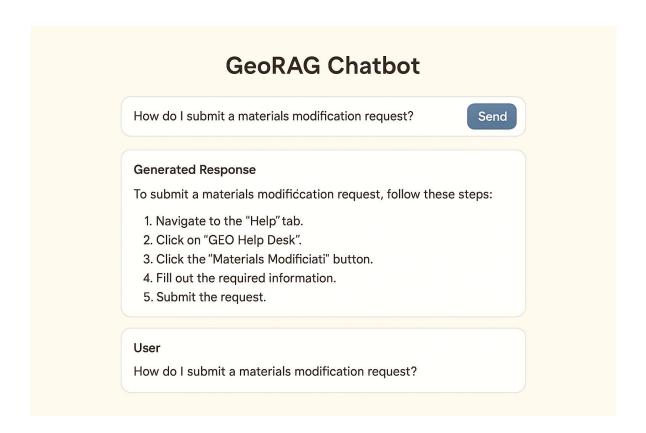


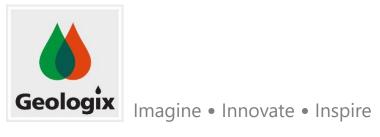
What's Next?



Future Improvements

- 1. Improve front-end UI layout.
- 2. Support querying across multiple documents.
- 3. Cache previous interactions to improve user experiences.
- 4. Integrate user feedback to improve the model's performance.





Thank You