Enhancing Access to Government Documents: A Retrieval-Augmented Generation Framework for Quezon City Ordinances and Resolutions

Alister Marc B. Domilies | Adviser: Miguel Francisco Remolona PhD

Background

The Quezon City Hall provides legal documents only as downloadable scanned PDFs, with limited search by exact titles or summaries. This makes retrieving specific ordinances slow and inefficient which hinders public access and legal transparency.

- **Static PDFs**
- **Limited Search**
- **Time-Consuming**

Objectives

1. **Create a RAG-based Chatbot:** Enable natural language queries for Quezon City legal documents and provide accurate, context-aware answers.
2. **Ensure Answer Accuracy:** Validate responses to guarantee they are relevant and grounded in the retrieved legal texts.
3. **Automate Document Indexing:** Develop a pipeline to extract, process, and index new ordinances and ensure the database remains up-to-date.

Architecture of Retrieval-Augmented Generation (RAG) System

1. Document Indexer

Optical Character Recognition (OCR) is applied to the scanned PDFs, segmenting the legal documents into Introduction, Preamble, Operative, and Signature categories, then storing them in a vector database.

2. Retrieval Decision Module

Determines whether a query requires external context; simple ones are answered directly by the language model, while knowledge intensive queries trigger document retrieval.

3. Query Transformation Module

The user's question is first refined and translated to English if needed, considering prior context. If the initial answer fails validation, the system runs additional trials using Hypothetical Document Embeddings (HyDE) in an attempt to retrieve a valid response.

4. Document Retrieval Module

A combination of dense and keyword search is used, re-ranked via Reciprocal Rank Fusion with equal weighting. The Top 7 results are retrieved and expanded with adjacent text chunks.

5. Answer Generation Module

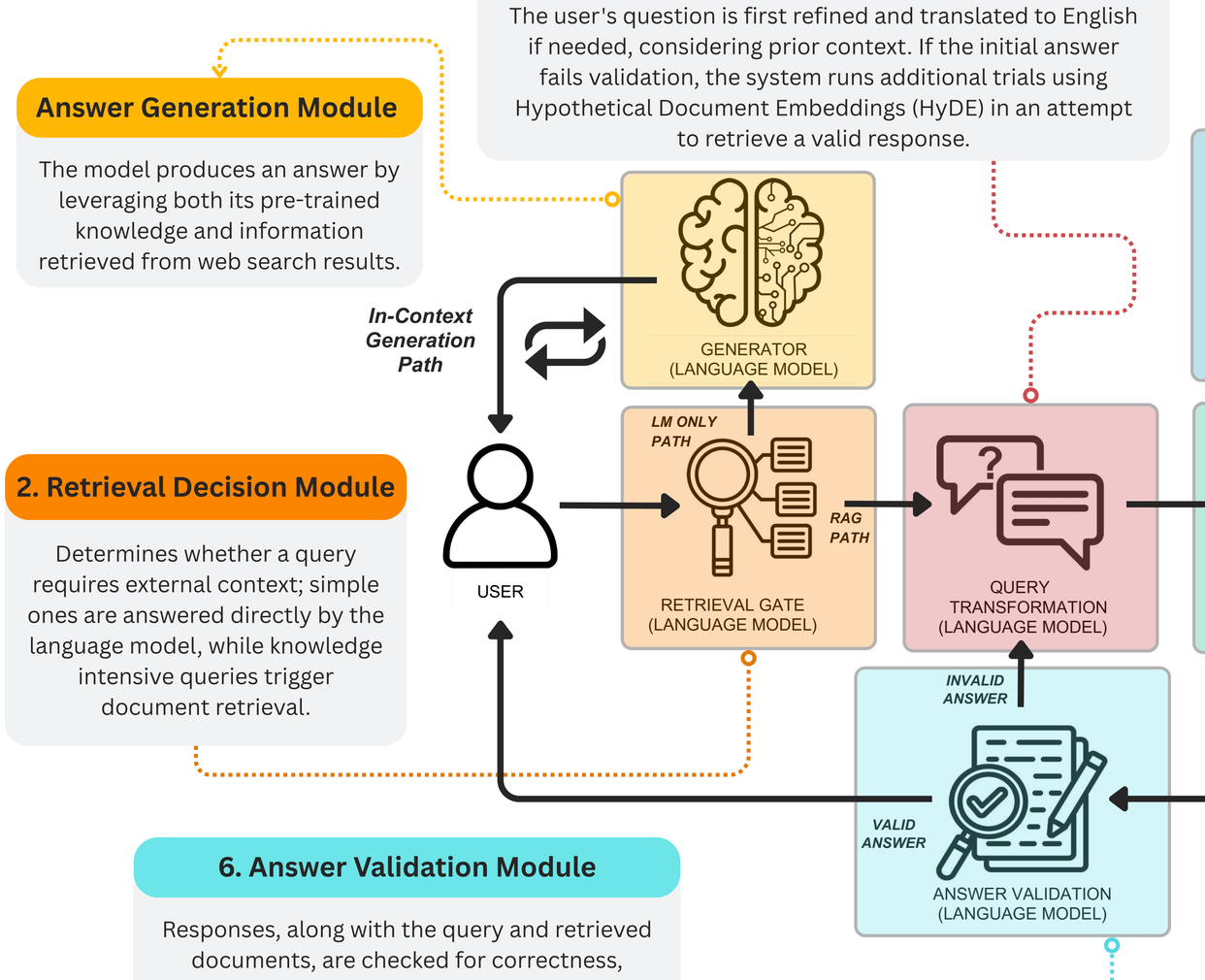
The retrieved documents, along with the query, are fed into the generator model to produce a legally grounded answer, including source citations.

6. Answer Validation Module

Responses, along with the query and retrieved documents, are checked for correctness, completeness, and honesty. Invalid answers retry the RAG process up to 3 times; valid ones are returned to the user.

Answer Generation Module

The model produces an answer by leveraging both its pre-trained knowledge and information retrieved from web search results.



Results

Figure 1. Valid Answer Rates Across Trials

Figure 1 shows increasing valid answer rates across the 3 trials, while Table 1 highlights strong performance of our RAG-enhanced model in RAGAS metrics - particularly in faithfulness, relevancy, similarity and correctness

The RAG-powered chatbot underwent testing and received approval from Quezon City Hall staff responsible for managing the official website. They **recognized its effectiveness in improving access to local laws and expressed confidence in its readiness**. The project is now moving forward toward deployment.

Table 1. Final Answer Metrics Summary

Language	Model	Faithfulness	Answer Relevancy	Answer Similarity	Answer Correctness
English (Direct LM Answers)	gpt-4o-mini	-	-	0.932	0.602
	gpt-4o-mini (web search)	-	-	0.930	0.589
English (Our RAG System)	gpt-4o-mini	0.892	0.945	0.958	0.786
	gpt-4o	0.922	0.953	0.955	0.779
Tagalog (Our RAG System)	gpt-4o-mini	0.865	0.937	0.951	0.741
	gpt-4o	0.921	0.952	0.956	0.766

Conclusion

This project leverages various RAG techniques to enhance retrieval accuracy and response relevance for Quezon City legal documents—resulting in more reliable answers, improved user experience, and increased public engagement.

Future Work

- Fine-tune RAG models with legal datasets for better retrieval precision.
- Add more legal documents including executive orders, transparency reports etc.
- Explore ways to further reduce latency and optimize system performance.

Acknowledgements

I sincerely thank my UPD AI professors, especially Sir Migs, my family and friends, and the Almighty for their support, strength, and guidance.