TAX INFORMATION & SERVICE CHATBOT FOR SMART CITY PUBLIC SERVICES

Atsadawut Surinrangsee, Juthamat Suwannakhajon,
Kallada Intaduang, Sedthanan Hankhompong, Thitichaya Sumethalangkarn
Advisor: Asst. Prof. Dr. Thanasak Mouktonglang, Dr. Suwika Plubin

ABSTRACT

This research develops a tax information chatbot by integrating a Large Language Model (LLM) with a Retrieval-Augmented Generation (RAG) architecture, aimed at reducing officers' workload and delivering accurate digital tax services within a Smart City context. The system leverages embeddings, a FAISS vector database, and an agent-based design to enhance retrieval quality and support multi-step reasoning.

Five configurations (M1–M5) were tested by varying the generator model, embedding model, and retrieval strategy. M5 uniquely employs a Recursive Language Model (RLM) framework, which performs iterative retrieval–reasoning cycles to strengthen logical consistency. Evaluation on a tax-related Q&A dataset revealed that while M4 achieved the highest scores, M3 was selected for deployment due to its strong performance and lower computational cost, offering a better balance between effectiveness and efficiency.

The system reduces repetitive inquiries handled by officers and provides a practical framework for applying LLM+RAG in Smart City public service systems.

INTRODUCTION

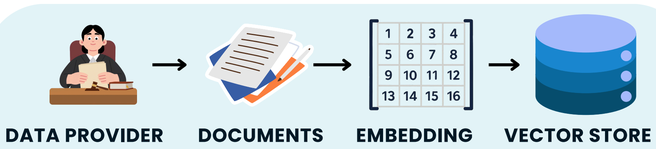
Tax information services currently face challenges such as service delays, limited personnel, and frequently updated tax regulations, making accurate information access difficult. This research applies artificial intelligence using a Large Language Model (LLM) integrated with Retrieval-Augmented Generation (RAG) to improve response accuracy. The system utilizes embedding techniques, a vector database (FAISS), and an agent framework to retrieve, analyze, and generate reliable responses from official documents.

OBJECTIVES

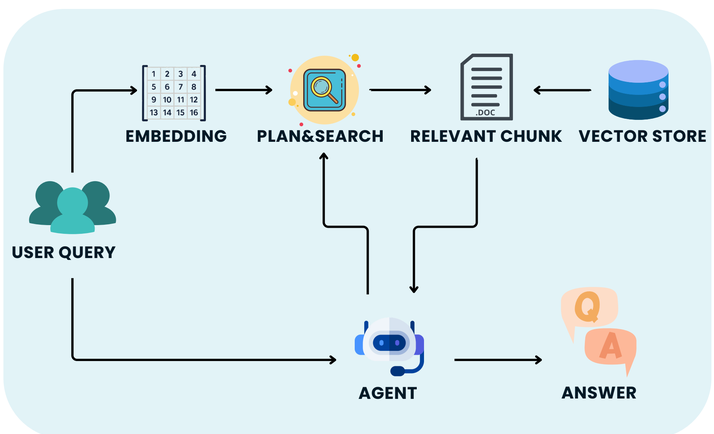
- To design and develop a tax information chatbot using LLM and RAG technologies.
- To develop a government document management and retrieval system using embedding techniques and vector databases to improve information accuracy.
- To evaluate system performance and support digital tax information services within a Smart City context.

METHODOLOGY

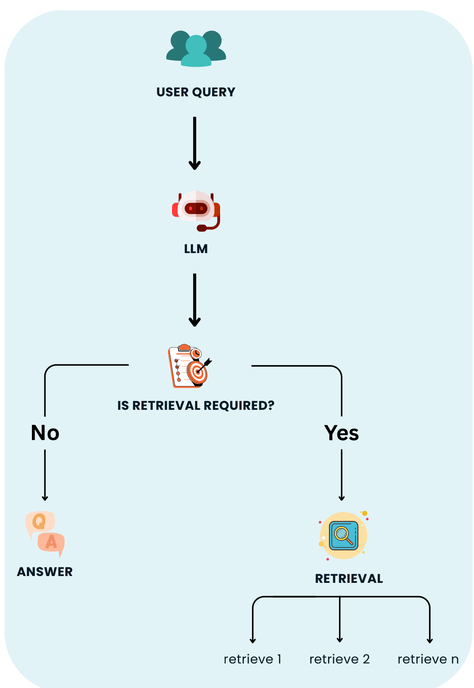
DATA PREPARATION



SYSTEM WORKFLOW



AGENT WORKFLOW



EXPERIMENT SETUP

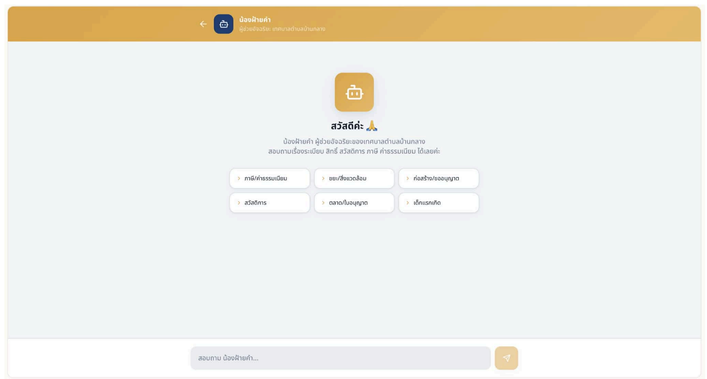
Method ID	Generator	Embedding	Retrieval	Framework
M1	Gemini-2.5-Flash	BGE-M3	Single	Standard
M2	Gemini-2.5-Flash	Qwen-0.6B	Single	Standard
M3	Gemini-2.5-Flash	BGE-M3	Multi	Standard
M4	Gemini-3-Pro	BGE-M3	Multi	Standard
M5	Gemini-2.5-Flash	BGE-M3	Multi	RLM

RESULTS

Method ID	Correctness	Completeness	Friendliness	Safety	Helpfulness
M1	2.80	2.96	4.24	2.89	3.51
M2	1.96	2.04	3.85	2.58	2.69
M3	3.11	3.18	4.02	2.82	3.71
M4	3.38	3.78	4.33	2.78	4.07
M5	3.24	3.24	4.02	3.02	3.69

CHATBOT SYSTEM INTERFACE







CONCLUSION

From the evaluation of five chatbot architectures, answer quality was mainly influenced by model reasoning capability and the integration of planning with multi-retrieve. Although Gemini 3 Pro + Planning + Multi-retrieve achieved the best performance, it required higher computational cost and latency. Therefore, this study selected M3: Gemini 2.5 Flash + Planning + Multi-retrieve due to its balanced trade-off between performance and efficiency, while still improving contextual completeness. Key challenges included incomplete retrieval, limited reasoning in smaller models, and increased cost with system complexity. Overall, an effective chatbot design requires balancing model capability, retrieval strategy, and resource efficiency.

REFERENCES

Zhang, Alex L., Tim Kraska, and Omar Khattab. "Recursive language models." arXiv preprint arXiv:2512.24601 (2025).
Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in neural information processing systems 33 (2020): 9459-9474.
Acharya, Deepak Bhaskar, Karthigeyan Kuppan, and B. Divya. "Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey." IEEe Access 13 (2025): 18912-18936.