



U.S. Department
of Transportation

Federal Highway
Administration

A Conversational Agent Framework for Multimodal Knowledge Retrieval

A Case Study in FHWA InfoHighway Web Portal Queries

School of Computing, George Mason University

Sai Surya Gadiraju
sgadira3@gmu.edu

Zijie He
zhe20@gmu.edu

Prof. Dr. Duoduo Liao
dliao2@gmu.edu



INTRODUCTION

The Federal Highway Administration (FHWA) manages vast multimodal infrastructure data through its InfoBridge and InfoHighway portals. Accessing this knowledge typically requires advanced technical skills, limiting its usability for non-engineering professionals.

To bridge this gap, we developed **InfoTech Assistant** a conversational agent framework that enables intuitive access to both structured (e.g., bridge databases) and unstructured (e.g., technical data). By integrating Large Language Models (LLMs) with **Retrieval-Augmented Generation (RAG)** and **Text-to-SQL**, our system supports natural language interactions that translate into actionable insights, democratizing data access across FHWA users and stakeholders.

METHODOLOGY

The InfoTech Assistant leverages a multimodal approach for retrieving infrastructure knowledge via conversational queries:

Data Collection: FHWA content was scraped and stored in JSON (text/images) and SQLite (tables).

Query Handling:

RAG: Semantic document search using embeddings and cosine similarity.

Text-to-SQL: SQL generation via schema-aware prompts and PRAGMA metadata.

Backend: Flask app routes user queries and returns results via HTTP.

Temperature Tuning: LLaMA model configured with $T = 0.7$ for balanced response diversity and precision.

SQL Evaluation Protocol: Level 1: Simple column retrievals; Level 2: Filters, thresholds, aggregates; Level 3: Multi-table joins and implicit semantics.

$$\text{Accuracy} = \frac{\text{Number of correct SQL queries}}{\text{Total number of queries executed}} \times 100\%$$

RAG Evaluation: Used cosine similarity and accuracy metrics on test queries.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

EVALUATION METRICS

Table 1: Sample Questions: Similarity and Accuracy Results for RAG-Based Semantic QA

Question	Similarity	Accuracy
What is Electrical Resistivity?	0.94	95%
What are the benefits of Hammer Sounding?	0.92	94%
Can you explain the Crack Propagation Gage (CPG)?	0.97	99%
How to do Screw Withdrawal Testing?	0.98	98%
Explain Transverse Vibration of Structural Systems.	0.96	96%
Why is coring considered best for visual inspection?	0.94	92%
Can you explain Stress Wave Timing?	0.94	96%

Table 2: Sample Questions: Accuracy Results for Text-to-SQL Module

Sample Questions	Level of difficulty	Accuracy per level
Show the year built for all bridges.	Level 1	92%
Retrieve the average daily traffic for all bridges.	Level 1	92%
Show bridges older than 80 years.	Level 2	88%
Retrieve all columns for bridges built in 2016.	Level 2	88%
List bridges with average daily traffic over 100,000 and built after 2000.	Level 2	88%
List congested bridges in Fairfax.	Level 3	86%
Show old bridges with high traffic.	Level 3	86%
Retrieve structurally deficient bridges built before 1970.	Level 3	86%

CONCLUSIONS

The **InfoTech Assistant** demonstrates a practical and scalable approach to multimodal knowledge retrieval for infrastructure data. By combining semantic search (RAG) and schema-aware SQL generation, it enables both technical and non-technical users to query FHWA datasets using natural language.

Achieving up to **95% accuracy in semantic tasks** and **88% in SQL queries**, the system improves data accessibility, transparency, and usability. Its modular design also makes it adaptable for broader government and industrial applications.

SYSTEM ARCHITECTURE

The **InfoTech Assistant** features a modular, high-performance architecture designed for multimodal knowledge retrieval.

Frontend Interface: A user-friendly web interface accepts natural language queries.

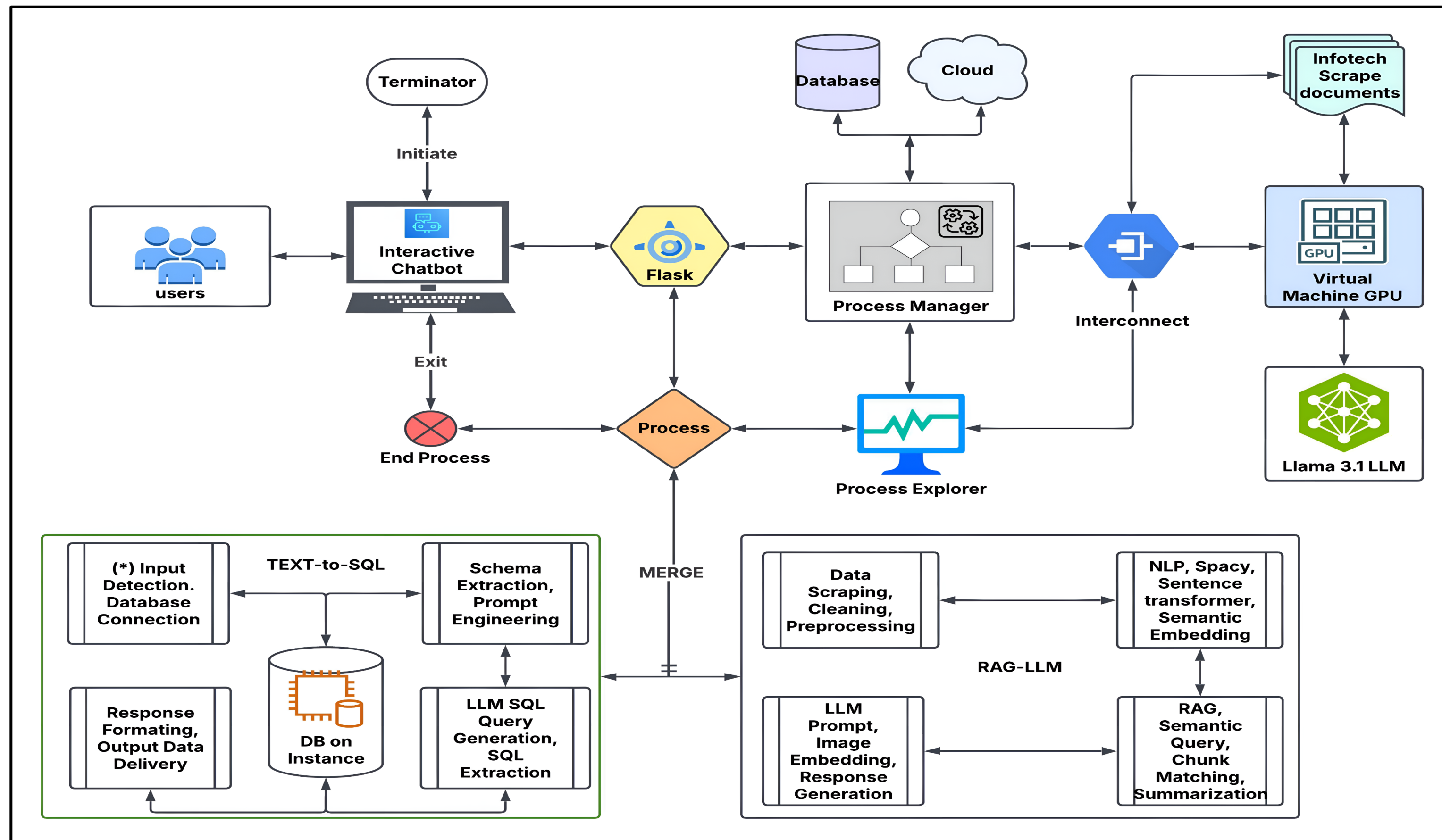
Flask Backend: Manages query routing through a process manager that distinguishes between semantic (RAG) and structured (Text-to-SQL) queries.

RAG Pipeline: Embedding models and semantic search to retrieve relevant content from FHWA documents.

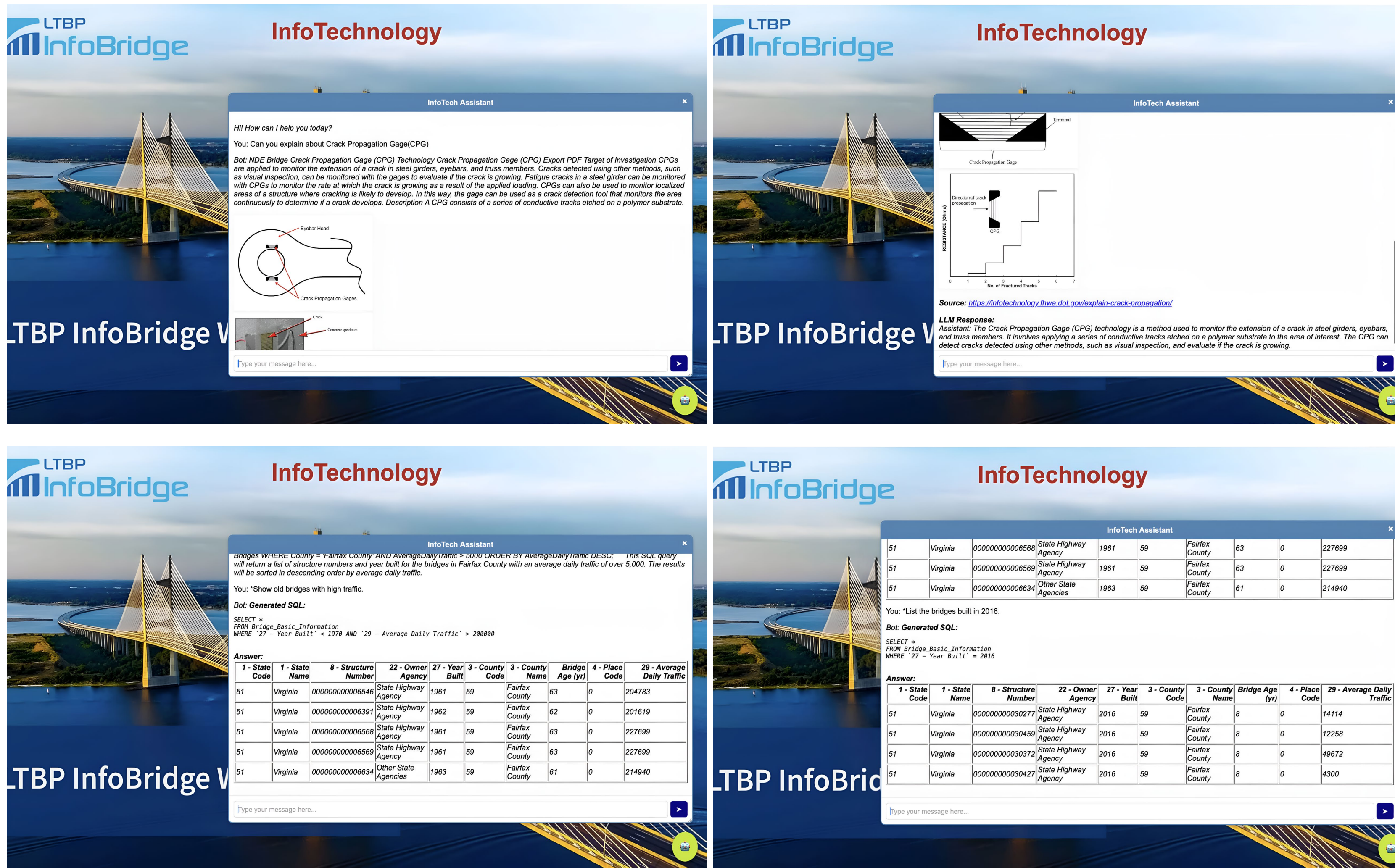
Text-to-SQL Engine: Generates accurate SQL queries for FHWA's structured databases using schema-aware prompts.

GPU-Powered Inference: Ensures low-latency processing using the LLaMA 3.1 8B model.

Monitoring & Session Control: Error handling and session re-initialization ensure robust, state-aware interactions.



EXPERIMENTAL RESULTS



ACKNOWLEDGMENTS

The authors thank the **Federal Highway Administration (FHWA)** for providing public datasets that were instrumental for system testing and evaluation. This research also utilized the **Hopper GPU cluster** (Office of Research Computing, 2025) and resources from the **Department of Information Sciences and Technology**, supported by the **George Mason University Online Research Computing Environment**.

REFERENCES

- Federal Highway Administration. FHWA InfoBridge: National Bridge Inventory and Inspection Data. 2024. <https://infobridge.fhwa.dot.gov/>
- Mohammadjafari, A. et al. From Natural Language to SQL: Review of LLM-Based Text-to-SQL Systems. *arXiv:2410.01066*, 2024.
- Meta AI. LLaMA 3.1 8B Instruct. <https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct>
- Siino, M. All-MPNET at SemEval-2024 Task 1: Semantic Relatedness with Sentence Transformers. *SemEval 2024*.