A Conversational Agent Framework for Multimodal Knowledge Retrieval A Case Study in FHWA InfoHighway Web Portal Queries											
U.S. Department of Transportation	School of Computing,	George Mase	on University	GEORGE MASON UNIVERSITY®							
Federal Highway Administration	Sai Surya Gadiraju sgadira3@gmu.edu z	Zijie He he20@gmu.edu	Prof. Dr. Duoduo Liao dliao2@gmu.edu								
	NTRODUCTION		SYSTEM ARTCHITECTURE								
infrastructure data through its	istration (FHWA) manages vast multimoda InfoBridge and InfoHighway portals. Accessing s advanced technical skills, limiting its usability ls.	for multimoda Frontend Inter Flask Backen	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								
framework that enables intuit databases) and unstructured Language Models (LLMs) with	ed InfoTech Assistant a conversational agent tive access to both structured (e.g., bridge (e.g., technical data). By integrating Large Retrieval-Augmented Generation (RAG) and orts natural language interactions that translate	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									

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.

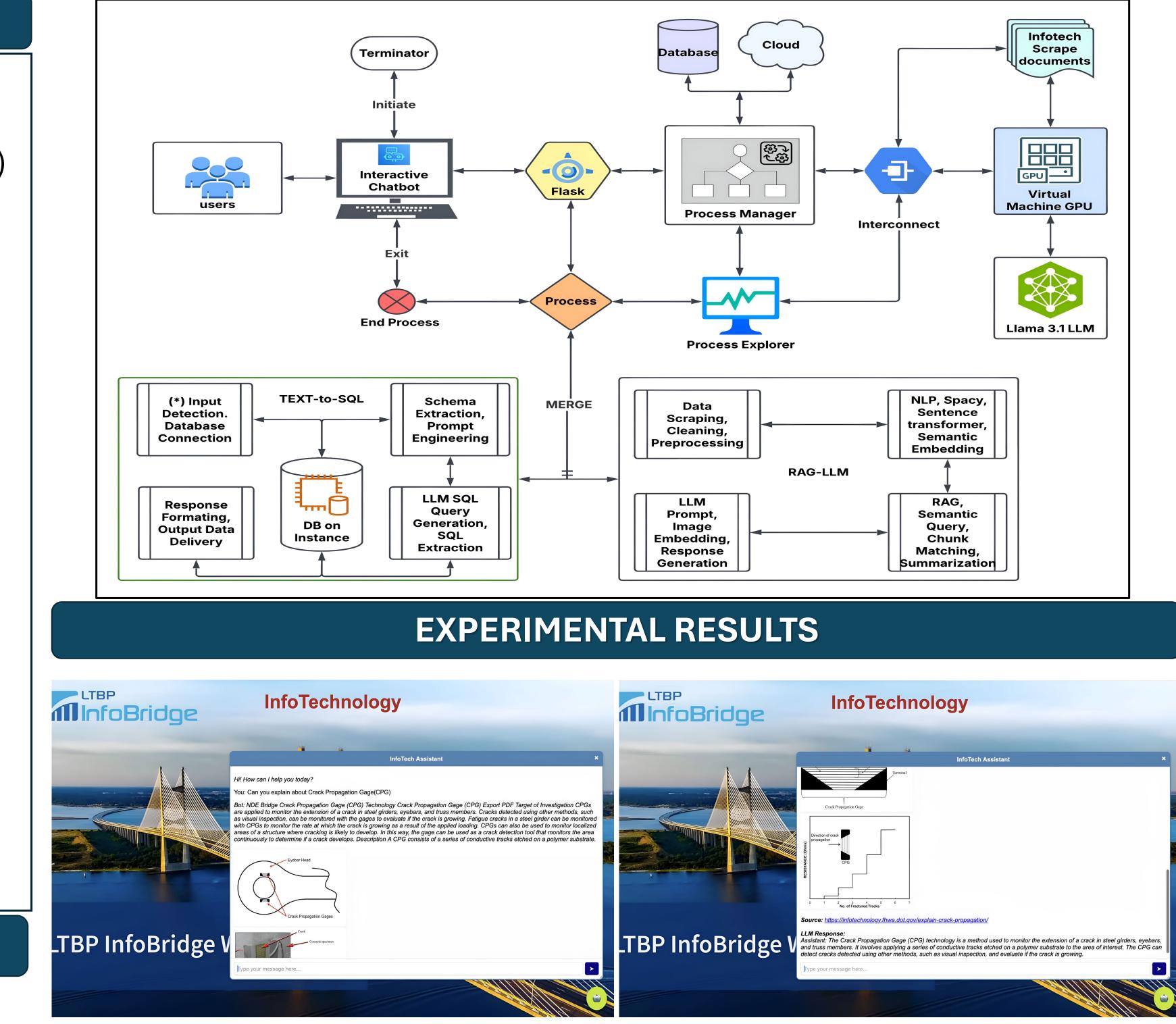
Temperature Tuning: LLaMA model configured with **T = 0.7** for balanced response diversity and precision.

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

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

model.

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



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

 $\cos(heta) = rac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = rac{\sum\limits_{i=1}^n A_i B_i}{\sqrt{\sum\limits_{i=1}^n A_i^2} \sqrt{\sum\limits_{i=1}^n B_i^2}}$

EVALUATION METRICS

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	Level 2	88%
built after 2000.		
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%

														_				_		_		
		<u> </u>	du														InfoTec	h Assistant				
	- HEIGGOD IA				InfoTech A					×			51	Virginia	00000000006568	State Highway Agency	1961	59	Fairfax County	63	0	227699
	will return	a list of structure	numbers and year order by average	built for the l	oridges in F	airfax County	with an avera	age daily t	traffic of ov	I his SQL query /er 5,000. The results			51	Virginia	000000000006569		1961	59	Fairfax County	63	0	227699
		w old bridges with		dany trame.									51	Virginia	00000000006634		1963	59	Fairfax	61	0	214940
	a starter and a starter at the start	rated SQL:	g								Marian -					Agencies			County			
	SELECT * FROM Bridge_Basic_Information WHERE `27 - Year Built` < 1970 AND `29 - Average Daily Traffic` > 200000										the bridges buil	t in 2016.										
								Thursday		Bot: Gene SELECT *	Bot: Generated SQL:											
	Answer:												FROM Bridg	ge_Basic_Infor – Year Built`	mation = 2016							
	1 - State 1 - State 8 - Structure 22 - Owner 27 - Year 3 - County 8 - County Bridge 4					4 04																
										ce 29 - Average	CALLE SAME	and the state of the										
	1 - Stat Cod 51	e Name	Number	Agency ate Highway	Built	Code	Name Fairfax			de 29 - Average Daily Traffic 204783		des P	Answer: 1 - Sta				r 27 - Yea		nty 3 - Coun		Age 4 - Pla	ce 29 - Average
	Cod	e Name Virginia 000	Number 000000006546 Sta Ag	Agency ate Highway gency ate Highway	Built 1961	Code 59	Name Fairfax County Fairfax	Age (yr		de Daily Traffic				de Name	Number	Agency State Highway			ode Nan Fairfax		Age 4 - Pla (yr) Co	29 - Average ode 14114
	51 Cod	e Name Virginia 000 Virginia 000	Number 000000006546 St. 000000006391 St. 000000006391 St. 000000006568 St.	Agency ate Highway aency ate Highway aency	Built 1961 1962	Code 59 59 59 59	Name Fairfax County Fairfax County	Age (yr 63		de Daily Traffic 204783			1 - Sta Co	de Name Virginia	Number	Agency State Highway Agency State Highway	/ Bui		ode Nan		Age 4 - Pla (yr) Co 0 0	ode
	51 51	e Name Virginia 000 Virginia 000 Virginia 000 Virginia 000	Number 000000006546 St. 000000006391 St. 000000006391 St. 000000006568 St.	Agency ate Highway lency ate Highway lency ate Highway lency	Built 1961 1962 1961	Code 59 59 59 59 59	NameFairfaxCountyFairfaxCountyFairfaxFairfaxFairfax	Age (yr 63 62		de Daily Traffic 204783 201619			1 - Sta Cod	de Name Virginia	Number 000000000030277 000000000030459	Agency State Highway Agency State Highway Agency	2016		ode Nan Fairfax County Fairfax County Fairfax		Age 4 - Pla (yr) Co 0 0 0 0 0 0	14114
TBP InfoBridge	Cod 51 51 51 51 51 51	e Name Virginia 000 Virginia 000 Virginia 000 Virginia 000	Number 000000006546 St. 000000006391 St. 000000006568 St. 000000006568 St.	Agency ate Highway lency ate Highway lency ate Highway lency ate Highway lency	Built 1961 1962 1961 1961	Code 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1 59 1	Name Fairfax County Fairfax County Fairfax County Fairfax	Age (yr 63 62		de Daily Traffic 204783 201619 227699 227699	LTBP Ir	ofoBrid	1 - Sta Con 51 51 51	de Name Virginia Virginia Virginia	Number 00000000030277 00000000030459 00000000030372	Agency State Highway Agency State Highway Agency State Highway Agency	2016 2016		ode Nan Fairfax County Fairfax County		Age 4 - Pla (yr) Co 0 0 0 0 0 0 0 0	14114 12258

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**.

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.

REFERENCES

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