The Center for Initiatives in STEM

Integrated Traffic Systems: Leveraging Data Processing, Deep Learning, and LLM Frameworks for Traffic Operation Support

Justin Tong, Samiul Hasan, Ph.D., Md. Mobasshir Rashid, Md Nafees Fuad Rafi, Jonah Nissan

Introduction

Context: Traffic operators manage roads to maintain safety and efficiency and must respond quickly to high-stake incidents and scenarios.

Problem: Current traffic management tools are frequently outdated, overly technical, and slow to utilize under pressure.

Objective: Our team aims to develop and implement modern, digital traffic tools that are easy to use and designed for traffic management to leverage.

Solution:
1. Data processing framework
2. Deep-learning traffic prediction models
3. AI-Powered Traffic Chatbot

1: Data Processing Framework

Purpose & Process: Raw data from traffic sensors come with necessary but flawed context. Our team developed a program to filter, impute, and organize large volumes of historical and real-time Florida traffic data. This data is stored and used for reference and model-training.

Feature Engineering: The dataset was enhanced by adding new spatial and temporal features based on internal features and external datasets to give each row more context and for analysis and future model training.

Hurricane Evacuation Focus: Hurricanes are a distinct, yet consistent occurrence in Florida. During evacuations, traffic patterns are anomalous and far more difficult to predict than regular periods. We decided to focus on this challenge for our data processing and model building.

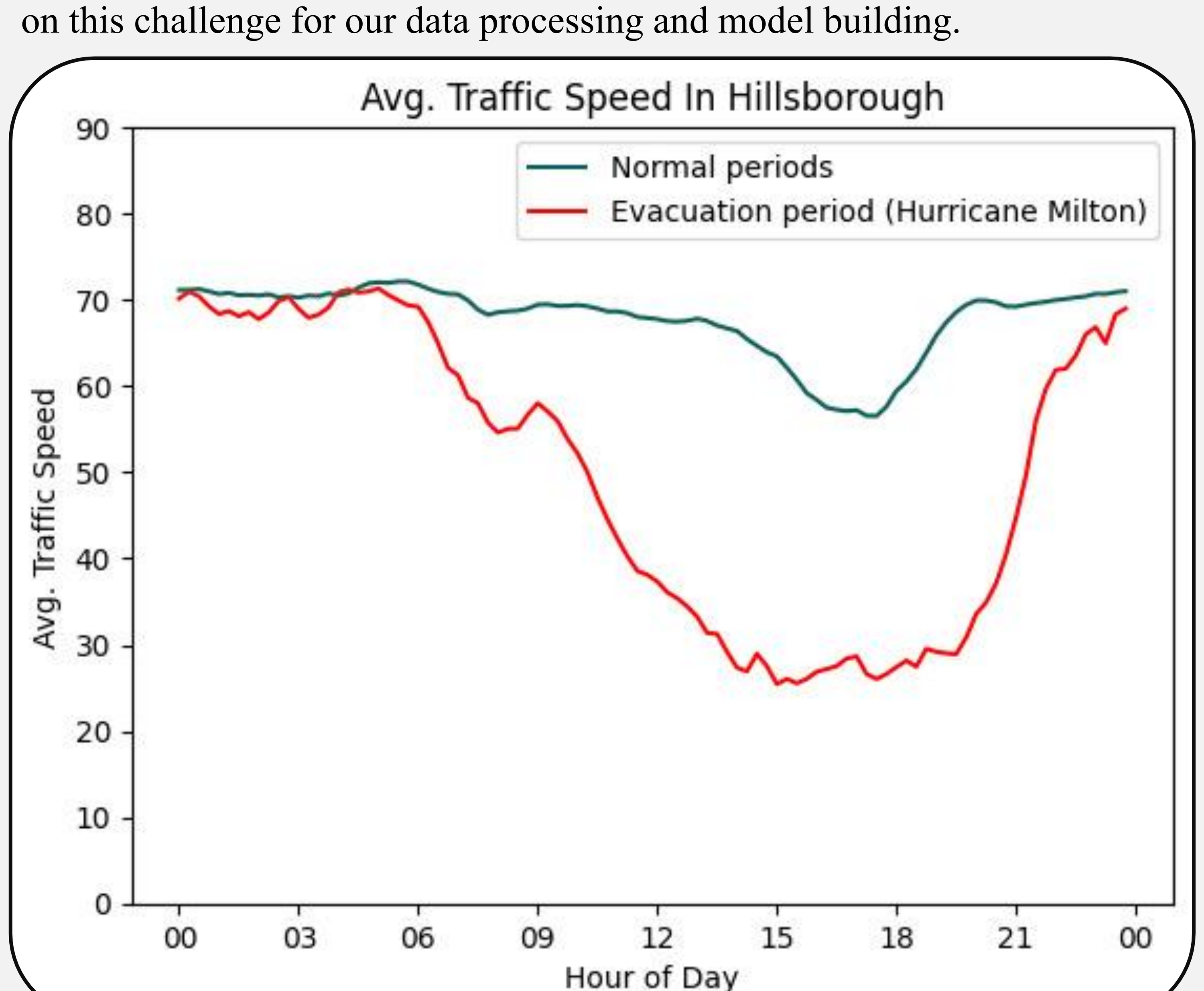

Figure 1: Illustrates traffic speed divergence in Hillsborough County highways between normal and evacuation periods for Hurricane Milton

2: Deep Learning Prediction Models

Goal: Developing traffic prediction models using machine learning provides a very powerful tool for traffic managers to forecast traffic accurately, even during volatile conditions.

Training: Development began by training a foundational LSTM algorithm, which was then expanded into advanced spatial-temporal models, training regressively on processed datasets of both regular and evacuation periods.

Testing: Iteratively retrained with various parameters reactively to test results to optimize model performance, achieving >97% accuracy for regular periods and >95% accuracy for evacuation periods.

Results: Tuned models were benchmarked in *Figure* 2. Our data-driven approach had successfully predicted regular and evacuation traffic conditions with high accuracy.

Model	Prediction Horizon	Regular Period			Evacuation Period		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE
LSTM	0-15 min	4.76	2.62	4.92	8.05	4.08	10.32
	15-30 min	4.78	2.63	4.96	8.14	4.13	10.33
	30-45 min	4.78	2.64	4.98	8.15	4.16	10.38
	45-60 min	4.82	2.64	4.99	8.22	4.16	10.42
CNN-LSTM	0-15 min	4.71	2.59	4.78	7.94	4.05	9.95
	15-30 min	4.76	2.6	4.79	7.94	4.09	9.97
	30-45 min	4.77	2.61	4.86	7.96	4.09	10.02
	45-60 min	4.79	2.62	4.88	7.98	4.1	10.09
GCN-LSTM	0-15 min	4.67	2.49	4.73	7.85	3.99	9.66
	15-30 min	4.68	2.5	4.74	7.86	4.02	9.72
	30-45 min	4.71	2.51	4.74	7.88	4.03	9.77
	45-60 min	4.73	2.53	4.76	7.9	4.05	9.81
GAT-LSTM	0-15 min	4.55	2.48	4.62	7.42	3.91	9.53
	15-30 min	4.57	2.5	4.63	7.49	3.94	9.57
	30-45 min	4.59	2.51	4.67	7.52	3.96	9.61
	45-60 min	4.61	2.52	4.69	7.65	3.98	9.66
GT-LSTM	0-15 min	4.42	2.41	4.52	7.12	3.87	9.1
	15-30 min	4.43	2.42	4.56	7.19	3.89	9.25
	30-45 min	4.45	2.44	4.59	7.27	3.9	9.28
	45-60 min	4.46	2.46	4.6	7.35	3.93	9.44

Figure 2: Charts various predictive models and their test accuracy scores. Darker-colored cells represent higher relative accuracy.

3: AI Chatbot

Goal: To enable non-technical traffic operators to use complex traffic data and forecasting models, we developed a custom chatbot application to query these tools with natural language.

Workflow: (*Figure 3*) By using RAG (Retrieval-Augmented Generation) and tool-calling, our chatbot pipeline can use sources to grab necessary context for answering user queries regarding traffic and road networks. Our prototyped and in-development RAG methods include full-data-loading, vector-embedding, SQL Generation, and Graph RAG.

Impact: By bridging the gap between technically-complex data/tools and human natural language, operators can receive vital traffic information immediately without requiring technical expertise.

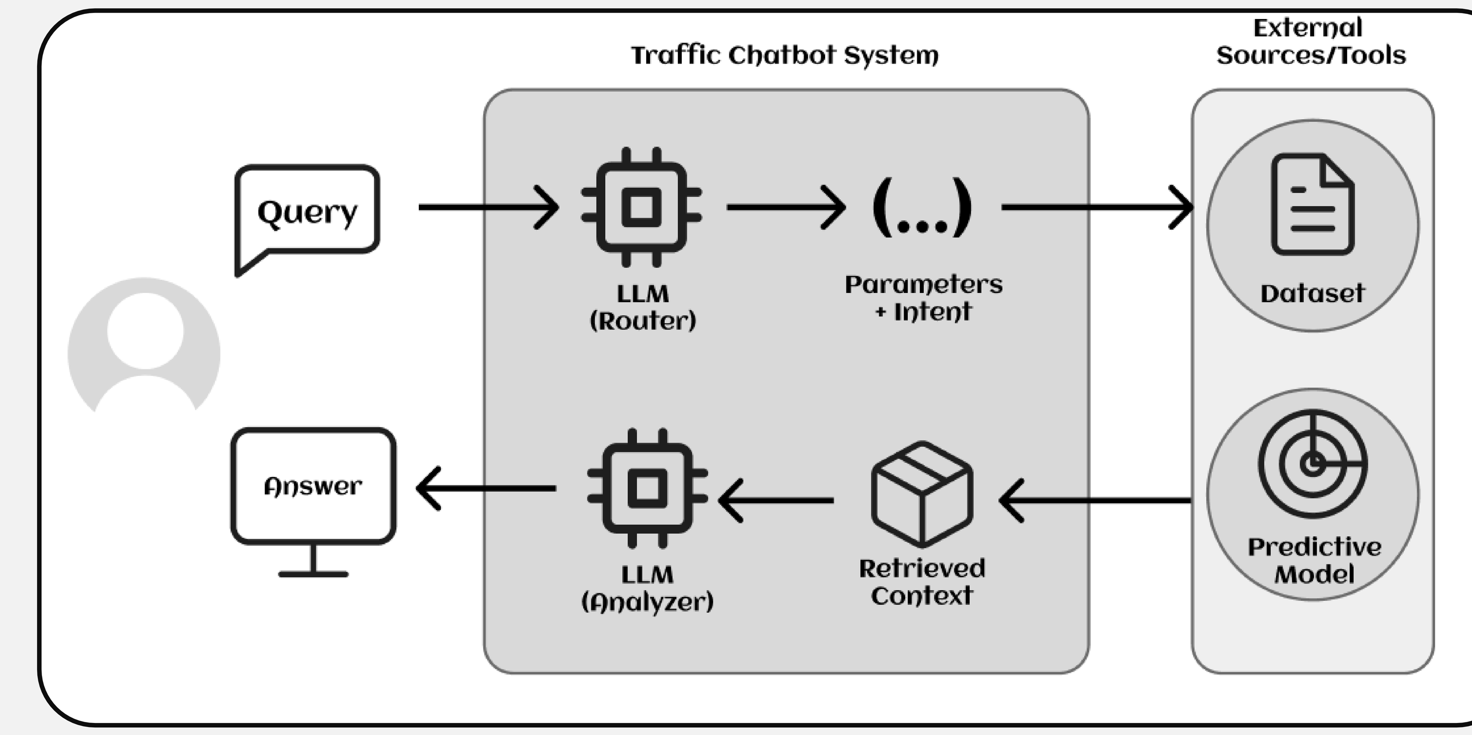

Figure 3: Graphic flowchart of Traffic Chatbot system architecture

---

Future Work

Data Processing Scaling: Our team is expanding the data processing framework to work with more traffic dataset formats. The system is designed to integrate real-time data for real, practical scenarios.

Knowledge Graph Expansion: A current in-development prototype of the Traffic Chatbot uses a knowledge graph powered by Neo4j, which maps relationships between network features to answer complex queries involving multiple spatial and temporal variables

Advanced RAG Methodology: In addition to the Knowledge Graph prototype, we hope to research and explore more techniques and test them against each other to optimize the chatbot's accuracy and speed.

Conclusion

Rich Datasets: Our team created a framework to process raw traffic data into feature-rich traffic datasets with machine-trainable quality.

Live Forecasting: Engineered, trained, and tested deep-learning models capable of forecasting traffic with high accuracy.

Intelligent Interface: As a hub for our frameworks, the Traffic Chatbot allows users to ask questions in natural language to leverage complex tools.

Real-World Application: This ecosystem of digital tools is well-researched and tested for proposal to traffic management, promoting traffic safety and efficiency.

References

[1] Rashid, M. M., Tong, J., & Hasan, S. (2025). *Transferability of Evacuation Traffic Prediction Models: Performance of A Spatiotemporal Graph Transformer Model Across Hurricanes.* Submitted for publication.