Disaster-Cast: Disaster-Aware POI Visit Forecasting with Multimodal LLM Fusion

Presented By:

Nurjahan

PhD Student
Department of Computer Science
Louisiana State University
nurja1@lsu.edu

Outline

- Introduction & Problem Statement
- Related Work and Research Gap
- Motivation
- Proposed Approach & Contributions
- Problem Formulation and Methodology
- Experimental Setup
- Results and Analysis
- Future Directions
- Conclusion

LSU

Introduction

- **Human mobility prediction** estimates future movement patterns using spatiotemporal data.
  - Under normal conditions, mobility is **highly regular and predictable** (~93% predictable; Song et al., 2010).
  - Daily visits to Place of Interests (POI) follow **stable temporal rhythms** (weekday/weekend cycles, seasonal patterns).
  - People repeatedly visit a small set of preferred places → **strong habitual structure**.
- **Natural disasters break this regularity**:
  - Disasters are increasing in frequency and severity.
  - They cause **abrupt disruptions** to normal mobility:
    - sudden business closures, evacuations, power outages,
    - **sharp drops** to near-zero activity for many POIs,
    - **uneven, spatially heterogeneous recovery** across neighborhoods and cities.
- Accurate mobility forecasting after disasters enables:
  - Emergency response & supply-chain continuity
  - Infrastructure and recovery planning

LSU

Problem Statement

- Design a forecasting framework that remains accurate **when standard assumptions break**.

- How can we model and forecast POI-level mobility **after disasters**, when:
  - Temporal regularity collapses,
  - Recovery differs across locations,
  - Recovery differs across business category

Conceptual Illustration of our Task

Place-of-Interest (POI)

7,12,7,9,2,0,1,10,5,0,0,4,0,0,7,6,7,13,6,0

Past Daily Visits at a **Park** at Universal Boulevard in **Orlando** **(Temporal+ Spatial +Semantic)**

→

Our multimodal Disaster-Cast System

→

Place-of-Interest (POI)

0

Next Day Prediction

Image: https://hiuniversal.com/theme-parks/

Literature Review: Numerical & Deep Learning

- Classical Models (ARIMA / SARIMA)
  - Assume smooth, stationary patterns
  - Fail under abrupt shocks
  - Require long histories
- Prophet
  - Additive trend + seasonality
  - Breaks under sharp collapses / rebounds
  - Short sequences confuse changepoint detection
- Deep Learning (LSTM / GRU / CNN)
  - Need dense, stable sequences
  - POI-level data is noisy + short
  - CNN extracts local patterns but lacks spatial/semantic reasoning

Literature Review: LLM-Based Mobility Forecasting

- **Mobility Forecasting as Language Generation**
  - Textualization of Mobility (SHIFT, WSDM’22; AuxMobLCast, Sigspatial’22)
    - Encode mobility as structured sentences
    - Forecasting = sequence-to-sequence translation
  - Prompt-Based Reprogramming (PromptCast, KDE’24)
    - LLM predicts future visits via prompts
    - No architecture change; prompt engineering only
  - Meta-Learned Prompts (Prompt Mining, Sigspatial’24)
    - Automatically discover effective prompt templates
    - Prompts act as learnable forecasting modules

Literature Review: Spatial–Temporal LLMs for Mobility Forecasting

- **ST-LLM (MDM, 2024)**
  - Approach: Spatial-Temporal Embedding + Partially Frozen Attention (PFA).
  - Uses three modalities:
    - Temporal values (traffic sequence)
    - Temporal context (hour/day embeddings)
    - Spatial embeddings (sensor-ID embedding)
- **Mobility-LLM (NeurIPS, 2024)**
  - POI ID, category words, and GeoHash(lat, lon)-> POI Point-wise Embedding Layer (PPEL)
  - Visiting Intention Memory Network (VIMN) to model temporal dynamics.
  - Human preference prompt tokens (HTPP) based on intention vectors.
  - Final LLM input = [VIMN tokens] + [user embedding] + [preference-prompt tokens]

LSU

Literature Review: LLMs for General Time-Series

- **Time-LLM (Approach (ICLR'24):**
  - Patch embeddings reprogrammed into LLM space
  - Natural-language prefix for task & stats
- **LLM4TS Approach (TIST'25) :**
  - Two-stage fine-tuning (alignment → forecasting)
  - Multi-scale temporal aggregation
- **TEMPO Approach (ICLR'24):**
  - Decomposition (trend, seasonality, residual)
  - Soft prompts & prompt pool for inductive bias

Research Gap

- Existing mobility research mostly focuses on
  - Next-location prediction,
  - Traffic forecasting,
  - Semantic reasoning
- Time-series LLMs (Time-LLM, Chronos, TEMPO, LLM4TS) assume
  - dense, stable, continuous data.
- No prior model utilize temporal encoder, H3 for spatial encoding
- **No prior model handles abrupt shocks, zero-heavy sequences, or uneven recovery across cities after hurricanes.**

LSU

Motivation

- **Hurricane impacts are complex and heterogeneous**
  - Recovery curves differ widely.
- **Business category strongly shapes recovery**
  - Essential services show early, sharp rebounds.
  - Non-essential categories stay near zero for days or weeks.
- **Spatial structure encodes local damage**
    - infrastructure outages,
    - evacuation zones,
    - neighborhood-level recovery conditions
- **LLMs carry rich external world knowledge**
    - **business functions** (e.g., hotels vs. gas stations),
    - **consumer behavior** (panic buying, evacuation return),
    - **typical disaster responses** (closures, supply shortages).
- A **purely numeric model cannot leverage this semantic context**.

Our Approach: Disaster-Cast (Diagnostic Multimodal Framework)

- Multimodal forecasting framework
  - Temporal encoder (CNN/GRU/LSTM/CNN+GRU) → visit dynamics
  - H3 embedding → spatial recovery patterns
  - LoRA-tuned LLM → POI semantics & instructions
  - Prefix fusion → unified multimodal input
  - Regression head → numeric prediction

- Evaluate: Temp-only → Temp+H3 → Temp+H3+LLM

Our Contribution

- First study to examine POI-level next-day forecasting in disaster settings over post-event horizons
- First **systematic diagnostic study** of LLM fusion under **disaster conditions**
- Insights into **when multimodal signals help or hurt**, guiding future disaster-aware model design.

LSU

Problem Formulation

- For each POI p, we observe past visits:

$$v_{k-n+1:k}^{(p)}$$

- Goal: predict future visit:

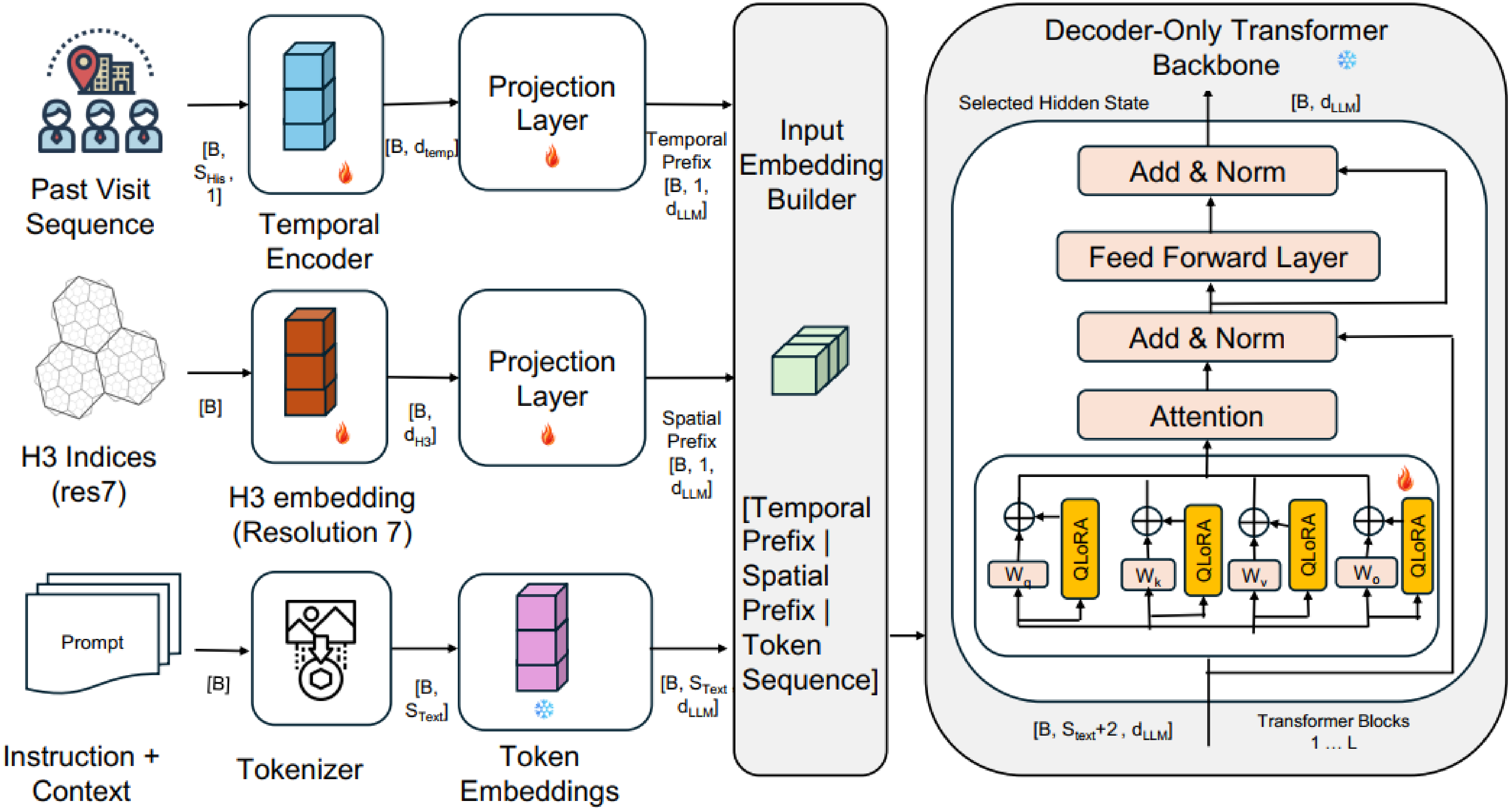
$$\hat{v}_{k+h}^{(p)}$$

- Mapping:

$$f\left(v_{k-n+1:k}^{(p)}, s_p, g_p\right) \rightarrow \hat{v}_{k+h}^{(p)}$$

where s_p =semantic features, g_p =H3 spatial index.

Methodology: Architecture of Disaster-Cast



LSU

The Disaster-Cast Framework

- Hybrid Architecture: Integrates three key components.
  - Temporal Encoder: Trainable encoder (CNN/GRU) for numerical history.
  - Spatial Embedding: Trainable H3 index for geographic context.
  - Semantic Encoder: LoRA-augmented LLM for structured hard prompts.
- Numeric visit sequences are excluded from the LLM prompt to improve interpretability and generalization

Technical Detail - Temporal Encoder

- Input: The recent visit history for each POI.
- Models: lightweight ID-CNN, GRU, and LSTM encoders.
- Purpose: Extracts short-term mobility dynamics by learning patterns such as drops, rebounds, and local fluctuations in visit counts.
- Output: A temporal feature vector representing recent behavioral trends.
- Integration: This vector is linearly projected into the LLM’s hidden space so it can be fused with semantic and spatial information.

Technical Detail- H3 Spatial Encoder

- H3 is Uber’s open-source hexagonal spatial indexing system.
- It partitions the earth into hexagons across 16 hierarchical resolutions (0 = largest, 15 = smallest).
- We use Resolution 7, where each hexagon covers ~5 km^2, providing neighborhood-scale spatial granularity.
- H3 embeddings are learned per H3 cell, not per POI.
- POIs inside the same H3-7 hexagon share the same spatial vector.
- The model generalizes to new POIs only if their H3 cell appeared in training.
- These embeddings provide spatial context, helping the model capture location-dependent recovery after Hurricane Ian.
- Project this spatial vector into the LLM space for multimodal fusion.

Technical Detail- LLM Fusion

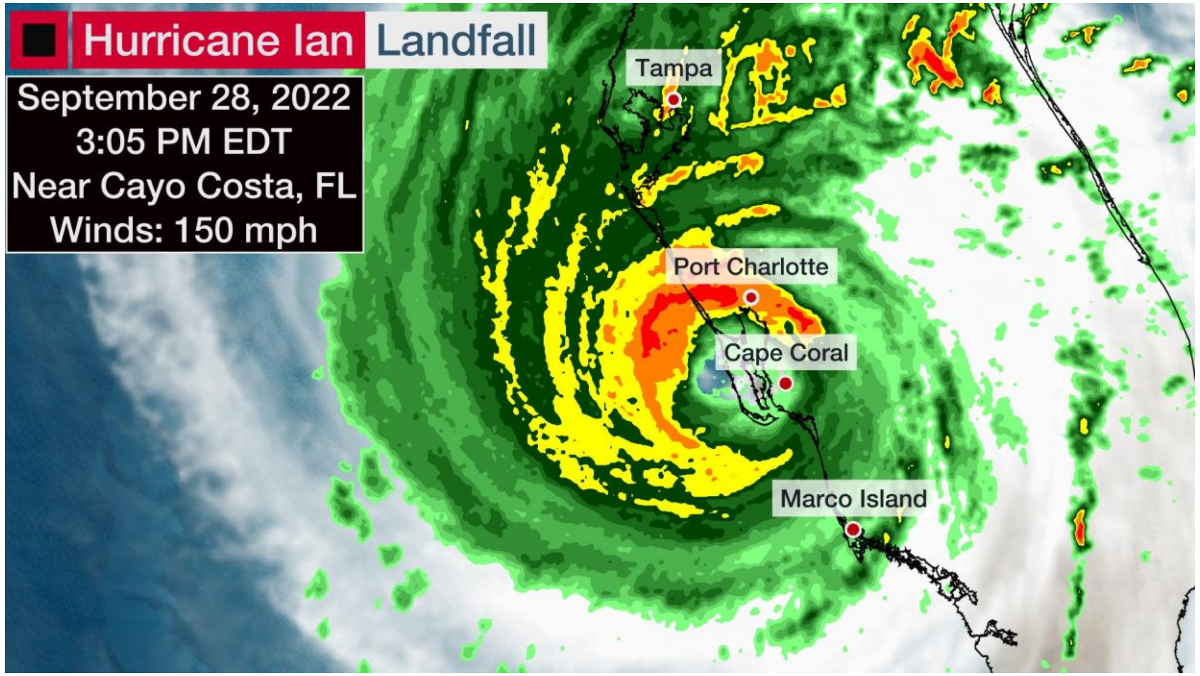
- We use structured prompts with category, city, and hurricane context.
- Temporal and spatial (H3) vectors are projected into the LLM's hidden space and added as a prefix to the prompt.
- The LLM (QLoRA-tuned) processes this multimodal prefix + text.
- A small regression head converts the LLM output into the final visit prediction.

Technical Detail - Training

- We fine-tune the LLM using QLoRA, which allows efficient training without updating the full model.
- The model is trained end-to-end, including the temporal encoder, H3 embedding, projection layers, and regression head.
- Training combines two objectives:
  - (1) Regression loss — encourages accurate numerical visit predictions.
  - (2) Language-model loss — helps the LLM stay consistent with the structured prompt format.
- A weighting parameter balances the numeric prediction and language modeling during training

Experimental Setup

- Event: Hurricane Ian in Florida.
- Dataset: 67,182 POIs from SafeGraph.
- Horizon: Day 14(4 days after landfall), Day 21 (11 days after Landfall)
- Series Date Started: Sep 19, 2022
- Landfall Date: Sep 28, 2022
- Series End Date: Oct 9, 2022
- City heterogeneity:
    - Severely impacted: Cape Coral
    - Mildly to Moderately Impacted: Orlando, Tampa, Miami
    - Not Significantly impacted: Jacksonville

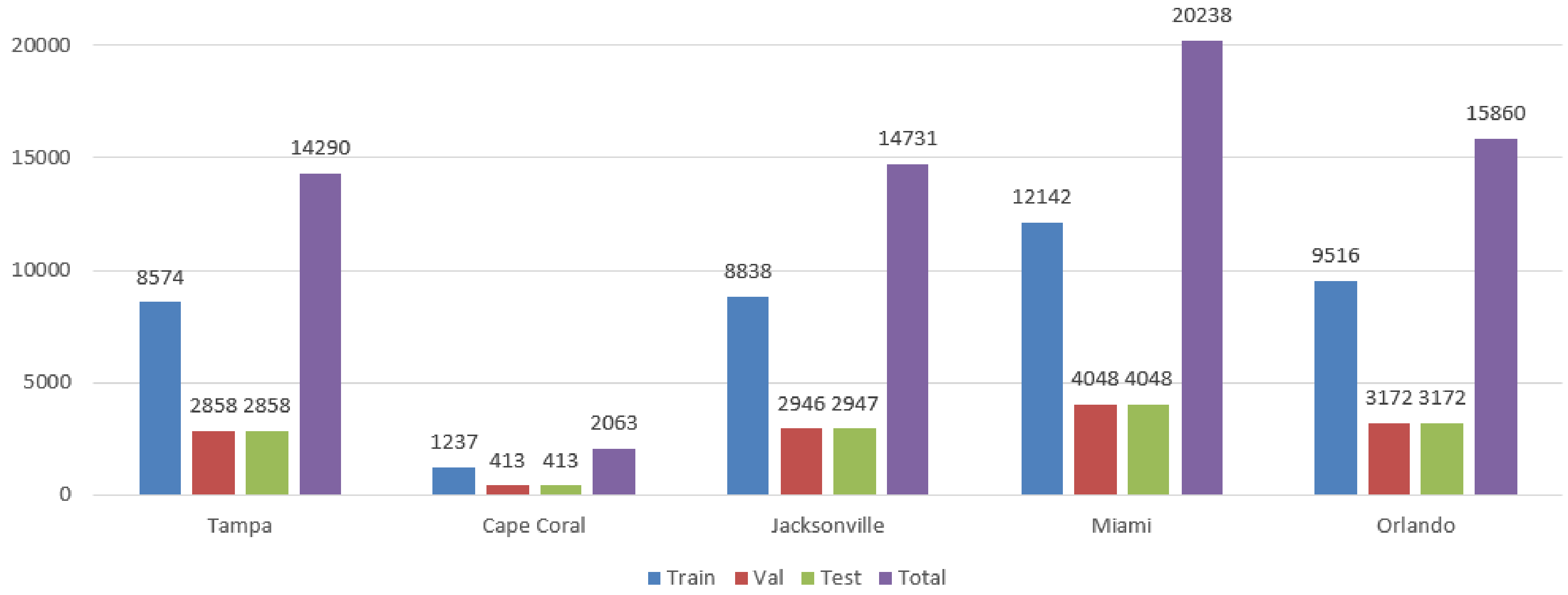


https://weather.com/storms/hurricane/news/2022-09-30-hurricane-ian-forecast-landfall-south-north-carolina-virginia

Data Distribution and POI based Split

	Day 14 Visit Ranges (%)									
City	0	1–5	6–10	11–20	21–50	51–100	101–200	201–500	501–1000	1000+
Cape Coral	41.8	42.3	7.6	4.5	2.3	1.0	0.5	0.0	0.0	0.0
Jacksonville	40.1	38.5	9.1	6.4	4.1	1.2	0.4	0.1	0.0	0.0
Miami	38.7	40.5	9.5	6.3	3.6	0.8	0.3	0.2	0.0	0.0
Orlando	34.8	38.3	9.9	7.7	5.9	2.0	0.9	0.3	0.1	0.1
Tampa	39.8	40.6	8.7	5.9	3.5	1.0	0.3	0.2	0.1	0.0

	Day 21 Visit Ranges (%)									
City	0	1–5	6–10	11–20	21–50	51–100	101–200	201–500	501–1000	1000+
Cape Coral	35.8	40.2	8.8	7.1	5.6	1.5	0.9	0.2	0.0	0.0
Jacksonville	39.9	38.3	8.9	6.7	4.2	1.3	0.4	0.1	0.0	0.0
Miami	39.7	41.0	8.8	5.9	3.3	0.7	0.3	0.2	0.0	0.0
Orlando	36.0	36.2	9.6	7.8	6.4	2.2	1.0	0.5	0.1	0.1
Tampa	40.3	39.2	9.0	6.1	3.7	1.1	0.3	0.2	0.0	0.0



Results & Analysis:

Performance Of Statistical, Neural, And Multimodal Models For Day-14 And Day-21 POI Visit Forecasting (Cape Coral, Orlando and Tampa)

City	Model	Day 14			Day 21		
		MAE	RMSE	RMSLE	MAE	RMSE	RMSLE
Cape Coral	Auto ARIMA	3.02	**6.89**	0.81	**3.05**	**6.08**	0.72
	Prophet	3.06	8.85	0.81	4.36	11.96	0.93
	CNN	2.89	8.16	0.71	4.42	8.55	0.94
	GRU	**2.83**	8.45	**0.70**	4.12	14.26	**0.66**
	GRU+CNN	2.96	8.17	0.73	3.83	9.78	0.78
	LSTM	3.00	9.01	**0.70**	4.60	15.82	0.75
	Disaster-Cast	2.90	7.57	0.72	4.74	13.49	0.85
Orlando	Auto ARIMA	4.66	23.44	0.75	**5.78**	**79.12**	0.75
	Prophet	4.76	26.91	0.69	8.99	166.09	0.76
	CNN	**3.72**	**13.47**	0.69	7.77	150.96	0.60
	GRU	5.73	71.26	0.65	24.57	831.23	0.63
	GRU+CNN	3.65	14.39	**0.64**	6.97	98.14	0.76
	LSTM	6.08	70.97	0.67	25.36	832.69	0.65
	Disaster-Cast	5.29	67.34	**0.64**	24.64	827.21	**0.59**
Tampa	Auto ARIMA	4.21	76.34	0.75	3.61	23.09	0.76
	Prophet	4.40	72.04	0.71	3.80	22.98	0.80
	CNN	**3.87**	64.94	0.70	3.50	**19.84**	0.69
	GRU	4.91	85.69	**0.65**	3.18	24.96	0.67
	GRU+CNN	3.89	**57.04**	0.70	**3.06**	22.11	**0.62**
	LSTM	4.96	86.20	**0.65**	3.35	25.63	**0.62**
	Disaster-Cast	4.67	82.70	0.71	3.46	23.75	0.79

Why Do Models Behave This Way?

- CNN → Best overall
    - Disaster sequences = short, noisy, 0–20 range
    - CNN captures local patterns without drift
    - Stable, low-variance predictions
- GRU / LSTM → Unstable
    - Long-horizon memory not useful here
    - Noise accumulation on heavy-tailed POIs
    - Large RMSE/RMSLE spikes
- Multimodal LLM Fusion → Mixed
    - Semantic signals are weak vs. temporal features
    - Prefix fusion can dilute strong temporal cues
    - Helps RMSLE (relative error)
    - Hurts MAE/RMSE in noisy, zero-heavy cities

Templates for Zero-shot LLM

Templ.	Prompt (Full Snippet)
A	`You are an expert time-series forecaster. Given the historical daily visit counts, predict the next day's visit count. Return ONLY: PREDICTION: <number>. DailyVisits: {values}. PREDICTION:`
B	`You are an expert in disaster-aware mobility forecasting. Predict the visit count for the target date using the provided historical daily visits and the hurricane landfall time. Do NOT explain your reasoning. Return ONLY: PREDICTION: <number>. Location: {location_name}. City: {city}. SeriesStart: {series_start}. HurricaneLandfall: {landfall}. TargetDate: {target_date}. DailyVisits: {values}. PREDICTION:`
C	`You are an expert in human-mobility forecasting after natural disasters. Predict the visit count for the target date by combining temporal trends, hurricane disruption, and POI semantics. Return ONLY: PREDICTION: <number>. === CONTEXT === LocationName: {location_name}. BusinessCategory: {business_category}. City: {city}. SeriesStartDate: {series_start}. HurricaneLandfallDate: {landfall}. TargetDateToPredict: {target_date}. PastDailyVisits: {values}. === END === PREDICTION:`

Zero-shot LLM Performance

- ❑ Template A works best; adding hurricane/semantic text (B, C) often reduces accuracy.
- ❑ Mistral > Llama on Template A because it naturally focuses on recent numbers.
- ❑ Llama's global attention makes it overly sensitive to extra text, causing large errors (especially in high-variance cities).
- ❑ Template C sometimes helps Llama when semantic business info is relevant.
- ❑ Overall, zero-shot LLMs struggle to merge textual context with numeric time-series.

City	Model	Temp	Day 14 MAE	Day 14 RMSE	Day 14 RMSLE	Day 21 MAE	Day 21 RMSE	Day 21 RMSLE
Cape Coral	Llama	A	7.32	22.04	1.23	8.36	33.68	1.17
		B	**4.52**	14.58	1.35	7.94	25.66	1.64
		C	5.00	15.50	1.36	10.17	29.79	1.71
	Mistral	A	4.91	**12.27**	**1.13**	**3.54**	**5.74**	**0.94**
		B	13.04	26.61	1.67	9.83	32.49	1.28
		C	5.01	17.17	1.27	7.11	22.06	1.60
Orlando	Llama	A	6.80	28.31	1.00	16.00	301.57	1.04
		B	10.44	79.29	1.69	27.52	772.49	1.80
		C	10.20	44.71	1.67	30.64	780.01	1.80
	Mistral	A	**5.69**	**26.95**	**0.94**	**7.28**	**64.28**	**0.97**
		B	14.06	119.43	1.50	14.09	231.24	1.18
		C	9.27	78.87	1.42	21.54	421.68	1.66
Tampa	Llama	A	6.58	75.88	1.04	6.25	35.30	1.01
		B	7.30	88.04	1.43	6.25	31.01	1.47
		C	8.24	89.89	1.46	7.90	41.99	1.49
	Mistral	A	**5.49**	**75.72**	**0.97**	5.91	34.05	**1.00**
		B	12.80	113.77	1.54	8.62	37.19	1.25
		C	7.41	89.23	1.30	**5.79**	**29.78**	1.38

Ablation study

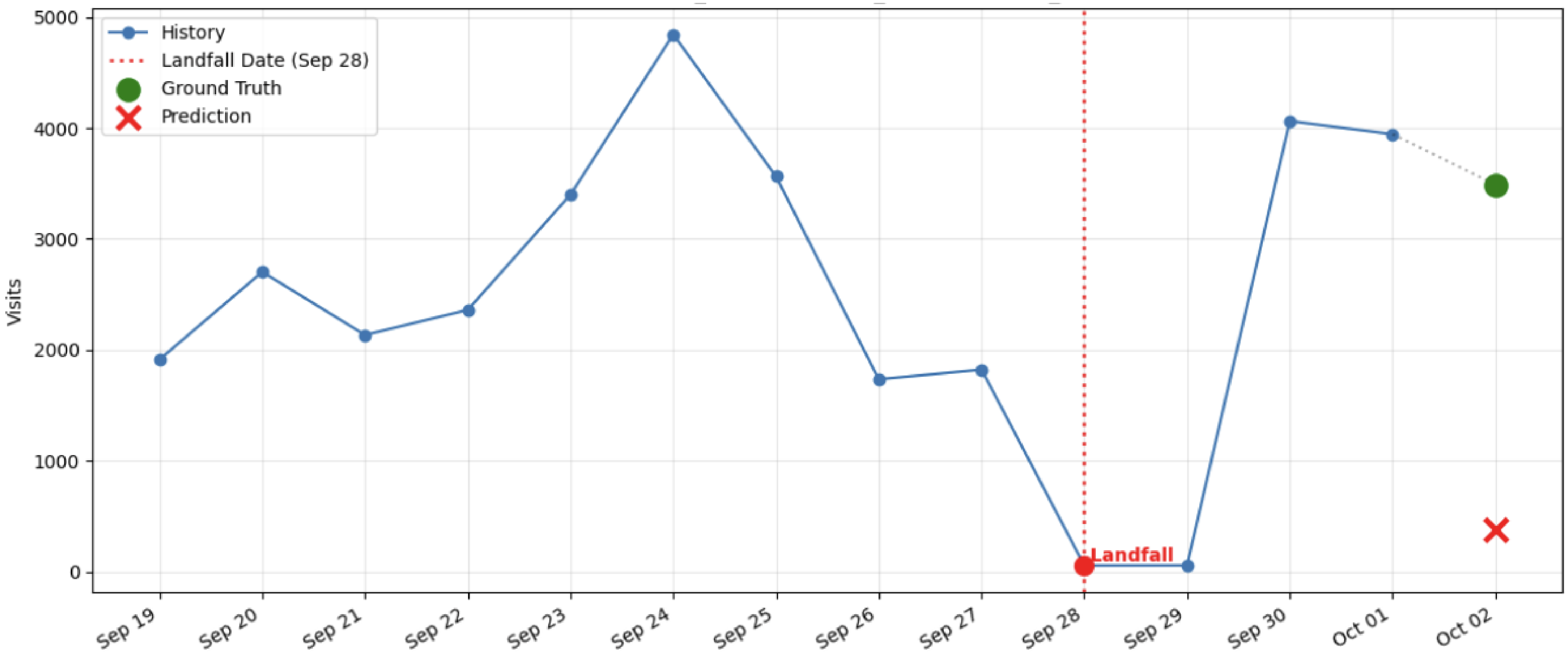
- ❑ CNN is the strongest model across most cities and horizons.
- ❑ H3 helps selectively, mainly where recovery patterns are spatially coherent.
- ❑ Adding LLM embeddings (Llama/Mistral) hurts performance and increases RMSE.
- ❑ LLM fusion is especially unstable in high-variance cities (Orlando, Tampa).

City	Model	Day 14 MAE	Day 14 RMSE	Day 14 RMSLE	Day 21 MAE	Day 21 RMSE	Day 21 RMSLE
Cape Coral	CNN	**2.89**	8.16	**0.71**	4.42	8.55	0.94
	CNN+H3	3.15	**6.51**	0.83	3.69	**6.93**	0.86
	CNN+H3+Llama	2.90	7.57	0.72	4.74	13.49	0.85
	CNN+H3+Mistral	2.92	6.55	0.76	**3.70**	8.81	**0.78**
Orlando	CNN	**3.72**	**13.47**	0.69	**7.77**	**150.96**	0.60
	CNN+H3	3.83	13.54	0.74	11.19	284.05	0.70
	CNN+H3+Llama	5.29	67.34	**0.64**	24.64	827.21	**0.59**
	CNN+H3+Mistral	5.85	65.74	0.83	25.17	828.19	0.67
Tampa	CNN	3.87	**64.94**	0.70	3.50	19.84	**0.69**
	CNN+H3	**3.82**	65.20	**0.67**	**2.82**	**16.56**	**0.69**
	CNN+H3+Llama	4.67	82.70	0.71	3.46	23.75	0.79
	CNN+H3+Mistral	4.59	81.09	0.72	3.54	23.58	0.85

Category Wise Error (Case Study: Orlando)

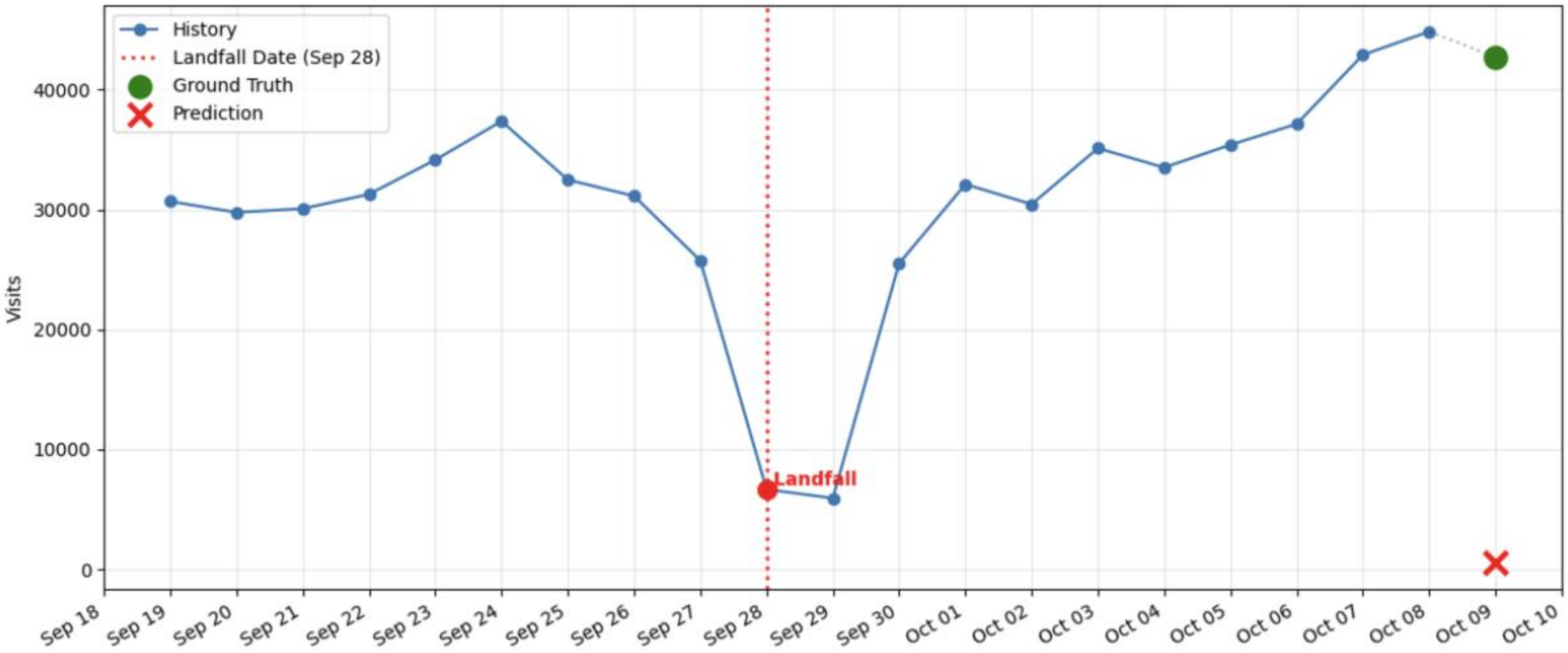
Business Category	count	mean	median	Percent
Amusement and Theme Parks	51	1238.82	17	1.61
Malls	31	34.48	17	0.98
Hotels (except Casino Hotels) and Motels	77	32.35	5	2.43
All Other General Merchandise Stores	29	10.28	5	0.91
Drinking Places (Alcoholic Beverages)	53	7.77	3	1.67

Worst Case (Case: Orlando; Day 14)



The prediction of our Disaster-Cast(CNN+H3+Llama) for the Day-14 horizon at *Disney Springs (Malls category): POI- zzw- 223@8fy-8kn-j35* in Orlando. The ground-truth value is 3,490; the model predicts 375, producing absolute errors of 3,115.

Worst Case (Case: Orlando; Day 21)



The prediction of our Disaster-Cast(CNN+H3+Llama) for the Day-21 horizon at *Walt Disney World Resort: POI-zzw-222@8fyfjg-b8v* in Orlando. Actual visits surge to 42,732 during the late-stage recovery period. Llama predicts 561, resulting in absolute errors of 42,171

Future Direction 1: Improve Accuracy of Multimodal LLM

- **Smarter fusion instead of simple concatenation**
- Replace naive fusion with a **learned gating or attention layer**:
  - Gate decides *how much* to trust temporal CNN vs H3 graph vs LLM output for each POI.
  - Example: For small stable POIs, rely more on CNN; for volatile heavy-tail POIs, LLM + H3 gets higher weight.
- Make the LLM predict corrections, not full values
  - Use the LLM to predict residuals on top of CNN/H3:
  - Base forecast = CNN/CNN+H3.
  - LLM learns to correct systematic biases (e.g., underestimation on stadiums, malls, airports).
  - This uses LLM capacity where classical models struggle instead of replacing them.

Future Direction 2: Better Signals, Data and Evaluation

- Richer disaster signals
  - Integrate more granular evacuation / closure / power-outage data.
  - Include pre- and post-hurricane multi-week context so model sees the full "collapse and recovery" shape.

- Cross-hurricane generalization
  - Extend beyond Hurricane Ian and design the model so it can generalize to future hurricanes and new cities.
  - Use the LLM to encode higher-level semantics (e.g., business category, narrative evacuation descriptions) for transfer

Research Direction 3: POI ID Embeddings

- Re-scope task: **fixed set of POIs**, generalize across **time**, not entities
- Use **trainable POI ID embeddings**:
  - capture typical scale & variability
  - capture POI-specific disaster sensitivity
- Combine with:
  - temporal encoder (history)
  - H3 spatial embedding
  - semantic LLM prompt
- Expected benefit:
  - **more stable training**
  - better modeling of **high-volume / special POIs** (airports, stadiums, malls)

LSU

Plan & Expected Outcomes

- Short-term
  - Implement residual + gated fusion architecture (CNN/H3 + LLM).
  - Run ablations to see when LLM improves RMSLE, especially for heavy-tailed POIs (Orlando, Tampa, Miami).
  - Re-scope task: fixed set of POIs, generalize across time, not entities
- Medium-term
  - Integrate additional disaster-related signals and test cross-hurricane settings.
  - Publish a paper focused on cost-aware, disaster-aware POI forecasting

Conclusion

- Introduced Disaster-Cast, a multimodal framework combining temporal, spatial (H3), and semantic (LLM) signals.
- Results highlight the current limitations of LLM fusion for POI-level forecasting.
- Provides new insights into when multimodal signals help or hurt, guiding future disaster-aware modeling.

References

- C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” Science, vol. 327, pp. 1018–1021, 2010.
- M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo, “A survey on deep learning for human mobility,” ACM Computing Surveys, vol. 55, no. 1, pp. 1–44, 2021.
- Y. Tang, H. Wang, X. Fan, and Y. Li, “Predicting human mobility in disasters via LLM-enhanced cross-city learning,” arXiv preprint arXiv:2507.19737, 2025.
- H. Xue, F. D. Salim, Y. Ren, and C. L. A. Clarke, “Translating human mobility forecasting through natural language generation,” in Proc. WSDM ’22, pp. 1224–1233, 2022
- H. Xue, B. P. Voutharoja, and F. D. Salim, “Leveraging language foundation models for human mobility forecasting,” in Proc. SIGSPATIAL ’22, pp. 1–9, 2022.
- H. Xue and F. D. Salim, “PromptCast: A new prompt-based learning paradigm for time series forecasting,” IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 11, pp. 6851–6864, 2024.[11] H. Xue, T. Tang, A. Payani, and F. D. Salim, “Prompt mining for language models-based mobility flow forecasting,” in Proc. ACM SIGSPATIAL ’24, 2024.
- H. Xue, T. Tang, A. Payani, and F. D. Salim, “Prompt mining for language models-based mobility flow forecasting,” in *Proc. ACM SIGSPATIAL ’24*, 2024.
- L. Gong *et al.*, “Mobility-LLM: Learning visiting intentions and travel preferences from human mobility data with large language models,” in *Proc. NeurIPS ’24*, 2025.
- C. Liu *et al.*, “Spatial-temporal large language model for traffic prediction,” in *Proc. IEEE MDM*, pp. 31–40, 2024.
- M. Jin *et al.*, “Time-LLM: Time series forecasting by reprogramming large language models,” in *International Conference on Learning Representations (ICLR)*, 2024.
- D. Cao *et al.*, “TEMPO: Prompt-based generative pre-trained transformer for time series forecasting,” *arXiv preprint arXiv:2310.04948*, 2023.
- C. Chang, W.-Y. Wang, W.-C. Peng, and T.-F. Chen, “LLM4TS: Aligning pre-trained LLMs as data-efficient time-series forecasters,” *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 3, 2025.
- T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “QLoRA: Efficient finetuning of quantized LLMs,” *arXiv preprint arXiv:2305.14314*, 2023.
- Meta AI, “Meta Llama 3 8B,” Hugging Face model card, 2024. (Accessed: 2025-10-10).
- Mistral AI, “Mistral 7B v0.1,” Hugging Face model card, 2023. (Accessed: 2025-10-10).
- SafeGraph, “SafeGraph data,” 2020. (Accessed: 2025-09-10).

Thank you!
Any Suggestions?

Appendices

Performance of All Cities

City	Model	Day 14			Day 21		
		MAE	RMSE	RMSLE	MAE	RMSE	RMSLE
Cape Coral	Auto ARIMA	3.02	**6.89**	0.81	**3.05**	**6.08**	0.72
	Prophet	3.06	8.85	0.81	4.36	11.96	0.93
	CNN	2.89	8.16	0.71	4.42	8.55	0.94
	GRU	**2.83**	8.45	**0.70**	4.12	14.26	**0.66**
	GRU+CNN	2.96	8.17	0.73	3.83	9.78	0.78
	LSTM	3.00	9.01	**0.70**	4.60	15.82	0.75
	Disaster-Cast	2.90	7.57	0.72	4.74	13.49	0.85
Jacksonville	Auto ARIMA	3.16	7.10	0.79	3.51	20.47	0.79
	Prophet	2.45	**6.14**	0.66	3.90	21.11	0.84
	CNN	**2.44**	6.38	0.62	3.14	19.68	0.71
	GRU	3.19	12.13	0.69	3.29	21.21	0.67
	GRU+CNN	3.00	7.44	0.77	**2.66**	**19.63**	**0.59**
	LSTM	2.98	11.90	0.67	3.29	21.85	0.67
	Disaster-Cast	2.63	9.57	**0.59**	3.14	20.46	0.65
Miami	Auto ARIMA	4.02	23.35	0.83	3.55	15.38	0.79
	Prophet	2.95	13.18	0.73	3.74	16.78	0.79
	CNN	**2.37**	**7.58**	0.63	**2.35**	**7.67**	**0.59**
	GRU	3.32	27.45	0.69	3.96	63.33	0.65
	GRU+CNN	2.55	13.82	**0.60**	2.77	16.39	0.61
	LSTM	3.16	26.69	0.62	3.87	63.62	0.61
	Disaster-Cast	2.92	22.83	0.62	3.52	60.25	**0.59**
Orlando	Auto ARIMA	4.66	23.44	0.75	**5.78**	**79.12**	0.75
	Prophet	4.76	26.91	0.69	8.99	166.09	0.76
	CNN	**3.72**	**13.47**	0.69	7.77	150.96	0.60
	GRU	5.73	71.26	0.65	24.57	831.23	0.63
	GRU+CNN	3.65	14.39	**0.64**	6.97	98.14	0.76
	LSTM	6.08	70.97	0.67	25.36	832.69	0.65
	Disaster-Cast	5.29	67.34	**0.64**	24.64	827.21	**0.59**
Tampa	Auto ARIMA	4.21	76.34	0.75	3.61	23.09	0.76
	Prophet	4.40	72.04	0.71	3.80	22.98	0.80
	CNN	**3.87**	64.94	0.70	**3.50**	**19.84**	0.69
	GRU	4.91	85.69	**0.65**	3.18	24.96	0.67
	GRU+CNN	3.89	**57.04**	0.70	**3.06**	22.11	**0.62**
	LSTM	4.96	86.20	**0.65**	3.35	25.63	**0.62**
	Disaster-Cast	4.67	82.70	0.71	3.46	23.75	0.79

Zero Shot Performance

TABLE IV: Zero-shot LLM performance using three prompt templates for Day-14 and Day-21 POI visit forecasting. Best values per city–horizon in **blue bold**.

City	Model	Temp	Day 14 MAE	Day 14 RMSE	Day 14 RMSLE	Day 21 MAE	Day 21 RMSE	Day 21 RMSLE
Cape Coral	Llama	A	7.32	22.04	1.23	8.36	33.68	1.17
		B	**4.52**	14.58	1.35	7.94	25.66	1.64
		C	5.00	15.50	1.36	10.17	29.79	1.71
	Mistral	A	4.91	**12.27**	**1.13**	**3.54**	**5.74**	**0.94**
		B	13.04	26.61	1.67	9.83	32.49	1.28
		C	5.01	17.17	1.27	7.11	22.06	1.60
Jacksonville	Llama	A	5.03	13.27	1.04	**5.51**	**25.84**	**1.05**
		B	5.72	20.04	1.49	6.08	27.26	1.48
		C	8.19	56.48	1.50	11.73	51.22	1.54
	Mistral	A	**4.37**	**12.64**	**0.57**	5.62	26.34	**1.05**
		B	10.36	29.79	1.50	7.99	35.75	1.23
		C	4.67	14.87	1.25	5.66	25.87	1.39
Miami	Llama	A	5.65	30.57	1.01	**5.46**	30.20	**0.96**
		B	5.86	34.29	1.42	6.36	65.83	1.41
		C	6.70	38.18	1.45	6.69	**28.39**	1.43
	Mistral	A	**5.12**	31.20	**0.96**	**5.46**	30.13	**0.96**
		B	10.60	34.81	1.51	7.62	31.37	1.18
		C	5.18	**24.91**	1.21	6.65	67.56	1.32
Orlando	Llama	A	6.80	28.31	1.00	16.00	301.57	1.04
		B	10.44	79.29	1.69	27.52	772.49	1.80
		C	10.20	44.71	1.67	30.64	780.01	1.80
	Mistral	A	**5.69**	**26.95**	**0.94**	**7.28**	**64.28**	**0.97**
		B	14.06	119.43	1.50	14.09	231.24	1.18
		C	9.27	78.87	1.42	21.54	421.68	1.66
Tampa	Llama	A	6.58	75.88	1.04	6.25	35.30	1.01
		B	7.30	88.04	1.43	6.25	31.01	1.47
		C	8.24	89.89	1.46	7.90	41.99	1.49
	Mistral	A	**5.49**	**75.72**	**0.97**	5.91	34.05	**1.00**
		B	12.80	113.77	1.54	8.62	37.19	1.25
		C	7.41	89.23	1.30	**5.79**	**29.78**	1.38

Ablation Study

City	Model	Day 14			Day 21		
		MAE	RMSE	RMSLE	MAE	RMSE	RMSLE
Cape Coral	CNN	**2.89**	8.16	**0.71**	4.42	8.55	0.94
	CNN+H3	3.15	**6.51**	0.83	3.69	**6.93**	0.86
	CNN+H3+Llama	2.90	7.57	0.72	4.74	13.49	0.85
	CNN+H3+Mistral	2.92	6.55	0.76	**3.70**	8.81	**0.78**
Jacksonville	CNN	**2.44**	6.38	0.62	3.14	19.68	0.71
	CNN+H3	2.84	**6.36**	0.74	**2.86**	**19.54**	**0.65**
	CNN+H3+Llama	2.63	9.57	**0.59**	3.14	20.46	**0.65**
	CNN+H3+Mistral	4.18	13.78	1.01	3.24	21.14	0.66
Miami	CNN	**2.37**	**7.58**	0.63	2.35	**7.67**	**0.59**
	CNN+H3	2.39	7.83	0.65	**2.32**	10.41	0.60
	CNN+H3+Llama	2.92	22.83	**0.62**	3.52	60.25	**0.59**
	CNN+H3+Mistral	4.85	26.30	1.18	3.49	59.78	**0.59**
Orlando	CNN	**3.72**	**13.47**	0.69	**7.77**	**150.96**	0.60
	CNN+H3	3.83	13.54	0.74	11.19	284.05	0.70
	CNN+H3+Llama	5.29	67.34	**0.64**	24.64	827.21	**0.59**
	CNN+H3+Mistral	5.85	65.74	0.83	25.17	828.19	0.67
Tampa	CNN	3.87	**64.94**	0.70	3.50	19.84	**0.69**
	CNN+H3	**3.82**	65.20	**0.67**	**2.82**	**16.56**	**0.69**
	CNN+H3+Llama	4.67	82.70	0.71	3.46	23.75	0.79
	CNN+H3+Mistral	4.59	81.09	0.72	3.54	23.58	0.85

Core Tools, Platforms, Hyperparameters Used in this Research

Category	Setting
Programming Environment	Python 3.12; `pyenv`
Frameworks	PyTorch; HF Transformers; PEFT (QLoRA)
LLM Backbones	Llama-3 8B (4-bit), Mistral 7B v0.1
Temporal Encoders	GRU / LSTM / 1D-CNN/ GRU+CNN
Spatial Modeling	Trainable H3 embeddings (64-dim)
Fusion Strategy	Prefix fusion into LLM hidden space
Optimizer	AdamW
Batch Size	4
Learning Rate	1×10^{-4}
Training Epochs	3
Quantization	QLoRA (NF4 4-bit)
Zero-Shot Decoding	Greedy; max_new_tokens = 30
Data Processing	NumPy, pandas, scikit-learn
Visualization	Matplotlib, Seaborn
Hardware	NVIDIA A100 (80 GB) GPUs (LONI HPC)

Performance Metrics

a) Mean Absolute Error (MAE).: MAE measures the average magnitude of prediction error and is robust to outliers:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|. \tag{14}$$

b) Root Mean Squared Error (RMSE).: RMSE penalizes larger deviations more strongly, making it sensitive to mis-estimation of high-traffic POIs:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}. \tag{15}$$

c) Root Mean Squared Logarithmic Error (RMSLE).: RMSLE measures error in log space and emphasizes accurate estimation of relative change:

$$\mathrm{RMSLE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(1+\hat{y}_i) - \log(1+y_i))^2}. \tag{16}$$

d) Coefficient of Determination (R^2).: R^2 quantifies the proportion of variance in y explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{17}$$

where $\bar{y}$ is the mean visit count.