Collective Narrative Grounding: Community-Coordinated Data Contributions to Improve Local AI Systems

Zihan Gao[1], Mohsin Y. K. Yousufi[2], Jacob Thebault-Spieker[1]

[1]University of Wisconsin-Madison | [2]Georgia Tech

Motivation: Local Blind Spots in Global LLMs

Global AI, Local Failure

- LLMs excel at general knowledge but struggle with information tied to specific communities or places.
- These "blind spots" create **Epistemic Injustice**: marginalized communities are undervalued, and misinformation fills the void.
- **Key Insight:** This is not just a technical gap, but a failure of training data priority.

Auditing the Gap

76.7% of Errors are Fixable

- We study LocalBench: 14,782 local QA pairs from 526 U.S. counties.
- Questions span census data, local news, and community discourse, across physical, cognitive, and relational localness.
- From 1,000 audited failures, four dominant error modes emerge:
  - Factual knowledge gaps (31.8%)
  - Cultural misunderstandings (23.4%)
  - Geographic confusion (12.4%)
  - Temporal misalignment (9.1%)
  - Examples in ***Figure (a) Local blind spots of global LLMs***
- These four together account for 76.7% of errors and are directly addressable with locally grounded narratives.

LocalBench

Model	Non-Numerical QA EM	ROUGE-1	Semantic	GPT Judge	Ans Rate	Numerical QA Accuracy	Ans Rate
GPT-4o	22.0	30.7	53.0	32.8	99.6	6.2	39.8
GPT-4.1	**32.2**	**52.5**	**74.1**	47.0	100.0	6.2	100.0
GPT-4.1+Web	13.5	27.9	43.2	35.6	92.9	**15.5**	92.0
Gemini-2.5-Pro	28.0	52.0	70.5	52.5	100.0	12.8	100.0
Gemini-2.5-Flash	31.1	46.0	67.6	43.2	100.0	7.5	100.0
Gemini-2.5-Pro+Grounding	21.9	50.1	66.0	**56.8**	91.7	12.8	100.0
Claude-Sonnet-4	23.4	38.5	64.0	39.7	100.0	7.1	97.3
Claude-Sonnet-3.7	21.7	42.5	65.5	43.7	100.0	8.4	91.2
Qwen3-235B-A22B	19.9	29.0	54.0	27.3	99.3	6.6	77.0
Qwen3-30B-A3B	20.5	29.6	54.9	28.0	99.7	2.2	100.0
Qwen3-32B	20.0	29.9	55.4	27.7	99.7	4.9	99.1
Qwen3-14B	19.5	29.8	55.4	27.5	99.6	4.0	100.0
Qwen3-8B	16.3	27.2	54.1	22.9	99.6	3.1	75.2

Performance Results Across Non-Numerical and Numerical QA

References

Gao, Zihan, Yifei Xu, and Jacob Thebault-Spieker. "LocalBench: Benchmarking LLMs on County-Level Local Knowledge and Reasoning. " AAAI 2026. https://arxiv.org/pdf/2511.10459

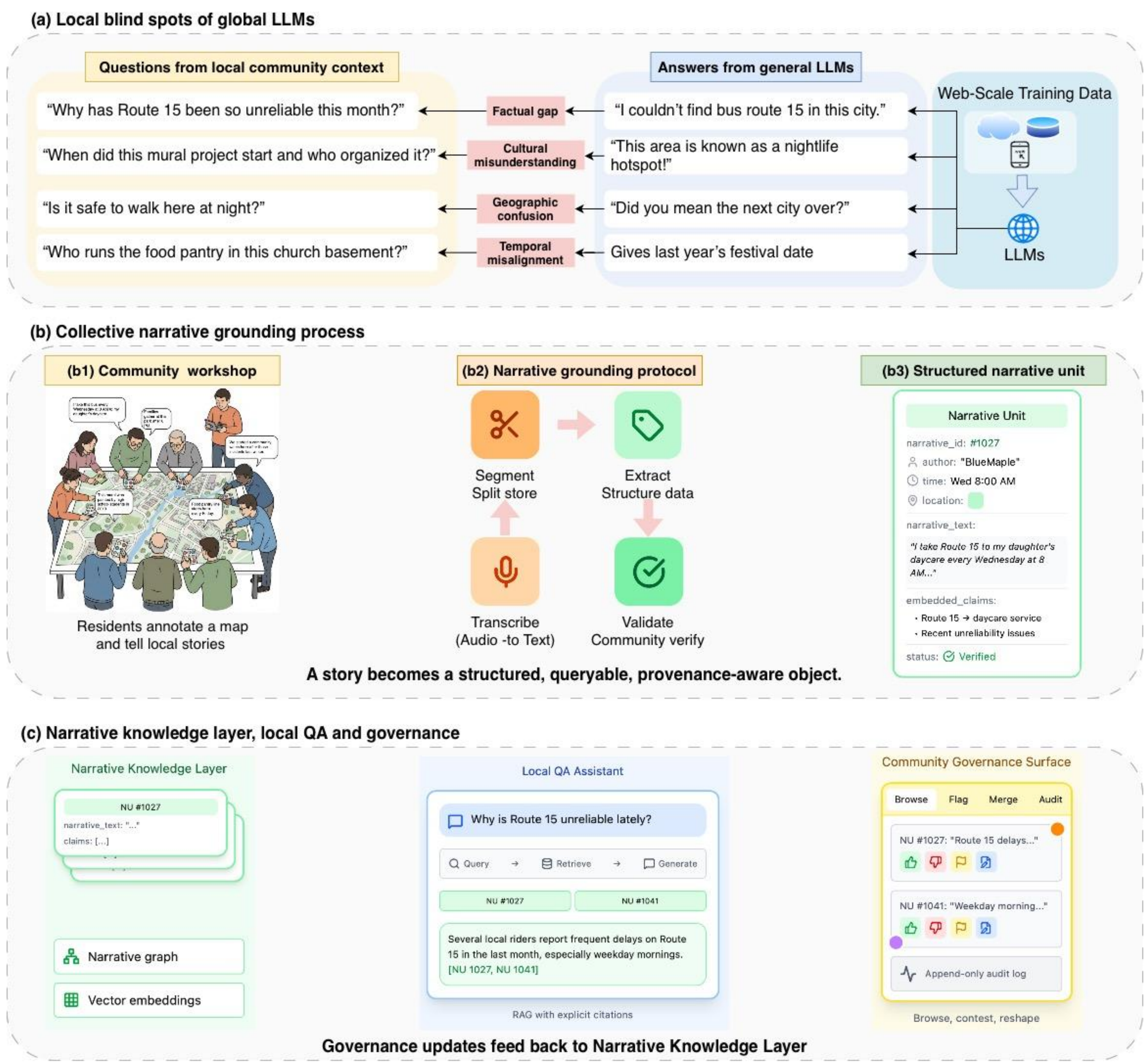
Yousufi, Mohsin YK, Charlotte Alexander, and Nassim Parvin. "Credibility Boosters as a Lens for Understanding Epistemic Injustice in Civic Tech: The Case of Heat Seek." Proceedings of the ACM on Human-Computer Interaction 9.7 (2025): 1-30.

Gao, Zihan, et al. "From Clips to Communities: Fusing Social Video into Knowledge Graphs for Localness-Aware LLMs." Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing. 2025.

The Narrative Grounding Protocol

From Stories to Structured Data

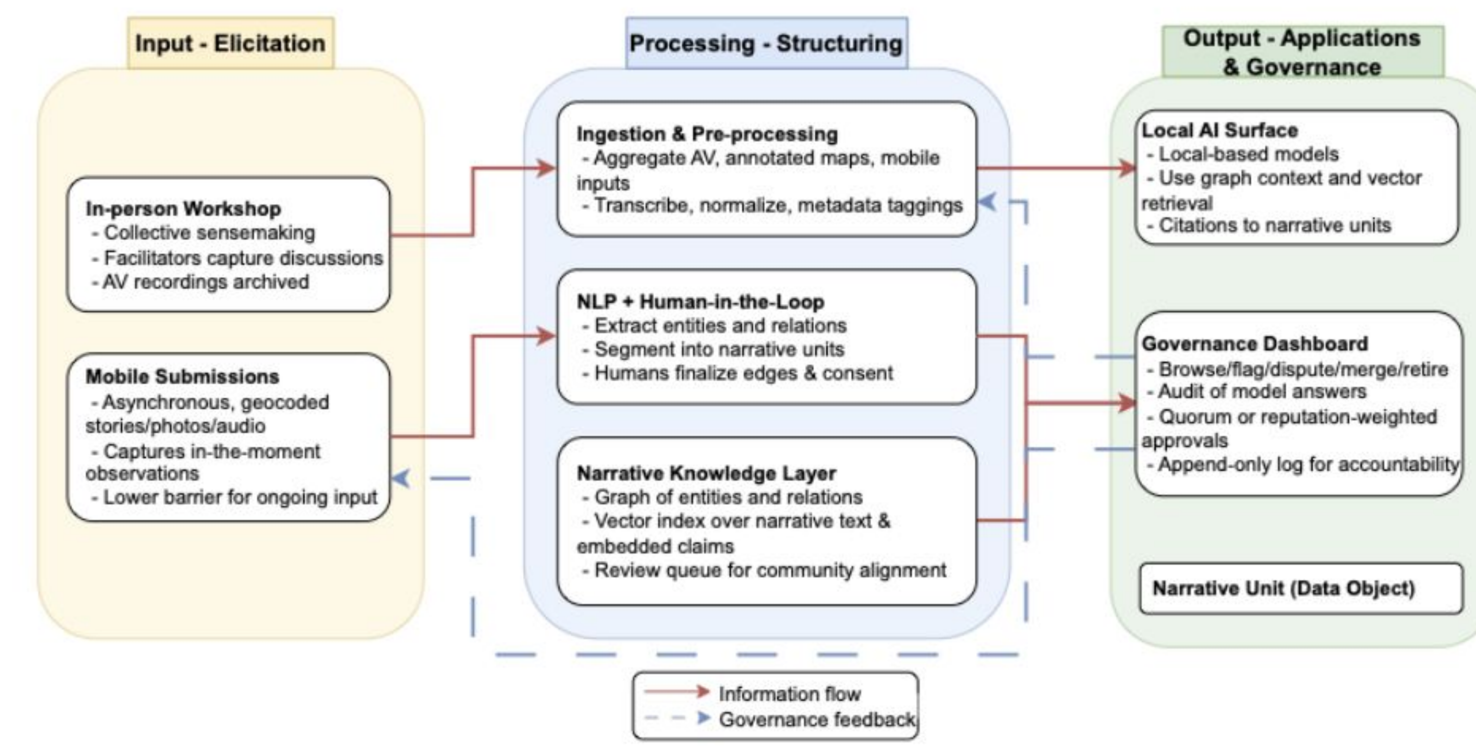
- We designed a participatory protocol involving $N=24$ community members in Atlanta.
- **Elicitation**: Used "Physical Scaffolding" (large paper maps) and asset-based framing to prompt rich storytelling rather than just complaints.
- **The Narrative Unit:** We convert oral stories into structured data (Entities + Time + Place + Provenance).
- See ***Figure (b) workshop progress***



Result: Narrative grounding mitigates factual, cultural, geographic, and temporal errors by injecting community narratives with explicit provenance at retrieval time.

The Narrative Knowledge Layer

- **Input**: In-person workshops + mobile submissions.
- **Processing**: Human-in-the-loop validation creates a "Narrative Graph".
- **Output**: A RAG system where AI answers are cited with specific community stories.



Closing the Gap

- **Baseline**: State-of-the-art LLM answered <21% of local questions correctly on its own.
- **The Fix**: In the majority of failures, the missing facts were present in our collected narratives.
- **Conclusion**: Narrative grounding directly addresses the dominant error modes (factual/cultural) identified in the audit.

Design Tensions & Governance

- **Representation**: **Who** speaks for the "**community**"? We must balance "local profiles" with internal dissent.
- **Privacy**: "**Hyper-local**" means **identifiable**. We balance utility with de-identification.
- **Control**: A **community governance** dashboard allows residents to dispute, merge, or retire data used by the AI

Takeaways

- **A Taxonomy of local LLM failures.**
- **A Protocol for transforming stories into AI-ready data.**
- **A Governance model for community ownership of AI knowledge**