CitySpeak: An AI-Based Mascot for Conversational Civic Engagement

Sabrina Kaiser

Advisors: Suvosree Chatterjee, Jenifar Selim, Sara Sevilla, Dr. Serena Hoermann, and Dr. Jason Hallstrom

FLORIDA ATLANTIC UNIVERSITY

NSF CENTER FOR SMART STREETSCAPES

BACKGROUND

Existing mechanisms of promoting civic engagement may not represent some voices, as there is an accessibility gap in interacting with local government.

- **74%** of Americans report multiple barriers to getting involved in community issues [1].
- **48%** of Americans say that digital tools may help break the barriers to their participation [2].
- In semi-structured interviews with **11** West Palm Beach residents, **8** reported never having attended a city council or neighborhood meeting.

Artificial intelligence (AI) could help close this gap. Large language models (LLMs) have been used to convert general citizen input into actionable ideas for officials [3]. In some cities, LLMs have been fine-tuned with specific municipal context to assist citizens [4].

PROJECT GOALS

1. Make civic participation convenient, engaging, representative, and spontaneous.
2. Provide useful data to city officials. Ultimately, the project is designed to increase citizen agency and representation.
3. Be a knowledgeable resource for citizens.
4. Maintain a conversational tone; avoid interviewing.

METHODS

Computer Vision (CV) Model Selection

Benchmark 3 CV models on detection and tracking performance (Fig. 3). Select RT-DETR .

Interaction Flow

Implement RT-DETR with multi-object tracker (ByteTrack) to invite pedestrians to interact after 5 seconds in the zone.

Knowledge Retrieval Development

Static knowledge (historical data, municipal structures, etc.) and live knowledge (automatic refreshes from trusted sources).

Conversation Design

Build conversational flow, policies, and norms. Compare OpenAI and Gemini conversation, transcription, and audio generation models. Select Gemini due to expected lower latency, iterate through natural tests, and refine behaviors. Run final qualitative tests, 8 test cases.

Feedback Capture

Create analytics dashboard containing AI-generated feedback summaries. Add filters and user navigation.

RESULTS

Conversational Quality

Appropriately adapts to subject changes.

Understands and remembers context.

Info retrieval accurate.

Mean conversational model latency: **2.32 s.**

Can ask excessive follow-ups.

Info can sometimes be technically related but contextually irrelevant.

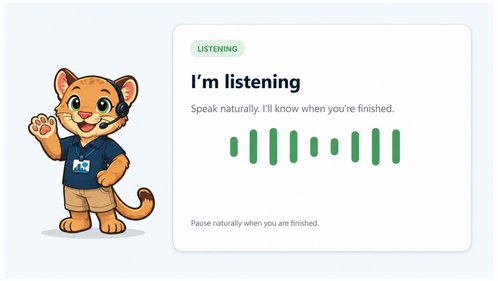


Fig. 1. Kiosk user interface, mid-interaction.

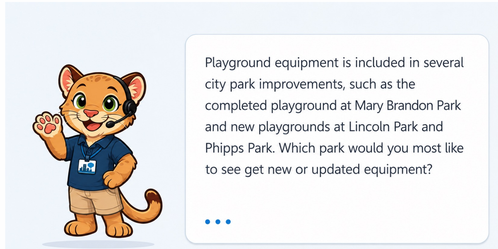


Fig. 2. Kiosk user interface, mid-interaction.

User response dashboard

Computer Vision Model Evaluation

Faster R-CNN (HOTA = 37.15) and YOLO11s (HOTA = 37.19) performed similarly.

RT-DETR-L (HOTA = 41.52) demonstrated a significantly higher aptitude for pedestrian identification and tracking.

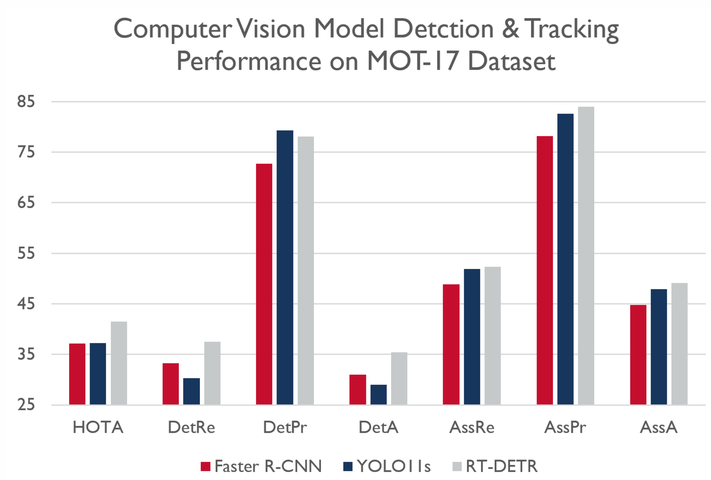


Fig. 3. Combined evaluation of CV models.

DISCUSSION

Since the models were evaluated on a representative dataset, not true CitySpeak footage, the quality assessment is partially limited



Fig. 4. Realistic kiosk video on Clematis Street, output of RT-DETR + ByteTrack.

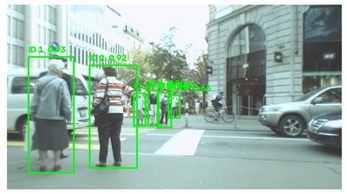


Fig. 5. Video from MOT-17 dataset, output of RT-DETR + ByteTrack.

RT-DETR's higher recall is especially essential, as it culminates in more invitations for citizens to participate. YOLO-11s threw fewer false alarms, though the difference was marginal.

CitySpeak has the potential to seed dialogue and transparency (through the public access feedback dashboard) between municipalities and their residents. While this prototype focuses on West Palm Beach, the system is adaptable to other municipalities.

FUTURE WORK

1. **Comprehensively test.** Evaluate LLMs and CV models rigorously on true CitySpeak inputs.
2. **Expand search capacity.** Add semantic re-ranking.
3. **Minimize latency.** Use streaming models.
4. **Add more safeguards.** Blur faces, prepare for users with nefarious intent. Request residency status from the user to ensure answer integrity.
5. **Maximize engagement.** Offer individuals the opportunity to use their own voice and likeness for the interface. Base the hook on visual stimuli.

ACKNOWLEDGMENTS

This work was supported through the NSF Engineering Research Center for Smart Streetscapes (CS3), funded through NSF Cooperative Agreement EEC-2133516.

REFERENCES

[1] Corning, A. (2025, September 30). *Disconnected but hopeful: What a new survey reveals about civic life in America*. National Civic League.

[2] Gallup, Inc., & Charles F. Kettering Foundation. (2026, March). *Democracy for All Project: The people's role in American democracy*.

[3] Domingues, M., Neto, M., Guerra, C., Pacheco, E., & Cordeiro, L. (2026). Supporting scalable e-participation through LLM-assisted proposal writing and summarisation. *AMCIS 2026 Proceedings*, 5.

[4] Williams, S., Beery, S., Conley, C., Evans, M. L., Garces, S., Gordon, E., Jacob, N., & Medina, E. (2024). People-powered Gen AI: Collaborating with generative AI for civic engagement. *An MIT Exploration of Generative AI*.