The Wizard in the Town Plaza: Voice-Based Interactive Storytelling in Public Spaces

Daniel DeFreez[1] and **Jackson Harrower**[1] and **Travis Wood**[1] and **Cynthia Salbato**[2]
[1]Southern Oregon University, Ashland, OR 97520 USA
[2]ARTNOW, Ashland, OR 97520 USA
defreezd@sou.edu, harrowerj@sou.edu, woodt1@sou.edu, secretstoryworld@gmail.com

Abstract

We present The Wizard in the Town Plaza: a field report on a system that generates narrative content using advanced voice cloning, a community-informed knowledge base, and a structured story world. In March 2025 (spring equinox), we deployed this system in a small tourist and college town in the USA as part of a seasonal event. A human actor served as a live intermediary between the AI system and visitors, relaying audience questions to the AI system which responded directly through loudspeakers with a cloned voice. This deployment revealed how AI-driven storytelling can operate in oral, performative, and social contexts, offering cultural and educational value. We detail the system architecture, including cloned voice, narrative knowledge graphs, and LLMs, while addressing ethical considerations and community response. We also outline its expansion into an interactive mural project designed to amplify local Indigenous voices and history through AI-supported storytelling.

Introduction

In March 2025, during a spring equinox theatre event in Ashland, Oregon, we deployed The Wizard as a live public installation. This approach allowed us to test how large language model-driven storytelling, grounded in a persistent knowledge graph, could function in real-time within a live, performative, and community-based context. A human actor served as The Wizard's representative, relaying visitors' questions into a microphone. The Wizard responded through a loudspeaker system with a rich, expressive voice that carried across the plaza, creating an immersive theatrical experience.

Our deployment location, Ashland, is a small tourist and college town with deep-rooted arts culture and home to the renowned Oregon Shakespeare Festival. We position this project as a cultural application of computational creativity, offering insights into system design, voice synthesis, knowledge graph user experience, along with a reflection on its social, ethical, and creative implications.

For this first public deployment, we chose to use a human liaison between the AI system and the audience as a risk mitigation strategy, given this was our initial test of the system in a public setting. Having a human intermediary allowed us to ensure smooth interactions and quickly address any technical issues that might arise during the live performance.

Our approach builds on and extends a growing body of research in AI-driven narrative systems. Branch et al. (Branch, Mirowski, and Mathewson 2021) developed a GPT-3-powered narrator for stage productions with human curation of outputs, while Shape of Story (Long, Jacob, and Magerko 2019) leveraged voice input for interactive storytelling. In contrast, our approach enables real-time conversational interactions without human content curation.

Góngora et al.'s PAYADOR system (Góngora et al. 2024) addresses the world-update problem in interactive storytelling by recasting it as an outcome-prediction task, enabling story worlds to evolve based on user interactions. PAYADOR uses a minimalist typed schema of characters, items, and locations, similar to the schema in our knowledge graph, though our current deployment operates in a read-only mode. PAYADOR's approach is relevant as it confirms that lightweight, structured grounding can effectively manage large models without heavy symbolic planners.

Commercial solutions like OpenAI's ChatGPT advanced voice mode (OpenAI 2023) offer impressive capabilities, but are completely dependent on their proprietary systems which, at the time of our project, did not allow voice customization. Alternative services like ElevenLabs had terms of service that did not satisfy legal requirements, compelling us to develop our own solution. We worked with a local voice actor who recorded a scripted set of wizard dialogues that we then used to condition a voice model at inference time. While we continued to use OpenAI's API for text generation, we implemented our own custom voice synthesis pipeline using the recorded voice samples and the open-weight TTS model Zonos (Zyphra 2023) instead of relying on proprietary text-to-speech services.

The contributions of this work include:

- A system that connects knowledge graphs to language models, enabling a variety of entities (human and non-human) to emerge from a shared knowledge base
- An architecture achieving real-time conversation with open-weight models allowing control of voice data
- A practical demonstration that AI characters with ethically-sourced voices can successfully engage public audiences

Knowledge Representation

One of the central challenges in AI-driven storytelling is narrative coherence, ensuring that characters, events, and timelines remain consistent over time and across interactions. Large language models (LLMs), while rich in generative capacity, are prone to losing track of details in extended conversations or across multiple sessions.

To solve this, our system incorporates a structured knowledge graph using Kanka (Owlchester 2024), a collaborative worldbuilding tool designed for non-technical users. Kanka serves as a narrative database that maps the relationships between characters, events, locations, and lore across our interactive narrative system.

Prior work (Wang et al. 2023; Blin 2022) supports the use of knowledge graphs to guide and constrain generative models, preventing contradictions and enabling more believable long-form narratives. Our integration of Kanka supports retrieval-augmented generation, ensuring that responses reflect accurate story context while allowing improvisational flexibility.

Our implementation accesses Kanka through its REST API, retrieving entities such as characters and quests from a structured campaign database. We chose Kanka over a traditional database approach specifically because of its user-friendly web interface, which allows system administrators and non-technical storytellers to collaborate on content creation without requiring custom UI development. While we could have implemented a standard database solution, this would have necessitated building a custom administrative interface—a significant additional development effort. Instead, Kanka provides an intuitive WYSIWYG editor with built-in relationship management tools that creators can use immediately, while our system programmatically accesses this shared narrative state through the API during interactions. Each entity in the knowledge graph maintains relationships to other entities, creating a rich network of connections that the narrative state machine traverses as it responds to user queries, ensuring responses are contextualized within the established mythology.

Narrative State Machine

To maintain state across interactions, we implemented a Narrative State Machine (NSM) that tracks a user's progression through a particular quest or story arc. This state machine (shown as a component in Figure 1) queries Kanka's structured knowledge base while maintaining its own internal state. Whereas Kanka provides the permanent narrative content and world knowledge, the NSM itself manages state transitions and conversation history in a separate Firestore database, storing the current node and interaction history for each user session. This separation of concerns ensures the wizard maintains consistent characterization across sessions and allows multiple users to progress through different narrative paths simultaneously. When participants ask questions that might contradict established lore or lead the character off-track, the NSM queries Kanka for canonical information.

Drawing from the theory of computation (Sipser 2013), we define our Narrative State Machine (NSM) as a formal

System Architecture

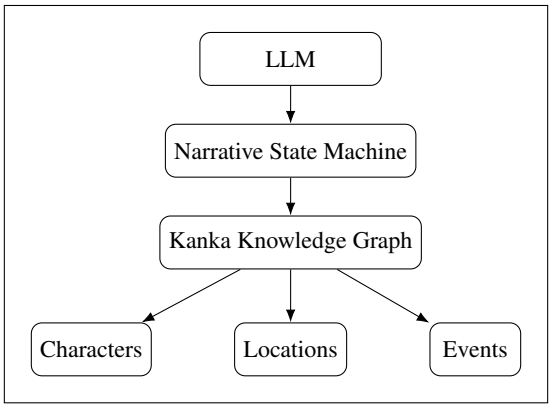

Figure 1: System architecture showing the relationship between the LLM, narrative state machine, and Kanka knowledge graph components used to maintain consistency across interactions.

model consisting of a set of states and a transition function augmented with an LLM matcher. Unlike traditional computational models, the NSM integrates natural language understanding with discrete state transitions, enabling narrative coherence across extended interactions.

Figure 2 shows the Narrative State Machine algorithm. Rather than using symbols from a finite alphabet to define transitions, the NSM transition function δ maps natural language descriptions (sentences written by the state machine author) to next states. For example, a transition might be defined as "User asks about the wizard's origins" or "User expresses disbelief in magic."

The transition function $\delta(q, \sigma, \mathcal{H}) \rightarrow (\text{response}, q')$ determines the next state by considering the current state q, user input σ, and conversation history $\mathcal{H}$. It returns a natural language response (response) and the new state q'. Within the function, the knowledge context K is retrieved based on the current state and user input. The conversation history $\mathcal{H}$ is maintained internally and updated. We also track node traversal history separately for analytics, but this is not represented as a formal parameter in the function signature.

The LLM is used in two distinct steps: first as LLM_δ to match user input to natural language transition descriptions, then as LLM_r to generate contextually appropriate responses based on the new state. The algorithm naturally terminates when it reaches a state with no outgoing transitions, representing an ending in the narrative. This approach leverages the LLM's semantic understanding to bridge formal state transitions with natural language interaction. The NSM allows end users to interact with the state machine using natural language, while enabling non-technical authors to define transitions using intuitive descriptions rather than formal symbols or code.

Voice Interface and Cloned Character Voice

Our motivation for creating a custom voice solution was twofold: to employ local talent and to achieve the highest

function $\delta(q, \sigma, \mathcal{H})$
    $\mathrm{K} \leftarrow \text{Knowledge}(q, \sigma)$
    $T \leftarrow \text{GetAvailableTransitions}(q)$
    if $T = \emptyset$ **then**
        return "End of story", q
    end if
    $q' \leftarrow \text{LLM}_\delta(q, \sigma, T, \mathrm{K}, \mathcal{H})$
    response $\leftarrow \text{LLM}_r(q', \sigma, \mathrm{K}, \mathcal{H})$
    return response, q'
end function

Figure 2: Narrative State Machine transition function δ where q is the current state, σ is user input, $\mathcal{H}$ is conversation history, response is the natural language output, q' is the next state, and K is knowledge context retrieved from Kanka. The function terminates at states with no outgoing transitions.

possible quality voice output. The wizard's voice is synthesized using locally hosted voice cloning system, leveraging the open-weight TTS model Zonos. Zonos v0.1 (1.6B parameters) was trained on over 200,000 hours of multilingual speech, delivering emotionally nuanced vocalizations that rival commercial TTS systems. It is our experience that Zonos successfully captures enough of the emotional range and expressiveness of the voice actor to be effective for our purposes (Zyphra 2023).

The voice recording session involved two separate recordings. First, the actor read a scripted wizard dialogue based on anticipated interactions. Second, the actor performed various passages with a range of emotional expressions—sad, excited, conversational, and more—to create a comprehensive vocal palette. In total, we recorded about five minutes of speech, though Zonos voice cloning only requires around 10 seconds of high-quality audio to condition the model at inference time. This approach of recording more material than needed and selecting optimal segments lets us choose the best emotional tone and voice quality. For this particular event, we sought a booming, authoritative wizard voice, so we selected specific segments that embodied this persona.

Voice cloning tolerates variability in the specific content recorded, but is highly sensitive to intonation, emotional range, and voice quality.

Because Zonos can only synthesize a few seconds of audio at a time, we incorporated NLTK sentence segmentation tools (NLTK Project 2023) to split wizard responses into manageable chunks. NLTK's sentence tokenization provided reliable boundary detection even with the wizard's archaic speech patterns and lengthy narrative descriptions. These chunks were rendered individually, then reassembled.

Audio can be streamed real-time using LiveKit, an open-source WebRTC platform. Deployed on a lightweight 2GB Digital Ocean droplet, it enabled low-latency, real-time interaction even in constrained computing environments (LiveKit 2023).

Character Factory A key feature of our system is that any entity in the storyworld can be engaged with directly. While the wizard served as our primary deployment persona, our "character factory" architecture supports the generation and deployment of multiple interactive entities that share the same knowledge graph. The system can dynamically create new characters through API requests that expand prompts using the Kanka knowledge base, allowing for rapid generation of additional NPCs like shopkeepers, artisans, or guides that remain consistent with the shared world.

This approach enables automated character scaling and allows not just human characters but also objects, plants, animals, and abstract concepts to be given voice and agency. Visitors can converse with a variety of entities within the shared narrative universe, each with appropriate interaction styles.

Multimodal Interaction Capabilities Beyond voice interaction, our system supports SMS and MMS text messaging, providing accessibility for users in low-bandwidth environments or those who prefer asynchronous communication. Using Twilio's messaging services integrated with Modal Labs' serverless compute functions, users can engage with characters via text messaging to continue their narrative journeys while on the move.

The system's multimodal capabilities are particularly powerful for place-based games and storytelling. Since generative AI can now process images alongside text, users can send photographs of real-world locations, landmarks, or objects which the system then incorporates into the narrative. For example, a visitor could photograph a distinctive tree in Lithia Park and the wizard might respond with lore about "Forest Guardians" specific to that location.

All user state is synchronized across interaction channels, maintaining narrative continuity whether engaging through voice, text, or image, which allows transitions between different modes of engagement.

Spring Equinox Festival

In March 2025, during a spring equinox festival in Ashland, we debuted the Wizard character as part of a larger community event.

We installed the experience in the heart of Ashland's downtown plaza. The actor welcomed visitors with a narrated opening ritual, then invited them to ask questions of the wizard. The AI-generated responses were voiced through our local voice cloning system and amplified through speakers, immersing the audience in a living dialogue with a mythical figure.

The technical setup relied on a single NVIDIA 3090 GPU, running local voice synthesis for minimal latency in a location with inconsistent internet. The decision to use offline voice generation ensured fidelity and resilience, allowing the wizard to remain responsive throughout the event. The system provided consistent character behavior throughout the event.

Audience Feedback

The audience consisted primarily of families with children, who showed particular enthusiasm for interacting with the wizard character. Young visitors asked questions about

magic and the wizard's backstory. Despite the mediated format, the audience was highly engaged with the experience, with the actor's theatrical presence helping to anchor attention while the AI's narrative fluency created compelling interactions. One confounding factor in our evaluation was the excellent performance by the human liaison, who skillfully drew visitors into conversation It remains an open question whether people would engage as readily with the AI system without a charismatic human intermediary to facilitate the interaction. Several attendees were surprised to learn they had been conversing with an AI-powered system. While this initial deployment did not include quantitative measurements, in future work we plan to conduct a formal study with surveys and metrics to assess audience engagement, comprehension, and satisfaction.

Ethical Considerations

Voice Cloning Ethics

AI-cloned voices carry growing ethical weight, especially in light of high-profile misuses in political and commercial arenas. These range from fraudulent Medicare scams featuring celebrities on YouTube (Koebler 2024) to political manipulation, as demonstrated by the recent $6 million FCC fine for AI-generated robocalls impersonating President Biden in a primary election (Naylor 2024). Our approach followed a clear ethical framework, starting with consensual voice capture: we used samples from a paid, local actor who explicitly approved the use of their voice for the AI wizard character.

Privacy was central to our design. All voice synthesis and transcription were performed locally on GPU-enabled hardware, ensuring that no personal audio data was transmitted to the cloud. This commitment to data sovereignty was both a technical necessity (given intermittent connectivity) and a cultural alignment with Ashland's community values of autonomy and transparency. Additionally, local processing avoided critical concerns about transferring sensitive voice data to third-party services, where ownership rights become ambiguous under terms of service agreements that often grant providers broad rights to use uploaded voice samples for their own model training or other purposes. By maintaining complete control of all voice data locally, we ensured clear ownership boundaries and prevented potential misappropriation of the actor's vocal identity.

Labor Considerations

The emerging tension between AI voice synthesis and professional voice actors has become a significant ethical issue, highlighted by the ongoing SAG-AFTRA video game strike that began in July 2024 (SAG-AFTRA 2024). The strike focuses specifically on establishing AI guardrails to protect voice actors from having their performances replicated without consent or compensation.

Our project acknowledges these legitimate labor concerns by implementing several ethical safeguards:

(1) Compensation: We hired and paid a local voice actor for his performance, ensuring compensation for his time and talent;

(2) Clear usage limitations: We established a clear agreement that the voice actor's recordings would be used exclusively for the Storyworld project and not for any other commercial or non-commercial projects;

(3) Data sovereignty: All voice data was processed locally and not shared with third parties, ensuring the actor maintained control over how his voice was used.

Rather than replacing voice actors with AI, we aimed to show how AI can extend and amplify human creative work while respecting artistic agency. While AI systems now make it possible to generate synthetic voices with minimal human involvement, this case demonstrates that community-centered AI projects can establish new standards for ethical labor practices rather than circumventing them. Human voice actors deserve both attribution and appropriate compensation when their vocal performances form the foundation of synthetic voices.

Future Work

Interactive Mural Application

The Wizard character project is the foundation for another undertaking: an interactive, AI-powered mural installation that would bring Ashland, Oregon's history to life through indigenous storytelling. This mural will depict significant historical and cultural scenes from the region, with a particular focus on indigenous perspectives that have often been overlooked in traditional historical narratives.

Our concept is to re-purpose the wizard infrastructure to tell authentic indigenous stories from the Shasta, Takelma, and other tribes indigenous to the Rogue Valley region. This would transition our project from fantasy narratives to real-world history through the same technical architecture, while amplifying indigenous voices that have been historically marginalized.

Initial conversations have begun with a local muralist and cultural organizations in Ashland about this potential collaboration. The proposed project would blend art, oral storytelling, and AI-driven interaction to transform a public wall into a living archive of indigenous voices and stories. We recognize that such a project requires extensive consultation with tribal representatives, and we have begun discussions with local indigenous knowledge keepers to ensure that if the project proceeds, it would be developed with appropriate cultural context and respect for tribal sovereignty and intellectual property rights.

If this project moves forward, visitors would be able to engage with interactive elements powered by the same voice cloning, large language models, and narrative knowledge graph infrastructure field-tested in our plaza deployment. Such a project would require careful attention to cultural authenticity, appropriate representation, and proper attribution of indigenous knowledge and stories. We acknowledge that any such project must address complex ethical questions around voice cloning of indigenous storytellers and the digitization of traditional knowledge systems.

Visitor Information Booth Application

We are also in discussions with Ashland city officials to deploy AI characters as virtual attendants for the town's visitor information booth, which is currently unstaffed. This application would allow tourists to access information about local attractions, events, and history through conversation with AI characters that reflect Ashland's cultural identity.

Visitor: hi
The Wizard: Greetings, traveler! You've found your way to the enchanted realm of Ashlandia. How may I assist you in your journey today? Whether you're seeking hidden treasures, magical experiences, or simply a delightful tale, I'm here to guide you.
Visitor: where can I get some children's books?
The Wizard: Ah, for enchanting tales and magical stories, TreeHouse Books is a splendid haven for young readers in Ashlandia. Established in 1978, this whimsical bookstore captures the imagination with its fairy-tale murals, cozy nooks, and delightful selection of children's literature. Alongside treasures for the young and young-at-heart, you'll find puppets, costumes, and art supplies to spark creativity. Whether you seek a gift or a new adventure tale for bedtime, TreeHouse Books awaits to welcome you into its enchanted world.

Figure 3: SMS/MMS response from our visitor information booth proof-of-concept, showing factual local information delivered with fantasy-themed presentation.

This interaction demonstrates how fantasy and utility can blend in AI-assisted tourism. Messages sent via SMS/MMS or voice calls receive identical information wrapped in the Wizard's distinctive voice, maintaining persona consistency while delivering accurate local knowledge.

Confabulation Detection for Information Services For both visitor assistance and indigenous stories, factual accuracy is paramount. To address the risk of LLM confabulation, also known colloquially as hallucinations, we are developing a citation-based verification system specifically for these applications. When providing tourist information, the AI character must cite specific sources from knowledge graph and a separate verification character checks these citations to ensure they exist and support the provided information. While double confabulation remains possible (where both the primary LLM and verifier get it wrong), the probability is reduced through this two-stage process.

Conclusion

We have presented a place-based, voice-powered interactive storytelling system designed as a showcase for computational creativity in public space. Our system integrates natural-sounding voice cloning with a knowledge-driven narrative engine for locally resonant storytelling. The deployment in Ashland demonstrated public willingness to engage with AI characters in playful dialogue and as companions in cultural exploration, showing potential to contribute to this artistic community's culture.

From a technical perspective, our work demonstrates that open-weight models can achieve production-quality voice synthesis and interactive storytelling when properly integrated. The combination of local voice models, knowledge graphs, and state tracking provides a robust framework for creating believable AI characters without dependence on proprietary cloud services.

From a computational creativity perspective, our project highlights the importance of embodiment (voice) and context (knowledge graph) in making AI creativity accessible, situated, and socially impactful. By combining these elements in real-time, performative environments within the distinctive cultural context of Ashland, we believe that generative AI can function as a facilitator of shared memory, imagination, and cultural exchange.

The contributions of this work, as outlined in the introduction, include: (1) a system that connects knowledge graphs to language models, enabling a variety of entities (human and non-human) to emerge from a shared knowledge base; (2) an architecture achieving real-time conversation with open-weight models allowing control of voice data; and (3) a practical demonstration that AI characters with ethically-sourced voices can successfully engage public audiences. These innovations show how AI-driven storytelling can thrive in oral, performative, and social settings. As we expand into applications like the interactive mural project, we continue to explore how our character factory approach can bring to life not just human characters but also objects, locations, and abstract concepts through API-driven character generation and AI-supported storytelling, while carefully considering the balance between creative interaction and systemic safeguards.

Author Contributions

Cynthia Salbato provided the overall vision and original concept for the project. Daniel DeFreez was the main architect of the software system and primary author of this manuscript. Travis Wood and Jackson Harrower assisted with research experiments, with Travis being the main proponent of voice cloning technology. All authors contributed to the writing of this manuscript.

Acknowledgments

We thank Travis Puntarelli for providing the voice of the Wizard, bringing the character to life with a delightful vocal performance. Thanks to Aurora Quinn Muller for serving as liaison between the audience and the Wizard during the spring equinox festival deployment, facilitating the interactive experience with theatrical flair. Special thanks to Elliot Glenn for initial work on an internal university poster project that was the foundation for Narrative State Machines. This work was generously supported by Southern Oregon University's STEMReX research for undergraduates program.

References

Blin, I. 2022. Building narrative structures from knowledge graphs. In *The Semantic Web: ESWC 2022 Satellite Events*, volume 13384 of *Lecture Notes in Computer Science*, 234–251. Springer.

Branch, B.; Mirowski, P.; and Mathewson, K. W. 2021. Collaborative storytelling with human actors and AI narrators. *CoRR* abs/2109.14728.

Góngora, S.; Chiruzzo, L.; Méndez, G.; and Gervás, P. 2024. PAYADOR: A minimalist approach to grounding language models on structured data for interactive storytelling and role-playing games. In *Proceedings of the 15th International Conference on Computational Creativity (ICCC)*, 101–106. Jönköping, Sweden: Association for Computational Creativity. System that grounds LLMs to minimal representations of fictional worlds for RPGs and interactive storytelling.

Koebler, J. 2024. Deepfaked celebrity ads promoting medicare scams run rampant on youtube. *404 Media*. Investigation of widespread AI-generated deepfake videos featuring celebrities in fraudulent Medicare scam advertisements, viewed over 195 million times.

LiveKit. 2023. Livekit: Open source infrastructure for real-time communication. LiveKit Project Website. WebRTC-based platform for building real-time audio, video and data applications. Accessed: 2025-04-26.

Long, D.; Jacob, M.; and Magerko, B. 2019. Designing co-creative ai for public spaces. In *Proceedings of the 2019 Conference on Creativity and Cognition*, 479–490. ACM.

Naylor, B. 2024. Fcc fines political consultant $6 million for ai-generated robocall impersonating biden. *NPR*. Report on FCC's first enforcement action against AI-generated voice cloning used for political manipulation in a New Hampshire primary election.

NLTK Project. 2023. Natural language toolkit. NLTK Website. Open-source Python library for NLP with over 50 corpora and lexical resources. Accessed: 2025-04-26.

OpenAI. 2023. ChatGPT can now see, hear, and speak. OpenAI Website. Announcement of multimodal capabilities in ChatGPT including voice conversations and image processing.

Owlchester. 2024. Kanka - online tabletop RPG campaign manager and worldbuilding tool. Web application. Community-driven platform for creating and organizing RPG campaigns with collaborative worldbuilding. Accessed: 2025-04-26.

SAG-AFTRA. 2024. Interactive media (video game) strike. Official SAG-AFTRA website. Labor strike involving voice actors and motion capture artists over AI protection concerns. Began July 26, 2024. Accessed: 2025-04-26.

Sipser, M. 2013. *Introduction to the Theory of Computation*. Boston, MA: Course Technology, third edition. Classic textbook covering automata theory, computability theory, and complexity theory.

Wang, Y.; Lin, J.; Yu, Z.; Hu, W.; and Karlsson, B. F. 2023. Open-world story generation with structured knowledge enhancement: A comprehensive survey. *Neurocomputing* 545:126091.

Zyphra. 2023. Zonos-v0.1: A leading open-weight text-to-speech model. GitHub Repository. Apache 2.0 License, trained on 200k hours of multilingual speech. Accessed: 2025-04-26.