AI as intermediary in modern-day ritual: An immersive, interactive production of the roller disco musical *Xanadu* at UCLA

Mira Winick
REMAP / Dept. of Theater
UCLA
Los Angeles, CA 90095
mirawinick@ucla.edu

Naisha Agarwal
Dept. of Computer Science
UCLA
Los Angeles, CA 90095
naishaa@g.ucla.edu

Chiheb Boussema
REMAP
UCLA
Los Angeles, CA 90095
chiheb@g.ucla.edu

Ingrid Lee
Dept. of CS / Theater
UCLA
Los Angeles, CA 90095
ingridlee@g.ucla.edu

Camilo Vargas
REMAP
UCLA
Los Angeles, CA 90095
cjvargas@g.ucla.edu

Jeff Burke
REMAP / Dept. of Theater
UCLA
Los Angeles, CA 90095
jburke@remap.ucla.edu

Abstract

Interfaces for contemporary large language, generative media, and perception AI models are often engineered for single user interaction. We investigate ritual as a design scaffold for developing collaborative, multi-user human–AI engagement. We consider the specific case of an immersive staging of the musical *Xanadu* performed at UCLA in Spring 2025. During a two-week run, over five hundred audience members contributed sketches and jazzercise moves that vision language models translated to virtual scenery elements and from choreographic prompts. This paper discusses four facets of interaction-as-ritual within the show: audience input as offerings that AI transforms into components of the ritual; performers as ritual guides, demonstrating how to interact with technology and sorting audience members into cohorts; AI systems as instruments "played" by the humans, in which sensing, generative components, and stagecraft create systems that can be mastered over time; and reciprocity of interaction, in which the show's AI machinery guides human behavior as well as being guided by humans, completing a human–AI feedback loop that visibly reshapes the virtual world. Ritual served as a frame for integrating linear narrative, character identity, music and interaction. The production explored how AI systems can support group creativity and play, addressing a critical gap in prevailing single user AI design paradigms.

1 Introduction

Mobile phones expanded the dominance of single-user interfaces for human-computer interaction from the "personal computer" era onwards. The emergent capabilities of LLMs and other foundation models to analyze and respond to a broad range of natural language and multimodal inputs (potentially from many sources at once) suggest new, AI-supported interface paradigms may be on the horizon and older, even ancient, concepts of human interaction may be more easily supported[1]. In this paper,

[1] There are notable explorations of group interfaces, from the tangible UIs of Ishii et al. (Ullmer and Ishii, 2000) to LLM-supported collaborative storytelling like Epic Saga Builder (Burglin, 2023), but there are far more single-user interfaces.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

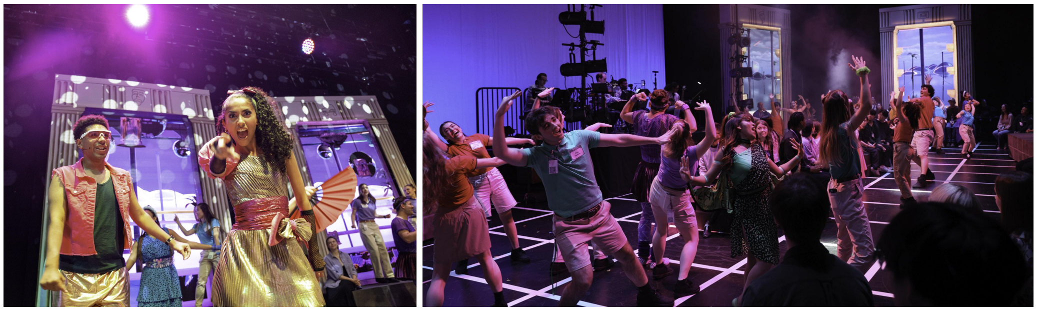

Figure 1: *Xanadu* was an immersive musical performed at UCLA in May 2025.

we describe the use of contemporary AI techniques to enable group participation in an immersive staging of the Broadway musical *Xanadu* at UCLA in Spring 2025[2] (see Figure 1). The production explored how AI-generated media from an audience could be integrated into designed extended reality (XR) scenery and sound. A key goal was to support collaborative creativity by audiences within an otherwise linear show that had a fixed narrative and musical spine. Conventional physical production elements (e.g., movable scenery, costumes, a live band, theatrical lighting) were integrated with their virtual XR counterparts (digitally realized scenery, characters and costumes, sound, lighting) rendered in real-time using Unreal Engine (Epic Games, 2024). The show invited audience members to actively participate by making music, drawing sketches, and dancing together, with AI observing and then transforming contributions into aspects of the show's evolving virtual world.

One of the team's design strategies proved particularly effective for integrating audience input and its AI-driven transformations. We conceptualized the show as a **contemporary ritual**, in which technology functioned as a supporting mechanism enabling the uninitiated audience to form groups and join experienced performers in the ritual. In this paper, we first introduce the show and its sources of inspiration. Then, we describe how ritual was used to design and understand group interactions, employing AI for transformation and synthesis of participation, and provide brief observations.

2 Production and research motivations

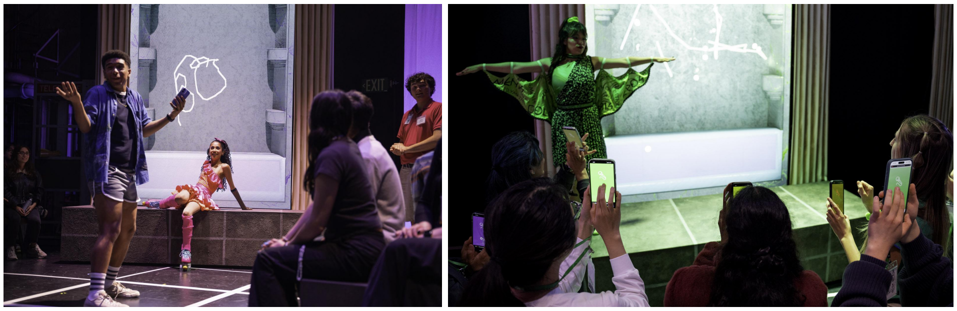

Figure 2: Audiences were shown by performers how to create drawings as offerings to the muses, which were transformed into images and objects in the muses' "shrines" by generative AI.

Xanadu (Beane et al., 2007) follows the Greek muse Clio, who comes to Venice, California in the early 1980's as "Kira", to inspire the mortal artist Sonny to pursue his creative dreams and open a roller disco. Over about a year, the UCLA Department of Theater and UCLA REMAP developed an immersive, participatory staging of the complete Broadway musical, performed for public audiences in May 2025. In this unique production, the audience was seated onstage, surrounded by seven thirteen-foot high, movable LED displays—*shrines*. Each shrine corresponded to one of the show's seven muses, including Kira / Clio. Audiences played an important role in Sonny and Kira's journey by interacting in person and contributing their own creativity to the virtual world and the resulting roller disco brought to life around them.

[2] Presented through special arrangement with Music Theatre International (MTI).

We[3] embarked on the production to investigate how AI and XR could create novel audience experiences. An early design decision was to frame the show as a modern-day ritual, paralleling aspects of Greek ritual, which is both well known and directly connected to the show's narrative. The team followed this frame to work coherently across the real-world and fictional interpretations of any given moment. Other theatrical traditions, particularly those involving participation and improvisation, such as Commedia dell'Arte (Oreglia, 1968), provided additional tools and perspectives. Greek rituals often feature a large gathering of lay participants guided by ritual "experts" to create and perform offerings for a non-human entity (Parker, 2011). *Xanadu* reproduced this structure: the audience became the ritual crowd, led by an ensemble of "acolytes" to participate with the lead performers, including in the creation of digital offerings presented to the muses. Participants began in a liminal space between the real world and the show's world, where acolytes "onboarded" them into muse groups. The acolytes acted as guides for the audience throughout the show, but also fluidly crossed the "fourth wall" to serve as a traditional ensemble, participating in the show's songs and choreography.

We elaborate on specific design strategies below. Engaging with collective participation explores a gap in current conceptualizations of user interaction with contemporary AI systems. It shifts the focus from predominantly individual interfaces to AI to emerging models of group-based interaction (Lee et al., 2025). The default focus on personal interfaces can be traced to early user experience (UX) research and cognitive psychology, which often emphasized individual-level performance metrics—an emphasis that continues to shape product design, human–computer interaction, and, more recently, human-centered AI (Lee et al., 2025). Frameworks for collective engagement with AI remain comparatively underdeveloped. A growing body of work recognizes the need for more research on interactions with AI in groups (Cui and Yasseri, 2024; Shiiku et al., 2025; Borghoff et al., 2025; Lee et al., 2025) as these present unique challenges, including issues of perceived ownership (Zhang et al., 2025), effective interaction modalities (Raees et al., 2024), user engagement and agency (Zhang et al., 2025; Raees et al., 2024), trust, group and power dynamics, the navigation of varying levels of experience, knowledge, preferences (Naiseh et al., 2024), as well as unpredictability-related trust issues and frustrations with non-determinism (Xu et al., 2023).

Engaging these challenges through artistic production, our work adopted a ritual framework that situates group-AI interaction within a theatrical performance that needed to operate in both fiction and reality at once, offering a means to explore collective participation, trust, agency, and play.

3 Designing for audience participation in a performance ritual framework

Embedding audience participation within a theatrical production that must follow strict script and timing constraints is not trivial. Some of the challenges that must be addressed involve (i) how to inscribe the audience's participation within the narrative flow, (ii) how to demonstrate the interaction modalities on-the-fly and in-story, (iii) how to get the audience engaged and willing to contribute in front of everyone else, and (iv) what to do with the aleatory and unpredictable aspects of generative AI. These accompany technical implementation challenges related to performance requirements and budget constraints, as well as a broad need to ensure generated media are aligned with the show's aesthetics. For the interested reader, we touch on the implementation challenges in Appendix A.

3.1 Input as offerings transformed by AI

A quintessential element of Greek ritual is the offering of gifts by mortals to the gods as an invitation for divine favor. This practice is exemplified by the festival of Chalkeia, where Athenian women collectively wove sacred robes as an offering to Athena, blending artistic labor with devotional intent (Parker, 2011). In *Xanadu*, all audience contributions, of drawing, music, and dance, were conceptualized as offerings, with AI as an intermediary translating these gifts into more polished representations in the virtual world inhabited by the show's gods. These offerings of phone-based drawing, sound-making, and skeletal tracking of movement (see Appendix A for more details) underscore the opportunity to expand interaction modalities in XR beyond speech, text, and gesture (Raees et al., 2024). Phone positional tracking enabled participants to draw on the shrine-canvases with their phones as wands, with generative AI transforming their sketches into polished images and 3D assets

[3] The authors include the faculty creator / producer (Burke), one of the faculty directors (Winick), the assistant choreographer (Lee), and AI / systems developers (Agarwal, Boussema, Lee, Vargas, Burke).

(see Figure 2). Skeletal tracking of dance and gesture, using computer vision, enabled group-level movement-based communion (see section 3.4).

The ritual aspect framed audience participation consistently with the plot and characters while establishing legible norms and conventions for design, rehearsal, and eventually audience contributions. *Improbotics* (Mathewson and Mirowski, 2018) used a similar approach, in which theater performers integrated lines from generative AI while performing. Lines were viewed as “offers" for an actor and, with the cast’s collective engagement, were integrated into scenes being performed. *Xanadu* worked in the reverse: performers prompted the audience, audience members contributed responses, and AI interpreted these to generate changes in the virtual world. Here, the human acolytes guided improvised exchanges with the generative AI, mirroring historical rituals where expert practitioners instructed participants in producing spontaneous offerings devoted to particular gods or ritual themes (Parker, 2011). Merging ancient models of guided offering with AI-driven improvisation, the show used ritual logic to coherently structure collective human–machine interaction.

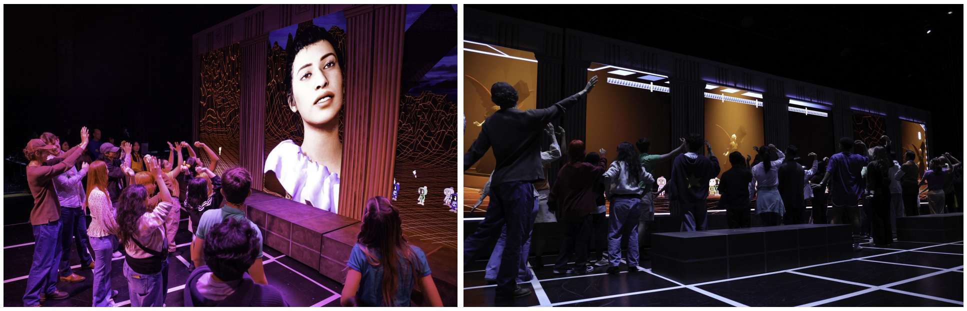

Figure 3: In Olympus, the audience’s movement ultimately brings Kira and Sonny back together.

3.2 Invitation to interact via performer demonstration

During the song “Magic”, Sonny first sketches Kira using a phone borrowed from an acolyte, and then invites the audience to sketch ideas for his roller disco. Following the ritual framework, acolytes drew first, acting as expert participants to demonstrate the process. This allowed them to show examples of valid inputs and model collaborative drawing, signaling to the audience that even rough scribbles were welcome. The focus on collective contribution over perfection lowered audience performance anxiety and instead encouraged collaboration. Since large language and diffusion models still struggle with contextual nuance and visual consistency (Avrahami et al., 2024), we chose to embrace these limitations instead of trying to hide them.

Through three successive cycles of drawing and generation, each modeled by the acolytes, the audience developed their own expertise with the drawing instrument. Generative AI pipelines, composed of vision-language and diffusion models (see A.2), converted each sketch into an image or a 3D mesh satisfying the show’s vaporwave aesthetic, muse-specific color palettes, and overall artistic direction. Similar to collaborative art installations like Dream Painter (Guljajeva and Canet Sola, 2022) and FRIDA (Schaldenbrand et al., 2022), collective input guided the AI system, which produced elements for a now audience-expanded virtual world. For instance, when a spectator sketched a roller skate, the AI (generally) generated a 3D asset of a skate rendered in the show’s vaporwave style. In cases where a generation misfired, a human-in-the-loop moderator could intercept the output. The acolytes explained remaining oddities as “mysterious insight" into the audience’s intent. This dramaturgical structure re-framed algorithmic opacity and unpredictability as part of the storyworld’s metaphysics. LuminAI (Trajkova et al., 2024) used similar tactics, treating an AI dance partner’s unexpected moves not as mistakes, but openings for improvisational responses and dialogue.

Furthermore, to strengthen collective identity, each audience group (about eight people plus two acolytes) was dedicated to one Muse. They had their own shrine with a unique iconography and color scheme. The AI output for each shrine reflected the perspectives and journeys of the associated group, shaped by the influence of their Muse. To achieve this, we took a two-fold approach. First, acolytes provided drawing cues aligned with their particular patron. For example, the Muse of Music’s group needed to collaboratively sketch musical instruments on their shared canvas. Second, the AI pipelines were internally directed towards each Muse’s aesthetics using designer-selected color palettes and reference imagery. Acting as a unifying force, the AI systems wove individual contributions into cohesive outputs embodying small group identity within the larger whole.

3.3 Stage machinery (including AI) as ritual instrument

Inspired by the Eleusinian mysteries, where initiates' song and dance summoned the gods' realm (Parker, 2011), *Xanadu*'s virtual world was conceived as a parallel sacred space accessed by the ritual participants through the shrine set pieces. The relationship between the physical and digital worlds was similar to Fiebrink's concept of the "meta-instrument", in which controllers are taught to convert sensed inputs into a bounded space of outputs, creating a new instrument to be played (Fiebrink, 2016). In that work, controllers are specific ML processes. *Xanadu* employed a broader network of relationships transforming the input of human physical actions into XR outputs via various AI-supported systems. For example, positional tracking of the shrines enabled performers to move them to change the LED screens' perspective into the virtual world, without technician intervention. A WebAR app on audience phones, acting as a ritual instrument, transformed gestures (made with the whole phone) into offerings. In groups, the audience aimed their phones at the shrines and, moving them as paintbrushes, drew colored poly-lines in the virtual world. (See A.1.) One of three custom AI pipelines, used at different moments in the show, processed the inputs. (A.2) A vision–language model, coupled with diffusion models, merged the audience sketches with designer reference images and actor portraits, producing 2D images or 3D meshes matching the show's aesthetics.

To develop the relationship between audience input and digital output, the team conducted weekly play-tests with invited participants during the show's development and rehearsals. Through this iterative process, designers adjusted gesture sensitivity, stroke width, emissivity, and color parameters, and refined model weighting to align the outputs with the production's aesthetics. By the performance, the system had a single gestural grammar for the audience, customized aesthetics for each Muse group, and different AI pipelines for specific moments. Actors prompted Muse image generation in one scene and 3D object creation in another, with audiences performing the same gestures for their own results. This gave the audience one interface to learn to participate in several aspects of the ritual.

In the final act, the Muses and the mortals–Sonny, the acolytes, and the audience–ascended to a virtual Olympus (see Figure 3). Computer vision (CV) tracked the mortals' position and pose using off-the-shelf facial and body detection with ML. This data drove Bomberman-inspired avatars (Konami, 2017), whose puny movements in the virtual world contrasted with the large area covered by the audience on stage. Actors playing gods stood on a narrow runway, visibly piloting larger-than-life virtual versions of themselves. Both muse and mortal bodies played the virtual avatars as instruments, with CV-based stage machinery translating their movement into an active manifestation of hierarchy and power within the virtual world. The integration of ritual logic with the meta-instrument framework helped the team to design a cohesive relationship between physical gestures, sensing, AI processes, and corresponding digital outputs, and a unified grammar for integrated audience participation.

3.4 Reciprocity: AI-directed audience choreography

Prior sections discuss a unidirectional structure: human action → AI/ML process → media output. With the final offering of dance, as shown in Figure 4, we reversed this flow. An AI component guided audience movement, creating a deliberate reciprocity with early scenes. This echoes role reversals in Greek rituals, which temporarily invert hierarchies (Parker, 2011). In the norms of Greek ritual, the gods do not speak directly to mortals. Messages pass through intermediaries who reveal divine instructions (Parker, 2011). An actor called the Oracle consulted an iPad with access to a vision–language model (VLM). The VLM analyzed that performance's sketch-driven, AI-generated media offerings and output a short poem. It used that poetry to describe three jazzercise-style moves selected from a curated set. The Oracle read the poem aloud and, with the Muse of Dance, taught the moves to the audience. With the acolytes enthusiastically demonstrating, the audience was invited to perform the choreography. A depth camera rig tracked skeletal poses and compared collective motion to a baseline. If the crowd's energy, timing, and alignment exceeded a threshold, the virtual environment could respond dynamically with lighting shifts and scene updates. In this way, we explored how immediate engagement of the whole audience could impactfully contribute to the show.

The choreography, presented as ritual, merged audience subgroups into a whole via the VLM analysis, turning spectators into co-performers. Research shows that synchronized motion, especially with moderate effort, improves trust, emotional intimacy, and group affiliation (Tarr et al., 2015). As participants danced together, movement became symbolic and communally significant. Shared movements became rituals of belonging, reinforcing audience group unity through embodied memory. The enactment effect, where physically performing gestures improves memory retention over passive observation (Roberts et al., 2022), further suggests how repeated choreography in *Xanadu* anchored story elements in shared experience. Informal feedback confirmed this: many audience members cited the participatory dance as a highlight, reflecting enjoyment and a sense of shared purpose. The Oracle scene thus illustrates how AI systems, when embedded in ritual dramaturgy, can foster group belonging and creativity, challenging the dominant single-user AI paradigm.

4 Limitations and societal impact

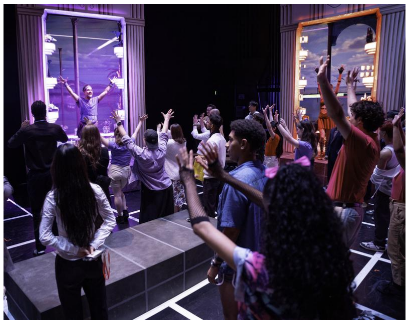

Figure 4: The audience's dance moves help build the *Xanadu* auditorium.

This work demonstrates how historical understandings of ritual could guide group interaction with AI, but several limitations remain. Notably, the approach was tailored to a specific musical, creative approach, and physical space. Reproducibility was not a focus. Assessment of audience experience was informal and observational, limiting evidence for any claims of generality. While considerable technical effort was made to reduce generative AI latency to meet the show's interactivity goals, group-AI improvisation, especially involving diffusion models, remains constrained by latency, quality, and budget. Finally, while framing AI not as a god but at least as divine machinery served the show's goals, it risks reinforcing problematic perceptions of AI as authoritative, inscrutable, or all-knowing. Human authorship and curation, while present, are explained within the fiction rather than directly, though we did offer various behind-the-scenes conversations to the public. Finally, although we aimed to support collective creativity and co-authorship, generative models introduce worrying homogenizing tendencies (Doshi and Hauser, 2024; Anderson et al., 2024). Their outputs often reflect dominant cultural aesthetics, tropes, and norms embedded in training data, which reduce diversity and nuance. The production tried to mitigate this through curated prompts, human moderation, and stylistic oversight, but we remain mindful and concerned about the erosion of creative diversity.

5 Conclusion: generalizations and future impact

Theater appears to mimic aspects of the "real world", so it can be tempting (and, at times, helpful) to think of it as a bounded testbed for new interfaces and system designs, tested within the microcosm of a given performance. Theater is also a unique human activity, with many different forms and historical antecedents. The types of ritual cited here are just a few of many from across the world. Each performance establishes and operates within its own rules for relationships among audiences and performers, fiction and reality, and humans, design, and technology. Strategies employed in performance find their way into film and television, gaming, and social media. They are also part of our everyday lives. Thus, this exploration of ritual framing to design and implement group interactions with AI may have utility for a range of collective group-AI interactions. In future work, we plan to examine more deeply the role of rehearsal as a site for training, experimentation, and refinement of human–AI interaction. We are also interested in how AI can support small group interactions within much larger wholes, by enabling tailored interactions with global-scale narrative worlds, scaled through decentralization of AI onto edge devices. Generally, we seek AI-supported methods that engage live audiences in new forms of collaborative fandom and collective meaning-making.

Acknowledgements

We thank the production's many contributors. Full show credits are at xanadu.remap.ucla.edu. Photos in this article are by M. Yepez, M. Beymer, and K. Liu. In-kind equipment and services were provided by Boxx, Mo-Sys, and 4Wall Entertainment. Funding was provided by the UCLA Department of Theater, UCLA DataX, the Amazon MGM Studios' Innovative Storytellers Initiative, and Qualcomm.

References

Anderson, B. R., Shah, J. H., and Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*, pages 413–425.

Anthropic PBC (2024). Claude 3.5 Sonnet. `https://www.anthropic.com/news/claude-3.5-sonnet`.

Avrahami, O., Hertz, A., Vinker, Y., Arar, M., Fruchter, S., Fried, O., Cohen-Or, D., and Lischinski, D. (2024). The chosen one: Consistent characters in text-to-image diffusion models. In *ACM SIGGRAPH 2024 conference papers*, pages 1–12.

Beane, D. C., Lynne, J., and Farrar, J. (2007). Xanadu. Directed by Mira Winick and Corey Wright. Stage musical premiered on Broadway at Helen Hayes Theatre, New York, July 10 2007. Music and Lyrics by Jeff Lynne & John Farrar; Book by Douglas Carter Beane.

Borghoff, U. M., Bottoni, P., and Pareschi, R. (2025). Human-artificial interaction in the age of agentic ai: a system-theoretical approach. *Frontiers in Human Dynamics*, 7:1579166.

Burglin, P. (2023). Epicsagabuilder. `https://github.com/pburglin/EpicSagaBuilder`. Accessed: 2025-08-06.

Cui, H. and Yasseri, T. (2024). Ai-enhanced collective intelligence. *Patterns*, 5(11).

Doshi, A. R. and Hauser, O. P. (2024). Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science advances*, 10(28):eadn5290.

Epic Games (2024). Unreal Engine 5. `https://www.unrealengine.com/`. Accessed: 27 Oct 2025.

Fiebrink, R. (2016). Machine learning as meta-instrument: Human-machine partnerships shaping expressive instrumental creation. In *Musical Instruments in the 21st Century: Identities, Configurations, Practices*, pages 137–151. Springer.

Guljajeva, V. and Canet Sola, M. (2022). Dream painter: an interactive art installation bridging audience interaction, robotics, and creative ai. In *Proceedings of the 30th ACM international conference on multimedia*, pages 7235–7236.

Huang, Z., Boss, M., Vasishta, A., Rehg, J. M., and Jampani, V. (2025). Spar3d: Stable point-aware reconstruction of 3d objects from single images. arXiv:2501.04689 [cs.CV].

Konami (2017). Super bomberman r. `https://www.konami.com/games/bomberman/r/us/en/`. Accessed: 2025-10-27.

Langford, A., Shah, A., Gupta, A., Bhatter, A., Goyal, A., Mathur, A., Mohanty, A., Kumar, A., Sethi, A., Komma, A., et al. (2025). The amazon nova family of models: Technical report and model card. *arXiv preprint arXiv:2506.12103*.

Lee, S., Hwang, S., and Lee, K. (2025). Beyond individual ux: Defining group experience (gx) as a new paradigm for group-centered ai. In *Companion Publication of the 2025 ACM Designing Interactive Systems Conference*, pages 357–362.

Lu, H., Liu, W., Zhang, B., Wang, B., Dong, K., Liu, B., Sun, J., Ren, T., Li, Z., Yang, H., Sun, Y., Deng, C., Xu, H., Xie, Z., and Ruan, C. (2024). Deepseek-vl: Towards real-world vision-language understanding. arXiv:2403.05525 [cs.AI].

Mathewson, K. and Mirowski, P. (2018). Improbotics: Exploring the imitation game using machine intelligence in improvised theatre. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 14, pages 59–66.

Naiseh, M., Webb, C., Underwood, T., Ramchurn, G., Walters, Z., Thavanesan, N., and Vigneswaran, G. (2024). Xai for group-ai interaction: towards collaborative and inclusive explanation. In *World conference for explainable artificial intelligence (17/07/24 - 19/07/24)*.

OpenAI et al. (2024). Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Oreglia, G. (1968). *The Commedia dell'Arte*. Hill & Wang, New York.

Parker, R. C. (2011). *On Greek Religion*. Cornell University Press.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv:2307.01952 [cs.CV].

Raees, M., Meijerink, I., Lykourentzou, I., Khan, V.-J., and Papangelis, K. (2024). From explainable to interactive ai: A literature review on current trends in human-ai interaction. *International Journal of Human-Computer Studies*, 189:103301.

Roberts, B. R., MacLeod, C. M., and Fernandes, M. A. (2022). The enactment effect: A systematic review and meta-analysis of behavioral, neuroimaging, and patient studies. *Psychonomic Bulletin & Review*, 30(1):1–24.

Schaldenbrand, P., McCann, J., and Oh, J. (2022). Frida: A collaborative robot painter with a differentiable, real2sim2real planning environment. *arXiv preprint arXiv:2210.00664*.

Shiiku, S., Marjieh, R., Anglada-Tort, M., and Jacoby, N. (2025). The dynamics of collective creativity in human-ai hybrid societies. *arXiv preprint arXiv:2502.17962*.

Stability AI (2024). Introducing Stable Diffusion 3.5. `https://stability.ai/news/introducing-stable-diffusion-3-5`.

Tarr, B., Launay, J., Cohen, E., and Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology Letters*, 11(10):20150767.

Trajkova, M., Long, D., Deshpande, M., Knowlton, A., and Magerko, B. (2024). Exploring collaborative movement improvisation towards the design of luminai—a co-creative ai dance partner. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–22.

Ullmer, B. and Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM systems journal*, 39(3.4):915–931.

Wang, Q., Bai, X., Wang, H., Qin, Z., Chen, A., Li, H., Tang, X., and Hu, Y. (2024). Instantid: Zero-shot identity-preserving generation in seconds. arXiv:2401.07519 [cs.CV].

Xing, P., Wang, H., Sun, Y., Wang, Q., Bai, X., Ai, H., Huang, R., and Li, Z. (2024). Csgo: Content-style composition in text-to-image generation. arXiv:2408.16766 [cs.CV].

Xu, W., Dainoff, M. J., Ge, L., and Gao, Z. (2023). Transitioning to human interaction with ai systems: New challenges and opportunities for hci professionals to enable human-centered ai. *International Journal of Human–Computer Interaction*, 39(3):494–518.

Yamer-AI (2023). Yamermix. `https://civitai.com/models/84040?modelVersionId=196039`. Accessed: 2025-10-27.

Ye, H., Zhang, J., Liu, S., Han, X., and Yang, W. (2023). Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. arXiv:2308.06721 [cs.CV].

Zhang, L., Rao, A., and Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. arXiv:2302.05543 [cs.CV].

Zhang, S., Wang, H., and Yi, X. (2025). Exploring collaboration patterns and strategies in human-ai co-creation through the lens of agency: A scoping review of the top-tier hci literature. *arXiv preprint arXiv:2507.06000*.

A Appendix/Supplemental Material

A.1 Brief overview of the AI-enabled show proceedings

We briefly describe technical details of how a typical performance of *Xanadu* proceeded, focusing on the key AI-enabled audience participation moments, taken in breaks between scripted material.

As audience members arrived, they were divided into seven groups, each group corresponding to one of the seven muses in the story. Two "acolytes" (performers) guided each group through the various interactions in the show. Upon arrival, audience members were prompted to scan a QR code to access our web app on their phones, before taking their seats in the center of the stage.

During the opening sequence, participants were invited to contribute to their muse's shrine by creating sketches at three key moments, each with a different creative task: (a) designing a background image for their muse, (b) drawing their muse in a pose of their choice, and (c) creating objects to populate the shrine. Phones acted as gestural interfaces; a WebAR app, built on the 8th Wall platform, translated 6DOF tracking and pose estimation into sketch strokes.

These sketches became the starting point for our generative AI pipelines. Background sketches were transformed into high-quality images with the muse composited as a "frieze" at the bottom of the shrine. Pose sketches generated fully clothed, realistically proportioned muse characters that preserved facial identity and matched the designer's aesthetic. Object drawings were turned into 3D assets, that were then placed within the virtual shrine environment. Further details on these pipelines can be found in A.2.

Beyond sketching, the show incorporated additional AI-driven segments that extended audience participation into new modalities. In the Oracle sequence, a vision-language model (VLM) analyzed a show's sketch-driven generated media, composed a short poem, and selected three dance moves from a curated set of low-impact, jazzercise choreography. In the final act, mortals and gods ascended to a virtual Olympus with the seven shrines being assembled into a continuous LED wall, and gods re-embodied as life-sized MetaHuman avatars driven by live motion capture (depth camera + iPhone LiveLink).

Code for these various modules can be found at our Github page `https://github.com/remap` and upon request.

A.2 Sketching: generative AI pipelines

To support the interactive sketching tasks, we combined contemporary AI methods, in particular multimodal large language models (MLLMs) and diffusion models, within a hybrid model hosting architecture. Some models were deployed directly on custom Amazon Web Services (AWS) SageMaker endpoints, giving us full control over model architecture, inference parameters, and generation logic. Others were accessed via AWS Bedrock, which provided faster inference for large models, albeit with limited configurability.

Across the three generative tasks, the pipelines followed a common structure. MLLMs interpreted audience inputs and generated prompts while local diffusion models, augmented with state-of-the-art control techniques, generated task-specific media outputs. Occasionally, Bedrock-hosted diffusion models were used for quality enhancements. Detailed description of each task follows.

Generation task #1: a scene for each muse

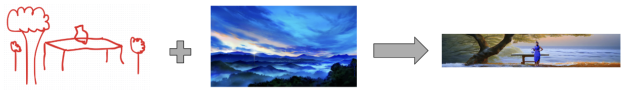

Figure 5: Generation Task 1, a scene for each muse

User Experience: In this generation task, audience members (in pairs) drew a sketch of a background they wanted their muse to be seen in. That sketch was then turned into a generated background with the muse composited on top, placing each muse in a new environment. Designers, in addition to crafted prompts that fit their aesthetic design, provided a per-muse scene reference image to steer both the color palette and the type of background each muse would appear in. Performers (both muses and acolytes) would adapt to the scene that was generated for them, letting audience members almost transport their character into a new world.

Implementation: The audience background sketch is passed into the DeepSeek Vision Language Model (VLM) (Lu et al., 2024) to generate a detailed text description of the image. We pass this text description, a designer style reference image, and the audience background sketch into Stable Diffusion 3.5 (Stability AI, 2024). Coupled with this model is IP-Adapter (Ye et al., 2023) that helps generate the background in the style of the reference image, and ControlNet (Zhang et al., 2023) that generates the background from the user sketch. Given the size of the model, we generate a low pixel count image (512x384) for a fast controllable first generation pass. This image is then used to guide an AWS Bedrock-hosted model (Amazon Nova Canvas) (Langford et al., 2025) for fast high-quality, high resolution image variation generation, which is rescaled to desired dimensions. This was a faster and more robust approach than generating the high resolution image directly from the local large Stable Diffusion model. The Muse is then composited at a random location along the bottom axis of the final image.

Generation task #2: our muses in custom poses & garments

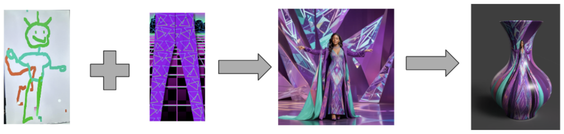

Figure 6: Generation Task 2, our muses in custom poses and garments

User Experience: In this generation task, audience members were invited to sketch a pose for their assigned muse. Each muse, depicted in the chosen pose, was dressed with a designer specific garment style, which is then layered as a texture on various objects in Unreal Engine. The styles of designer garment (including colors, motifs, and patterns) were provided by the costume design team to ensure the garment each Muse was generated in was appropriate and fitting to the costumes the actors were seen in onstage. Acolytes guided audiences through this task, while the muses offered suggestions for poses they would like to be seen in.

Implementation: We designed a multi-agent system to handle both audience and designer inputs, powered by Claude 3.5 Sonnet (Anthropic PBC, 2024). The first agent generates a detailed text description about the designer garment style. The second generates a text description of the pose from the audience sketch. The third one converts this text description into numerical pose keypoints. The latter two agents used few-shot learning in their prompting. The generated keypoints were processed to ensure body proportions were accurate to the muse. We then use the Instant ID framework (Wang et al., 2024), which consists of the YamerMIX-8 model (Yamer-AI, 2023), a Stable Diffusion XL (SDXL) checkpoint (Podell et al., 2023), as the base diffusion model, Identity Net (IN) for facial preservation, and a Pose Control Net (PCN) (Zhang et al., 2023) to generate the image. The pose keypoints are passed into the PCN, the muse image is passed into IN, and the garment text description is passed directly into SDXL. In order to represent poses and faces as accurately as possible, we had the IN active throughout the entire generation, and the PCN start right after and end midway in the generation, capturing the essence of the desired pose while avoiding deleterious effects on facial features. For pose sketches that are too abstract, we created a custom pose library to randomly select a pose for the muse to appear in. From here, we generate an image of the muse in the specified garment and pose. This image is converted into a texture to be overlayed on various vases in the Unreal environment.

Generation task #3: 3D object offerings for the muses

User Experience: In this generation task, audience members drew sketches of various objects as offerings to the muses. The final output is a 3D asset of the object rendered back into the Unreal-powered world. The inputs for this generation are the audience sketch, along with a designer style reference. Designers provided vaporwave aesthetic reference images to ensure the final generations

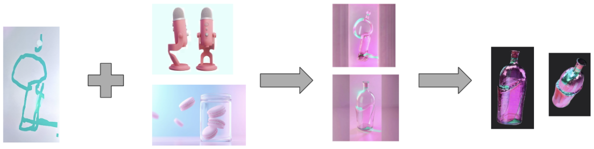

Figure 7: Generation Task 3, 3D object offerings for the muses

fit into *Xanadu*'s 1980s retro vaporwave style. As objects rendered back onto the stage, actors improvised pointing out particularly interesting generations to audience members.

Implementation: The audience sketch is first passed into a VLM to generate a text description of the object. This description, along with the user sketch and a designer style reference, is passed to the SDXL diffusion model (Podell et al., 2023) augmented with a content-style disentanglement module (Xing et al., 2024), enabling the generation of a cleaner and more refined sketch that captures the original user sketch in the designer style. This then guides a larger diffusion model hosted on AWS Bedrock to rapidly generate a high quality image. Finally, the image is converted to a 3D mesh using an image-to-3D diffusion model (Huang et al., 2025) and rendered back on stage.

A.3 System implementation

In order to scale these generations to audiences of up to sixty-five every night, we developed a modular and scalable system architecture using Amazon Web Services. Drawings were first made on textures in the Unreal Engine virtual world through the show's custom WebAR application, which sends data via Google Firebase to the game engine. Local Python code exported the drawings, and uploaded them to AWS Simple Storage Service (S3), an object storage service. This then notifies our serverless orchestration code hosted on AWS Lambda, a compute service that runs code in response to events without any external server provisioning.

Notifications are sent via Amazon Simple Notification Service (SNS) that publishes messages to Amazon Simple Queue Services (SQS), which, under the orchestration of a Lambda Helper Function, routed the incoming sketches to one of three generation modules used in the show. Each module generated media based on the drawings it received, invoking both hyperscale foundation models on AWS Bedrock, a service with access to multiple large language or image models, and smaller faster models deployed as real-time endpoints on Amazon SageMaker, AWS's platform for building, training, and deploying models. The resulting generations were stored in S3 and fetched for human-in-the-loop moderation before being loaded into the Unreal Engine virtual world experienced by the audience. This process, from sketch to display, completes within 30-60 seconds, which fit within the show's timing to ensure seamless audience participation.

Twenty-four SageMaker AI endpoints handled inference requests, eight for each generation task. We used 8 g6.12xlarge (48 vCPUs, 4x24GB Nvidia L4 GPUs) for generation tasks 1 and 3, and 16 g6.4xlarge (16 vCPUs, one 24GB Nvidia L4 GPU) instances for generation task 2. Each instance hosted a customized stable diffusion model and included support from Amazon Bedrock for average inference times of about 20-30 seconds for generation tasks 1 and 3, and about 40-60 seconds for generation task 2.

A.4 Oracle: generated choreography

We developed a multimodal computer vision pipeline that integrates LLMs and pose estimation to generate and evaluate dance choreography. Initiated during live performances, the pipeline employed the GPT-4o model (OpenAI et al., 2024), to process a randomly selected subset of audience-generated sketches collected each evening. A predefined list of twelve dance moves, derived from the musical number performed in that scene, provided the movement vocabulary for the interaction. Each night, the AI model selected three moves from this list and generated a short poem inspired by those selections, integrating the chosen gestures into its language to inform the Oracle's choreographic interpretation.

The second phase of the pipeline involved pose estimation and skeletal tracking using Stereolabs' ZED cameras. During playtesting, we used the ZED SDK's built-in 2D skeletal tracking system and evaluated its performance against ground-truth data captured from the Oracle's dance movements. The system tracked individual movement, allowing for comparative analysis across three key metrics: movement accuracy, timing, and dancing energy. Movement accuracy and timing were quantified using Object Keypoint Similarity (OKS) and Dynamic Time Warping (DTW), which were computed by analyzing 2D skeletal frames captured in real time and comparing them to the Oracle's ground-truth choreography. Dancing energy was inferred from the velocity of joint movements over time, with greater joint velocity indicating higher levels of engagement and intensity.

To make this data expressive and actionable within the performance, the computed metrics were normalized to a 0–1 scale and integrated into Unreal Engine. This mapping drove visual and environmental feedback, such as pulsing lights and visible progress in constructing the mythical *Xanadu* auditorium, across both the virtual world and the physical set. In this way, this real-time feedback loop created a dramaturgical bridge between audience participation and narrative progression.

While the full pipeline demonstrated strong performance during playtesting with small audience groups, we encountered scalability limitations with the ZED SDK's 2D skeletal tracking in larger crowd settings, where tracking fidelity declined. As a result, the pose estimation component was replaced with human-driven visual evaluation for production. Nevertheless, the pipeline provided valuable insights into the relationship between audience motion and theatrical response, offering a promising foundation for future work in responsive performance design and audience-aware dramaturgy.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes, the abstract details the main contributions this paper makes, including a new ritual logic framework of audience input as offerings, a meta-instrument design pattern for a theater performance, and a reciprocal loop. Overall, we propose new contributions for group interfaces versus single user paradigms.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Yes, please refer to 4.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This paper does not include theoretical results of the types referenced in this item.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: While this paper does describe an experiment–the production of *Xanadu*–it is not designed as or intended to be reproducible. Playtests, rehearsals, and proof-of-concepts were similarly bespoke.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: We have included the link to our Github repo in our appendix. This has majority of the code with additional modules' code available upon request. However, as mentioned previously, *Xanadu* is not intended to be reproducible, hence users may encounter varying results from the ones presented here.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setting is presented in the paper with the description of various generative AI pipelines used for the sketching interaction in the show as well as the technology behind the computer vision pipelines powering the Oracle sequence. Please refer to A.2 and A.4 for further details on these workflows.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

Justification:The paper does not include experimental results that can be analyzed statistically.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Refer to A.3 for information regarding compute.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research follows the NeurIPS Code of Ethics. Regarding potential harms in the research process, in particular with direct interactions with human participants, all audience members were informed when purchasing a ticket for *Xanadu* about the tech elements at work in the show. The tech team worked with performers to ensure all audience members felt comfortable and not obliged to participate in the various interactions in the show if not interested/able. In regards to data, no dataset was directly used to train any models; all models used are released publicly and open for usage.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Please refer to 4 for details on societal impact of our work (both positive and negative).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: All models used in this paper are pre-trained language models and diffusion models. During the show, operators in the crew managed generated content through a human in the loop system where images were checked before appearing on stage. In cases where images were deemed inappropriate, a fallback image was displayed instead. Please refer to 3.2 for more details.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Yes, all models and services used are cited (please refer to A for details on each of these). There was no fine-tuning, and the use of other assets was consistent with typical approaches and license requirements for our productions.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets are released in this paper; *Xanadu* is primarily an art piece that uses various AI components together in the context of a live performance.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Please see also the note below regarding human subjects research. If appropriate for the Creative AI track, we can provide details on information provided to the audience ahead of the production for the camera ready version.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our theatrical productions are not typically subject to IRB review and we do not use data collected in the show's systems to draw generalizable conclusions about human behavior. Discussion about audiences in this paper are based on our observations of public performances in which the audience has been informed of technology use and could opt out of each type of participation without consequence.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: Yes; the usage of LLMs is described as a part of the experimental procedures in this paper and as a part of the core methodology. Please refer to A for more details on which LLMs were used and how they were used in the various pipelines in the show.