Human Factors in Communication of Design, Vol. 177, 2025, 13–23
https://doi.org/10.54941/ahfe1006327

Interaction Modes and User Experience in Generative AI-Based New Media Art

Zihan Mei, Yifei Feng, Zixuan Zhang, Jia Wang, Yurun Han, Ziling Wang, and Ying Long

School of Art and Design, Wuhan University of Technology, Wuhan, HB 430070, China

ABSTRACT

This paper aims to explore how Generative Artificial Intelligence (GAI) can empower new media art creation that leverages different interaction modes, offering users a novel artistic experience in the context of AI development. In recent years, significant breakthroughs in artificial intelligence have had a profound impact on social development. In particular, since the emergence of generative large models such as ChatGPT, Midjourney, and Stable Diffusion, GAI technology has become more widely applied across various scenarios, redefining innovative pathways in artistic creation. Therefore, to better understand the methods of GAI-based new media creation, this paper primarily adopts a case analysis approach. Within a theoretical framework, it analyzes the feasibility of integrating new media art with GAI based on their developmental trajectories. It also delves into the specific applications of GAI in the field of new media art and examines the diverse experiential effects produced for users through different interaction modes. Subsequently, the research findings indicate that GAI not only expands the creative dimensions of new media artists but also leverages various interaction modes to provide users with more intelligent and multimodal artistic experiences, aligning with the inevitable trends in artistic creation in the digital age. Finally, this paper incorporates the author's creative output based on GAI, proposing new design strategies and methods for creation. It clarifies the application value of GAI in new media art creation and, from a design perspective, provides new insights into how artificial intelligence can better enrich human artistic experiences, offering new research perspectives and ideas for the development of art and technology.

Keywords: GAI, New media art, Interactive modes, User experience

INTRODUCTION

Generative Artificial Intelligence (GAI) has emerged as a transformative force in various industries, especially in the realm of new media art. These advancements are not only reshaping traditional artistic workflows but also redefining the relationship between humans and machines in the creative process (Chui et al., 2021).

New media art, which blends digital technology with traditional artistic practices, provides a unique interdisciplinary foundation for the application of GAI in artistic creation (Duan et al., 2019). By combining GAI with body-based interaction, voice-based interaction and biofeedback-based interaction,

new media artworks are able to deliver dynamic and immersive experiences. These interaction modes offer audiences a deeper sense of engagement and personalization, allowing users to influence and interact with the artwork in meaningful ways. This convergence of GAI and interactivity creates a multimodal artistic experience that aligns with the growing demand for intelligent and adaptive systems in the digital age (Kamar, 2016).

The primary aim of this paper is to investigate how GAI can empower the creation of new media art, focusing on its integration with various interaction modes and the impact of these modes on user experience. To achieve this, the study adopts a case analysis approach, examining the feasibility and effectiveness of combining GAI with new media art through their developmental trajectories. Additionally, the paper analyzes the experiential effects of GAI-powered artworks on users, providing insights into how AI can deepen user engagement and enhance creative possibilities.

Finally, this paper incorporates the author's own creative output based on GAI-driven interactive art, presenting new design strategies and creative methods for GAI-based new media art. This contribution highlights the application value of GAI in enriching artistic creation and provides a fresh perspective on the intersection of artificial intelligence, art, and user interaction. By addressing both theoretical and practical aspects, this study aims to offer a comprehensive framework for the future integration of GAI in new media art (Kamar, 2016).

METHODS

This study aims to explore how the integration of GAI with new media art interaction modes affects user experience. A case analysis approach is adopted as the primary research method. By conducting an in-depth analysis of representative new media art cases, this study examines the specific applications of GAI in body-based interaction, voice-based interaction and biofeedback-based interaction, as well as their impacts on user experience. The methodology framework is outlined as follows:

Case Selection Criteria

To ensure the relevance and scientific validity of this study, the selected cases meet the following criteria:

1. Technological Sophistication: Cases should demonstrate the practical application of GAI technologies, such as generating visual, linguistic, or audio content using tools like MidJourney or ChatGPT.
2. Interaction Diversity: Each case focuses on a specific interaction mode (e.g., body-based or virtual reality), allowing for a comparative analysis of user experiences across these modalities.
3. User Engagement: Cases should reflect user behavioral and emotional responses during the interaction process, such as recorded data on interaction depth and willingness to participate.

Evaluation Metrics

User experience is evaluated using the following dimensions:

1. Technological Sophistication: Cases should demonstrate the practical application of GAI technologies, such as generating visual, linguistic, or audio content using tools like MidJourney or ChatGPT.
2. Interaction Diversity: Each case focuses on a specific interaction mode (e.g., body-based or virtual reality), allowing for a comparative analysis of user experiences across these modalities.
3. User Engagement: Cases should reflect user behavioral and emotional responses during the interaction process, such as recorded data on interaction depth and willingness to participate.

Based on the case analysis, this study further extracted design strategies and creative methods suitable for GAI-enabled new media art and validated these strategies through the researcher's own creative practices. Through this process, the study ultimately derived insights into the application value of GAI in different interaction modes and its implications for user experience design.

CASE STUDIES

The integration of GAI with interactive media art has redefined how users engage with digital artworks. Unlike traditional new media art, which primarily relies on pre-programmed interactions, GAI leverages advancements in generative models—from early Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) to contemporary diffusion frameworks—to enable real-time, dynamic content generation. This evolution reflects a shift from static rule-based systems to adaptive, user-driven artistic processes. Among the various interaction modes, body-based, voice-based and biofeedback-based interactions have emerged as key modalities in AI-driven interactive art. These methods allow users to engage with AI-generated content through natural, intuitive, and personalized interfaces. This section reviews representative works that leverage GAI within each interaction mode, analyzing their implementation, artistic impact, and research challenges.

Body Interaction

Body interaction within GAI-driven new media art has evolved beyond simple gesture-based inputs, enabling full-body motion to drive dynamic dialogues between participants and digital installations. A prime example is AI Forest, designed by Hu Haijie and the VAVE Studio team, located in Shanghai's Xintiandi. In this installation, visitors navigate a constructed urban "forest" where pathways built atop architectural ruins lead to interactive wells composed of screens and mirrors that simulate reflective water surfaces. As participants gaze into these wells, an AI-driven system captures their facial features, transforming them into digital animal avatars and generating a personalized 30-second video accessible via QR code. This innovative process

challenges traditional self-perception by embedding human identity within an ecological narrative and prompting users to reconsider their relationship with nature and their environmental responsibilities (Slater & Wilbur, 1997; Isbister, 2016).

Beyond its aesthetic and narrative achievements, AI Forest also illustrates significant technical challenges inherent in GAI-driven body interaction. The system must process facial data in real time, ensuring that any delay or inaccuracy does not disrupt the immersive experience. This need for precise, low-latency performance underscores the high computational demands of real-time visual synthesis and highlights the importance of optimizing motion detection algorithms (Bolton et al., 2015; Ramesh et al., 2021). In essence, while AI Forest successfully integrates body interaction into digital art, it also sets the stage for future innovations that balance technical precision with artistic expression, ultimately enhancing interactive storytelling in new media art.

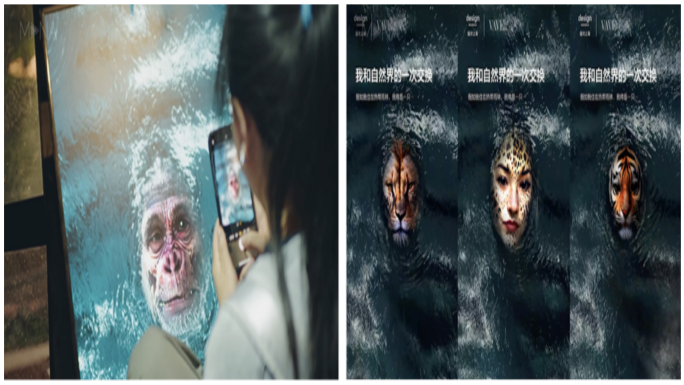

Figure 1: AI Forest facial recognition process.

Voice-Based Interaction

GAI-driven voice-based interaction art fuses artificial intelligence, human speech, and real-time computational creativity to transform abstract verbal commands into concrete visual images. By leveraging AI-powered speech recognition and generative models, this art form deciphers spoken language into structured prompts that guide image generation. The resulting visuals are dynamic and responsive, continuously adapting to audience input. This interplay between the inherent ambiguity of language and the precision of AI not only enriches the aesthetic experience but also expands cognitive and perceptual engagement, challenging traditional notions of authorship, creativity, and meaning-making.

A notable example is the voice interaction art demo developed by Annie and the Xinghuan Laboratory, which employs TouchDesigner as its central processing platform to seamlessly integrate speech recognition, image generation, and real-time visual effects. In this demo, the system accurately interprets spoken input and converts the extracted semantic meaning into

structured text prompts that drive an AI-based image generation model, resulting in unique visual compositions dynamically shaped by participants' words. Following the initial image creation, the artwork further evolves as the generated 2D image is transformed into a three-dimensional point cloud model, adding depth and spatiality. Within the real-time rendering environment, the point cloud is continuously adjusted in response to additional auditory cues and environmental inputs, altering its form, density, and motion. This dynamic transformation bridges the gap between verbal expression and digital aesthetics, offering an immersive platform that invites viewers to engage in an ongoing dialogue with the evolving digital landscape.

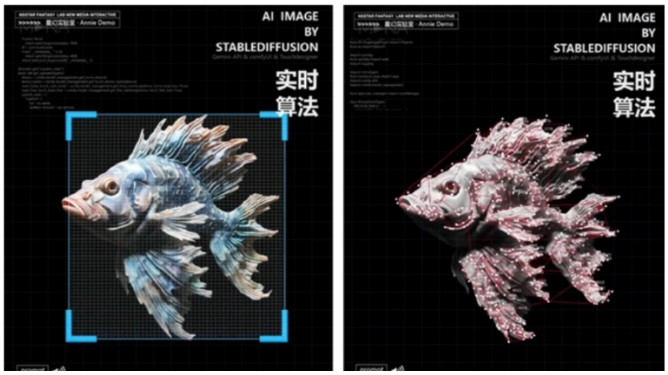

Figure 2: Voice-to-image generative process in Xinghuan Laboratory's GAI-driven art demo.

Biofeedback-Based Interaction

Biofeedback-based interaction, however, takes the form of interactive artworks to the next level. Biofeedback-based interaction, in form, eliminates cumbersome command-based interactions, whether they rely on gestures, language, or physical sensations,relying solely on participants' biological attributes, such as respiratory patterns, cardiac rhythms, or bioelectrical signals. This gives participants higher liberty to explore and reflect during the immersive experience. The audience's biological characteristics shape the generation of the artwork, while the artwork, in turn, elicits introspection, thereby exerting a reciprocal influence on their biological state. In a way, biofeedback-based interaction blurs the boundaries between the audience and the artwork. The following case studies illustrate how GAI empowers biofeedback-based interaction in interactive art while also highlighting technical challenges and artistic implications.

Intention Flower, a real-time interactive artwork created by Annie, leverages real-time brainwave monitoring and GAI-generated imagery to create a deeply personalized and immersive experience. Equipped with an EEG (electroencephalography) device, participants engage in a dynamic interaction where their neural activity directly influences the visual output; the system captures fluctuations in brainwave patterns and translates them

into algorithmic instructions that guide an AI-driven painting process. Notably, only until the participant's level of focus reach a predetermined threshold and a verbal command is issued does the generation process begins, thereby ensuring a deliberate and conscious engagement with the installation. Furthermore, the generated image will be transformed into a particle-based model in TouchDesigner to render and manipulate the visual effects. The more concentrated participants are, the more particulate and variegated the image is.

Figure 3: EEG-based neural interaction in Intention Flower.

This adaptive mechanism not only externalizes cognitive states into tangible artistic forms but also invites participants to explore the interplay between mental concentration, verbal input, and digital aesthetics, bridging the domains of neuroscience, artificial intelligence, and generative art.

ANALYSIS AND DISCUSSION

Our case studies—ranging from AI Forest's body-based interaction to voice-driven and biofeedback-based installations—demonstrate how GAI reshapes user participation in new media art. For instance, AI Forest transforms visitors' physical movements into digital avatars that merge ecological narratives with real-time visual synthesis (Slater & Wilbur, 1997; Isbister, 2016). Similarly, voice-based interactions convert spoken language into dynamic visual compositions, while biofeedback systems capture physiological signals to externalize users' inner states. Each mode, though innovative, reveals inherent technical and experiential challenges such as latency, limited interaction modalities, and difficulty in preserving dynamic outputs.

Our analysis also highlights several critical issues. First, sensor-dependent interactions tend to offer a singular mode of engagement, limiting creative diversity; this can be alleviated by integrating text-to-image generation methods. Second, artworks built on predetermined themes restrict users' creative freedom—leveraging large models to let users define their own themes can enrich the experience. Third, the inherent variability of real-time interactions makes it challenging to archive and share the content; combining these systems with a web-based platform can ensure that dynamic outputs are saved and disseminated. These insights underscore the necessity of balancing technical reliability with artistic expression, as noted in studies by Oviatt (1999), Norman (2013), and others.

These findings pave the way for our subsequent project design. By addressing the identified challenges—through the adoption of multimodal interaction techniques, lightweight AI models, and modular frameworks that support cross-modal data synergy—we aim to develop an interactive system that enhances user freedom and ensures real-time content preservation and shareability. This new approach not only builds on the strengths of existing case studies but also sets a clear roadmap for overcoming current limitations, ultimately fostering a more inclusive and adaptive human-AI collaborative ecosystem.

PROJECT DESIGN

This project is designed based on the multimodal characteristics of GAI technology, aiming to explore innovative forms of expression in the field of new media art. With the advancement of GAI technology, its applications in artistic creation have gradually expanded from single-content generation to complex multimodal interactive experiences. Through multi-stage technological integration and process optimization, this project has developed a comprehensive system that integrates speech, text, image, and speech synthesis. This system systematically demonstrates the potential of GAI in optimizing user experiences and enhancing artistic creation efficiency, while also expanding the boundaries of new media art. By combining technology and art, this project provides users with a novel digital art experience.

The design philosophy of this project emphasizes user-centric creation, aiming to reduce the impact of technological complexity on the creative process, making GAI an intelligent assistant rather than a technical barrier in artistic creation. Specifically, the system features an intuitive interactive interface that allows users to participate directly in the creative process through voice or text descriptions. The multimodal characteristics of GAI models are fully utilized in the system design, integrating technologies such as speech recognition, large language models, diffusion models, and speech synthesis to provide users with a multisensory creative experience. Additionally, the system design adopts modularization as its core logic. By using TouchDesigner and ComfyUI to build the technical framework, the workflow's stability and flexibility are enhanced, while allowing room for future scalability and customization.

The technical pathway of this project encompasses four stages, from text processing to content distribution, with each stage achieving seamless functionality through precise design and optimization.

The first stage is the text processing phase, which primarily uses Automatic Speech Recognition (ASR) technology to convert speech into text. OpenAI Whisper and Belle-whisper-large-v3-turbo-zh models were selected for this project, combined with a Voice Activity Detector (VAD) to enhance the accuracy and continuity of speech input, providing a high-quality semantic foundation for the generation phase.

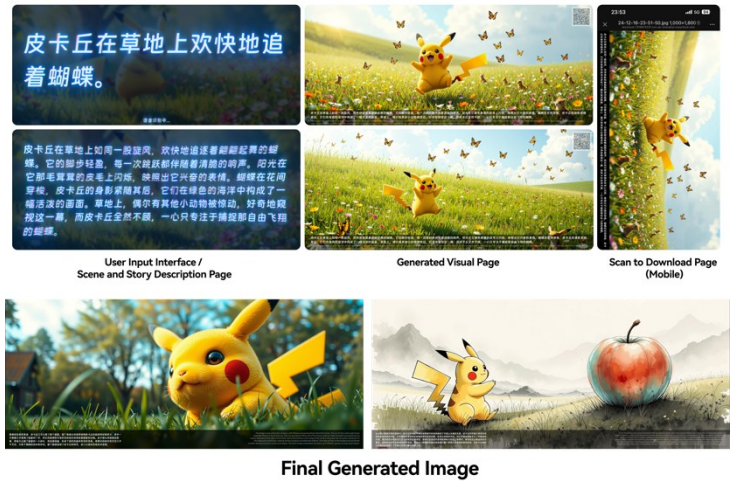

Figure 4: User interaction flow from input to content generation.

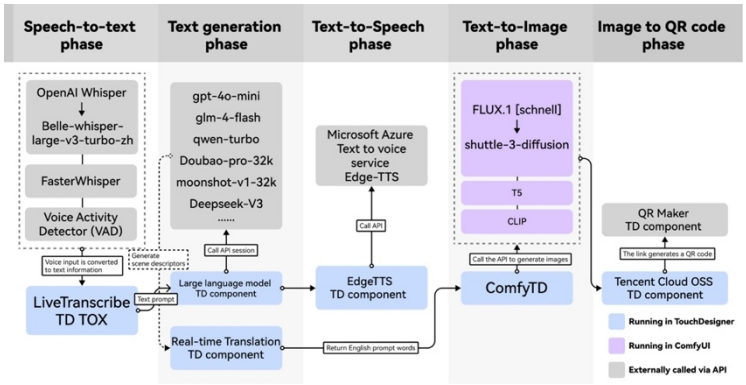

Figure 5: Multimodal system architecture.

The second stage involves content generation, where the system uses models like GPT-40-mini and Doubao-pro-32k to generate text content in real time. It also supports multilingual translation to meet the needs of users in different language environments. The third stage focuses on image and speech generation. Image generation is based on the diffusion model (Shuttle-3-Diffusion) and CLIP text encoding model (Radford et al., 2021), with fine-tuned optimizations to enhance the consistency between generated images and input semantic descriptions, while ensuring high-quality details in output images. Speech generation uses Microsoft Azure Cognitive Services' Edge-TTS technology to convert text into natural and fluent speech, providing users with an immersive audiovisual experience.

Finally, the content distribution phase employs Tencent Cloud Object Storage to host the generated content and includes a QR code generation module to enable easy downloading and sharing of content. This complete workflow not only enhances the smoothness of user experience but also validates the technological advantages of GAI in terms of real-time performance and multifunctionality.

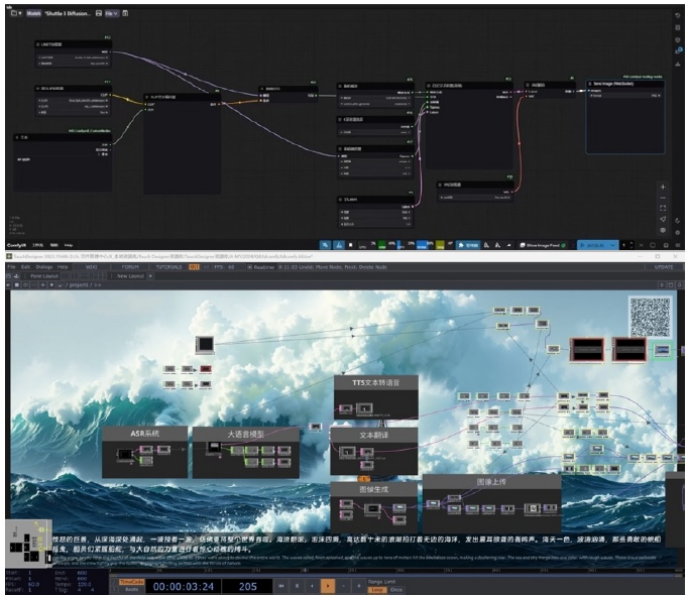

Figure 6: Modular workflow in TouchDesigner and Shuttle-3-Diffusion.

In practical applications, this project successfully achieved synchronized improvements in artistic creation efficiency and content generation quality. Taking the generation of a virtual image of "Pikachu chasing butterflies in the grass" as an example, it only takes a few seconds from the user's input description to the final high-quality image generation. The generated image is not only visually vivid but also highly consistent with the user's semantic input. This outcome demonstrates the collaborative effect of the diffusion model—grounded in the denoising probabilistic framework (Ho et al., 2020) and surpassing GANs in synthesizing photorealistic details (Dhariwal & Nichol, 2021)—and the CLIP text encoding model (Radford et al., 2021). The transition from GANs to diffusion frameworks addresses historical limitations in training stability and output diversity, enabling more reliable artistic expression. The generated content is further made accessible through cross-platform sharing via QR code links and Tencent Cloud storage, providing a reliable solution for the practical application of GAI in digital content creation. With the modular design of TouchDesigner and ComfyUI, each functional module achieved efficient collaboration, ensuring workflow stability and responsiveness in high-concurrency generation tasks.

This project not only showcases the innovation of GAI in technical implementation but also reveals its significant potential in optimizing user experience and enhancing artistic creation efficiency in application scenarios. Through multimodal interaction design, users can experience the complex generative process with simple operations, realizing the democratization of artistic creation. The research outcomes of this project provide a new technical pathway for new media art creation and explore broader possibilities for the artistic application of GAI.

CONCLUSION

This study systematically explores the technical implementation, user experience optimization, and future design strategies of GAI in new media art through interaction mode analysis. Findings reveal that body-based, voice-based, and biofeedback interaction expand artistic expression through GAI's dynamic generation across three dimensions: bodily expression, linguistic boundaries, and physiological data. Technically, optimizing real-time computational efficiency and modular design are key to enhancing system reliability; experientially, improving immersion and emotional resonance requires prioritizing interaction fluidity.

Case studies inform clear technical pathways for subsequent project design. For example, lightweight models and toolchain integration optimize generation efficiency, while cross-modal synergies enhance participatory diversity. In creative practice, the virtual image generated through voice input, validates technical feasibility, with its semantic consistency and efficient distribution serving as a practical application paradigm.

Future project design should further integrate the technical advantages of multimodal interaction, deepening the human-AI collaborative ecosystem through dynamic feedback mechanisms and real-time adaptation to user behavior. By integrating theory and practice, this study provides systematic guidance for the technological integration and design strategies of GAI in new media art, emphasizing that technology must serve the essence of artistic expression to advance a more inclusive and creative human-AI collaborative ecosystem.

REFERENCES

Bolton, A., Salmon, J. and Ross, D. (2015). Enhancing interaction fluidity in immersive virtual environments. Journal of Human-Computer Interaction, 30(4), 231–249.

Boden, M. A. (2019). Artificial intelligence and art: Toward computational creativity. Philosophy & Technology, 32(2), 195–209.

Chui, M., Intezari, A., Taskin, N. and Pauleen, D. (2021). Cognitive biases in developing biased artificial intelligence recruitment system. In: Proceedings of the 54th Hawaii International Conference on Systems Science.

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat GANs on image synthesis. Advances in Neural Information Processing Systems, 34, 8780–8794.

Duan, Y., Edwards, J. S. and Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, 48, 63–71.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.

Hassenzahl, M. and Tractinsky, N. (2006). User experience – a research agenda. Behaviour & Information Technology, 25(2), 91–97.

Ho, J., Jain, A. and Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.

Isbister, K. (2016). How games move us: Emotion by design. Cambridge, MA: MIT Press.

Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. Microsoft Research.
Norman, D. A. (2013). The design of everyday things. New York, NY: Basic Books.
Oviatt, S. (1999). Ten myths of multimodal interaction. Communications of the ACM, 42(11), 74–81.
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning.
Ramesh, A., Pavlov, M., Goh, G., et al. (2021). Zero-shot text-to-image generation. In: ICML.
Slater, M. and Wilbur, S. (1997). A framework for immersive virtual environments. Presence, 6(6), 603–616.