softserve

GENERATIVE ARTIFICIAL INTELLIGENCE IN 3D:

Exploring AI Capabilities in 3D Scene Generation

Abstract

The advent of generative artificial intelligence (Gen AI) has revolutionized creative processes across multiple domains, including 3D scene generation. This paper focuses on the transformative potential of Gen AI in supporting remote education and professional training, particularly in fields requiring practical, immersive experiences, such as architecture, design, education, and employee onboarding. By leveraging models like diffusion and transformer-based approaches, these systems enable the creation of photorealistic and customizable 3D environments with minimal human intervention, fostering accessible and scalable virtual learning spaces.

User interaction studies reveal that such technologies not only bridge geographical barriers but also provide hands-on experience in otherwise resource-intensive or hazardous environments (Bahroun et al., 2023). Moreover, the integration of Gen AI into educational and corporate workflows can streamline processes like onboarding and skill development, offering adaptive, personalized training modules (Batista et al., 2024).

However, challenges persist, including usability limitations, computational demands, and ethical considerations such as biases and intellectual property concerns. This work proposes a useful tool to the growing discourse on the role of generative systems in shaping the future of virtual education and training, emphasizing their potential to democratize access to experiential learning and improve workforce readiness.

Key Words

Generative AI 3D | Interactive | Spatial computing | Generative AI | AI-Driven 3D Scene Generation | Gen AI 3D

1. Introduction

Gen AI is reshaping creative processes by enabling rapid production of complex, photorealistic content. Applications such as gaming, virtual reality, architectural visualization, and film production have particularly benefited from these advancements. Traditional 3D content creation is often resource-intensive, requiring skilled professionals, substantial time, and significant costs. However, Gen AI offers a paradigm shift, speeding up many of these processes while maintaining creative flexibility and with scalability in mind.

At the core of this transformation are advanced machine learning models, such as diffusion models and transformer-based architectures, which can generate high-quality 3D objects and scenes with minimal input. These systems interpret user prompts, automate asset creation, and streamline workflows, enabling designers and developers to focus on higher-level creative decisions. Despite their promise, deploying Gen AI for 3D scene generation is not without challenges. Issues such as ensuring usability, providing adequate user control, and addressing biases in generated outputs remain critical areas for improvement.

This paper investigates the capabilities of Gen AI in the realm of 3D scene generation, focusing on its ability to augment human creativity and optimize workflows. Through a detailed analysis of existing models and an evaluation of a novel AI-driven system, we explore how these tools can balance automation and user agency. Furthermore, this research examines ethical and technical concerns, including intellectual property issues, computational demands, and the inclusivity of generative models. By bridging the gap between AI capabilities and human-centric design, this study aims to contribute to the development of intuitive, collaborative, and efficient systems for 3D content creation.

The paper is structured as follows: a literature review outlines existing challenges in 3D content creation and explores AI-based solutions and their role in 3D; the methodology section details the proposed system and its operational workflow; the results section showcases results and provides a comparative analysis of cost and time efficiency. We conclude with an evaluation of the broader implications of AI-driven 3D scene generation across various industries. By addressing current limitations, this research aims to contribute to the development of AI-driven solutions that will shape the digital world into a vibrant and accessible digital space with feasible future scalability.

2. Background and Related Works

The evolution of generative artificial intelligence has significantly impacted the creative industries, particularly in 3D scene generation. This section provides an overview of key contributions from recent research, focusing on advancements, methodologies, and challenges in using AI for automated 3D content creation. Studies have explored various approaches, including diffusion models, Generative Adversarial Networks (GANs), and transformer-based architectures, which have demonstrated the ability to generate photorealistic and complex virtual environments. However, gaps remain in areas such as usability, scalability, and ethical considerations, necessitating further exploration of how these technologies can be optimized for real-world applications. By critically examining the state of the art, this review establishes the foundation for the proposed innovations in AI-driven 3D content generation.

2.1 Theoretical Background

Generative artificial intelligence (Gen AI) has rapidly emerged as a transformative tool in the creative domain, particularly in 3D content generation. This section reviews the theoretical foundations behind the AI models that drive such advancements, focusing on the underlying mechanisms of generative adversarial networks, diffusion models, and transformer-based architectures. These technologies, while distinct in their operation, share a common goal of automating and enhancing the creative process, offering significant potential for applications in virtual environments, gaming, architecture, and film production.

2.1.1 Generative adversarial networks (GANS)

Introduced by Goodfellow et al. in 2014, GANs consist of two neural networks—a generator and a discriminator—that work in opposition to create realistic data samples (Goodfellow, I. et al, 2014). The generator aims to produce new data that closely resembles real data, while the discriminator's job is to distinguish between generated and real data. The adversarial process encourages the generator to produce increasingly realistic outputs, making GANs particularly suitable for applications like 3D object generation and texture creation (Abdellatif, Elsheikh and Menke, 2024; Schnepf, Vasile and Tanielian, 2023; Sheremetieva et al., 2023). In recent years, GANs have been applied to the generation of photorealistic images and 3D models, demonstrating their potential for creative industries such as film and gaming (Karras et al., 2018; Schnepf, Vasile and Tanielian, 2023).

2.1.2 Diffusion models

Diffusion models, another powerful approach in generative AI, simulate the process of transforming data into noise and then reverse it to generate new data. Unlike GANs, which rely on adversarial training, diffusion models work by gradually denoising a random input to create meaningful outputs (Ho, Jain and Abbeel, 2020). In 3D content generation, diffusion models have shown remarkable capabilities in producing high-quality textures and complex scene structures, while often requiring less computational power compared to GANs (Wang, Li and Jiang, 2024). These models have recently gained popularity for their ability to generate diverse, photorealistic results, especially for creating 3D environments for virtual and augmented reality (Kim et al., 2023).

2.1.3 Transformer-based architectures

Transformers, initially designed for natural language processing tasks, have also found applications in generative 3D content creation. Transformer models, such as GPT (OpenAI, 2024), have proven effective at understanding and generating complex sequences, making them suitable for interpreting user prompts and generating corresponding 3D objects. When applied to 3D scene generation, transformer-based models can process large amounts of data, enhancing the automation of scene composition, object positioning, and texture assignment (Po and Wetzstein, 2023; Huang et al., 2023). These models are also highly adaptable, allowing for fine-tuning based on specific creative inputs, which is crucial for applications in industries like gaming and architecture.

2.1.4 Interplay of models in 3D content creation

The integration of these advanced models (GANs, diffusion models, and transformers) enables the creation of high-quality 3D content with the ability for human-design-abilities interactive augmentation (Kazi et al., 2017; Zhang et al., 2024). While each model has distinct strengths, combining their capabilities can lead to even more powerful systems that automate object generation, scene composition, and asset management, accelerating workflows in industries where 3D content is crucial. The synergy between these models is central to the ongoing evolution of AI-driven design tools, pushing the boundaries of what is possible in 3D content creation.

2.2 State-of-the-art

Generative artificial intelligence (Gen AI) has rapidly emerged as a transformative tool in the creative domain, particularly in 3D content generation. This section reviews the theoretical foundations behind the AI models that drive such advancements, focusing on the underlying mechanisms of generative adversarial networks, diffusion models, and transformer-based architectures. These technologies, while distinct in their operation, share a common goal of automating and enhancing the creative process, offering significant potential for applications in virtual environments, gaming, architecture, and film production.

2.2.1 Advancements in AI for 3D content generation

Recent advancements in generative artificial intelligence (Gen AI) have significantly transformed 3D content creation. A notable breakthrough has been the use of diffusion models and generative adversarial networks (GANs) for generating photorealistic 3D scenes and objects. Diffusion models, such as DALL-E 2 from OpenAI, have shown promise in generating highly detailed and diverse 3D objects from textual descriptions. These models enable users to specify the design of an environment or object in natural language, drastically reducing the time and effort required to manually craft these assets.

3D model generation has been further enhanced through innovative algorithms and frameworks. OpenAI's PointCloud Models are a prime example, allowing for high-fidelity 3D reconstruction from images and text, improving the realism and accuracy of digital assets (Shi et al., 2023). These advancements are complemented by 3D CLIP search (Song et al., 2024), which leverages text-image matching to retrieve and generate relevant 3D objects from large datasets. The ability to retrieve 3D models based on semantic understanding of a scene has opened new possibilities for automatic scene generation, reducing the reliance on human-curated asset libraries.

Additionally, the application of Procedural Texturing and other texture generation methods has been revolutionized by AI. Procedural texturing algorithms allow for the creation of complex surface textures without the need for manual intervention, ensuring that textures scale dynamically with 3D models. Techniques like style transfer and neural network-based texture synthesis (Zhang et al., 2024) are now widely used to enhance the visual quality of 3D environments, enabling real-time applications in virtual worlds and simulations.

Furthermore, Structured Domain Randomization (SDR) has been applied to improve object placement and scene generation. SDR techniques randomize scene variables such as object placement, lighting, and background to train AI models in diverse scenarios, enhancing the robustness of 3D environments for applications like robotics and autonomous vehicles (Prakash et al., 2020).

2.2.2 Application across industries

The potential applications of AI in 3D content creation are vast, with industries like gaming, virtual reality (VR), architecture, and film production leading the way. In gaming, AI-driven 3D generation tools enable developers to automate world-building, creating expansive, interactive environments at a fraction of the traditional cost. Services such as Animate Anything (based on the CLIP model) have introduced AI-powered character and environment animation directly from text prompts, empowering creators to generate entire gaming worlds dynamically (Zhang et al., 2024; Po & Wetzstein, 2023).

In virtual reality, AI-based 3D content generation facilitates the creation of immersive worlds tailored to individual users. By combining procedural generation techniques with AI-driven texturing and asset generation, VR developers can quickly populate virtual spaces with realistic, interactive elements. This scalability is crucial as VR applications demand high-quality environments with constant updates.

Architecture and design also benefit significantly from AI-based 3D modeling. Tools like Meshy (a 3D asset generation service) or Edify leverage AI to rapidly generate realistic models for buildings, interiors, and landscapes, helping architects visualize concepts more efficiently and accurately (Meshy.ai, 2024; NVIDIA et al., 2024). Additionally, AI-assisted design platforms allow architects to automate aspects of the building design process, such as room layout, facade generation, and environmental integration, using AI-driven algorithms (Onatayo et al., 2024).

Finally, in film production, AI-driven 3D scene generation is transforming the process of creating visual effects. Generative models are employed to create intricate 3D environments for films with the ability for interactive human creativity augmentation (Zhang et al., 2024; Kazi et al., 2017). This is particularly valuable in film production, where the demand for detailed, customizable 3D models and scenes is often overwhelming.

2.3 Contribution

In this paper, inspired by advancements in Generative artificial intelligence (Gen AI) and their application to creative industries, we examine how transformer-based and diffusion models can enhance 3D scene generation processes. Building on prior research on AI-assisted creativity and educational technology, we explore how these tools can address the challenges of creating immersive, photorealistic environments. Specifically, we focus on leveraging AI for remote education systems and workforce onboarding, where hands-on, visual training is often critical.

Our work investigates how Gen AI systems can generate highly realistic and customizable virtual environments to simulate real-world scenarios, providing learners and professionals with interactive spaces that bridge theoretical knowledge and practical experience. By integrating findings from usability studies and adapting user-centric design principles, we aim to balance automation with creativity, enabling AI-driven systems to meet the demands of diverse educational and training contexts. Furthermore, our research highlights how these advancements contribute to reducing resource constraints, improving accessibility, and fostering collaboration in both creative and professional development domains. Taken together, this study emphasizes the transformative potential of Gen AI in creating scalable, efficient, and intuitive solutions for 3D content generation and practical education.

2.4 System setup

The system setup for the AI-driven 3D scene generation platform includes a high-performance workstation equipped with an NVIDIA RTX 4090 GPU to handle intensive computational tasks related to generative modeling and real-time rendering. A dual-monitor setup provides one screen for user interaction through a graphical interface and another for displaying the generated 3D scenes.

The input device is a tablet for precision sketching and object manipulation. The workstation runs a custom software environment built on PyTorch, integrated with pre-trained diffusion and transformer-based models.

To streamline the user experience, the system employs voice recognition through a microphone, enabling natural-language prompts to guide the generation process. This multi-modal setup ensures adaptability for various creative workflows, from architectural visualization to gaming environment design.

The software environment runs within NVIDIA Omniverse USD Composer, leveraging the powerful capabilities of the NVIDIA ecosystem. The system is built upon the following components:

NVIDIA Asset Management Server: This backend manages a vast library of 3D assets and associated metadata, supporting efficient storage and retrieval.

Deep Search: A semantic search feature that allows users to find assets based on textual descriptions or keywords.

Custom LLM: A Mixture of Experts (MoE) large language model, optimized for scene generation tasks. The model can be optionally quantized to Q4K for reduced storage requirements, albeit with slight degradation in output quality.

RAG (Retrieval-Augmented Generation) Pipeline: The pipeline integrates with the LLM to provide contextually rich asset data. It retrieves and processes metadata, such as object names and specific properties, from the Asset Management Server, ensuring precise and efficient scene assembly.

This combination of advanced hardware and NVIDIA's Omniverse platform provides a robust, scalable environment for generating high-quality, interactive 3D scenes. We've initially built the demo inspired by NVIDIA's SIGGRAPH presentation (NVIDIA, 2022).

Additionally, in Figure 1, the setup architecture is displayed, which represents a practical evolution of the system described, incorporating advanced capabilities like RAG-based asset management, AI-driven scene assembly, and seamless cloud integration for deployment at scale.

To enhance scalability and meet enterprise requirements, this architecture provides an alternative approach for implementing the solution described earlier. By leveraging a cloud-based modular system with an integrated LLM and asset management pipeline, it offers a reliable and extensible framework for scaling 3D scene generation to support a growing number of users and complex scenarios.

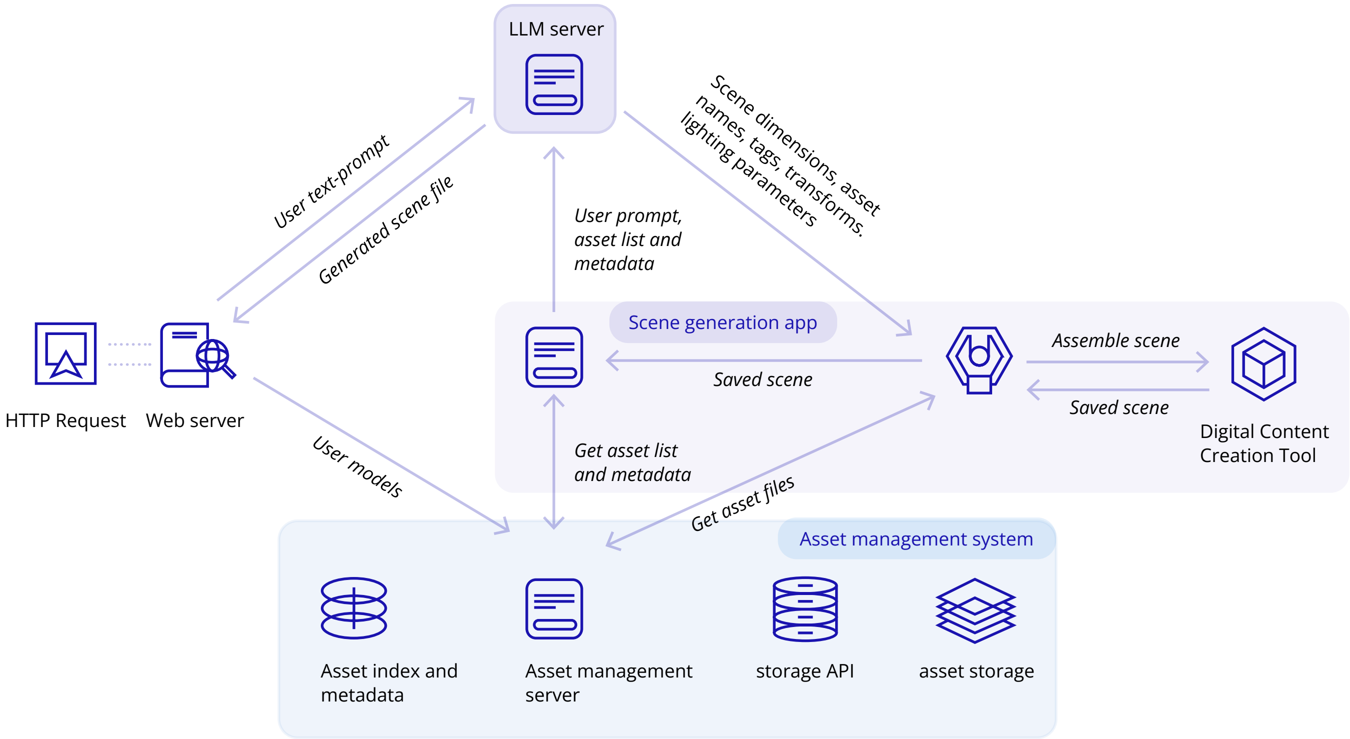

Figure 1. Proposed scalable system architecture.

The workflow presented in Figure 2 illustrates the system's process for generating interactive 3D scenes based on user prompts. The experience begins when the user submits a text prompt describing the desired scene. The system generates a virtual ground plane for the assets to be placed on. Next, the system retrieves metadata about required assets, including names and parameters, from a database.

This information is processed by a Retrieval-Augmented Generation (RAG) pipeline, which compresses and organizes asset data for efficient retrieval.

The RAG-enhanced data and user prompt are passed to a language model, which creates a detailed scene layout by interpreting the prompt and incorporating asset metadata. The system then identifies and downloads the necessary files for rendering the scene. After assets are gathered, they are positioned on the virtual ground plane according to the layout generated by the language model. Lighting parameters are applied to enhance realism.

Finally, the system renders a preliminary view of the scene, allowing users to preview it. The fully assembled scene file is made available for download, completing the interaction. This streamlined pipeline ensures the efficient generation of customized 3D environments while leveraging preexisting asset databases.

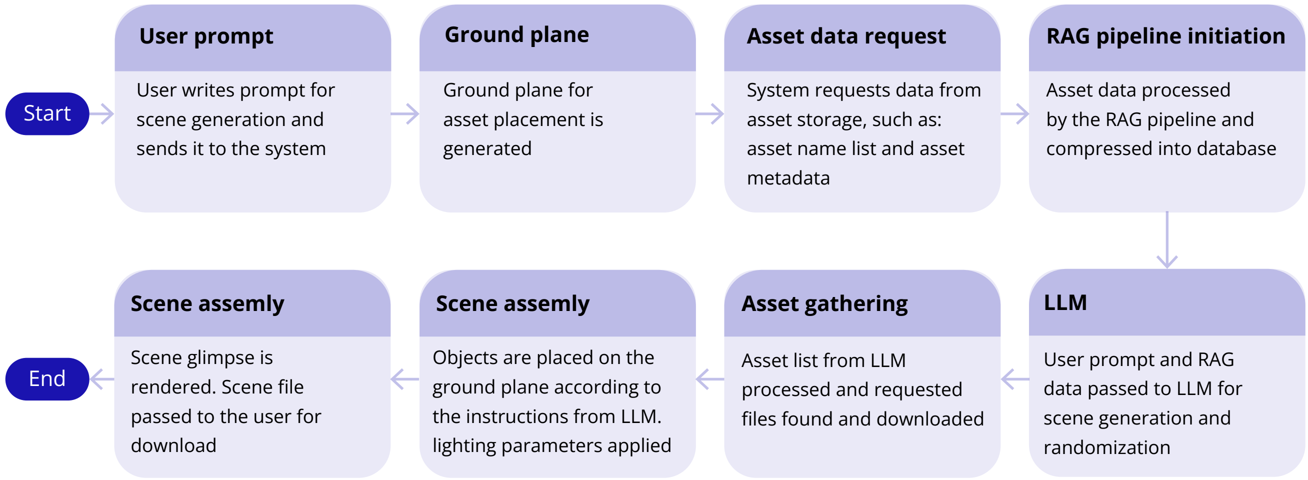

Figure 2. Flowchart of interactive experience.

2.5 Implementation

2.5.1 Architecture

The proposed system architecture, shown in Figure 1, consists of several interconnected components that facilitate scalable 3D scene generation using AI-based tools and asset management systems. The architecture integrates user interaction, asset indexing, and large language model (LLM) inference to automate the assembly and rendering of 3D environments. Further implementation details are described below.

1. **Web Server :** Acts as the central entry point for users, handling HTTP requests with user text prompts and returning the generated 3D scenes as responses. Integrates with downstream components, ensuring seamless communication between the user interface and backend services.

2. **Scene Generation Application :** Orchestrates the scene-creation workflow by processing user prompts and interfacing with other modules, such as the LLM Server and the Asset Management Server. Sends user-generated prompts and asset metadata to the LLM Server (3) for processing. Receives structured scene instructions, including asset names, positions, dimensions, and lighting parameters.

3. **LLM Server :** Runs a custom Mixture of Experts (MoE) LLM, which generates instructions for scene composition based on user prompts. Integrates a Retrieval-Augmented Generation (RAG) pipeline to query the Asset Management Server for relevant asset metadata, ensuring contextual accuracy and optimized object placement. Optionally supports model quantization to reduce computational overhead, enabling deployment in resource-constrained environments.

4

Asset Management System : Manages a database of 3D assets and their metadata, such as object dimensions, tags, and categories. Interfaces with a Storage API (5) to retrieve physical asset files for use in the generated scenes. The metadata, indexed for fast retrieval, enhances the efficiency of LLM scene composition.

5

Digital Content Creation Tool : NVIDIA Omniverse USD Composer to assemble, render, and finalize the generated 3D scenes. Receives structured scene parameters (e.g., asset placement, lighting configuration) from the Scene Generation App. Saves the final 3D scene for user access via the web server.

2.5.2 Detailed workflow

The list below provides a step-by-step workflow description.

1. User input and interaction:

a. Users submit a text prompt describing the desired 3D scene through the web interface.

b. The web server forwards this input to the Scene Generation Application for further processing.

2. Scene Composition with LLM and RAG:

a. The Scene Generation Application queries the LLM Server with the user prompt and available asset metadata.

b. The LLM Server, powered by a Mixture of Experts (MoE) model, generates structured outputs containing scene details (e.g., asset names, dimensions, positions, lighting).

c. The RAG pipeline ensures accuracy by retrieving relevant asset metadata from the Asset Management System, enriching the LLM's responses.

3. Asset Retrieval and Assembly:

a. The Scene Generation Application uses the generated instructions to query the Asset Management System for relevant assets.

b. The Asset Management System retrieves asset files from the Storage API and sends them to the Scene Generation App.

c. The Scene Generation App passes the assembled scene data to the 3D Software for final rendering.

4. Rendering and Output:

a. The NVIDIA Omniverse USD Composer processes the scene data, applying object transformations, lighting parameters, and other visual effects.

b. The finalized 3D scene is saved and returned to the web server for user access.

2.5.3 Detailed workflow

This architecture is designed for scalability, enabling deployment in cloud-based environments or on local servers. Key considerations include the following:

Modular Design:

Each component can be scaled independently to handle increased loads. For example, the LLM Server and Asset Management System can be hosted on separate instances for high availability.

Efficient Retrieval:

The use of a pre-indexed asset metadata database accelerates query times and reduces latency during scene generation.

Optimized AI Models:

The MoE LLM supports optional quantization (e.g., Q4K) to minimize resource usage, making the system adaptable for various deployment environments.

Cloud Integration:

The architecture supports integration with cloud-based storage APIs, enabling elastic scaling for asset retrieval and storage.

By following this implementation approach, the system provides an efficient, extensible, and scalable pipeline for AI-driven 3D scene generation at scale.

2.6 Discussion

The integration of advanced generative AI systems into 3D scene creation workflows marks a transformative shift in user experience (UX). While earlier generative models primarily focused on producing static assets or images, current approaches increasingly serve as generative scene-orchestration agents. These agents move beyond passive content creation to actively assist users in tasks such as object placement, stylistic adjustments, and scene configuration. Instead of viewing AI solely as a tool, this paradigm enables a more dynamic, collaborative relationship in which designers, educators, and other creative professionals guide and co-create with the AI.

A key facet of this evolution is the transition from traditional, manual inputs (e.g., mouse and keyboard) toward more intuitive, language-based interactions. By specifying desired outcomes through conversational prompts — such as requesting a hyper-realistic nighttime setting with vintage-inspired textures — users can focus on conceptual and aesthetic decision-making. This approach aligns with broader trends in HCI research, where advancements in natural language processing and prompt-based interfaces lower technical barriers and facilitate rapid, iterative exploration. As a result, experts can focus on high-level creative direction, while the AI efficiently generates variations and refines ideas.

Notably, the growing sophistication of these systems does not diminish the importance of human expertise. On the contrary, the role of the designer becomes even more critical for evaluating outputs, making nuanced creative judgments, and ensuring that generative results align with ethical and cultural values. Human oversight remains essential, especially given the expanding societal impact of automated systems. In tandem, developing industry guidelines, ethical frameworks, or professional certification programs can help ensure that these tools are deployed responsibly and inclusively.

Inclusivity and accessibility are likewise central to this new UX landscape. Future generative AI platforms should be designed to accommodate users with diverse abilities — such as those with visual, auditory, or motor impairments — through multimodal interfaces (e.g., voice commands, haptic feedback, and screen reader compatibility). This inclusive approach can democratize 3D content creation, enabling a broad range of users to collaborate with AI systems and contribute meaningfully to creative workflows.

Finally, as designers learn to navigate this environment, they will cultivate a new hybrid skill set that blends prompt engineering techniques with established domain expertise. For instance, familiarity with relevant artistic principles enhances a user's ability to obtain desirable results from generative tools. At the same time, the iterative feedback loop between humans and AI fosters mutual learning: users gain mastery of these tools' capabilities and constraints, while the systems themselves improve through exposure to expert guidance.

In sum, this transformation in 3D scene generation reflects a broader trend where generative AI evolves into a creative partner, not merely a productivity tool. By simultaneously expanding creative possibilities, reinforcing the importance of human judgment, and emphasizing inclusivity, these generative scene orchestration agents set the stage for a future in which designers devote less time to technical minutiae and more effort to conceptual vision, narrative integrity, and socially responsible innovation.

3. Conclusion

We have reviewed the up-to-date capabilities in 3D scene generation, described our exploration, and outlined our contribution. The proposed system design optimizes workflows, democratizes access to immersive experiences, and bridges gaps in usability, scalability, and ethical design.

The approach and the implemented system clearly demonstrate the speed-up and expansion of possibilities enabled by GenAI in the creative process in 3D. The LLM's spatial awareness still has some limitations, including misplacement and overlaps. The implications of using such systems are rather controversial. On one hand, we have a speed-up; on the other, we have not yet estimated the potential social impact.

We invite you to discuss the current capabilities and promises of the research field.

Refrences

1. Bahroun, Z. et al. (2023) 'Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis', Sustainability, 15(17). Available at: https://doi.org/10.3390/su151712983.
2. Batista, J., Mesquita, A. and Carnaz, G. (2024) 'Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review', Information, 15(11). Available at: https://doi.org/10.3390/info15110676.
3. Goodfellow, I. et al. (2014) 'Generative adversarial nets', in Z. Ghahramani et al. (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.
4. Schnepf, A., Vasile, F. and Tanielian, U. (2023) '3DGEN: A GAN-based approach for generating novel 3D models from image data', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2312.08094.
5. Abdellatif, A., Elsheikh, A.H. and Menke, H.P. (2024) 'Local padding in patch-based GANs for seamless infinite-sized texture synthesis', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2309.02340.
6. Sheremetieva, A. et al. (2023) 'What's my future: A multisensory and multimodal digital human agent interactive experience', in Proceedings of the 2023 ACM International Conference on Interactive Media Experiences. New York, NY, USA: Association for Computing Machinery (IMX '23), pp. 40–46. Available at: https://doi.org/10.1145/3573381.3596161.
7. Karras, T. et al. (2018) 'Progressive growing of GANs for improved quality, stability, and variation', arXiv [cs.NE] [Preprint]. Available at: http://arxiv.org/abs/1710.10196.
8. Ho, J., Jain, A. and Abbeel, P. (2020) 'Denoising diffusion probabilistic models', arXiv [cs.LG] [Preprint]. Available at: http://arxiv.org/abs/2006.11239.
9. Wang, Z., Li, D. and Jiang, R. (2024) 'Diffusion models in 3D vision: A survey', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2410.04738.
10. Kim, S.W. et al. (2023) 'NeuralField-LDM: Scene generation with hierarchical latent diffusion models', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2304.09787.
11. OpenAI et al. (2024) 'GPT-4 technical report', arXiv [cs.CL] [Preprint]. Available at: http://arxiv.org/abs/2303.08774.
12. Po, R. and Wetzstein, G. (2023) 'Compositional 3D scene generation using locally conditioned diffusion', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2303.12218.
13. Huang, S. et al. (2023) 'Diffusion-based generation, optimization, and planning in 3D scenes', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2301.06015.
14. Kazi, R.H. et al. (2017) 'DreamSketch: Early stage 3D design explorations with sketching and generative design', Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST 2017), pp. 401–414. Available at: https://doi.org/10.1145/3126594.312666
15. Zhang, L. et al. (2024) 'CLAY: A controllable large-scale generative model for creating high-quality 3D assets', arXiv [cs.CV] [Preprint]. Available at: https://doi.org/10.48550/arXiv.2406.13897.
16. Shi, R. et al. (2023) 'Zero123++: A single image to consistent multi-view diffusion base model', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2310.15110.
17. Song, D. et al. (2024) 'MV-CLIP: Multi-view CLIP for zero-shot 3D shape recognition', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2311.18402.
18. Prakash, A. et al. (2020) 'Structured domain randomization: Bridging the reality gap by context-aware synthetic data', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/1810.10093.
19. Meshy.ai (2024) 'Meshy.ai'. Available at: https://www.meshy.ai (Accessed: 27 November 2024).
20. NVIDIA et al. (2024) 'Edify 3D: Scalable high-quality 3D asset generation', arXiv [cs.CV] [Preprint]. Available at: http://arxiv.org/abs/2411.07135.
21. Onatayo, D. et al. (2024) 'Generative AI applications in architecture, engineering, and construction: Trends, implications for practice, education, and imperatives for upskilling—A review', Architecture, 4(4), pp. 877–902. Available at: https://doi.org/10.3390/architecture4040046.
22. NVIDIA (2022) 'As far as the AI can see: ILM uses Omniverse DeepSearch to create the perfect sky'. NVIDIA. Available at: https://blogs.nvidia.com/blog/ilm-omniverse-deepsearch/.

About US

SoftServe is a premier IT consulting and digital services provider. We expand the horizon of new technologies to solve today's complex business challenges and achieve meaningful outcomes for our clients. Our boundless curiosity drives us to explore and reimagine the art of the possible. Clients confidently rely on SoftServe to architect and execute mature and innovative capabilities, such as digital engineering, data and analytics, cloud, and AI/ML, robotics and physical AI.

Our global reputation is gained from more than 30 years of experience delivering superior digital solutions at exceptional speed by top-tier engineering talent to enterprise industries, including high tech, financial services, healthcare, life sciences, retail, energy, and manufacturing. Visit our **website**, **blog**, **LinkedIn**, **Facebook**, and **X (Twitter)** pages for more information.

AUSTIN HQ

201 W. 5th Street, Suite 1550
Austin, TX 78701
+1 866 687 3588 (USA)
Toll Free: +1 866 687 3588

LONDON

30 Cannon Street
London EC4 6XH
United Kingdom
+44 203 807 01 41

info@softserveinc.com
www.softserveinc.com

softserve