sigradi2024
biodigital_intelligent_systems

iBAG
UIC
BCN

Osmosis: Designing immersive urban environments using generative Artificial Intelligence

Iason Paterakis[1], Nefeli Manoudaki[1]

[1]University of California, Santa Barbara, Santa Barbara, USA

iason@ucsb.edu; nefeli@ucsb.edu

Abstract. This paper explores advancements in immersive space design and generative Artificial Intelligence (AI) by demonstrating a methodology for creating AI-driven immersive urban environments. "Osmosis" is a series of large-scale, real-time urban installations related to transmodal design in Extended Reality (XR) using generative AI. "Osmosis" blends real-time AI-generated visuals with urban settings, envisioning virtual elements merging with building façades and blurring the lines between tangible and virtual. These generative layers enhance collaboration, allowing data to move through and influence them, presenting a new paradigm in immersive urban design. This paper invites reflection on the evolving relationship between XR technology and AI-generated art in urban environments by showcasing AI's creative potential.

Keywords: Artistic Practices and Creative Industries, Generative Artificial Intelligence, Architectural Creativity, Extended Reality, Immersive Design

1 Introduction & Literature Review

The rise of immersive-mediated urban environments has gained significant popularity. Starting with iconic spaces like Times Square in New York City, we now see architectural-scale displays emerging in various city centers. High-tech interactive façades, projection-mapped buildings, and immersive museums with LED walls are shaping the future of augmented cities.

1.1 Emergence of urban media façades

Times Square, as described in Samuel R. Delany's "Times Square Red, Times Square Blue," evolved from a working-class neighborhood to a connected urban space where people can see themselves on high-resolution LED displays (Barron, 2004; Delany, 1999). As early as 1907, Times Square included large-scale electrified advertisements, making it one of the first augmented urban centers. Today, it has over 55 LED displays, highlighting content transformation into digital property within this iconic location. Other urban landscapes, such as Shibuya Crossing in Tokyo, COEX K-Pop Square in Seoul, Fremont Street Experience in Las Vegas, and Piccadilly Circus in London, are also influenced by media-driven surfaces. Globally, large-scale architectural structures adorned with LEDs are becoming more common, with the "Sphere" in Las Vegas standing out as a prime example (Keats, 2023).

1.2 Media Arts and Media Architecture

Times Square, as media are being creatively and experimentally showcased at International Light Festivals. The renowned Fête des Lumières festival, held in Lyon since 1852, has incorporated projection mapping into its displays since the early 2000s. After the Fête des Lumières in France, the Festival of Lights in Berlin, Germany, Light Night Leeds in the UK, Tokyo Lights in Japan, and the Amsterdam Light Festival in the Netherlands stand out as some of the most prominent festivals that showcase creativity through engaging light installations and projection mapping in urban settings.

Architecturally, there's a disparity between light installations and permanent non-permeable structures on building façades. In Times Square, the architectural charm of its historic buildings is overshadowed by black LED displays that rely on electricity. While captivating during operation, these urban areas could become alienating without power. The challenge lies in integrating modern display technologies into existing structures. Newly constructed buildings seamlessly incorporate displays, but the concern remains about what happens when the display technology becomes obsolete.

From a content perspective, urban centers have become advertising platforms controlled by private companies focused on profit. In contrast, cultural events and museums showcase artistic content, emphasizing aesthetics, creativity, diversity, equity, inclusion, and ethics. Although advertisements can be compelling, scaling material produced for personal displays to building-sized screens often overlooks the medium's scale and location. This lack of choice in visual aesthetics can disconnect users from the architectural integrity and identity of the space.

1.3 Artificial Intelligence and Extended Reality in Architecture

Over the past few years, digital content creation has evolved through several stages to today's generative AI-enhanced aesthetics. In June 2021, we participated in the production of "Desire for Freedom," one of the most expansive cultural projection mapping installations in Europe (Liapi et al., 2024). This installation was showcased on 18 historically significant buildings, including the Greek Parliament (Figure 1). The narrative, created by analyzing, reconstructing, and animating over 300 paintings from international galleries, was tailored to the distinct façades of each building.

If we were to undertake this project today, using current Generative AI models, the creative process would have been significantly different. The integration of Generative AI and modern Graphics Processing Units (GPUs) has radically transformed speculative design processes.

Figure 1. The 18 buildings underwent 3D scanning, 3D modeling, and projection mapping for "Desire for Freedom". Source: Institution Hidden.

AI is widely used as a creative and design tool in arts and architecture (Leach, 2021). The first stage of ideation and speculation is ideal for applying AI tools, particularly text-to-image and image-to-image algorithms (Bastos, 2023). Tools like Stable Diffusion, MidJourney, and DALL·E are favored by creators due to their extensive interfaces and output control, and platforms such as "Hugging Face" offer numerous choices for downloading entire AI models and running them locally on a personal computer.

In architecture and the arts, it's crucial to be able to train generative AI models. Pix2Pix and Generative Adversarial Network (GAN) algorithms, including CycleGAN and StyleGAN, were among the first used in architectural composition, followed by Generative Diffusion models (Hassab et al., 2021). “Deep Himmelblau” is an evolutionary research project integrating AI with the projects of Coop Himmelblau. Deep Learning enables the speculation and prediction of new architectures based on training with existing ones (dPrix et al., 2022).

Similar expressions exist in the space of Media Arts and Extended Reality. Refik Anadol studio is among the first to explore large databases and the latent space of Digital Arts and Machine Learning. Gaudi embraced generative design and procedural techniques and Refik expressed that through the work “Living Architecture : Casa Batlló” (Anadol, n.d.). Gaudi's embrace of generative design and procedural techniques is expressed through Refik Anadol's work “Living Architecture: Casa Batlló.” Casa Batlló served as a platform for generative data expressions in this project. Refik Anadol Studio utilized the intricate relationship between Gaudi's marine-inspired façade, the surrounding urban environment, and real-time weather data to fuse XR with physical architectural composition.

The intersection of Generative AI and XR is redefining speculative design and virtual architecture. Immersive mediums like projection mapping have advanced generative and procedural design methodologies. Recently, Generative Diffusion Models (GDMs) have revolutionized virtual content creation (Wang, He, & Peng, 2024). This rapid evolution has had an immediate impact on architectural design and ideation.

XR technologies have evolved and integrated into public spaces through high-fidelity LED displays, projections, and recently AR devices like the Apple Vision Pro. The fusion of AI and XR suggests that AI-generated morphologies will be tangibly integrated into future urban façades. Although hybrid AI spaces have been presented through projection mapping and interior installations, the computational demands have often removed interaction and real-time collaboration from the process.

2 Methodology

The "Osmosis" series investigates the symbiosis between generative AI content creation algorithms and immersive urban environments, presenting a novel biomimetic approach to architectural projection mapping. Inspired by the biological process of osmosis—where solvent molecules traverse a semipermeable membrane along a concentration gradient—we've developed a methodological framework that conceptualizes the urban facade as a dynamic interface between digital and physical realms. This framework posits AI-generated visual content as the 'solute' and the built environment as the 'solvent', with the projection surface acting as a semipermeable membrane. By algorithmically modulating the 'osmotic pressure' of visual elements, we facilitate a controlled diffusion of digital information into the architectural space. This bio-digital methodology enables a gradient-driven flow of AI-generated content, allowing for the emergence of hybrid spaces that challenge traditional notions of architectural boundaries. Notably, this osmotic concept extends beyond the visual metaphor to inform our technical implementation, with the software pipeline and inter-device communication protocols designed to emulate the selective permeability and directional flow characteristic of biological osmosis. This approach not only metaphorically emulates osmotic processes but also provides a quantifiable model for integrating generative AI with the built environment, contributing to the evolving discourse on bio-digital architecture.

This paper describes three generative projection mapping installations publicly exhibited at the Santa Barbara Center for Arts, Sciences, and Technology (SBCAST) in California in 2023 and 2024. Each iteration experimented with generative text-to-video, image-to-video, and video-to-video AI models, projection mapping elements, and alternative system setups. The results were presented as a series of real-time generative artworks, showcasing different AI-generation techniques. The evolution of the design process was demonstrated through public projection mapping installations on the same building facades.

Our research explored the integration of Artificial Intelligence and Architectural Synthesis with Media Arts and XR. The primary objective is to explore the use of Diffusion AI models for immersive virtual content in urban spaces. We selected Machine Learning Diffusion AI models for their ability to operate with natural language, making them more accessible and comprehensible to a wider audience compared to lower-level programming languages. This makes LLMs an effective form of communication and interaction. In this series of projects, we expand this communication method to immersive urban spaces, enabling users to modify the environment using input devices and natural language descriptions/prompts. The proposed pipeline (Figure 2) outlined in this paper serves as a framework for generating architectural content using Generative AI and Extended Reality mediums.

Research Questions:

-How can a real-time, AI-driven content pipeline be developed to create immersive architectural façades?

-How can real-time collaboration be achieved to produce transmodal XR layers tailored to the architectural features of buildings?

2.1 AI Hybrid Content Generation Pipeline

Diffusion models create data by progressively refining a sample from an initial distribution, such as Gaussian noise. When combined with transformers, which handle sequential data and capture long-range dependencies, diffusion models exhibit improved performance. The self-attention mechanism in transformers allows these models to understand and generate intricate data patterns. This collaboration results in highly detailed and contextually accurate content, making it particularly effective for tasks like image and video synthesis (Ho, Jain, & Abbeel, 2020).

In "Osmosis," we experimented with various AI models across different iterations. This allowed us to distinguish the results based on visual complexity and implement alternative approaches to real-time interactivity. All three installations utilized Diffusion Models, which need text, image, or video as input and generate outputs based on these inputs (Figure 2). We fed procedurally generated videos of the building and AI-generated images into these models. The decision to use projection mapping required adapting the content to the building's characteristics. The content design choices focused on the intricacy and texture of the forms presented, as well as their relationship to the façades and the surrounding space. The complexity of the visuals was influenced by the AI text prompts describing diverse natural formations, large-scale galaxy clusters, and microscopic images of microorganisms and bacteria.

Blender was utilized to create a procedural model of the building, integrating various effects through Blender's geometry nodes, and Python interface. Each building volume was designed with a virtual architectural frame containing doors, windows, and other characteristics. Fluid simulations and particle systems were incorporated in two cuboids -one for each building volume-, functioning as virtual containers for the simulations. Virtual cameras were strategically positioned for rendering, simulating projectors in physical space. This technique enhanced the sense of depth, influenced by the viewer's field of view, the vanishing points, and the building's perspective.

After generating procedural content, multiple videos and image series were exported for AI model input. Dall-E and MidJourney AI played pivotal roles in producing AI images using the procedurally generated videos of the building. These images depicted diverse abstract formations in space, nature, and the microcosmos. The Blender procedural videos, and AI images were then fed into the final AI video generation process.

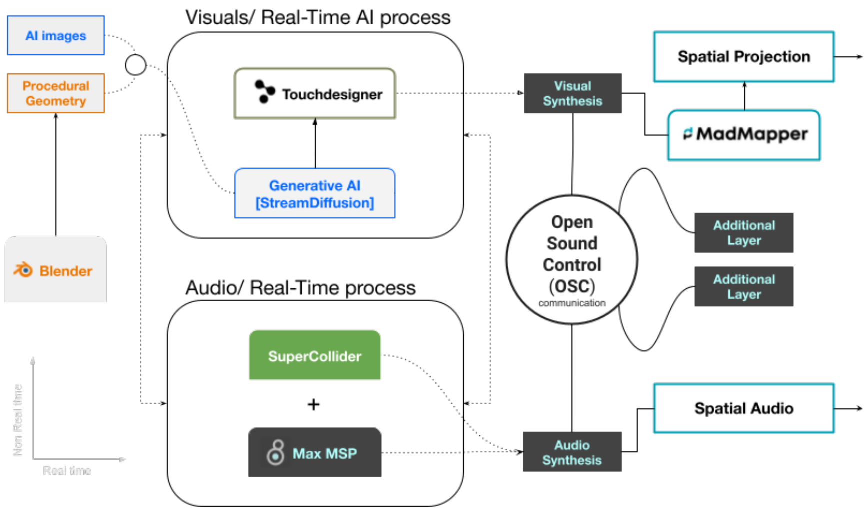


Figure 2. The proposed pipeline of the process for real-time AI-generated transmodal immersive façades. transLAB, MAT, UCSB. Source: Authors, 2024.

In the initial two installations, we proceeded with RunwayML's Gen-1 image-to-video model (Esser et al., 2023). We noted that the subsequent Runway Text-to-video models offered enhanced graphics but reduced control, limiting our projection mapping purpose. With Gen-1 we generated various video outputs based on the input AI images and procedural videos, maintaining the architectural characteristics of the building (Figure 3).

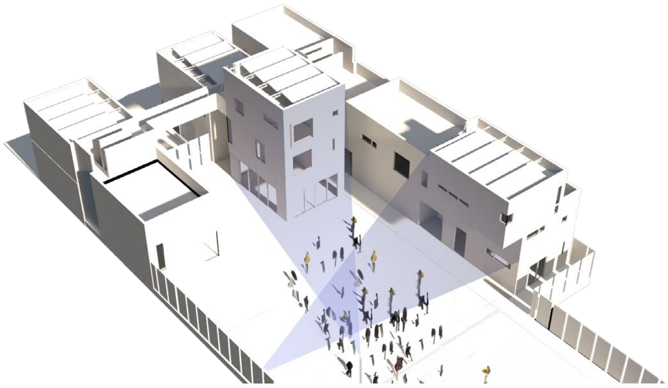
Figure 3. A collection of AI-generated façades overlaying the building's form with text-image, image-video, and video-to-video generations. Source: Authors, 2024**.**

However, these were linear pre-rendered videos, with limited length and repetitive patterns. We overlayed multiple layers of Perlin noise and blending effects using the visual development platform TouchDesigner, to allow for interactivity. The platform's flexibility and efficiency facilitated the seamless integration of the previously generated videos with real-time layers.

In the third installation, we integrated SteamDiffusion (Kodaira et al., 2023), a real-time interactive AI generation pipeline, directly into TouchDesigner, allowing for real-time AI video generation. This approach eliminated the need for additional real-time layers, as the AI video could be rendered and manipulated instantaneously. Utilizing our previously generated Blender videos enabled us to showcase the building's features and effects without recreating architectural frames and simulation effects within TouchDesigner. However, we encountered limitations in computing additional effects due to StreamDiffusion's high GPU utilization, which exceeded 80%. Additional layers of Perlin noise were employed as an input in StreamDiffusion, to increase the complexity of the AI generations.

2.2 Technical Implementation Pipeline

We utilized two distinct volumes within the SBCAST complex for the installations: one three-story and the other two-story, positioned at a 90-degree angle from each other (Figure 4). The positioning of the projected surfaces relative to the audience's placement around the central courtyard was crucial for perceiving depth. Performers were located at the center of the courtyard, facing the audience, while the audience, in turn, stood centrally in front of the performers, directly facing the projection-mapped volumes. Each performance had a duration of around 20 minutes.



.

Figure 4. Axonometric diagram of the installation layout of “Osmosis” SBCAST, transLAB, MAT, UCSB. Source: Authors, 2024**.**

One machine was connected to two large venue projectors and handled projection mapping through MadMapper, while the other streamed real-time AI video content to the first machine using Network Device Interface (NDI) over IP. Communication between the three computers involved data exchange related to various parameters through an Open Sound Control (OSC) network. OSC is a prominent networking protocol that facilitates message exchange among diverse devices, including synthesizers and computers.

Figure 5. A picture from the third installation of "Osmosis". SBCAST, transLAB, MAT, UCSB. Source: Authors, 2024**.**

Figure 6. A picture from the third installation of "Osmosis". SBCAST, transLAB, MAT, UCSB. Source: Authors, 2024**.**

3 Discussion

Urban spaces have been dominated by profit-driven, linear deterministic content. The emergence of Machine Learning Art and the impact of higher dimensional forms in architecture offer an opportunity to transcend the boundary of observable space through stochastic processes (Novak, 2002). This work leverages AI tools and contributes to the genre of avant-garde AI virtual architecture that can be experienced in physical space through XR. The three iterations of “Osmosis” illustrate that while Diffusion AI models are primarily used for conceptualization and speculation, their fusion with XR technologies provides a robust framework for immersive allourban spaces. This fusion introduces complex forms and real-time data-driven collaboration, creating a harmonious immersive ecosystem where urban space, AI-driven data, and its inhabitants can operate cohesively. This approach fosters creativity and collaboration through merging layers of AI and human-generated art. While digital and virtual design was once the primary focus, the AI era is the coming wave.

3.1 Comparison

When evaluating the three setups it is evident that current AI models have limitations. Although the installations occurred at different times, the Diffusion AI algorithms underwent significant changes that increased their speed, efficiency, and output resolution. The Runway Gen-1 produced non-real-time videos, whereas StreamDiffusion generated instant real-time videos. This evolution altered the pipeline, eliminating the need to add layers for interactivity. The AI-video generation became interactive by modifying the prompts and the seed of the generations in real time.

Furthermore, these models are still in an early stage of development. StreamDiffusion is considered one of the fastest models and can produce nearly 30 frames per second at 512x512 resolution inside TouchDesigner, although it is claimed to be capable of exceeding 90fps in general. In our case, the output ran below 24fps due to the 800x600 resolution. Output resolution is a limitation and requires additional high-resolution layers and effects overlaid for users to perceive an overall higher-resolution composition. The distance between the observer and the building also plays a role. It is important to note that the resolution of the generated image impacts the complexity of the result. Overlaying additional Perlin noise with the input videos provides a more complex result. The limitation of the resolution and the computationally expensive generations will be overcome shortly with the introduction of more efficient algorithms and faster GPUs with more memory.

3.2 Future Directions

In future research, it is important to consider incorporating more types of stimuli, produced by haptic and olfactory interfaces. Computation and data communication over the network is crucial, allowing artists to share their intellectual property and transmit data in real time. This system offers numerous potential for extension. Another research avenue extends from the concept of 'NeuroNanoBioTectonics,' which explored the fusion of nanotechnology, biotechnology, and computational methods to create self-organizing, adaptive architectures that blur the boundaries between the biological and the technological (Novak, 2003). Our most recent research project, “Organoid_Protonoesis 1”, demonstrates the potential for integrating real-time generative AI with biological systems. It invites participants to explore and contemplate the evolving boundaries of cognition, signal processing, and contemporary media. This project introduces a real-time interface that connects spontaneous brain organoid activity between lab hidden and lab hidden with external stimuli, enabling unique interactions between human intellect and emerging self-organizing neural connectivity. These real-time interactions operate with the AI pipeline of “Osmosis” and can extend through XR introducing n-dimensional morphologies in the urban landscape.

Acknowledgements. The authors extend their sincere thanks to SBCAST - Santa Barbara Center for Art, Science & Technology and Alan Macy for their generous support of these studies, as well as to the transLAB, MAT, UCSB. Additionally, the authors wish to acknowledge the valuable contributions of Ryan Millett and Sabina Hyoju AHN for their collaborative work during the performances and their contribution in the sound component of the performances using machine learning and biodata.

References

Delany, S. R. (1999). Times Square Red, Times Square Blue. New York University Press.

Barron, J. (2004, April 8). 100 Years Ago, an Intersection's New Name: Times Square. The New York Times. Retrieved June 25, 2024, from https://www.nytimes.com/2004/04/08/nyregion/100-years-ago-an-intersection-s-new-name-times-square.html

Keats, J. (2023, May 25). The Sphere is Here: Are We Ready for More High-Tech Architecture? Smithsonian Magazine. Retrieved June 25, 2024, from https://www.smithsonianmag.com/innovation/the-sphere-is-here-are-we-ready-for-more-high-tech-architecture-180983077/

Liapi, M., Manoudaki, N., Paterakis, I., Christoulakis, M., Ioannidis, M., & Oungrinis, K.-A. (2024). Interfacing Heritage and Technology: The 'Desire for Freedom' Project as a Paradigm of Extended Reality in Architecture. AMPS 2024 "Urban Futures Cultural Pasts" Conference, Universitat Politècnica de Catalunya (UPC), UCL Press. (forthcoming)

Leach, N. (2021). Architecture in the Age of Artificial Intelligence. Bloomsbury Publishing.

Bastos, B. C. A. (2023). Artificial Intelligence Applied to Ideation in Design. SIGraDi 2023.

Hassab, A., Abdelmohsen, S., & Abdallah, M. (2021). Generative Design Methodology for Double Curved Surfaces using AI.

dPrix, W., Schmidbaur, K., Bolojan, D., & Baseta, E. (2022). The Legacy Sketch Machine: From Artificial to Architectural Intelligence. Architectural Design, 92, 14-21. https://doi.org/10.1002/ad.2808

Anadol, R. (n.d.). Living Architecture: Casa Batlló. Retrieved June 25, 2024, from https://refikanadol.com/works/living-architecture-casa-batllo/

Wang, X., He, Z., & Peng, X. (2024). Artificial-Intelligence-Generated Content with Diffusion Models: A Literature Review. Mathematics, 12(7), 977. https://doi.org/10.3390/math12070977

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 33, 6840-6851. https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca2c-Paper.pdf

Esser, P., Chiu, J., Atighehchian, P., Granskog, J., & Germanidis, A. (2023). Structure and Content-Guided Video Synthesis with Diffusion Models. arXiv preprint. https://arxiv.org/abs/2302.03011

Kodaira, A., Xu, C., Hazama, T., Yoshimoto, T., Ohno, K., Mitsuhori, S., Sugano, S., Cho, H., Liu, Z., & Keutzer, K. (2023). StreamDiffusion: A Pipeline-level Solution for Real-time Interactive Generation. arXiv preprint. https://arxiv.org/abs/2312.12491v1

Novak, M. (2002). Speciation Transvergence Allogenesis Notes on the Production of the Alien. Architectural Design, 72(3), 64-71.

Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M. R., Groh, M., Schroeder, H., Smith, A., Akten, M., Fjeld, J., & Farid, H. (2023). Art and the science of generative AI: A deeper dive. arXiv preprint. https://arxiv.org/abs/2306.04141

As, I., Pal, S., & Basu, P. (2018). Artificial intelligence in architecture: Generating conceptual design via deep learning. International Journal of Architectural Computing, 16(4), 306-327. https://doi.org/10.1177/1478077118800982

Xiao, X., Wang, J., Shu, Y., & Tan, J. (2024). Creativity and perception: Unveiling the role of cross-modal audiovisual integration. The Journal of Creative Behavior. Advance online publication. https://doi.org/10.1002/jocb.668

Novak, M. (2003). Neuro~, Nano~, Bio~: New Atomism and Living Nanotectonics. ©2003 Marcos Novak.