Generative AI for Wellness Applications via User-Generated Immersive Virtual Environments

RAYNA NEY, Mobiliar Lab for Analytics, ETH Zurich, Switzerland
AYDIN JAVADOV, Mobiliar Lab for Analytics, ETH Zurich, Switzerland
MELANIE BAUMGARTNER, Mobiliar Lab for Analytics, ETH Zurich, Switzerland
RAPHAEL WEIBEL, Mobiliar Lab for Analytics, ETH Zurich, Switzerland
FLORIAN VON WANGENHEIM, Chair of Technology Marketing, ETH Zurich, Switzerland
JOSEPH OLLIER, Mobiliar Lab for Analytics, ETH Zurich, Switzerland

Generative AI (genAI) allows for the enhancement of personalization in immersive virtual environments (IVEs) for relaxation and wellness purposes. Unlike traditional customization methods, genAI enables users to generate environments via natural language prompts. This study evaluates the feasibility of genAI IVEs through a usability study where three participants created and experienced personalized environments using SkyboxAI and a Meta Quest 3 headset. Content analysis revealed themes of control, realism, engagement, and misalignment with expectations. While genAI offers creative freedom, limitations in dynamic elements and scene accuracy remain. Future research should refine genAI-driven customization to enable users to bring the landscapes they envision into virtual reality.

1 INTRODUCTION

Immersive virtual environments (IVEs) delivered via virtual reality (VR) have been widely applied in wellness interventions as a tool to improve mental health and resilience, for example, by offering breathing (e.g., box breathing [2] or relaxation (e.g., progressive muscle relaxation [6]) exercises, particularly in settings where access to preferred relaxation spaces is limited (e.g., during in-patient treatment [5] or in urban environments [4]). While research to date has underscored the importance of personalization in IVEs, such as increasing relaxation during guided full-body relaxation exercises [6], a new paradigm has emerged with the advent of generative AI (genAI).

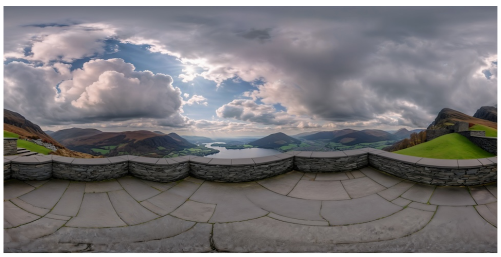
By providing highly detailed and customizable environments, generated near instantaneously, genAI enables users to adopt IVEs to their ongoing preferences, moods, and needs, moving far beyond the personalization afforded by static (i.e., menu-based) interfaces where users select from a limited number of pre-generated environments. For vulnerable populations with highly limited mobility, such as children confined to hospitals due to weakened immune systems, the incorporation of rich and varied AI-generated virtual environments to wellness interventions can help provide benefits comparable to real-world nature exposure [3].

Extant research has proposed that mechanisms behind the effectiveness of immersive virtual environments include increased control, engagement, sense of presence, immersion, and reminiscence — where aligning the virtual scenario with personal memories enhances the overall experience. For example, Pardini et al. implemented personalization using a drop-down menu, allowing participants to control the time of day, select from three realistic environments (mountain,

Authors' addresses: Rayna Ney, rayney@ethz.ch, Mobiliar Lab for Analytics, ETH Zurich, Switzerland; Aydin Javadov, aydin.javadov@ethz.ch, Mobiliar Lab for Analytics, ETH Zurich, Switzerland; Melanie Baumgartner, mbaumgart@ethz.ch, Mobiliar Lab for Analytics, ETH Zurich, Switzerland; Raphael Weibel, raweibel@ethz.ch, Mobiliar Lab for Analytics, ETH Zurich, Switzerland; Florian von Wangenheim, fwangenheim@ethz.ch, Chair of Technology Marketing, ETH Zurich, Switzerland; Joseph Ollier, jollier@ethz.ch, Mobiliar Lab for Analytics, ETH Zurich, Switzerland.

marine, countryside), and choose whether people were present in the scene. However, genAI eliminates the need for pre-defined options, with users able to simply describe their vision aloud. This raises questions as to the validity of the mechanism proposed in extant, non-genAI IVE research. For example, whether the relative simplicity of stating a sentence and then receiving a rich, detailed, and near-instantly generated world increases immersion and presence, and whether this process (which omits the slower and more controlled menu based interactions in extant research) comes at the expense of decreased control.

In the current paper, therefore, we aim to assess the feasibility of using genAI for IVEs and explore the mechanisms underlying user reactions to such environments using a small usability study.



Participant 1 Prompt: English Lake District with small distant peaks, slate-grey buildings, during the daytime

Participant 2 Prompt: Desert with sand dunes which goes directly into the ocean. The sun is going down in the sky, sunset time. The ocean is very calm, and the sky is vibrant and changing

Participant 3 Prompt: Bryce Canyon (Utah) on a nice spring morning, clear beautiful day with barely any clouds. Cedar trees

Fig. 1. The three different prompts from the three different participants and the corresponding output.

2 CASE STUDY ON USER-GENERATED IVES

As generative AI introduces new possibilities for personalizing IVEs, it is crucial to evaluate how effectively these AI-generated environments align with user expectations and support relaxation experiences. Traditional IVE customization methods rely on pre-defined options, which may limit personalization and user engagement. By contrast, genAI offers a more flexible, user-driven approach, enabling individuals to create environments purely from verbal descriptions. However, the extent to which these AI-generated environments meet user expectations and provide a sense of immersion remains an open question.

This case study was conducted to assess the feasibility of using genAI for the creation of IVEs intended for relaxation. Specifically, it aimed to explore whether users felt in control of the generation process, whether the resulting environments matched their envisioned settings, and whether any unexpected or misaligned outputs disrupted immersion. By understanding these factors, this study provides insight into the potential of genAI IVEs as a tool for virtual relaxation interventions.

2.1 Technical Specifications

The immersive experience was rendered on a head-mounted display using Unity 3D game engine and the Meta Quest 3. SkyboxAI was utilized for IVE generation [1]. SkyboxAI uses a Latent Diffusion model, LDM3D, which generates large 2D images with depth characteristics to simulate a 3D effect, and then uses autocomplete-like inference to generate missing data, enabling fast, high-quality scene creation for a 3D context [7]. Depth maps provide a 3D structure to 2D images with minimal computation time. SkyboxAI allows for a selection of 30 distinct styles, including digital painting, low poly render, fantasy, and Neo Tokyo. For our purposes, we used photo realism.

2.2 Participants and Procedure

A sample size of three participants was chosen for this study as an exploratory, qualitative investigation into the feasibility and user experience of generative AI-driven immersive virtual environments (IVEs). Given the novelty of genAI for IVE personalization, the primary goal was to gather in-depth insights into user expectations, perceived control, engagement, and misalignment rather than achieve broad generalizability. This approach aligns with case study methodologies and usability research, where small-scale studies are commonly used to identify key themes, usability challenges, and directions for future, larger-scale investigations. Participants were first briefed on the study's objective. They were told that they would be creating an immersive virtual environment that "would serve as their ideal place to relax, meditate, or be mindful in." It was explained that this place could be somewhere that really existed in the world, or completely fabricated from their imagination. Then, once ready, they described their envisioned setting in words. The prompt was then typed onto a computer screen, closely matching the participant's description with only minor edits to remove redundancy. The prompts and corresponding virtual environment can be seen in Fig. 1. The participants were able to edit the prompt before it was sent to SkyboxAI for generation. Once the IVE was generated, participants viewed it through the Meta Quest 3 HMD for the full immersive experience. While still immersed in the environment, an open-ended interview was conducted to explore participants' experiences, including their sense of control over the process and alignment with expectations of the output.

2.3 Content Analysis

The content analysis of participant interviews revealed four key themes: perceived control, engagement, misalignment, and relaxation. Quotes within these themes can be seen in Table 1. First impressions were overwhelmingly positive, with two of the participants initially expressing a "wow factor" upon entering their generated environments. One participant described their immediate reaction as being "almost exactly like it," indicating an initial sense of realism and immersion. The visual fidelity and scale of the environments contributed to these strong first impressions, with several participants remarking on details such as the sky, slate textures, and atmospheric conditions that closely resembled their expectations. However, as participants spent more time in the environment, minor discrepancies in realism and interactivity became more apparent, leading to more critical reflections.

Main Theme	User Quotes
Perceived Control	"I could easily re-tweak it with a few more prompts to be perfect.“ (Participant 1) "Quite a lot [of control], 80%.“ (Participant 2) "Also cool that it integrated random grass into the dunes, so I wouldn't have expected or didn't prompt. But I liked it.“ (Participant 2) "So I think I had enough control, but the results were maybe not as I expected them to be.“ (Participant 3) "Knowing what I know now, I would have just not said anything about adding trees.“ (Participant 3)
Engagement	"If I look behind me, that's really kind of... I have like the slate kind of farmhouse here. And then yeah, I can just screen out the urban parts.“ (Participant 1) "It looks just like I'm part of a very nice drawing or picture. Like a piece of art.“ (Participant 2) "I feel like I've looked at everything already.“ (Participant 3)
Relaxation	"For me, relaxation is nature. And I wouldn't really associate an urban environment with relaxation.“ (Participant 1) "I would say 4 [out of 5 for level of relaxation]. If it would have some gentle movement, it would be perfect.“ (Participant 2) "Probably 3 to 4 [out of 5 for level of relaxation], let's say.“ (Participant 3) "If I had a task, it would be a nice environment to do something in, but not just chilling.“ (Participant 3)
Misalignment	"So you can kind of see like the towns are way too big and grey that are around here like they should not be like that, it almost reminds me a bit of China.“ (Participant 1) "There is for some reason there is a mirror glass looking thing in the sky, I guess where the VR environment joins together." (Participant 1) "The ocean is not moving at all or looks frozen.“ (Participant 2) "It looks more like a picture. But I can imagine with moving objects, it would be more realistic.“ (Participant 2) "The trees that are very close to me are very big and now I can notice that even though it has bark, it looks like a tree. But if I look close, it's actually made out of stone or rock.“ (Participant 3)

Table 1. Themes and quotes from participants.

Participants generally appreciated the ability to generate environments with minimal constraints, with one describing the process as offering “quite a lot of control” (80%). However, some noted that while they technically had control over input descriptions, the generated outputs did not always align with their expectations, particularly regarding unintended AI interpretations (e.g., trees appearing as stone formations).

Engagement was influenced by the extent to which participants felt immersed in their environments. While some reported aesthetic appreciation, describing the environments as “like a piece of art,” others noted that the lack of interactivity and dynamic elements (e.g., frozen water, static landscapes) reduced long-term engagement. Additionally, several participants highlighted misalignment issues, where AI-generated elements disrupted realism, such as “mirror-like” seams in the sky or an urban presence that felt incongruous with their relaxation goals.

Despite these challenges, participants generally found the environments conducive to relaxation, with ratings ranging from 3 to 4 out of 5 on a subjective relaxation scale. Participants emphasized the importance of natural elements, reinforcing prior research on the restorative effects of nature-inspired IVEs. However, several expressed that adding subtle motion and ambient sound would improve immersion and realism.

3 DISCUSSION

The initial response to the generated IVE may depend on whether the user selects a real-world location or an imagined one. Two participants chose to recreate familiar locations and expressed initial excitement at how well the virtual environment aligned with their real-world experiences. In contrast, the third participant designed an entirely new location without a specific sense of place and responded with less enthusiasm toward the initial result.

This difference in reaction may stem from the way users compare their envisioned space to the generated output. When a user selects a real location, their expectations are anchored in reality, making it easier for the AI to generate familiar elements if it has sufficient training data on that location. This can lead to a stronger sense of recognition and satisfaction. Conversely, when a user creates a completely original environment, the AI lacks specific real-world reference points, increasing the likelihood of unexpected or less cohesive results, which may lead to disappointment. Further research with a larger participant pool is needed to determine whether this trend holds across different users and scenarios.

As participants spent more time exploring their generated IVEs, they began to notice more visual artifacts and inconsistencies that were not immediately apparent upon first entering the environment. Examples included a pane of mirror-like reflections in the sky, which appeared to be a rendering artifact, and a stone pillar that abruptly transformed into a cedar tree, likely due to the AI misinterpreting structural elements. These anomalies disrupted the sense of immersion and presence, as they reminded users of the artificial nature of the environment, pulling them out of the experience. Rather than remaining fully engaged with the IVE, participants began to shift their focus toward analyzing the generation errors and considering how they might modify their prompts to avoid similar issues in the future. This suggests that while generative AI provides a powerful tool for IVE customization, unexpected artifacts can become a source of distraction, preventing users from fully immersing themselves in relaxation or mindfulness exercises. To address this limitation, future studies should explore real-time re-prompting mechanisms, allowing users to refine and adjust the generated scene iteratively. By enabling participants to correct errors as they appear, researchers can assess whether this approach leads to a more seamless and satisfying experience, where users can focus on relaxation rather than being preoccupied with visual inconsistencies.

Overall, the feasibility of using genAI for IVEs is promising, as it enables highly personalized virtual spaces with relatively low effort as the only input is the ability to craft a prompt. Even with relatively basic prompting, the outputs were conducive to a relaxing background. With further refined prompting techniques, repeated prompts, and more user input, it is expected that the results will better match what has been envisioned. One of the biggest limitations as pointed out by the three participants, is the lack of dynamic elements. There is no animation for elements that typically move, like the ocean in the second participant's IVE. This contributed to the IVE feeling more like a drawing than an environment and hindered the sense of presence. Similarly, since the generated IVEs are just skyboxes, or images on the face of a cube, participants are not able to move through the environment or interact with anything they are seeing. This is one of the biggest drawbacks of using an image generation model to create the entire environment: it is completely static. Therefore, the generated IVE should serve as the backdrop for the wellness intervention. Additionally, an embedded task within the IVE can help keep participants engaged while remaining in a stationary setting. To enhance the experience, participants should receive guidance on how to effectively prompt the model, ensuring they avoid incorporating elements that imply motion, such as ocean waves or birds in flight.

4 CONCLUSION & NEXT STEPS

Generative AI presents a new opportunity for personalized VR wellness experiences. However, its effects on immersion, emotional engagement, and user perception require empirical study. Future work will explore how SkyboxAI can be better utilized to better handle nuances in user descriptions and incorporate interactive features to enhance engagement in relaxation-focused virtual spaces. A planned study will investigate how users interact with generative AI by refining their environments through repeated prompting, selecting stylistic preferences, and applying best practices for effective prompt construction within the VR environment itself. This will be done to further investigate whether having a more active role in the creation of the IVE leads to a sense of ownership within the space, and therefore leads to a sense of empowerment. Feeling empowered could greatly enhance wellness outcomes by fostering a deeper sense of agency, engagement, and connection to the virtual environment. Ultimately, such an empowerment-focused approach could substantially improve the transformative potential of VR wellness interventions, leading to more personally meaningful virtual experiences.

REFERENCES

[1] 2025. Skybox AI. https://skybox.blockadelabs.com/

[2] Saad A Alhammad. 2024. Advocating for Action: Exploring the Potential of Virtual Reality in Breathing Exercise–A Review of The Clinical Applications. *Patient preference and adherence* (2024), 695–707.

[3] M. H. E. M. Browning, K. J. Mimnaugh, C. J. Van Riper, H. K. Laurent, and S. M. LaValle. 2020. Can simulated nature support mental health? Comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. *Frontiers in Psychology* 10 (2020). https://doi.org/10.3389/fpsyg.2019.02667

[4] Zhe Ding, Akapong Inkuer, Yi Gu, and Muhammad Shahid Khan. 2024. Virtual Reality and Multi-Sensory Interactions: Enhancing Emotional Wellbeing of Urban White-Collar Workers in Shanghai. *Revista de Gestão Social e Ambiental* 18, 7 (2024), 1–22.

[5] Laura Kolbe, Abhishek Jaywant, Alka Gupta, W Michael Vanderlind, and Gina Jabbour. 2021. Use of virtual reality in the inpatient rehabilitation of COVID-19 patients. *General Hospital Psychiatry* 71 (2021), 76–81.

[6] Susanna Pardini, Silvia Gabrielli, Silvia Olivetto, Francesca Fusina, Marco Dianti, Stefano Forti, Cristina Lancini, and Caterina Novara. 2023. Personalized, Naturalistic Virtual Reality Scenarios Coupled With Web-Based Progressive Muscle Relaxation Training for the General Population: Protocol for a Proof-of-Principle Randomized Controlled Trial. *JMIR Research Protocols* 12, 1 (April 2023), e44183. https://doi.org/10.2196/44183 Company: JMIR Research Protocols Distributor: JMIR Research Protocols Institution: JMIR Research Protocols Label: JMIR Research Protocols Publisher: JMIR Publications Inc., Toronto, Canada.

[7] Gabriela Ben Melech Stan, Diana Wofk, Scottie Fox, Alex Redden, Will Saxton, Jean Yu, Estelle Aflalo, Shao-Yen Tseng, Fabio Nonato, Matthias Muller, and Vasudev Lal. 2023. LDM3D: Latent Diffusion Model for 3D. https://doi.org/10.48550/arXiv.2305.10853 arXiv:2305.10853 [cs].