



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2026 **Issue:** onferend **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.82966>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Serenity Scape- A Virtual Mental Health Assistant in VR using AI

Dr. Mahendra Sawane, Bhavesh Yadav, Ayush Singh, Dev Rai, Aditya Narayan Prasad
Department of Computer Engineering G. H. Raisoni College of Engineering and Management Pune, India

Abstract—Mental health disorders have become a growing concern worldwide, and existing care pathways are often poorly equipped to meet the demand—particularly where cost, stigma, and geographic isolation place treatment out of reach for many people. This paper presents SerenityScape, a VR-based mental health assistant that brings together immersive virtual environments and AI-driven conversational therapy in a single accessible system. The platform relies on three core components: Google Cloud Speech-to-Text for capturing and transcribing user speech, Google Gemini 2.5 Pro for generating contextually appropriate and empathetic responses, and AWS Polly for converting those responses into natural-sounding voice output. Together, these technologies allow users to hold genuine-feeling therapeutic conversations with a virtual AI doctor within calming, carefully designed VR spaces. This work explores how the combination of multimodal AI and VR immersion can lower the barrier to mental health support, reduce anxiety, and contribute to emotional recovery.

Index Terms—Speech Recognition, Conversational AI, Mental Health, Virtual Reality (VR), Digital Health, Meta Quest 2, Unity, Text-to-Speech (TTS), Gemini 2.5 Pro

I. INTRODUCTION

Mental health sits at the core of how people think, feel, and navigate daily life. The World Health Organization estimates that roughly one in eight people globally lives with some form of mental disorder—a figure that makes the lack of affordable, accessible care not just an inconvenience but a genuine public health problem. Traditional services continue to struggle with a familiar set of obstacles: a shortage of qualified professionals, prohibitive costs, the enduring weight of social stigma, and the simple fact that in many parts of the world, help is not nearby. For millions of people, especially in low- and middle-income regions, these barriers add up to no care at all. Recent progress in both Virtual Reality (VR) and Artificial Intelligence (AI) has opened up new possibilities for digital mental health support. VR can place a user inside a calming, fully controlled environment in a way that no screen can replicate, while AI systems have grown capable enough to recognize emotional tone, maintain coherent therapeutic dialogue, and respond with something approaching genuine empathy. These are not separate developments—they are, when combined, stronger together.

SerenityScape emerges from this intersection. The system pairs conversational AI with VR-based therapy to create a private, stigma-free space where users can speak freely and receive meaningful feedback in real time. A virtual doctor avatar engages the user in natural spoken conversation, with Google Cloud Speech-to-Text handling recognition, Gemini 2.5 Pro generating the responses, and AWS Polly rendering them as lifelike speech. The entire experience runs in Unity on a Meta Quest 2 headset. The goal is not to replace trained clinicians—rather, it is to make early-stage emotional support broadly available in a way that is engaging, scalable, and, above all, accessible.

II. LITERATURE REVIEW

The growing global burden of mental illness has pushed researchers toward digital solutions, and in recent years, the convergence of AI and VR has emerged as one of the more promising directions. Traditional therapy remains valuable, but its structural limitations—expense, scarcity of providers, social stigma—have created a real appetite for alternatives. The literature reflects a gradual but sustained effort to develop technology-driven interventions that can supplement, or in some cases stand in for, conventional therapeutic encounters.

A. VR-Based Mental Health Interventions

Virtual reality has been studied as a therapeutic medium for some time, and the evidence base, while still developing, is reasonably encouraging.



A randomized controlled trial by Edwards et al. (2013) [12] found that Virtual Reality Exposure Therapy (VRET) produced meaningful reductions in social anxiety symptoms—an early signal that immersive environments could do more than simply entertain. Kim et al. (2017) [11] surveyed VR applications across various psychiatric settings and found that immersive simulations were effective at driving behavioral change and deepening emotional engagement.

More recently, Spiegelet al. (2023) [6] validated a mixed-reality AI-enabled program (MAIA) for anxiety and depression, demonstrating both therapeutic potential and solid user acceptance. Taken together, these studies make a reasonable case that VR is more than a novelty in clinical contexts—it can function as a legitimate vehicle for psychological intervention.

B. Conversational AI and Mental Health Support

Chatbots and AI dialogue agents have found a growing role in digital mental healthcare, largely because they can respond instantly, maintain availability around the clock, and—when well designed—simulate something close to empathetic listening.

Systematic reviews by Milne-Ives et al. (2020) [10] and Li et al. (2023) [5] both conclude that AI-based conversational systems can improve user engagement and support symptom monitoring. Sharma et al. (2022) [9] took a different angle, showing that human-AI collaboration in text-based peer support settings produces conversations with noticeably greater emotional depth. The caveats in this literature are real, though: questions around safety, transparency, and what happens when a user expresses acute distress have not been fully resolved. The consensus appears to be that conversational AI needs to be not just intelligent but emotionally grounded and anchored to some form of clinical guidance.

C. LLMs and Virtual Patient Systems

Large language models have begun making inroads into healthcare simulations. Gutierrez Maquilon et al. (2024) [4] demonstrated that GPT-powered virtual patients could provide clinicians with realistic communication training scenarios—a finding with direct relevance to the design of AI-driven therapeutic agents. That said, the paper also flags familiar weaknesses: hallucinations, inconsistent emotional tone, and a tendency to drift from context. Approaches like retrieval-augmented generation (RAG) and human-in-the-loop validation are increasingly being proposed as ways to make these systems more reliable and clinically appropriate.

D. Embodied Avatars and Conversational Agents in VR/AR

Progress in avatar technology has made it genuinely feasible to build AI therapy companions that feel lifelike and emotionally expressive. Trempe et al. (2025) [2] proposed animated conversational avatars capable of responding through coordinated speech and gesture—a meaningful step toward the kind of immersive presence that promotes sustained engagement. Open-source efforts in this space, such as those documented by Yin et al. (2023) [7], have added reproducibility and scalability to the picture. The persistent challenge is synchronization: getting speech, animation, and contextual awareness to cohere in real time, without perceptible delay, remains technically demanding. SerenityScape addresses this head-on by treating low-latency multimodal integration as a primary design requirement.

E. Speech Technology for VR Mental Health Systems

Voice is the most natural mode of interaction in a therapeutic context, and the underlying technologies have matured considerably. Guo et al. (2023) [8] presented end-to-end TTS models specifically optimized for conversational tone and emotional expressiveness—a development that moves voice synthesis beyond robotic monotony. Real-time STT systems, for their part, have reached a level of accuracy sufficient for use in acoustically complex VR environments. SerenityScape makes direct use of Google Cloud STT and AWS Polly, both of which are well-suited to the demands of immersive, hands-free interaction.

F. Feasibility and Ethical Dimensions

The National Institutes of Health (2024) [3] have emphasized that combining spatial computing with AI for mental health purposes is technically feasible, but that inclusivity, ethical governance, and data privacy must be treated as central concerns rather than afterthoughts. Qualitative validation research, including Spiegel et al. (2023) [6], adds to this picture by underscoring the importance of user consent, data security, and safe crisis response protocols. SerenityScape has been designed with these considerations in mind, with particular attention to maintaining appropriate boundaries and psychological safety.

G. Research Gaps and the Case for SerenityScape

Despite meaningful progress across these research areas, several gaps remain. Few studies have used standardized metrics or followed users over extended periods to assess safety, empathy, and therapeutic impact [4]–[7]. Real-time synchronization of speech, gesture, and emotion recognition continues to limit both realism and user engagement [8]. Ethical concerns—particularly around privacy and crisis response—demand ongoing attention [3, 6, 9]. SerenityScape is designed as a direct response to these gaps: an emotionally intelligent, ethically grounded VR mental health assistant built on Gemini 2.5 Pro, AWS Polly, and Google Cloud STT.

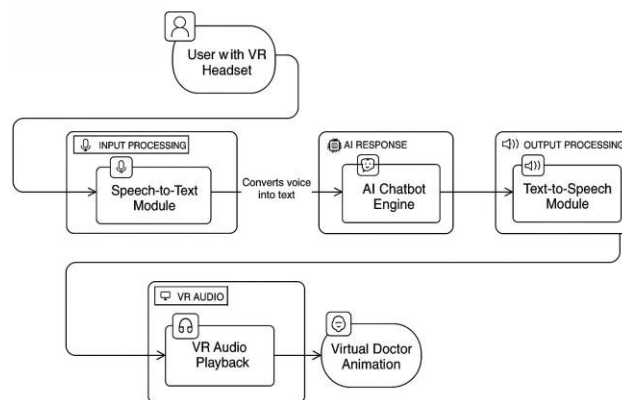
III. METHODS

This section describes how SerenityScape was conceived, designed, and built. The development process spanned conceptualization, technology selection, integration, and testing, all oriented toward producing an immersive therapeutic experience that feels natural to use and is genuinely supportive for people dealing with stress and emotional difficulty.

A. System Architecture

SerenityScape is built around a pipeline of cloud-based services coordinated through a VR application running on the Meta Quest 2. At the input end, the user speaks through the headset's built-in microphone. That audio is transcribed in real time via Google Cloud STT, formatted as JSON, and forwarded to Gemini 2.5 Pro, which interprets the content and generates an empathetic conversational response. That response is then sent to AWS Polly for synthesis into speech, which is played back in Unity in sync with the lip movements and gestures of a virtual doctor avatar. The whole loop—speech in, AI response, speech out—operates through RESTful API calls, balancing low latency with high realism.

The virtual environment itself is a calming 3D therapy room built in Unity 2022.3 LTS, using soft ambient lighting, natural color palettes, and background audio to create an atmosphere conducive to relaxed conversation. The avatar was sourced from Mixamo and configured in Unity's Animator Controller to move naturally in response to conversational events—talking, listening, occasionally gesturing.



B. Implementation Phases

Development proceeded through a set of structured phases, each building on the last:

- 1) Environment Design – A peaceful, immersive 3D space—drawing on therapy room and nature-inspired design principles—was created in Unity with attention to lighting, texture, and ambient sound.
- 2) Virtual Avatar Integration – The Mixamo avatar was imported, rigged with talking and listening animations, and connected to Unity's Animator Controller to respond dynamically during conversations.
- 3) Speech Processing Module – User audio is recorded via the Meta Quest 2 microphone and sent to Google Cloud STT using Unity WebRequest. The resulting transcription is returned in JSON format.
- 4) Conversational Response Generation – Gemini 2.5 Pro processes each transcription, reading user intent and emotional cues to produce contextually appropriate, empathetic replies.
- 5) Speech Output and Synchronization – AWS Polly synthesizes the AI response into audio, which is played back through Unity while the avatar's lip movements and gestures are synchronized in real time.

C. Integration Workflow

Every interaction follows the same cycle. The user speaks; the microphone captures the input; Google Cloud STT transcribes it. Gemini 2.5 Pro then reads the transcript, assessing intent and emotional context, and generates a natural response. That text goes to AWS Polly, which produces the synthesized voice output. Unity handles playback while simultaneously driving the avatar's animation. The cycle then repeats. In practice, this creates a conversation that feels fluid and responsive, with minimal perceptible latency. The modular design also means individual components can be swapped or upgraded without disrupting the rest of the pipeline.

D. Testing and Evaluation

System testing was carried out using the Meta Quest 2 in real conditions. Google Cloud STT was assessed for transcription accuracy across varying noise environments. Gemini 2.5 Pro was evaluated on the relevance and emotional alignment of its responses. AWS Polly was examined for the clarity and naturalness of its output. Unity playback was tested for animation smoothness and overall timing. Beyond these technical measures, informal user testing was conducted with fifteen faculty members and peers, with feedback collected on comfort, usability, and emotional engagement.

E. Technology Stack

Unity 2022.3 LTS served as the core development environment, with the Meta XR SDK ensuring compatibility with the Meta Quest 2 hardware. Google Cloud STT handled real-time transcription; Gemini 2.5 Pro managed conversational intelligence; Amazon Polly produced voice output. C# scripts in Unity managed front-end interaction and API communication, while Python was used on the backend for REST API integration. This combination allowed seamless coordination across the AI, speech, and VR layers of the system.

IV. RESULTS

A fully functional SerenityScape prototype was completed and deployed to a Meta Quest 2 as a standalone APK. The full interaction cycle—from the user's spoken input through AI processing to voice output—consistently completed in three to five seconds, a response time that informal testers generally found acceptable for ordinary conversation. Fifteen faculty members and peers participated in informal testing sessions, and the feedback was encouraging. Most participants reported feeling genuinely immersed and described the virtual environment as peaceful. The AI responses were widely regarded as relevant and empathetic, and the voice-driven interface struck testers as more intuitive and natural than text input would have been. Every tester agreed that the system held real potential as a tool for stress reduction and emotional self-reflection.

The system also produced substantive conversational exchanges. For instance, when a user said:

Message: "I've been feeling constantly anxious lately, and it's hard to focus on anything. What should I do to calm my mind?"

Response (Google Gemini 2.5 Pro):

"It sounds like you're going through a period of heightened anxiety, which can be overwhelming. One helpful step is to take slow, deep breaths and bring your attention to the present moment. If possible, try grounding yourself by noticing what you can see, hear, and feel around you. Small breaks, physical activity, or talking with someone you trust can also help. Remember, you're not alone in this—many people experience similar feelings, and reaching out for support is a sign of strength."



V. DISCUSSION

The results from SerenityScape's testing sessions broadly confirm that combining VR and AI can produce a meaningful mental health support experience. Participants who engaged with the simulated therapy sessions reported feeling more emotionally comfortable and open compared to conventional non-immersive settings—a finding that aligns with the broader VR therapy literature and suggests that immersion itself does useful psychological work. The integration of Gemini 2.5 Pro, Google Cloud STT, and AWS Polly held up well under real-world conditions, enabling conversations that felt smooth and genuine. Network-dependent latency in speech processing was occasionally noticeable, and the absence of real-time emotion recognition remains a ceiling on how responsive the system can be to subtle emotional shifts. Neither issue undermined the core experience, but both point toward meaningful directions for future development. On the whole, these results suggest that SerenityScape represents a viable and practically useful framework for delivering mental health support that is accessible, private, and empathetically grounded—a meaningful complement to, rather than a replacement for, traditional therapy.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

SerenityScape demonstrates what becomes possible when immersive VR and conversational AI are redesigned to work together around mental health use cases. The system provides natural, voice-driven therapeutic conversations within a calming virtual environment built in Unity 3D, powered by Google Speech-to-Text, Gemini 2.5 Pro, and AWS Polly. The result is a space where users can speak freely, without fear of judgment, and receive responses that feel genuinely attentive to what they have said. What the project ultimately shows is that immersive technology, thoughtfully deployed, can do real work in lowering barriers to care. SerenityScape offers privacy, accessibility, and emotional safety in a single package—qualities that are hard to guarantee in traditional care settings. It also points toward a broader vision of digital therapeutic that can sit alongside conventional counseling rather than compete with it, filling gaps in access and availability while the larger mental health system works to keep pace with demand.

B. Future Work

There are several directions in which SerenityScape could be meaningfully extended:

- 1) Emotion Detection and Sentiment Analysis – Integrating real-time recognition of emotional cues from speech tone, facial expressions, and behavioral signals would substantially deepen the system's contextual awareness, allowing the avatar to respond with greater precision to the user's actual state.
- 2) Personalization—Future versions could adapt responses based on session history, mood patterns, and progress over time, creating a more individually tailored therapeutic experience that evolves with the user.
- 3) Multilingual and Cultural Adaptation—Expanding the system to support regional languages and culturally appropriate conversational norms would make it significantly more inclusive, particularly for users in linguistically diverse contexts.
- 4) Integration with Mental Health Databases – Linking the system to validated psychological frameworks such as Cognitive Behavioral Therapy (CBT) would strengthen its clinical grounding and improve the quality of its guidance.
- 5) Offline and Privacy-Focused Deployment – Lightweight, on-device AI models would reduce dependence on external servers, improve performance in low-connectivity settings, and offer stronger data privacy guarantees.
- 6) Expanded Evaluation Metrics – Larger participant samples, longitudinal satisfaction studies, and assessment by trained therapists would provide a more rigorous picture of the system's real-world therapeutic impact.

REFERENCES

- [1] K. N. and U. J., "MediBot: Healthcare assistant on mental health and wellbeing," *International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS)*, vol. 7, no. 3, pp. 1–8, Mar. 2025.
- [2] Trempe, A., Dubois, M., Leblanc, S., Patel, S., and R. Moreau, "Personalized mental health support through VR/AR conversational AI with animated avatars," *TechRxiv*, pp. 1–15, Jan. 2025.
- [3] National Institutes of Health, "Feasibility of combining spatial computing and AI for mental health," *NIH Public Health Reports*, vol. 139, no. 2, pp. 1–12, Mar. 2024.
- [4] R. Gutierrez Maquilón, J. P. Sánchez, L. M. Rodríguez, and T. K. Breideband, "Integrating GPT-based AI into virtual patients to facilitate communication training," *JMIR Formative Research*, vol. 8, no. 1, p. e5862, Jan. 2024.
- [5] H. Li, R. Zhang, Y.-C. Lee, R. E. Kraut, and D. C. Mohr, "Systematic review and meta-analysis of AI-based conversational agents for mental health," *npj Digital Medicine*, vol. 6, no. 1, Art. no. 186, 2023.



- [7] B. M. Spiegel, J. A. Turner, E. L. Alpert, and C. M. Bulik, "Qualitative validation of mixed reality artificial intelligence assistant (MAIA): An AI-enabled VR program for anxiety and depression," *JMIR Mental Health*, vol. 10, p. e49822, 2023.
- [8] Yin, M., Roccabruna, G., Azad, A., and Riccardi, G., "Open-Source VR Conversational Agents for Digital Health," in *Proc. ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2023.
- [9] H. Guo, S. Zhang, F. K. Soong, L. He, and L. Xie, "Conversational End-to-End TTS for Voice Agents," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1234–1245, 2023.

- [10] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, "Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support," *arXiv*, arXiv:2203.15144, Mar. 2022.
- [11] N. Milne-Ives, C. de Cock, E. Lim, M. H. van Velthoven, and E. Meinert, "The effectiveness of artificial intelligence conversational agents in health care: Systematic review," *J. Med. Internet Res.*, vol. 22, no. 10, p. e20346, Oct. 2020.
- [12] S. J. Kim, B. O. Rothbaum, J. L. Maples-Keller, and
- [13] B. E. Bunnell, "The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders," *Harvard Review of Psychiatry*, vol. 25, no. 3, pp. 103–113, May–Jun. 2017.
- [14] S. M. Edwards, M. A. Obasaju, S. K. Schmertz, E. Zimand, M. R. Calamaras, P. L. Anderson, and M. Price, "Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial," *Journal of Anxiety Disorders*, vol. 27, no. 6, pp. 605–619, Aug. 2013.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)