This is the authors' version of the paper.

Final publication is available from Mary Ann Liebert (Sage):
https://doi.org/10.1177/21522715261454719

VR-SOCRATES — An Immersive Conversational Agent for Substance Use Disorder Support

Fabio Frisone, PhD[1,2], Christian Corrò, MSc[3], Luca Chittaro, Prof[3], Giuseppe Riva, PhD[1,4]

Cyberpsychology, Behavior, and Social Networking, 29(7), 2026, pp. 514-516.

[1] Humane Technology Laboratory, Catholic University of the Sacred Heart, Largo Gemelli, 1, 20121 Milan, Italy

[2] Sophia University Institute, Via di S. Vito, 28, 50064, Figline e Incisa Valdarno, Florence, Italy

[3] Human-Computer Interaction Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Via delle Scienze 206, Udine 33100, Italy

[4] Applied Technology for Neuro-Psychology Laboratory, IRCCS Istituto Auxologico Italiano, Via Magnasco 2, 20149 Milan, Italy

Introduction

Generative artificial intelligence (GenAI) is rapidly transforming the delivery of psychological support. Until recently, digital mental health tools were largely limited to static psychoeducational materials, symptom-monitoring applications, or rule-based chatbots with limited conversational adaptability. The emergence of large language models (LLMs) has significantly reshaped this landscape by enabling the development of conversational agents (CAs) capable of sustaining open, context-sensitive, and emotionally attuned exchanges. These systems do not simply provide information; they can actively engage users by asking questions, reframing personal narratives, facilitating emotional expression, and guiding users through meaningful experiential processes [1].

The VR-Socrates project was developed within this emerging paradigm and extends it through the use of virtual reality (VR). Socrates is a GenAI-based conversational agent, embodied in VR, designed to provide psychological support through reflective dialogue informed by psychological, clinical, and philosophical knowledge. Its aim is not to replace clinicians, but rather to offer an accessible and structured dialogical environment in which users can explore thoughts, emotions, and personal meanings. In previous work, Socrates was evaluated as a chatbot in a usability study with young workers. The findings indicated that even a brief interaction with the chatbot could be perceived as meaningful, intuitive, and emotionally supportive [2].

The present article focuses on a subsequent and clinically more relevant stage in the development of Socrates: its integration into VR. The novelty of this approach does not lie solely in the use of a conversational agent for psychological support, but in embedding such an agent within an immersive

environment, where dialogue becomes spatially situated, embodied, and experiential. VR-Socrates does not merely generate text responses on a screen; it appears as a virtual agent within a three-dimensional space, embodied through a visual and vocal avatar.

This embodiment transforms the user experience. Users no longer engage with a simple flow of text, but instead interact with a perceived life-sized presence situated within the virtual environment. The voice of the agent, modulated through intonation, pauses, and natural rhythm, together with the avatar's movements and expressions, contributes to a heightened sense of co-presence compared with conventional screen-mediated communication.

VR-Socrates represents a collaborative research initiative involving the Humane Technology Lab at Università Cattolica del Sacro Cuore in Milan, the Human-Computer Interaction Lab at the University of Udine, and the drug rehabilitation center "Le.L.A.T." in Messina, Italy. The project was designed to evaluate VR-Socrates with patients with substance use disorder (SUD), with the aim of assessing its capacity to provide psychological support to a highly vulnerable clinical population. The integration of clinical and academic contexts enabled the convergence of expertise from clinical psychology, cognitive science, and human-computer interaction.

What changes when the dialogue with a Conversational Agent happens in Virtual Reality

The most innovative development of the VR-Socrates project lies in its transition from a text-based chatbot to an immersive CA. This shift fundamentally alters the phenomenology of the interaction. When dialogue occurs through a screen, the user exchanges messages with an interface. In contrast, when dialogue occurs in VR, the user can experience the CA as a presence located in the same virtual space. In this context, the interaction becomes not only linguistic, but also spatial, sensory, and embodied. The main strengths of VR reside in its capacity to enhance the sense of presence [3], which is the subjective impression of being inside a simulated environment, as well as to support co-presence, namely the perception of sharing that environment with another entity. When Socrates is represented by an avatar and communicates through voice, the conversation can acquire a richer temporal and relational structure, characterized by rhythm, proximity, and interpersonal attunement. Elements such as pauses, intonation, gaze direction, interpersonal distance, and the organization of the virtual setting may all contribute to the way the user experiences the dialogue [4].

For patients with SUD, this immersive dimension may be particularly salient. Addiction is not merely a behavioral or neurobiological condition; it also involves emotional regulation, autobiographical memory, interpersonal vulnerability, craving, shame, avoidance, and impaired capacity to maintain reflective awareness with one's internal states [5]. A VR-based CA may offer a protected environment in which patients can pause, verbalize their experience, and engage in psychological reflection without the immediate pressure of face-to-face exposure. The artificial nature of CAs may reduce perceived judgment, while the immersive nature of VR may increase attention and emotional involvement. At the same time, VR must be used with caution. Immersion can increase engagement, but it can also intensify discomfort if the environment is inadequately calibrated. The design of the virtual setting, the appearance of the avatar, the intensity of sensory stimulation, and the pacing of the dialogue must be adapted to the clinical population. The goal is not to impress patients with technological novelty, but to create a stable, safe, and meaningful context for self-reflection. In this sense, VR is not merely an accessory added to Socrates; it is a clinical and experiential medium capable of transforming the quality of the interaction.

First Exploratory Observations in Patients with Substance Use Disorder

A first exploratory longitudinal study of VR-Socrates was conducted with eight patients with substance use disorder (SUD) recruited from the drug rehabilitation center "Le.L.A.T.". The intervention consisted of four 20-minute conversational sessions with VR-Socrates, delivered once per week over a one-month period. This design allowed for repeated interaction with the immersive conversational agent (CA) and was intended as a preliminary exploration of feasibility, engagement, emotional impact, and perceived relational quality.

The longitudinal structure was particularly relevant because patients in rehabilitation settings often require continuity, repetition, and time to develop trust in supportive tools. Conducting four sessions over four weeks made it possible to examine whether interaction with VR-Socrates remained acceptable and emotionally beneficial beyond a single exposure. Perceived stress was assessed using the Subjective Units of Distress Scale. Following the sessions with VR-Socrates, patients reported reduced stress levels. Although these findings should be interpreted cautiously because of the small sample size and exploratory nature of the study, the observed reduction in subjective distress suggests that immersive dialogue with Socrates may have a short-term regulatory effect. The VR environment may have contributed to this effect by focusing attention, reducing external distractions, and creating a psychologically contained space for dialogue.

At the end of the four sessions, patients completed the Working Alliance Inventory. Therapeutic alliance is traditionally understood as a relational construct based on collaboration, trust, and agreement on goals and tasks. Participants reported a positive connection with VR-Socrates, particularly on the Bond subscale. The emergence of a positive bond with a conversational agent does not suggest that VR-Socrates can replace the human therapeutic relationship. Rather, it indicates that patients may perceive the interaction as sufficiently coherent, supportive, and goal-oriented to foster meaningful engagement.

User engagement was also assessed after the four sessions using the User Engagement Scale. Patients showed generally high levels of engagement, with particularly elevated scores on the Perceived Usability subscale. These findings support the hypothesis that VR may enhance the motivational and experiential dimensions of chatbot-based support. In clinical populations, engagement is not a secondary outcome; rather, it is a key factor in determining whether a tool is tolerated, accepted, and integrated into a broader care pathway. Overall, these preliminary observations suggest that VR-Socrates may represent a promising direction for further research in SUD rehabilitation.

Conclusion

The use of VR-Socrates should be understood within the broader context of rehabilitation needs. Drug rehabilitation centers often aim to support emotional awareness, reflective functioning, relapse prevention, interpersonal trust, and the reconstruction of personal meaning. However, clinical staff may have limited time and resources, and patients may not always feel prepared to disclose sensitive experiences in direct interpersonal encounters. Therefore, an immersive CA could provide an additional space for psychological reflection between clinical sessions. This does not mean that GenAI should become a substitute for therapists, educators, or rehabilitation staff; rather, it is more appropriate as a complementary element in a clinically responsible framework. VR-Socrates may help patients prepare for human sessions, clarify what they want to discuss, regulate acute stress, or rehearse the expression of difficult emotions. In addiction care, the body plays a central role. Craving,

anxiety, shame, and emotional dysregulation are often experienced physically before they are verbalized [6]. VR may help bridge this gap by placing dialogue within an embodied environment. In this context, the user is not merely typing thoughts but inhabiting a space in which reflection can be associated with posture, presence, voice, and sensory context. This could make the dialogue more concrete and meaningful, especially for patients who struggle with abstract introspection.

However, the implementation of GenAI-based tools in psychological support requires constant ethical attention. This consideration is particularly critical when such tools are used with clinical populations including individuals with SUD. Patients may present emotional vulnerability, comorbid symptoms, trauma histories, or acute crises. For this reason, VR-Socrates should always be embedded within a clinically supervised framework. The central principle is clear: the CA may support reflection, but clinical responsibility remains with human professionals. Clinicians must define the aims of use, select appropriate patients, monitor potential risks, and determine when the tool is beneficial or when it should be suspended. They must also understand the basic functioning of GenAI systems, including their strengths and limitations [7]. A clinician who uses VR-Socrates responsibly should not treat it as a mysterious black box or as an autonomous therapeutic authority, but rather as a structured instrument that can be integrated into care under professional judgment.

VR-Socrates represents a significant example of how generative artificial intelligence (GenAI) can be adapted for psychological support through careful design, clinical grounding, and clearly defined ethical boundaries. Its previous development as a chatbot demonstrated the potential of conversational agents (CAs) to facilitate meaningful and reflective exchanges. The integration of Socrates into virtual reality (VR) represents a further step in this process, transforming dialogue into an immersive experience.

The first exploratory observations involving eight patients with substance use disorder (SUD) suggest that repeated weekly interactions with VR-Socrates may be associated with reductions in perceived stress after each session, as well as satisfactory levels of relational involvement after one month of use. These findings remain preliminary and require confirmation through larger controlled studies. Nevertheless, they point to a promising direction for the use of immersive conversational agents in addiction rehabilitation settings.

The core innovation of this project lies in the convergence of three dimensions: GenAI, Socratic dialogue, and VR. GenAI provides conversational flexibility, the Socratic method offers a structured reflective framework, and VR contributes presence, embodiment, and experiential depth [8]. Together, these elements may create a new form of digital support. This should not be understood as psychotherapy delivered by a machine, but rather as an immersive environment that helps patients pause, speak, reflect, and prepare for deeper clinical work with human professionals.

The aim of this work is not to demonstrate that GenAI can replace human care. Rather, it is to examine how immersive conversational agents can responsibly extend the reach, continuity, and accessibility of psychological support in rehabilitation centers. The reflections of Chiara Pupillo, psychologist and AI PhD student at the University of Pisa, whose research explores the application of AI to psychological issues, help clarify the broader implications of this work. Pupillo notes that “experiments like this, conducted in a real clinical setting, help us observe the potential and limitations of VR-AI instruments, which, like VR-Socrates, can become GenAI tools to support clinical practice”. However, she also emphasizes that this is possible “only with appropriate

methodological caution and if it remains very clear where the support of a CA must end and where the careful management of clinical experts must begin".

Pupillo further suggests that the future integration of GenAI into psychological support will depend less on technological enthusiasm than on professional competence: "These systems can support reflection, engagement, and emotional expression, but they must be used by clinicians who understand how GenAI works. Only clinicians with this awareness can recognize both the limits and the potential of these tools, which may vary substantially across individuals and clinical populations".

These considerations highlight the central message of the VR-Socrates project: GenAI can become a valuable ally to psychological care, but only when it is guided by human expertise, ethical responsibility, and clinical judgment. In this sense, the main challenge ahead is not merely technological, but clinical. It concerns the careful development of these tools as safe, meaningful, and genuinely human-centered extensions of care.

Bibliography

1. Bigman, Y. E., Briker, R., & Langer, M. (2026). Human–AI interaction research needs to be embedded in psychological theory. *Nature Reviews Psychology*, 1-2. Available from: https://doi.org/10.1038/s44159-026-00551-4
2. Frisone, F., Pupillo, C., Rossi, C., & Riva, G. (2025). SOCRATES. Developing and evaluating a fine-Tuned ChatGPT model for accessible mental health intervention. *Cyberpsychology, Behavior, and Social Networking*, *28*(5), 366-368. https://doi.org/10.1089/cyber.2025.45510.cyeuro
3. Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 82-96. https://doi.org/10.1089/cyber.2017.29099.gri
4. Frisone, F., Pupillo, C., Rossi, C., & Riva, G. (2026). Toward clinical integration of generative AI in mental health: personalization, multimodality and inter-entity experience. *Frontiers in Public Health*, *14*, 1603238. Available from: https://doi.org/10.3389/fpubh.2026.1603238
5. Stellern, J., Xiao, K. B., Grennell, E., Sanches, M., Gowin, J. L., & Sloan, M. E. (2023). Emotion regulation in substance use disorders: A systematic review and meta-analysis. *Addiction*, *118*(1), 30-47. Available from: https://doi.org/10.1111/add.16001
6. Sayette, M. A. (2016). The role of craving in substance use disorders: theoretical and methodological issues. *Annual review of clinical psychology*, *12*(1), 407-433. Available from: https://doi.org/10.1146/annurev-clinpsy-021815-093351
7. Zhao, X. (2026). Psychologists must be involved in building conversational AI chatbots. *Nature Reviews Psychology*, 1-3. Available from: https://doi.org/10.1038/s44159-026-00564-z
8. La Rocca, S., Brighenti, A., Tosi, G., & Daini, R. (2020). No country for old men: Reducing age bias through virtual reality embodiment [Article]. *Annual Review of CyberTherapy and Telemedicine*, *18*, 127-131.

Address correspondence to:

Fabio Frisone, Ph.D.
Humane Technology Lab, Catholic University of Milan
Largo Gemelli 1, 20123
Milan, Italy

*E-mail:*fabio.frisone@unicatt.it

Christian Corrò, MSc.
Human-Computer Interaction Lab, University of Udine,
Via delle Scienze 206, Udine 33100, Italy

E-mail: corro.christian@spes.uniud.it

Luca Chittaro, Prof.
Human-Computer Interaction Lab, University of Udine,
Via delle Scienze 206, Udine 33100, Italy

E-mail: luca.chittaro@uniud.it

Prof. Giuseppe Riva, Ph.D.
Humane Technology Lab, Catholic University of Milan
Largo Gemelli 1, 20123
Milan, Italy

*E-mail:*giuseppe.riva@unicatt.it