From Code to Care: Building AI virtual humans for accessible and ethical healthcare support

S. Mozgai[1], R. Shilling[2], B. Hale[2], P.A. Palmieri[2], D. Freyberger[2] & A.S. Rizzo[1]

[1]Institute for Creative Technologies, University of Southern California, East Waterfront Dr., Los Angeles, California, USA | [2]Sidekick Platform, Los Angeles, California, USA

[1]mozgai@ict.usc.edu, [2]rs@sidekickwellness.com, [2]bh@sidekickwellness.com, [2]df@sidekickwellness.com, [2]Patrick.Palmieri@uhhospitals.org, [1]rizzo@ict.usc.edu

DOI: 10.54337/aau.icdvrat26.01

Abstract

This paper presents the evolution of clinical virtual reality (VR) systems toward AI-driven virtual human (VH) agents for healthcare support. Building on decades of research, we describe the progression from early VR exposure systems to conversational VH platforms such as SimCoach, Battle Buddy, and the Sidekick platform. We highlight user-centered design efforts, mobile deployment strategies, research, and ethical challenges, including safety, bias, and clinical scope. The work demonstrates how VH systems can improve access to care while emphasizing the need for responsible design and human oversight.

1. Introduction

Over the past three decades, clinical virtual reality (VR) has evolved from experimental laboratory systems into scalable tools for assessment, training, and intervention across a wide range of healthcare applications. Early work focused on immersive environments for clinical targets including exposure therapy targeting anxiety disorders and post-traumatic stress disorder (PTSD), cognitive and motor assessment/rehabilitation for orthopaedic/central nervous system dysfunction, acute/chronic pain management, and clinical training (Bell et al., 2024; Rizzo et al., 2023). However, recent advances in artificial intelligence (AI) – particularly large language models (LLMs) – have accelerated the development of conversational virtual human (VH) agents capable of engaging users in naturalistic dialogue (Rizzo et al., 2026).

This shift represents a fundamental transition in clinical VR: from immersive environments as the primary therapeutic medium to intelligent agents that can operate both within VR and on non-immersive platforms such as mobile devices. These agents are increasingly capable of delivering psychoeducation, guiding structured self-help strategies, supporting clinical training, and facilitating access to care. As such, VH systems are now positioned as scalable tools to address persistent gaps in healthcare access, particularly in mental health domains.

This paper provides an overview of this evolution, drawing on a lineage of systems developed over nearly two decades, including SimCoach, Battle Buddy, and the Sidekick platform. We describe the design principles, user-centered findings, research strategies, and technological inflection points that enabled this progression. We also examine the ethical, clinical, and regulatory challenges associated with deploying AI-driven VH systems at scale.

2. Evolution of clinical virtual reality and virtual humans

2.1. Early clinical VR foundations

Clinical VR emerged in the mid-1990s as a promising approach for creating controlled, immersive environments for behavioral healthcare. Early systems targeted domains such as exposure therapy for phobias, cognitive assessment, physical rehabilitation, and pain distraction. Despite technological limitations at the time, these systems demonstrated the value of immersive simulation for eliciting ecologically valid responses while maintaining experimental control. Advances in computing power, graphics rendering, motion tracking, and interface devices – driven in part by the gaming industry – enabled increasingly sophisticated VR environments. These developments laid the groundwork for integrating virtual humans into clinical applications. Virtual humans during this period were mainly instantiated as props to add to the context of a VR simulation (e.g., virtual audience members for fear of public speaking exposure therapy) but rarely for natural language interaction.

2.2. From scripted agents to interactive virtual humans

Thus, early virtual humans were limited in their interactivity, often functioning as static or scripted elements within VR environments. Over time, research in embodied conversational agents introduced systems capable of perceiving user input (e.g., speech, facial expressions) and generating contextually appropriate responses. These systems began to incorporate multimodal sensing, allowing VH agents to analyze vocal prosody, gaze, and gesture to infer psychological states. This marked a transition from VHs as environmental props to active participants in clinical interaction, enabling applications in clinician training, social skills development, and patient engagement.

2.3. The AI inflection point

The recent emergence of LLM-based AI has dramatically expanded the capabilities of VH systems. Modern VH agents can now engage in fluid, natural language dialogue, maintain contextual awareness, and generate responses that approximate human conversational patterns. This has enabled deployment beyond laboratory settings into mobile and web-based platforms, supporting "anytime, anywhere" access to healthcare information and support. Despite these advances, VH systems remain at an early developmental stage. While they demonstrate impressive fluency and adaptability, they are still prone to inaccuracies, overconfidence, and occasional confabulation. This duality – high capability paired with persistent limitations – defines the current state of AI-driven VHs and underscores the importance of responsible design.

3. Case studies in VH system development

3.1. SimCoach: Reducing barriers to care

SimCoach was developed in 2009-2011 to address barriers to mental healthcare among military service members and veterans, including stigma, access limitations, and concerns about confidentiality. The system provided an anonymous, VH-mediated interface through which users could explore mental health concerns, complete screening measures, and receive information about treatment options. (see Figure 1).

User-centered evaluations demonstrated that individuals were more comfortable disclosing sensitive information to a VH compared to traditional web-based tools. However, the system's lack of persistent memory limited its ability to build longitudinal engagement and trust. While Sim-

Coach improved user experience and engagement, it did not significantly increase help-seeking behavior, highlighting the importance of continuity and personalization in VH interactions (Meeker et al., 2015).

3.2. Battle Buddy: Mobile VH for Veteran safety planning and mental health psychoeducation

Battle Buddy represents a transition from static, web-based VH systems to mobile, user-initiated interactions. Designed as a prototype mHealth application, Battle Buddy supports suicide prevention through structured safety planning. The system guides users in identifying warning signs, coping strategies, and sources of support, emphasizing preparation rather than real-time crisis intervention (see Figure 1).

A key design principle was minimizing interaction burden. Sessions were intentionally brief, modular, and repeatable, aligning with real-world mobile usage patterns. The system also leveraged existing smartphone features (e.g., music, photos, messaging) to support coping strategies, reducing friction and increasing usability during periods of distress. Development followed a user-centered rapid assessment process, incorporating feedback from veterans, clinicians, and researchers. Findings emphasized the importance of predictable interaction flow, non-judgmental tone, and clear system boundaries (Mozgai et al., 2023).

3.3. Sidekick platform: Scalable VH deployment

The Sidekick platform (which evolved from the Battle Buddy concept) represents the current generation of VH systems, designed for scalable deployment across multiple domains. It provides a cloud-based infrastructure that allows subject-matter experts to author, curate, and manage VH interactions without requiring deep technical expertise. (see Figure 1).

Key features include:

- Role-based VH agents ("Sidekicks") tailored to specific domains
- Expert-curated knowledge sources and constrained response parameters
- Multi-agent coordination for complex user needs
- Integration of multimedia resources to support actionable outcomes

The platform emphasizes a human-in-the-loop approach, enabling experts to define system scope, enforce boundaries, and ensure alignment with clinical and organizational standards. Applications being developed thus far include veteran resource navigation/clinical access, caregiver/special needs parent support, and mental health resource access and management for police and other first responder populations.

In a recent evaluation conducted within the United States Department of Veterans Affairs, 1,005 Veterans were given early access to a non-clinical version of the Sidekick/Battle Buddy application (results to be presented in the full paper and conference presentation). Findings from this study were highly encouraging: 92.5% of participants indicated they would recommend Sidekick/Battle Buddy to a friend or use it regularly when available; 85.1% reported improved ability to *identify* VA resources; 83.7% reported improved ability *to connect to* appropriate resources; 80.5% reported improved ability to *access* VA staff and services; and 78.4% expressed satisfaction with the navigation experience.

Participants also reported high levels of engagement with the virtual human (VH), noting the value of conversational, personalized guidance and reduced frustration when navigating multiple VA systems. Although derived from an early-stage prototype, these findings suggest that the platform represents a scalable, patient-facing AI support approach that can enhance Veterans' awareness of and access to resources – while complementing, rather than replacing, live human clinicians.

4. User-centered design and accessibility implications

Across all systems described, user-centered design has been critical to success. Iterative testing consistently revealed preferences that shaped system design, including:

- Clear privacy protections and transparency
- Reduced authority cues (e.g., avoiding military uniforms in VH design)
- Customization of agent characteristics (e.g., gender, age)
- Informal, conversational language
- Low-friction interaction models

These findings are particularly relevant for disability and accessibility contexts. VH systems offer unique advantages for individuals who face barriers to traditional care, including those with mobility limitations, social anxiety, or neurodivergent communication preferences. For example, some users prefer text-based interaction or simplified language structures, highlighting the need for adaptable communication modalities. Co-design with target populations remains essential to ensure accessibility, usability, and cultural relevance.

5. Ethical and clinical considerations

5.1. Scope and role definition

A central principle in VH design is that AI systems should augment – not replace – human clinicians. VH agents are best suited for structured support functions such as psychoeducation, screening, and skill-building. They should not perform diagnosis, treatment planning, or crisis management independently. A rich detailing of these ethical and professional guidelines can be found in (Rizzo et al., 2025).

5.2. Safety and crisis response

Ensuring safe responses to users in distress is a critical challenge. Current AI systems can detect explicit indicators of risk but often struggle with indirect or ambiguous signals. Effective design requires conservative escalation strategies, clear pathways to human support, and avoidance of autonomous risk management.

5.3. Privacy and data protection

VH systems collect sensitive data through user interaction and, in some cases, multimodal sensing. This raises significant concerns regarding privacy, consent, and data governance. Compliance with regulatory frameworks (e.g., HIPAA, FERPA) is necessary but not sufficient, as these frameworks were not designed for AI-driven systems.

5.4. Bias and inclusivity

AI systems may reflect biases present in training data or design decisions. This can disproportionately impact underserved populations, including individuals with diverse linguistic, cultural, or cognitive profiles. Addressing bias requires diverse datasets, inclusive design practices, and ongoing evaluation.

5.5. User attachment and dependency

VH agents can foster strong user engagement and self-disclosure. While this can reduce stigma and increase access, it also raises concerns about over-reliance or substitution for human relationships. Design strategies should encourage connection to real-world support networks.

6. Discussion

The progression from early VR systems to AI-driven VH platforms reflects a broader shift in healthcare technology toward scalable, user-centered solutions. VH systems offer a promising approach to addressing global mental health needs, particularly in contexts where access to care is limited. However, the transition from research prototypes to real-world deployment introduces new challenges. The increased fluency and realism of AI systems can lead to overestimation of their capabilities, necessitating careful design of scope, transparency, and safety mechanisms. Lessons from both clinical VH systems and consumer AI applications highlight the risks of deploying systems without adequate guardrails (Rizzo et al., 2026). The most effective approach appears to be hybrid design models that combine generative AI with structured constraints, expert oversight, and curated knowledge sources. This allows systems to leverage the strengths of AI while maintaining control over safety-critical interactions.

7. Conclusion

AI-driven virtual humans represent a significant advancement in clinical VR, enabling scalable, accessible support for a wide range of healthcare applications. Systems such as SimCoach, Battle Buddy, and the Sidekick platform illustrate the potential of VH agents to enhance engagement, reduce barriers to care, and extend the reach of expert knowledge. At the same time, these technologies must be developed and deployed with careful attention to ethical, clinical, and regulatory considerations. The future of VH systems will depend not only on technological progress but also on the establishment of robust frameworks for safety, accountability, and human oversight. When designed responsibly, VH agents can serve as powerful tools for bridging gaps in care – supporting, rather than replacing, the human relationships at the core of healthcare. These elements will be discussed in more detail in the final paper and conference presentation, if accepted.

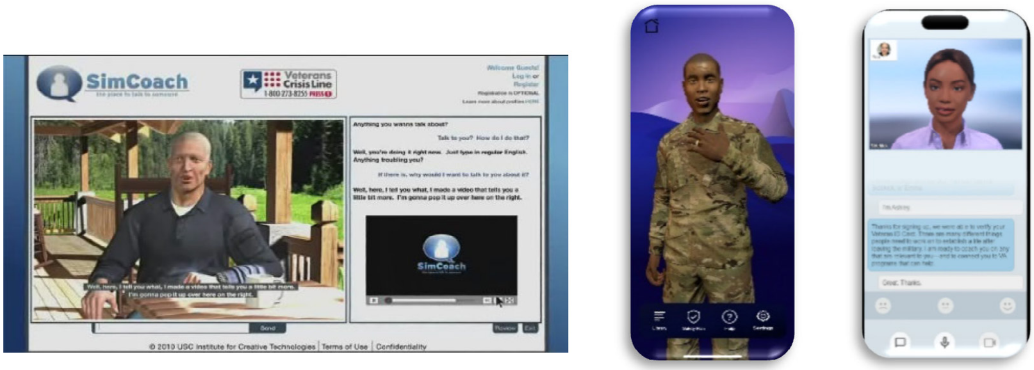

Figure 1: Simcoach, Battlebuddy, Sidekick.

References

Bell, I H, Pot-Kolder, R, Rizzo, A, Rus-Calafell, M, Cardi, V, Cella, M, Ward, T, Riches, S, Reinoso, M, Thompson, A, Alvarez-Jimenez, M and Valmaggia, L (2024), Advances in the use of virtual reality to treat mental health conditions, *Nature Reviews Psychology*, 3, 8, pp. 552–567.

Meeker, D, Cerully, J L, Johnson, M, Iyer, N, Kurz, J and Scharf, D M (2015), *SimCoach evaluation: A virtual human intervention to encourage service-member help-seeking for posttraumatic stress disorder and depression,* RR-505-OSD, RAND Corporation, Santa Monica, CA.

Mozgai, S, Rizzo, A and Hartholt, A (2023), Human-centered design for a virtual human led mHealth intervention for suicide prevention, *Human Systems Engineering and Design (IHSED2023),* 112, pp. 98–106.

Rizzo, A, Koenig, S and Lange, B (2023), Clinical virtual reality: The state of the science, in G G Brown, B Crosson, K Y Haaland and T Z King (Eds), *APA Handbook of Neuropsychology, Vol. 2: Neuroscience and Neuromethods*, pp. 473–491, American Psychological Association, Washington, DC.

Rizzo, A, Mozgai, S, Sigaras, A, Rubin, J E and Jotwani, R (2025), Expert consensus best practices for the safe, ethical, and effective design and implementation of artificially intelligent conversational agent (i.e., chatbot/virtual human) systems in health care applications, *Journal of Medical Extended Reality*, 2, 1, pp. 209–222.

Rizzo, A S, Mozgai, S, Hale, B and Shilling, R (2026), The evolution of virtual human agents in clinical virtual reality: From birth to late-stage adolescence, in D D Luxton (Ed), *Artificial Intelligence in Behavioral and Mental Health Care*, 2nd ed., pp. 127–188, Elsevier.