"Poker with Play Money": Exploring Psychotherapist Training with Virtual Patients

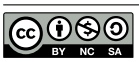
CYNTHIA M. BASEMAN, School of Interactive Computing, Georgia Institute of Technology, USA
MASUM HASAN, Department of Computer Science, University of Rochester, USA
NATHANIEL SWINGER, School of Interactive Computing, Georgia Institute of Technology, USA
SHEILA A. M. RAUCH, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, USA; Mental Health Service Line, Joseph Maxwell Cleland Atlanta VA Healthcare System, USA
EHSAN HOQUE, Department of Computer Science, University of Rochester, USA; Health Services, Ministry of Defense, KSA
ROSA I. ARRIAGA, School of Interactive Computing, Georgia Institute of Technology, USA

Role-play exercises are widely utilized for training across a variety of domains; however, they have many shortcomings, including low availability, resource intensity, and lack of diversity. Large language model-driven virtual agents offer a potential avenue to mitigate these limitations and offer lower-risk role-play. The implications, however, of shifting this human-human collaboration to human-agent collaboration are still largely unexplored. In this work we focus on the context of psychotherapy, as psychotherapists-in-training extensively engage in role-play exercises with peers and/or supervisors to practice the interpersonal and therapeutic skills required for effective treatment. We provide a case study of a realistic "virtual patient" system for mental health training, evaluated by trained psychotherapists in comparison to their previous experiences with both real role-play partners and real patients. Our qualitative, reflexive analysis generated three themes and thirteen subthemes regarding key interpersonal skills of psychotherapy, the utility of the system compared to traditional role-play techniques, and factors which impacted psychotherapist-perceived "humanness" of the virtual patient. Although psychotherapists were optimistic about the system's potential to bolster therapeutic skills, this utility was impacted by the extent to which the virtual patient was perceived as human-like. We leverage the Computers Are Social Actors framework to discuss human–virtual-patient collaboration for practicing rapport, and discuss challenges of prototyping novel human-AI systems for clinical contexts which require a high degree of unpredictability. We pull from the "SEEK" three-factor theory of anthropomorphism to stress the importance of adequately representing a variety of cultural communities within mental health AI systems, in alignment with decolonial computing.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *Natural language interfaces*; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → *Collaborative learning*; *Psychology*.

Additional Key Words and Phrases: Human-AI Collaboration, Virtual Agent, Virtual Patient, Role-Play, Interpersonal Skills, Psychotherapy, Mental Health, Mental Health Training, Humanness

Authors' Contact Information: Cynthia M. Baseman, School of Interactive Computing, Georgia Institute of Technology, Georgia, USA; Masum Hasan, Department of Computer Science, University of Rochester, New York, USA; Nathaniel Swinger, School of Interactive Computing, Georgia Institute of Technology, Georgia, USA; Sheila A. M. Rauch, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Georgia, USA; and Mental Health Service Line, Joseph Maxwell Cleland Atlanta VA Healthcare System, Georgia, USA; Ehsan Hoque, Department of Computer Science, University of Rochester, New York, USA; and Health Services, Ministry of Defense, KSA; Rosa I. Arriaga, School of Interactive Computing, Georgia Institute of Technology, Georgia, USA.



ACM 2573-0142/2025/11-ARTCSCW269
https://doi.org/10.1145/3757450

ACM Reference Format:
Cynthia M. Baseman, Masum Hasan, Nathaniel Swinger, Sheila A. M. Rauch, Ehsan Hoque, and Rosa I. Arriaga. 2025. "Poker with Play Money": Exploring Psychotherapist Training with Virtual Patients. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW269 (November 2025), 29 pages. https://doi.org/10.1145/3757450

1 Introduction

Role-play exercises are a ubiquitous training technique across a multitude of scenarios. Such exercises may be familiar to university students [54, 80], managers [12, 51] and their staff [18, 66], law enforcement [120], soldiers [110], and medical professionals [43, 63, 84] alike. Whether with a fellow peer or a supervisor, role-play enables those in training to hone their interpersonal and domain-specific skills in a safe learning environment. In clinical settings, role-play is crucial for establishing communication skills and learning reflective practices.

In the context of psychotherapy, the focus of the clinician is building interpersonal skills: they must build rapport with a patient[1] while managing their own emotions and remaining professional. To practice interpersonal and therapeutic skills, training frequently involves role-play exercises [63, 95, 113]. In these exercises, a psychotherapist-in-training (a trainee) may practice delivering therapy with a peer or supervisor without putting a real patient at risk. Both parties (and sometimes peer observers) then collaboratively reflect on the experience and what could have been done differently. The development of these interpersonal skills is considered vital: some clinical training contexts even employ human standardized patients (i.e., live actors) to ensure that a trainee's role-play partner is able to demonstrate realistic behaviors and symptoms in support of practicing differential diagnosis or other targeted learning goals [95]. However, role-play has a number of shortcomings including low availability, prohibitive cost, and lack of diversity. The incorporation of AI-powered virtual agents (in this context, virtual patients) into psychotherapeutic training may provide engaging, cost-effective, culturally representative, and reliable exercises with immediate feedback for trainees [31, 116]. Previous explorations of virtual patient avatar-based simulation for mental health training have seen promising results, but were limited by the realism of the avatar and simulation scenarios.

As we look to AI-powered technologies such as virtual patients to augment traditional role-play techniques, the implications of shifting this human-human collaboration to human-AI collaboration are still largely unexplored. New AI technologies such as large language models (LLMs) have the capability to simulate human-like behaviors [1, 85] and even role-play [73, 106], but researchers haven't yet fully investigated the potential implications of LLM-driven virtual avatars for psychotherapeutic training. Further, there have been calls in HCI and CSCW to investigate the social impacts of AI-based healthcare technology [86]. Suggesting that an AI agent act as a psychotherapy patient goes a step beyond other human-AI collaboration use cases, as the perceived "humanness" and emotionality of the agent may be crucial for interpersonal and therapeutic skill development.

To explore the social impacts of virtual patients within mental health training, we engaged 12 trained psychotherapists in interactive demos of a virtual patient system and semi-structured interviews. We utilize the open-source virtual avatar platform developed by Hasan et al. [49], which incorporates an LLM and a 3D game engine to provide a more natural and flexible, and therefore realistic, experience. This case study of a virtual patient system provides an opportunity to examine the perceived humanness of a system that incorporates these AI technologies in a specialized context. In this work, we therefore provide the following three contributions:

[1]Referred to as the "patient-client issue" [121], some psychotherapists use the term "patient" to refer to those seeking mental health treatment, while other psychotherapists prefer the term "client." In this empirical study, we utilize the term "patient" in alignment with the majority of our participants.

- A case study of a virtual patient system for mental health professional training, evaluated by trained psychotherapists
- An analysis of the virtual patient's utility in bolstering three key interpersonal skills of psychotherapy (building rapport, emotional regulation, and navigating flexible therapy delivery), leveraging the Computers Are Social Actors (CASA) framework
- A discussion of eight factors that impacted perceived "humanness" of the virtual patient, compared to psychotherapists' previous experiences with both real role-play partners and real patients

Content Warning: This study includes psychotherapists' descriptions of individuals seeking mental health treatment, including references to trauma and suicidal ideation.

2 Related Work

2.1 Human-AI Collaboration for Interpersonal Skill Development

Previous CSCW and HCI research in human-AI collaboration for interpersonal and social skill development has spanned a wide variety of domains [19]. Interactive systems utilizing AI have demonstrated utility for practicing and bolstering social skills such as teamwork [130], job interviewing [53], LGBTQIA+ advocacy [88], and emotionally intimate self-disclosure [28]. Ubiquitous, interactive technologies have offered opportunities for training in emotional regulation [97, 108], as well as social emotional learning and collaboration for children [25, 42, 117]. Much HCI literature focuses on how interactive technologies may assist children and adults with autism with social communication skills [129], from virtual reality [60] to conversational agents [5, 42, 44, 118].

Increasingly, technological advancements more closely align such systems with the Computers Are Social Actors (CASA) framework [81]. CASA has been leveraged extensively to better understand how humans interact socially with technologies [68, 72, 89, 101, 105]. In this view, interactions between a user and a computer are fundamentally social: users have the tendency to enact learned social and/or anthropomorphizing behaviors towards technology. The CASA framework provides a theoretical basis for interpersonal skill development with AI: although users may not view the AI system as a social actor the way they would another human, users still behave as though they do. In our discussion, we leverage the CASA framework to comment on the extent to which (human-AI) collaboration may bolster rapport building as an interpersonal skill.

Throughout varying domains, introducing AI to train interpersonal and social skills has engaged users and increased their confidence. However, from an HCI perspective, the potential of AI for interpersonal training has not yet been widely explored within a psychotherapeutic context. In particular, there has been a call in CSCW to better understand how technology could bolster interpersonal skill development for counselors [109]. Due to recent advancements in AI (e.g., large language models) and computer graphics (e.g., 3D game engines), the stage is set to answer this call with an interactive, high-fidelity prototype. In this study, we explore the utility and psychotherapist perceptions of an AI-powered virtual patient for therapy training. We explore how collaborating with an AI system could bolster interpersonal skills in this context, addressing the call of [109] and translating previous work in human-AI collaboration to a psychotherapeutic training context.

2.2 Technology within a Psychotherapeutic Context: Virtual Patients and AI For Health

Role-play is a very commonly implemented training exercise for mental health professionals, whether with peers, a supervisor, or human standardized patients (i.e., live actors) [63, 95, 113]. In addition to post-exercise descriptive feedback and trainee reflection, competency assessment tools may include grading rubrics to assess skills specific to a therapist's specialty, e.g., adherence to a theory-driven treatment protocol [84, 95]. However, because traditional role-play relies on other

humans, weaknesses include low availability, prohibitive cost, high turnover rates, and lack of cultural diversity [31, 116]. Incorporating virtual patients into clinical training may provide engaging, cost-effective, and reliable exercises with standardized, immediate qualitative and quantitative feedback [31, 116]. Because of these potential benefits, virtual patient training tools have been in development for over 15 years, continuously integrating new breakthrough technologies to aid robustness, adaptability, and flexibility [92, 116]. Previous explorations of avatar-based simulation for clinical training have seen promising results with a variety of conditions including PTSD [94, 116], conduct disorder [116], and Alzheimer's [13]. Within mental health nursing education on bedside manner, virtual patient training improved trainees' learning attitudes towards communication skills, increased their perceived self-efficacy, and supported reflexivity and ethical behaviors [45, 107].

While virtual patients for clinical training have been explored for years, these systems were originally rigid, with preset responses [57], and technical limitations reduced their utility in practice [114, 116]. Further, the implications of these systems regarding realism and the extent to which the agent is perceived as human-like (see Section 2.3) could not be examined in depth due to these technological limitations. Now, with recent advancements in AI, we begin to view virtual patients as autonomous social actors with which a trainee may collaborate to simulate an effective psychotherapy session and develop their skills. While Ali et al. [6] have previously explored a virtual patient for targeted practice of a specific healthcare scenario (i.e., conversing with a late-stage cancer patient), our study takes a more flexible approach by allowing customization of the virtual patient and scenario (as described in Section 3.1).

In alignment with the calls in HCI and CSCW to investigate the social impacts of AI-based healthcare technology [86], recent literature has explored many facets of the integration of AI into healthcare. Researchers have suggested design principles for AI-powered clinical decision support tools [20, 26, 122, 124], studied the perception of AI technologies and the trust placed in them by various clinical stakeholders [26, 82], and investigated the tensions of autonomy between patient, clinician, and AI in health decisions [46, 58]. Our study contributes a commentary on the social impacts of the potential shift from human-human collaboration to human-agent collaboration within mental health training role-play. While there exists work on designing virtual agents as patient-facing therapists or tools which support self-awareness and health education [77, 83, 93, 138], our study instead explores a virtual patient system as a relatively low-risk method for psychotherapist skill development. Further, the virtual patient system [49] aims to create a more realistic simulation by incorporating an LLM and MetaHumans. We discuss the challenges and opportunities of prototyping novel AI systems within a psychotherapeutic context.

2.3 The "Humanness" of Affective Technology and Simulation

A considerable amount of HCI and CSCW literature has aimed to determine what factors influence the perception of interactive technologies as human-like. In particular, previous research has studied the anthropomorphism of conversational agents by modulating factors such as response speed [91], message contingency and responsiveness to the user [74, 91], conversational and empathetic language [56, 99], and identity [48]. Most of this work utilized quantitative methods, such as surveys or log data. While these methods allow for scalability, they miss the deeper nuance that may be examined through a qualitative approach to understanding the perceived humanness of technology.

Similarly, research on voice assistants has investigated how human-like voice (i.e., showing empathy, not sounding robotic, offering natural conversation) can facilitate human connection and relationship development [105, 136]. The socioemotional context of use (e.g., feeling lonely) may also impact the social role of voice assistants and user perception [89].

Explorations of embodied agents within virtual environments have provided insights into how visuals and nonverbals (e.g., gesturing, facial expressions, posture, mirroring) influence communication, expression, and social presence [37, 39, 59, 125]. Zargham et al. found that users related more to an embodied virtual assistant with which they had demographic qualities in common, and that they desired a virtual assistant who performed some autonomous actions (e.g., changing their hairstyle unprompted) [135]. Similarly, there is a rich body of work exploring the realism and social behaviors of non-player characters (NPCs) [27, 36, 132], including recent attempts to enhance NPC social presence through LLM-powered agents [30] and to increase the realism of AI navigation behaviors [75]. Humanness has also been implicitly studied through research on the uncanny valley, first introduced by Masahiro Mori [76]. The uncanny valley has been connected not only to visual aspects of virtual agents, but also to auditory aspects (i.e., speech) [38] and perceived social cognitive abilities of the agent [112].

Most previous work in CSCW and HCI, however, has investigated the "humanness" of these interactive technologies only in support of other functional desires. This work has primarily sought to better understand and/or influence factors such as user self-disclosure [56, 91, 134], relationship building (e.g., perceived warmth, social presence) [52, 74, 105], or trust [48, 102, 136]. This view utilizes "humanness," anthropomorphism, or personification as an instrumental objective in the ultimate goal of creating a trustworthy and/or engaging system. A human system is viewed to foster trust just as an inhuman system (such as one which falls into the uncanny valley) would deter potential users. In this study, we instead explore "humanness" in its own right, aligning with HCI's psychological underpinning. Within social psychology, Epley et al. introduced a theory of anthropomorphism (known as "SEEK") which details three psychological determinants [40]: sociality motivation, effectance motivation, and elicited agent knowledge. Each of these determinants may be influenced by four variables (dispositional, situational, developmental, and cultural), supporting a better understanding of why users are likely to anthropomorphize interactive technologies (e.g., [33, 70]).

In this study, we provide a reflexive analysis of eight factors which influenced psychotherapist perception of the "humanness" of a virtual patient. We concentrate on the domain of psychotherapy because of its high reliance on interpersonal skills, which introduces the question of whether a virtual patient must closely simulate human-like behaviors. Because the use case involves deployment within a training workflow, the question is not whether trainee therapists would use the system; they may be required to. Our focus instead is whether the system is perceived as "human," and to what extent this influences its utility for context-specific skill development. We pull from SEEK to discuss the importance of adequately representing a variety of cultural communities within the virtual patient system, in alignment with decolonial computing [4, 7].

3 Methodology

To study the potential of virtual patients and the application of AI-based technologies in mental health training, we conducted interactive demos of a virtual patient system and semi-structured interviews with 12 psychotherapists. Below, we first describe the virtual patient system and then detail our study methodology.

3.1 Virtual Patient System Interaction

In this study, we utilize an open-source large language model-driven virtual agent platform [49]. The platform allows a face-to-face video call with a real-time virtual agent. The platform offers a diverse list of virtual avatars (see **Figure 1**) with customizable identities and personalities. Alongside turn-based conversation, the avatars can express basic emotions through facial expression and voice based on the conversation context. The background and premise of the conversation can be

changed via text input. On top of the open-source project, we developed a specific interface for mental health applications where our study participants can tailor the background of the virtual patient (e.g., mental health profile, personality) and conversation premise (see **Figure** 2).

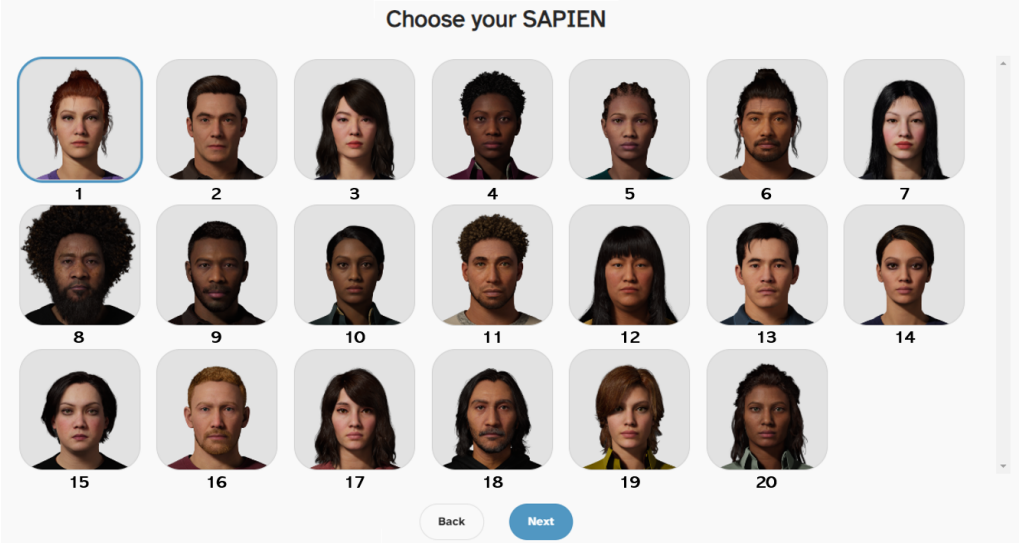


Fig. 1. The twenty virtual patient avatar options

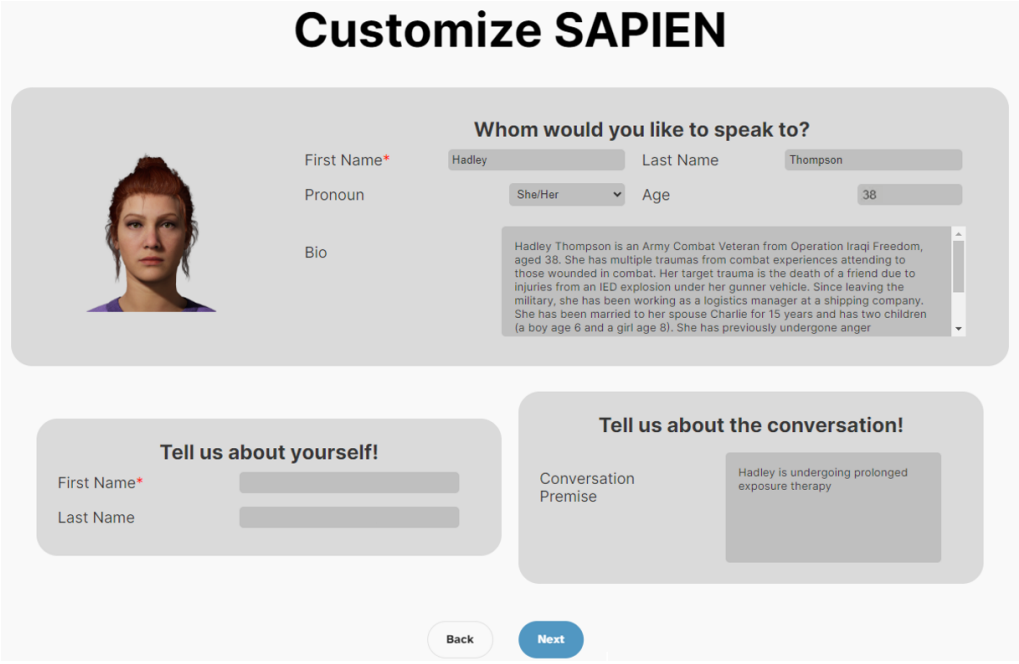


Fig. 2. Web interface for customizing the virtual patient's background and the conversation premise

To interact with the virtual patient, the participant holds down the space bar and speaks though their computer microphone. The speech is then transcribed into text and sent to a large language model (LLM), `gpt-3.5-turbo`, via API calling. Based on the background and premise, the LLM is prompted to respond to the participant as the chosen avatar. Once the LLM response is generated, the response is sent to a text-to-speech model to synthesize a realistic audio response, which is played through the participant's browser. The back-and-forth conversation works in near real-time on a reliable internet connection. The exact LLM prompt used in this study is provided in **Figure 3**. For a more in-depth system architecture description, see Hasan et al. [49].

Language model prompt

You are `patient firstname` `patient lastname` (Pronoun: `patient pronoun` `age`) and you are having a conversation with Dr. `clinician firstname` `clinician lastname`, who is an expert in prolonged exposure therapy and treating `patient firstname` `patient lastname`.
`patient firstname`'s narrative `patient narrative`
Premise: `conversation premise`

Fig. 3. Language model prompt for the virtual patient system. The highlighted variables were replaced with pre-set input (first demo) or the participant's input (second demo).

3.2 Study Recruitment and Participation

To study the potential of virtual patients and AI-based technologies in mental health training, our qualitative study aimed to gain holistic feedback on the virtual patient system from trained psychotherapists.

A university IRB approved the study protocol and participants provided consent. Twelve participants were recruited via purposeful sampling, through email, word of mouth, and snowball sampling. To be eligible for the study, participants were required to have prior experience working with those who seek mental health treatment. We compensated each participant with a $20 Amazon gift card for their individual demo and interview session, lasting between 30 and 45 minutes in total.

Seven participants self-identified as women, four as men, and one as non-binary. Ten of the twelve participants identified as non-Hispanic white, one identified as Asian, and one identified as Latino. Nine of the participants hold a PhD, two are at the master's level, and one holds a BA. All participants held their highest degrees within the fields of psychology or counseling. Participants' experience with those seeking mental health treatment varied. Most participants work at a large academic and medical institution as a clinician, researcher, and/or professor (PhD level). Our participants also included private practice psychotherapists and clinical psychologists (PhD or MS/MA level) and a participant who has experience with counseling (BA level). The duration of clinical practice varies across participants from one year to over thirty years. See **Table 1** for an overview of the participants.

3.3 Study Design

Participants first engaged in two 5-minute interactive demos of the virtual patient via a website. The first interaction demo was seeded with a standardized prompt and avatar. The second demo was a "free play," with an avatar freely customized by the participant. After completing both interaction demos, each participant engaged in a semi-structured interview. We provide an overview of our study design in more detail below.

Table 1. Participant Overview

	Gender	Race or Ethnicity	Highest Degree Completed	Clinical Practice (Years)
P1	Female	White	PhD	10-20
P2	Non-Binary	White	PhD	5-10
P3	Female	White	MA	20-30
P4	Male	White	PhD	10-20
P5	Female	Asian	BA	0-5
P6	Male	White	PhD	10-20
P7	Female	White	PhD	30-40
P8	Female	White	MS	5-10
P9	Male	White	PhD	10-20
P10	Female	White	PhD	5-10
P11	Female	White	PhD	5-10
P12	Male	Latino	PhD	10-20

3.3.1 First Demo: Standardized. The first 5-minute interactive demo was seeded with a standardized prompt, provided by our clinical co-author to ensure clinical validity. The fictionalized prompt is adapted from materials provided to an actor in preparation for the taping of an improvised training case video. Our participants interacted with "Hadley Thompson" (See **Figure 4**), a virtual patient with post-traumatic stress disorder (PTSD). PTSD may develop in those who have been exposed to a traumatic or stressful event, including but not limited to death or sexual violence [9]. The bio used for Hadley Thompson (she/her) is below:

> "Hadley Thompson is an Army Combat Veteran from Operation Iraqi Freedom, aged 38. She has multiple traumas from combat experiences attending to those wounded in combat. Her target trauma is the death of a friend due to injuries from an IED [improvised explosive device] explosion under her gunner vehicle. Since leaving the military, she has been working as a logistics manager at a shipping company. She has been married to her spouse Charlie for 15 years and has two children (a boy age 6 and a girl age 8). She has previously undergone anger management treatment to get family conflict under control, but she is still bothered by intrusions and avoidance. She avoids people, crowds, and driving long distances. She is an overachiever and is having difficulty keeping it all together. Her spouse, Charlie, is a Marine veteran, unable to work due to PTSD with VA service connection related to infantry service and heavy combat."

The "Conversation Premise" for this interaction was "Hadley is undergoing prolonged exposure therapy." While there are many psychotherapies which target PTSD, the gold standard treatment is Prolonged Exposure (PE) therapy [96] due to its efficacy as an evidence-based practice [16, 90]. PE therapy is a cognitive behavioral therapy which helps patients gradually approach and confront their trauma memory and trauma-related stimuli in order to decouple these stimuli from their feelings of distress [41]. PE therapy is also classified as a manualized therapy, which requires adherence to a strict session manual or protocol [32]. For this first interactive demo, participants were asked to "please treat this like it is a call with a real patient who you're meeting for the first

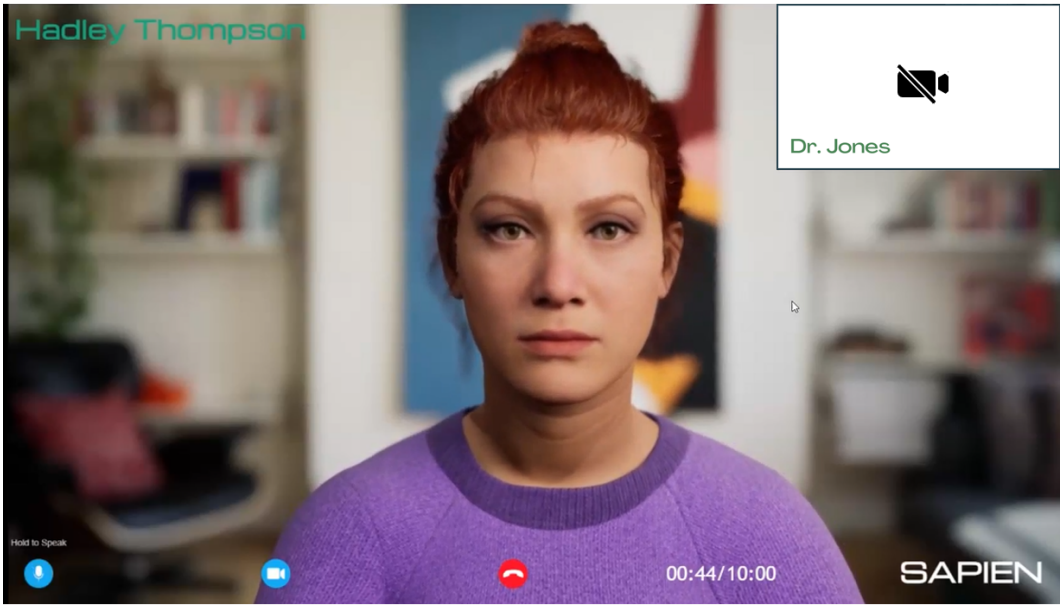


Fig. 4. Screenshot of "Hadley Thompson," one of the virtual patients in our study

or maybe second time for prolonged exposure therapy for PTSD." While there was wide variety in how participants interacted with the virtual patient system during their two demo interactions, most attempted to engage in relatively realistic conversations as requested by this prompt.

3.3.2 Second Demo: Customizable "Free Play". Due to a technical interruption and time constraints, P1 only completed the first demo interaction. All other participants engaged in a second, "free play" 5-minute interaction. Participants were asked to pick any avatar from the 20 options (see **Figure 1**) and freely customize their name, age, bio, and the conversation premise (see **Figure 2**). In cases where a participant did not practice prolonged exposure therapy and/or treat PTSD, they were explicitly encouraged to more closely align the virtual patient to their specific expertise through the bio and conversation premise. **Table 2** provides details on the avatar (chosen from the twenty options) and mental health condition (freely customized via the bio and conversation premise) for each participant's second demo interaction. Of the 11 participants who engaged in this "free play," seven chose to again focus on PTSD. Five participants chose a feminine-presenting avatar and six chose a masculine-presenting avatar.

3.3.3 Semi-Structured Interview. After these two demo interactions, the participants engaged in semi-structured interviews [103] of approximately 20 to 30 minutes. These interviews included characterization of their clinical training experience, their evaluation and thoughts on the virtual patient system, comparisons of the system to familiar training methods (i.e., role-play) and real patients, and use cases for the system within the clinical training process.

3.4 Data Collection and Analysis

All sessions were held over video call (Zoom) and recorded in support of later transcription and analysis. The audio from the demo interactions and interviews were automatically transcribed and subsequently manually checked for accuracy by the first author.

Table 2. Details of "Free Play" Demo

	Avatar (See Fig. 1)	Conversation Focus
P1	(N/A due to a technical interruption and time constraints)	
P2	14: "Natalia Petrova"	PTSD - Military Combat
P3	3: "Danielle Rodriguez"	PTSD - Sexual Assault
P4	8: "Hudson Johnson"	PTSD - Military Combat
P5	3: "Danielle Rodriguez"	Borderline Personality Disorder
P6	9: "Jesse Carter"	PTSD - Military Combat
P7	13: "Myles Wright"	PTSD - Military Combat
P8	3: "Danielle Rodriguez"	Parent Guidance Therapy (Autism)
P9	9: "Jesse Carter"	Relational Therapy
P10	9: "Jesse Carter"	PTSD - Military Combat
P11	17: "Pia Andersson"	PTSD - Military Sexual Assault
P12	18: "Seneca Mitchell"	Suicidal Ideation

We analyzed the transcripts using bottom-up, reflexive thematic analysis [21–24]. This approach encourages reflective researcher engagement through discussions of positionality and interpretive influence on our findings [22, 127]. The first three authors familiarized themselves with the transcripts. The first author then coded all of the transcripts, and the third author independently coded half of them. The first and third authors met frequently to discuss important ideas we identified in the data and candidate themes. The first author created an initial set of themes and subthemes, followed by collaborative discussions and refinement. This resulted in a final set of 3 themes and 13 subthemes, which are discussed in our findings.

4 Findings

Twelve psychotherapists engaged in interactive demos of the virtual patient system and were interviewed about their clinical training experience, their thoughts on the virtual patient system, comparisons of the system with familiar training methods (i.e., role-play) and their real patients, and use cases for the system. We now detail the three themes and thirteen subthemes distilled by our reflexive analysis. We first discuss three key interpersonal therapeutic skills which participants viewed as crucial for effective psychotherapy delivery. We then discuss the potential of virtual patients within the mental health training process and our participants' views on how the system compares to traditional role-playing methods. Finally, we provide an overview of eight factors which influenced psychotherapist-perceived "humanness" of the virtual patient.

4.1 Developing Three Interpersonal Skills For Psychotherapy Delivery

When discussing their experiences with the mental health training process and providing feedback on the virtual patient system, psychotherapists revealed three key interpersonal skills for delivering psychotherapy: building rapport, emotional regulation, and navigating flexible therapy delivery. While psychotherapy effectiveness relies on mastering a wide variety of competencies and abilities, participants referred to the development of these three skills as a particular challenge.

4.1.1 Building Rapport: Fostering a Relationship with a Patient Requires Responding to Subtle Cues. Participants noted that building a relationship with the patient is crucial. While maintaining

rapport is required throughout the entire therapy, building a "safe space" (P3) is perhaps most important at the start of the therapeutic intervention. Psychotherapists must develop their skills in "helping people feel safe, trusting, and willing to open up and talk" (P9). Building rapport may also require incorporating a patient's culture or background. The APA ethical code recommends that psychologists utilize a Cultural Formulation Interview (CFI) [9] with new patients to enhance understanding of cultural factors and context. P4 noted that "sometimes there's just this really rich, either cultural or historical background that people have that could be incorporated into psychotherapy. And sometimes it's hard to find that out, at least in the very beginning of a relationship." Psychotherapists must also take into account the patient's current knowledge, as "the job of the therapist is to not only do the therapy, but make it understandable to someone who doesn't understand the theoretical background" (P6).

While some participants saw an opportunity for rapport building with a virtual patient, especially for new psychotherapist trainees to get comfortable delivering the therapy and listening actively, most participants were skeptical about the utility of practicing rapport with the system. This was primarily because "there's so many nuances with rapport" (P10). P12 noted "there's a lot when people walk into the room for the first time that you pick up on nonverbally [...] All these different things that I think inform where the person is." Participants also emphasized that building rapport with the virtual patient may require feeling an interpersonal connection. For example, P3 stated that rapport "was challenging for me, because I felt so disconnected from this avatar who I was supposed to just have a relationship with and apply my skills to. Like, that's kinda out of the blue."

4.1.2 Emotional Regulation: Delivering Effective Psychotherapy Requires That You Hold Your Feelings. Psychotherapists must also develop skills of emotional regulation, de-centering their own emotions to ensure an effective therapy session. They must maintain their composure, professionalism, and empathy, despite the patient's actions or emotional response. P8 said that psychotherapists hear "some intense things. And that's part of what is so important about doing good internships and doing good practice is just [...] learning how to regulate yourself in those stressful situations." P10, who has worked with multiple patients who have trauma from sexual assault, noted that part of emotional regulation is desensitization to talking about trauma. Participants also expressed the challenge of holding their feelings at the start of therapy, especially when activating a trauma response, instead of immediately trying to help or "fix" their patient's distress. On exposure-based therapies, P2 said, "I think that there's a lot of preconceived notions when we enter the field, [...] that we should alleviate stress, that they're not coming to us to feel distressed. And so getting over like the mental hurdle that this is actually what's going to help them was probably the biggest mental thing."

Participants felt that virtual patients could bolster emotional regulation skills by allowing trainees to get used to talking about trauma and get comfortable with intense or stressful situations. P8 noted that the virtual patient could help trainees "practice like, none of this is about me. It's all about them [the patients]. I'm allowed to hold my feelings, but I don't have to let those feelings take over my professionalism or anything like that."

4.1.3 Navigating Flexibility: Even When Delivering a Protocol, You Need to Be Flexible. Even manualized therapies, which require strict adherence to a protocol, can't be delivered too rote or scripted. Participants identified the challenge of "doing the science of the therapy, but being human while you're doing it" (P6). P2 noted that "that's the biggest feedback I hear from trainees now is just being able to go through the manualized pieces without feeling rote. Or feeling like they're reading from a script." While a psychotherapist practicing a manualized therapy may decide to deviate from the protocol in some situations, they must also be able to get the session back on track when needed. For example, even engaged patients with PTSD may display "avoidance behaviors," or

methods to avoid trauma-related stimuli or emotional processing [41]. These avoidance behaviors pose a challenge, as they "are kind of pushes to do the exact opposite of our treatment" (P11). A psychotherapist must "be able to respond where the patient's going, and then be able to kind of bring that back to whatever the goal is for that day" (P2).

Virtual patients may be able to assist with practicing the flexible delivery of therapy, especially by encouraging trainees to repeat the protocol many times to get more comfortable with small deviations and "finding your way of saying it" (P10). Participants praised the virtual patient's utility for practicing flexibly thinking on their feet to determine how to respond, for example if a "patient asked me a question that's not what the manual says I should say next" (P10) or does something unexpected such as mention suicidality (P2). P6 even suggested that the virtual patient system could provide feedback "to figure out like, do you sound like robot? Right, are you just reading this protocol?" by analyzing sessions over time to evaluate variation, if you're "meeting each patient where they are, or are all your recordings the same and you're saying the exact same thing?"

4.2 The Potential of Virtual Patients for Clinical Training

With an understanding of some of the challenges psychotherapists faced during their training, we asked participants to imagine use cases for the virtual patient system and where it should (or shouldn't) fit into the training process. Below we discuss the psychotherapists' comparison of the virtual patient system to traditional methods of role-play. In particular, psychotherapists highlighted two key benefits of the system: practicing in a safe space, and customizing or tailoring the virtual patient.

4.2.1 Comparing Virtual Patients to Mental Health Role-play: Freeing and Safer For "Baby Clinicians," But Lacking Nuance. While some had "learned on real patients" (P7) or "really, truly, trial by fire" (P3), most of our participants had extensively utilized role-play exercises with a peer, supervisor, and/or a family member or friend. Compared to this traditional role-play, one of the biggest perceived benefits of the virtual patient system is that it could offer "baby clinicians" a safe space to practice without fear of risk, judgment, or negative repercussions. P6 said that virtual patients "could be a good first step, right, where it's like, I'm not throwing you any curve balls. [...] Like, just get in the rhythm of what an ideal session should be in the perfect world and have a shared language."

Practicing with a virtual patient was seen as low-risk "because you don't actually have a person that you could harm, right? Do no harm" (P3). P4 echoed this sentiment: "I noticed that it felt very safe, too. [...] That felt very kind of freeing actually, to not have to worry about a human." P12 emphasized the "sense of freedom to be across from a machine, knowing that what you're gonna do is not gonna have any greater repercussions than, you know, whatever feedback you're gonna get."

Participants also noted that role-play "with somebody who's actually like sitting in front of you with expectations can be really daunting" (P2), and can have "a natural level of awkwardness to it" (P12). Role-play with a supervisor may further elevate feelings of evaluation or judgement. Role-play with a virtual patient was therefore described as "a low-threat, low-risk way to practice having these difficult conversations" (P2), a way "you can make mistakes, like really honest mistakes, and still get feedback on it, cause the virtual patient's gonna respond" (P10).

Because the virtual patient is knowledgeable about mental health content, the system was viewed as preferable to practicing with a peer, family member, or friend, who sometimes "have limited knowledge of what a PTSD patient would say" (P7). In addition, P11 (who identified as a white woman) noted that the virtual patient system could represent more diverse patients, as "peer to peer [role-play], a lot of times they look very similar to me." P10 also noted that role-playing some intense situations, such as aggressiveness or inappropriate behavior (to practice emotional regulation and asserting boundaries), would be safer with a virtual patient than with a supervisor.

However, others expressed that interacting with real people during role-play was the key to skill transfer and generalization. In particular, they stressed rapport building, nonverbal queues, and reflexive feedback discussions. P9 said that role-play with the virtual patient is "almost like learning how to play poker with play money. [...] I mean, you might be learning some skills. But it's not, you're not learning the way it actually operates in a real game or, in this regard, in real therapy."

4.2.2 Customizing and Tailoring the Virtual Patient for Wider Utility Within a Clinical Training Context. Comparing the virtual patient to other training methods, participants also widely discussed the ability to alter the system to meet the needs of individual trainees. This positioned the virtual patient as a standardized, repeatable, and customizable training method which could be tailored to trainees at multiple access points throughout the training process, "instead of just having a one size fits all" (P12). The ability to customize the virtual patient would also greatly broaden system utility by offering use cases beyond new trainees. As P4 stated, "there's never a time in a psychotherapist's career where they wouldn't benefit from any type of training, especially experiential training or role-playing."

Participants stressed the need to be able to interact with many different kinds of patients and "a diverse set of training scenarios" (P5). Participants identified many factors for potential customization, including: patient background (e.g., comorbidities, previous psychotherapy experience, vocabulary, language, age, gender, cultural background), patient personality and level of engagement (e.g., readiness to change, emotionality, talkativeness, manner of dress), patient mental health disorder and presentation (e.g., avoidance behaviors, trauma type), and the context of the session (e.g., type of intervention or protocol, progress within the intervention).

By customizing the virtual patient, one could also tailor the system to different types of psychotherapists, even those who are more experienced (e.g., if learning a new kind of psychotherapy or taking a Continuing Education Unit). P3 said that as psychotherapists, "we are always learning about new theories and seeing how they're applied. So it would be cool to be able to say, like, 'okay, I'm gonna teach you IFS [Internal Family Systems Therapy]' [...] and then you can do that with this avatar and practice it for 20 minutes or whatever, and see how that feels."

4.3 Perceived Humanness of the Virtual Patient

While psychotherapists favorably compared the virtual patient to other role-play partners, they also expressed that the utility for bolstering interpersonal and therapeutic skills largely hinged on the ability of the virtual patient to simulate a real patient, i.e., to seem sufficiently "human." Below we outline eight reflexive subthemes of humanness as discussed by participants in comparison to their real patients: attitude, predictability, showing affect, culture, competence, message contingency, latency, and simulation context.

4.3.1 Attitude: Why Does the Virtual Patient Want to Be In Therapy So Badly? Participants described the virtual patient's general demeanor as too positive and forthcoming, too easy and mild, and "a really compliant patient" (P5). P9 stated that the virtual patient provided content and opened up too fast, "kind of like a scripted television show," and that "they offered everything perfectly on a silver platter." Participants were quick to note that this is not the case with real patients. P5 said that the virtual patient "is like, the perfect world, how it will work. But a lot of times you have patients who are like not responding to you [...] or start crying a lot in the session or who are fighting you." P6 stressed the importance of practicing with patients who push back or express conflicts within their motivations for psychotherapy: "...for a lot of vets [veterans] especially who are male, right, 'I'm coming in because my wife made me' or 'My fiance made me because I'm intolerable to live with.' " P6 described these as "the vagaries of being human." Because the virtual patient was seen as having an inhuman attitude towards being in therapy, the system was relegated to a tool for new

psychotherapists to practice delivering therapy once the patient is fully ready to engage, rather than to develop the rapport-building skills necessary to help patients feel comfortable and open up.

4.3.2 Predictability: The Story of the Lovely Unpredictable Human and the Broken AI. The virtual patient was "like the perfect client [...] like, wow, they're saying exactly what you would want them to say, so in some ways that's not realistic" (P9). While P6 felt that this predictability could be a good thing (because it indicated that the virtual patient was informed by "scientific writing or the protocol"), they also noted that "patients don't talk like that and you never know what they're gonna say." Participants described humans as more spontaneous and nuanced. P10 noted that after multiple sessions of practicing PE with a patient, they might be successfully working with a patient "and then all of a sudden, there's like a new avoidance structure."

Some participants were unsure about the capability of AI to simulate human unpredictability. P8 noted that "you get a lot of curve balls with humans" and P10 stated that "I don't know if you can ever really program [so much unexpectedness]." P6 asserted that "we know from the literature we're terrible at predicting who attempts suicide and who doesn't," suggesting that some aspects of patients may be impossible to simulate because we're not able to adequately model those behaviors.

Because humans are expected to be unpredictable only within patterns and/or an expected scope, peculiarities of the virtual patient were not interpreted as desirably human but instead as less realistic and inhuman. A patient doing something surprising was described as a key part of being human, while the virtual patient doing something unexpected (e.g., answering a prompt incorrectly, displaying an unusual facial expression) was interpreted as decidedly non-human.

4.3.3 Showing Affect: Why Are You Here If You Feel Fine? Participants stressed that the virtual patient must be able to show both big and subtle affect. P10 expressed the importance of displaying "the full range of emotional responses, like full body sobs, flipping me off..." This is especially vital for practicing emotional regulation: "There's a lot of silence and quiet and crying and emotion. And the therapist needs to be comfortable sitting with that and holding that. And I think in this situation... I just didn't know where to, like, how to really connect with this avatar, I guess" (P3).

More nuanced signs of affect may be even more important, for "seeing like a flicker of an emotion and knowing to dive in deeper" (P10). Participants noted that cues such as a patient's eye contact, tone of voice, and even the way they are dressed "can be a lot of information for us, too, as we're trying to understand kind of what their experience is" (P1). However, participants struggled to pick up on these subtle cues with the virtual patient system, and were unsure if a virtual agent would be capable of showing subtle and nuanced affect. P9, for example, said "I mean, I really don't know. Cause real resistance, rooted in some sort of fear [...] I don't know if AI has those capabilities" (P9).

However, all affect is not created equal. The lack of displayed negative (rather than positive) affect had a larger impact on perceived humanness and utility for skill development. P1 found it difficult "to sort of believe that the avatar was really upset about their trauma." Similarly, P9 only perceived positive affect in the virtual patient and "didn't perceive much struggle or strife. [...] It felt like an expected vignette, you know. Like this is how, you know, this is what people feel. But this isn't a person." Two participants also noted that effectively showing negative affect could enhance feedback on rapport; this feedback could come directly from the virtual patient through affective responses or by saying, as suggested by P7, "When you said it this way, it shut me down."

4.3.4 Culture: Ethical Training Requires Representation and Cultural Competency. Therapy delivery relies heavily on detecting subtle cues of affect, engagement, and avoidance, while maintaining rapport with patients. Training with a system that is not representative of real patients could lead to under-training, and ultimately poor therapy effectiveness. P9 noted the potential for "an empathic miss," if trainees are taught to "deal with the presentation of affect in a certain way" and

are therefore unable to recognize a wide variety of emotional presentations. P8 also expressed concern regarding a virtual patient with unrealistic responses, "then I, as a brand new counselor, then go into a real therapy setting and my real human doesn't react the same way that the AI did."

Cultural considerations were particularly intrinsic to "humanness" and the ethical representation of patients. Participants highlighted the importance of "representing patients accurately, representing cultural differences" (P7) so that trainees don't "understand how to do therapy with only a segment of the population" (P4). At the same time, P12 expressed the importance of ensuring that minoritized communities are not portrayed in a stereotypical and/or offensive manner. P12 chose to interact with a virtual patient who they presumed to be Indigenous and said: "I was kinda interested to see if some of that would come through. But I know you put in the prompts and it basically spits out what you put in." Especially in prolonged exposure therapy, which may involve completing exposure exercises in the real world (e.g., in a grocery store, playground), P2 emphasized the need to consider a patient's culture and identity, "because what is safe for one group of people isn't always going to be safe for all groups of people." P2 stressed the importance of having "diversity considerations of what is safe for you. What do other people in your community feel comfortable doing, so [...] that it allows for this conversation of what's really going to be helpful, but also what's genuinely, objectively safe for you and your community."

4.3.5 Competence: Where'd You Learn That? The Virtual Patient Knows Too Much. Participants felt that the domain-specific content of the virtual patient's responses (i.e., the LLM output) "sounded pretty good, pretty accurate" (P7) and was "pretty similar to what I would hear" (P2). P9 noted that the responses realistically followed from the provided patient biography (as in **Figure 2**). However, the level of psychotherapy knowledge demonstrated by the virtual patient was unrealistic, especially because of the "kind of textbookiness of it" and that it was interpreted as "based on scientific writing or the protocol" (P6). P10 said that the virtual patients "are pretty high level, like they're saying 'intrusive memories.' I usually don't have patients say words like that. That's like, very doctor speak. [...] when they do, I'm a little skeptical like, where did you learn this word, 'intrusive memory?"' Beyond not being realistic, this domain-specific knowledge could indicate to a psychotherapist that their patient has been exposed to dubious information or misinformation (perhaps, as P10 suggested, from TikTok or other social media). This may cause a trainee to redirect their approach to address this unexpected knowledge rather than focusing on other aspects of the therapy protocol.

4.3.6 Message Contingency: It's Not Just What You Say, But How You Say It. Message contingency, or how related or dependent the LLM's responses are to the user's inputs, is important for conversational grounding [29]. P7 said the high level of responsiveness (as opposed to the virtual patient simply reading off a script) was pivotal to their immersion in the role-play: "I wasn't just holding it [the spacebar] down, and you know, they're waiting for someone to talk, and then going to go on with their script. So when I realized that, then yes, I got more invested and acted like a regular role-play." Despite this immersion, P12 highlighted that the virtual patient's "sequence of responding to the questions is somewhat unnatural." When asked a compound question, a real patient will respond to the more serious prompt first, e.g., prioritizing a question about suicidal ideation. The virtual patient did not respond to the pragmatic aspect of clinical prompts, instead responding sequentially. This suggests that message contingency must manifest not only in the response itself, but also in the emotional priority implied by how it is presented (i.e., content ordering). This would help avoid confusion ("like, okay, do you mean, yeah suicide or no?" (P12)), but is especially important for the emotionally intensive use case of psychotherapy, which relies on a therapist's ability to identify nuanced affect.

In addition, the system must react appropriately to messiness or humanness *from* the user, in order to seem more human *to* the user. When P9 provided the virtual patient a prompt it could not

effectively parse, the system responded with a general self-disclosure from the input biography. P9 expressed that they would instead prefer that the virtual patient ask for clarification: the way the virtual patient "anticipated the next [...] properly phrased question from a therapist [...] made me think that they're not really responding to me, anyway." Trainee therapists will not always deliver perfect therapy and should not learn to over-rely on a patient to get the session back on track. Especially for training contexts, the virtual patient should respond realistically to clinicians who are not yet experts.

4.3.7 Latency: Hadley Can You Hear Me? Hadley Can You See Me? Participant comments on latency varied, but a time delay was not necessarily associated with a lack of realism. While P4 was "pretty impressed with the latency" and P5 praised the "real-time feedback," others felt that the delay in responding to the user took away from perceived humanness. Notably, P6 commented that the delay in itself wasn't necessarily making it inhuman, but that the virtual avatar was missing "the human factors in there when it's taking a while to respond. So patients can take a while to respond, but they're like kinda hemming and hawing, or like thinking, or like getting uncomfortable. So I'd say that was a pretty noticeable difference." Rather than aiming to simply reduce system latency, intentional variation in delay may make the virtual patient seem more "human" to trainee therapists; however, P6's comments highlight the importance of verbal and nonverbal behaviors which suggest deliberation, discomfort, or other cognitive processes during these periods.

4.3.8 Simulation Context & Immersion: I Believe, I Believe. It's Silly, But I Believe. Although some of the previously discussed factors negatively influenced the perceived humanness of the virtual patient, multiple participants were surprised by their level of immersion. Two participants expressed that it felt "weird" or "rude" to hang up on the virtual patient without saying goodbye. P4 said that, "at times I actually was interacting with it as if it was a human, like trying to be polite and not, you know, speak while they're speaking. I was kind of surprised that I kind of so quickly responded to it as if it was a human." Participants who weren't as immersed in the interaction with the virtual patient still felt that they could "show empathy to them" and "approach them in the same way I would a real patient" (P10). P11 further said: "I think there'd always be something in the back of my head, knowing that it wasn't [real]. But [...] I think the practice on that would definitely, definitely translate to like a real person." Participants noted that while they would not necessarily "care" about a virtual patient, they could still get invested in their own performance: "...we get a satisfaction from seeing a therapy progress correctly. And I feel like I'd have that same satisfaction with this, you know, artificial agent. Just knowing that what I'm doing is, it's working" (P4).

A factor positively impacting perceived humanness was priming: participants made the decision to engage with the virtual patient, suspending disbelief. P4 said, "I just kind of reminded myself, oh, just act as if it was real. Or you know, don't *pretend* it's real, but *act* as if it was real." Similarly, P12 noted that they easily got immersed in the interaction, "but I also had to make the decision prior that I'm gonna- I'm not gonna take it as a joke, I'm not gonna laugh if they say something, you know, kinda off or weird. I'm gonna like, really take it like it's a person in front of me."

Two participants also praised the simulated video call's close alignment with the increasingly ubiquitous paradigm of telehealth, especially after the COVID-19 pandemic [14, 128]. P12 noted that "a lot of therapy's being done virtually so that kind of fits that model," and P5 said that "since the pandemic, a lot of therapy has moved to be online, to be virtual. So clinicians need to be prepared to conduct both virtual, remote therapy and in-person."

5 Discussion

Role-play exercises are a ubiquitous training technique across many domains [43, 51, 66, 110]. Although AI-powered technologies such as virtual patients offer the promise of more engaging,

representative, and reliable exercises, the implications of augmenting role-play techniques with human-AI collaboration have not been fully explored by previous work. As a case study of training which relies highly on the interpersonal, we investigated a virtual patient system intended for therapists-in-training. Our evaluation of the system with 12 trained psychotherapists provided insights into three interpersonal skills required for effective psychotherapy delivery (building rapport, emotional regulation, and navigating flexible therapy delivery) and the potential of virtual patients within a psychotherapeutic training process. In addition, our reflexive analysis distilled eight factors which influenced perceived "humanness" of the virtual patient. In this section we discuss the implications of our findings in more detail.

We first leverage the Computers Are Social Actors (CASA) framework [81] to comment on a cognitive dissonance suggested by psychotherapists' feedback, and we explore rapport-building with a virtual patient, as a use case in which the AI is expected to be less receptive or even adversarial. We then discuss challenges of prototyping novel human-AI systems for clinical contexts in which a high degree of unpredictability is required. Finally, we leverage Epley et al.'s three-factor theory of anthropomorphism [40] to contextualize our eight factors of humanness and stress the importance of representing diverse cultural communities within mental health AI systems, in alignment with decolonial computing.

5.1 Human–Virtual-Patient Rapport

5.1.1 Cognitive Dissonance, Rapport, and "Humanness". The Computers Are Social Actors (CASA) [81] framework suggests that users have the tendency to enact learned social and/or anthropomorphizing behaviors towards technology. This may be the case even when users do not actually view the technology as a human or social actor, and may even believe that their behaviors towards the technology are inappropriate. This leads to a cognitive dissonance as users struggle to rationalize their social behaviors towards technologies [105]. Multiple psychotherapists were surprised by the level to which they were immersed in the virtual patient interaction. While there were flaws within the system (e.g., facial expression bugs) that hindered participants from using the virtual patient to its fullest potential, these flaws did not entirely impede participants' immersion. Of course, suspension of disbelief is required to some degree with any role-play, even with human role-play partners; however, participants also revealed a potential case of cognitive dissonance regarding rapport-building with the virtual patient system, as we discuss below.

Rapport has been defined as "a harmonious relationship underpinned by effective communication," and it may be supported by factors such as humor, commonalities, positivity, and empathy [61]. Non-linguistic cues and nonverbal communication in particular are important in facilitating intimacy and building rapport with a virtual agent [28, 73, 125]. In clinical practice, rapport refers to a mutually favorable state which facilitates a patient's sense of safety and therefore self-disclosure [56, 64]. After their demo interactions, participants commonly cited rapport as the skill they most needed to practice and would most like to practice with a virtual patient. At the same time, they indicated that they largely did not feel it would be possible to improve rapport skills by using the system. Practicing rapport may not necessarily require two social actors, as is made evident by many well-known methods for practicing rapport individually, e.g., practicing phone calls with no one else on the line. We must ask, then, why our participants felt that the "humanness" of a role-play partner would impact their ability to practice rapport. "Humanness" may just be a scapegoat to resolve the cognitive dissonance of users when they are asked to more deeply explain their interactions with a virtual agent.

5.1.2 Designing for Rapport Receptiveness. Our findings, especially the investment and immersion reported by psychotherapists, suggest that it may be possible to practice rapport with a virtual

agent. The system aimed to create a highly realistic simulation, providing high-fidelity avatars and dynamic responses from a large language model. Participants revealed, however, that the system must better simulate the behavior of real patients. Previous work [2, 56, 126] has explored rapport-building during human-AI collaboration in support of the design of health chatbots. These chatbots were designed to drive rapport, intentionally building a mutually beneficial relationship with a user. Wang et al. correlated rapport-building with specific linguistic attributes such as emotional tone, positive tone, and affective processes [126]. Our findings suggest, however, that a virtual patient who exhibited these attributes would be seen as far too forthcoming, with an inhuman attitude towards therapy (see Section 4.3.1). While previous work has explored a virtual agent's ability to build and drive rapport, less is known about designing an AI agent to play the complementary role, i.e., to be reliant on a user to build that rapport. This would require a virtual agent to be receptive to the user's attempts to build a relationship, or even adversarial (e.g., if a patient is not engaged in therapy or is exhibiting avoidance behaviors). Our study acts as an early indication that previously documented benefits of human-AI collaboration for interpersonal skill development (e.g., user engagement [97], self-confidence [88]) would translate to the clinical training context. Future work is required, however, to better understand how to simulate the behaviors of patients in response to a psychotherapist's attempts to build rapport.

5.2 Imagining and Prototyping Unpredictable Virtual Patients for Mental Health Training

Psychotherapists were optimistic about the utility of the virtual patient system, especially for practicing general psychotherapy skills and key interpersonal skills at an introductory level. However, there was much variety in participants' feedback and opinions of the system, dependent on their specific session. As shown in previous work on human-AI interaction, there exists a number of design and prototyping challenges for human-AI interaction [69], stemming from two main sources of complexity: capability uncertainty and output complexity [131]. Because virtual patients must be particularly unpredictable (see Section 4.3.2), the psychotherapy use case may further enhance these challenges.

Participants interacted with the virtual patient system for only approximately ten minutes total (due to limited availability and prioritization of clinical tasks, i.e., treating patients), but whatever occurred in their session was understood to be reflective of the capabilities of AI. For example, if the system lagged due to their internet connection, or the LLM provided an overly technical response, it was interpreted as evidence that AI is too slow, or AI knows too much about therapy. Understandably, participants seemed to extrapolate from their short interaction to form or update opinions on what the system (and AI more generally) is capable of. Our findings, however, suggested that virtual patients must offer a high degree of unpredictability, offering unexpected situations (see Section 4.3.2), and diverse patient profiles (see Sections 4.2.2 and 4.3.4). Because the system aims to simulate the behavior of individuals diagnosed with a mental disorder, this introduces a need for variability above and beyond traditional human-AI collaboration use cases. Designers will soon need techniques to quickly present simulated unexpected behavior without participants interpreting this simply as a broken AI (see Section 4.3.2).

In addition, our study builds upon prior work [17, 98] which emphasizes the need for effective partnerships with domain experts for interdisciplinary health design. Our findings suggest that psychotherapist involvement is crucial to ensure clinical validity and representation of real patient behavior. Designers must therefore provide intuitive, useful, and usable interfaces and thoughtful back-end implementations so that domain experts and supervisors (i.e., expert psychotherapists) can effectively customize the virtual patient system with little to no system training. This challenge, too, is heightened when the system aims to be inherently unpredictable, as designers must balance

the need for unexpected virtual patient behavior, while still aligning with expert psychotherapist intention. In support of the human-centered design of these interfaces, researchers should be cognizant of the need for human-AI partnerships [119]: virtual patients should be viewed as an empowering tool to help trainees practice their skills such that they can make the most out of the human-human role-play exercises in which they do partake.

5.3 What It Takes To Be "Human"

Concerns of "dehumanizing medicine" convey an existential threat to humanness in healthcare [35, 104, 119]. This threat also underlies our participants' concerns, from delivering the therapy with the patient in mind, to worries of representativeness and the potential for an "empathic miss" (P9). However, it is not clear that "humanness" is required for effective interactive technologies. Rather than relying on anthropomorphism, some work instead suggests that *non*-anthropomorphic agents garner higher user trust because machines are seen as more objective and knowledgeable than humans [71, 102]. Users may still perceive some degree of consciousness in non-anthropomorphic technologies: Scott et al. showed that consciousness may be perceived in even everyday technologies such as a vacuum cleaner [100], and some readers may fondly recall training their "Pet Rock" in 1975.

Previous studies of virtual agents suggest that their visual aspect (compared to a purely text-based agent) positively influences perceived humanness [11]; being able to "see" an AI via some visual presence increases the trust placed in the AI, such that it may be considered an autonomous teammate rather than simply a tool [50]. Similarly, previous work on voice assistants suggests that a human-like voice facilitates human connection [105, 136]. However, our participants focused more on the content provided by the LLM. If the look of the avatar, its bodily movements, or the voice generation system were mentioned by participants, it was as something that they could easily get past or ignore. The content of responses posed more of an obstacle to being able to practice with a virtual patient as if it were a real human.

In **Table 3** we provide implications for the design of virtual patients which aim to be perceived as human-like, inspired by each of the eight factors of humanness outlined in our findings. Many of these eight factors of humanness align with Epley et al.'s theory of anthropomorphism ("SEEK") [40]. For example, the view that virtual patients should have similar "predictability" to humans is aligned with the desire for uncertainty avoidance while interacting effectively with others (cultural effectance motivation). Importantly, the elicited agent knowledge of the SEEK theory explains that the basis for anthropomorphism is existing knowledge of other humans and their behavior, including one's own self-knowledge. As an extension, we assert that perceived humanness of a virtual patient is also inherently based on the corpus of patients with which psychotherapists have previously interacted. This finding was echoed by Zhang et al. [137] in their study of expectations of AI teammates within a multiplayer online game, highlighting that expectations for "human-like behaviors and performance" were largely formed throughout their participants' many years of gaming experience. This suggests that our participants' views of how a patient should talk and act is heavily influenced by the patients they have seen in the past. It therefore may be necessary not only to interrogate psychotherapist-perceived humanness, but also to understand how diverse patients interact in therapeutic settings to better simulate their experience.

Further, attempts to adequately represent a diversity of patients and cultures within a virtual patient system may offer an opportunity to align the system with decolonial computing [4, 7] and specifically decolonial mental health [87]. The influence of culture on mental health, as expressed by our participants, may be thoughtfully reflected in the system to mitigate the decentralization and erasure of minoritized experiences, propagated by how evidence-based psychotherapeutic treatments understand mental health. As P6 noted, the virtual patient seemed to be informed by

"scientific writing or the protocol." We must acknowledge that while considered the gold standard within clinical contexts, utilizing these materials for training a virtual patient may not be inclusive to marginalized presentations of mental health or distress [111]. In fact, basing the definition of "humanness" only off of individuals who have sought therapy with a PhD qualified psychotherapist in the United States, although compelling in this work, may propagate the historical inequities of those considered less than human by colonial forces and other forms of power. In Section 4.3.4, we presented the need for the virtual patient system to understand what would be safe for the simulated identity in order to respond appropriately to a psychotherapist. A homework assignment to sit near a playground for an hour, while safe for some, would likely introduce risk for a Black American or a visibly queer individual in the United States. This demonstrates a need for heightened cultural sensitivity within the virtual patient system; however, LLMs and generative AI have extensively demonstrated stereotypes and biases correlated with gender, race, religion, and nationality [34, 55, 62, 67, 79]. Chat-GPT in particular has been shown to frequently lack depth of authenticity [73], cultural sensitivity, and self-awareness of biases [3]. A more equitable avatar system for mental health professional training would require participatory methods which embrace diverse patients and perspectives [8], and center longstanding power relations and structural factors, even if this requires demonstrating a critical stance or adversarial responses to a psychotherapist user. To this end our study is only a first step, and we call for HCI and CSCW researchers to explicitly explore how views of anthropomorphism and "humanness" of interactive technologies may be reframed through a decolonial lens.

6 Limitations & Future Work

This paper has a number of limitations and has introduced open questions which provide avenues for future work.

Although it was not the focus of the psychotherapists in this study, future work should center the ethical implications of human-AI collaboration in a psychotherapy use case. Especially when exploring the representation of patients and cultures, it will be vital to involve a wide diversity of voices, including those from communities which have been historically marginalized within mental health care. While we utilized purposeful sampling and attempted to recruit a wide variety of psychotherapists via cold emailing, most of our participants were ultimately recruited through word of mouth or snowball sampling. These recruiting methods face a variety of challenges and potential biases [10, 65, 78]. The participants in this study were therefore largely homogeneous along the lines of race/ethnicity and education: the majority identified as white, hold a doctorate, are affiliated with large medical and research institutions, and have a background of evidence-based psychotherapy. Additionally, previously documented biases in medical and psychiatric diagnoses on the basis of race, ethnicity, and gender [115] threaten to further marginalize those who already face substantial health disparities. In addition to incorporating more diverse perspectives from psychotherapists, future studies could explore whether a virtual patient system offers an opportunity to create a quasi-controlled environment in which psychotherapists may be trained to better serve these communities.

As this study is an early exploration of this virtual patient system, our work is also limited in its ecological validity. Traditional role-play exercises occur within a highly specialized training context, with peers and/or supervisors. Future work should aim to assess the ecological validity of our findings by exploring the virtual patient system within a more contextualized, long-term study, e.g., a pilot deployment within a larger training program. This may help explicate the impacts of priming and context of use, as well as provide an opportunity to offer more targeted simulations with specific learning goals. This pilot deployment could also take a mixed method approach,

Table 3. Our Eight Factors of Humanness & Associated Considerations for the Design of Human-Like Virtual Patients

"Humanness" Factor	**Design Implication to Increase Perceived "Humanness"**
Attitude (Section 4.3.1)	A virtual patient should not be too immediately forthcoming. Disclosure should be gradual and layered, in alignment with simulated patient engagement, avoidance behaviors, and identity.
Predictability (Section 4.3.2)	Virtual patients should aim to be surprising and offer "curve balls" (i.e., unexpected or surprising events), but must attempt to closely align *the ways in which they are unpredictable* with real human behavior.
Showing Affect (Section 4.3.3)	Virtual patients should be capable of displaying both overt and subtle affect. Negative and nuanced affect attribute more heavily to perceived "humanness," as identifying this nuanced affect is crucial for psychotherapy training.
Culture (Section 4.3.4)	Virtual patients must be both representative of, and sensitive to, diverse cultural communities, including reacting appropriately to psychotherapists who are not culturally competent. Virtual patients should attempt to simulate situated knowledge [47], but without propagating stereotypes of minoritized communities.
Competence (Section 4.3.5)	Both the quality and quantity of (domain-specific) knowledge displayed by a virtual patient must mimic that of the human it wishes to simulate. In some cases, an incorrect or mistaken virtual patient will be a more human-like patient.
Message Contingency (Section 4.3.6)	A high level of message contingency is crucial, not only in the content of LLM responses, but also in the emotional priority implied by the response via its ordering. High message contingency may also require showing the interaction's seams [123] and acknowledging a breakdown in communication with the user.
Latency (Section 4.3.7)	A virtual patient should not simply aim to respond to a user as quickly as possible. To simulate human-like behavior, designers should allow for lower latency but incorporate the appropriate nonverbals, e.g., "hemming and hawing" (P6) to simulate the cognitive processes common during psychotherapy.
Simulation Context (Section 4.3.8)	A simulated video call aligns with increasingly prevalent telehealth; however, in support of immersion and suspension of disbelief, special attention must be paid to how psychotherapists are primed to interact with the virtual patient.

including quantitative measures such as clinically validated pre- and post- assessments to explore the system's pedagogical utility. Our findings outlined only three skills important to the delivery of psychotherapy, and future studies could uncover additional skills which participants did not highlight during our interviews. For example, a longitudinal study could track a trainee's improvement in communication skills or empathy. In addition, a user's trust in AI as a colleague may be associated with its perceived capability of self-improvement without direct human supervision [50].

Longitudinal deployment studies could investigate the temporal aspects of perceived humanness by exploring perceived growth of the virtual patient over time.

Because we rely on domain expert feedback in order to better understand system utility and implications, we must find ways to facilitate these experts' understanding such that they can correctly imagine the system in use. Future work should continue to explore domain experts' perceptions of AI capabilities and unearth potential methods to effectively conduct both speculative design and prototyping studies of complex and intentionally unpredictable human-AI systems.

AI technologies for mental health must not only be human-centered in regards to the end user (e.g., psychotherapists), but also to the patients whose data is collected in order to build the system. Yoo et al. [133] stress the value of patient perspectives, although AI researchers may feel uncomfortable working directly with patients (due to considerations of patient safety). Future research must thoughtfully explore how to best obtain realistic and representative training data for virtual patient systems, while prioritizing the privacy of patient stories and experiences.

Finally, other open questions include how our "humanness" factors might transfer to other role-playing scenarios outside of the psychotherapeutic context, as well as a hierarchy of these factors. We have provided an overview of eight aspects which influenced our participants' perception of virtual patient humanness, however we did not attempt to understand the relative importance of these factors. Some aspects may be crucial or negligible depending on the specific context and/or function (e.g., to practice rapport building vs. emotional regulation). Further, researchers could explore how other details of simulation context impact immersion and perceived humanness. For example, personalizing a virtual agent and the subsequent feeling of ownership may positively influence acceptance of the agent [15]. The "Customize SAPIEN" screen (see **Figure 2**) may therefore have positively impacted participants' views of the second demo interaction (i.e., the "free play").

7 Conclusion

Role-play is a widely utilized training exercise for practicing interpersonal and domain-specific skills across personal and professional life. To explore human-AI collaboration for interpersonal skill development, this study has provided a case study of a novel virtual patient role-play system evaluated by 12 trained psychotherapists. We outlined three key interpersonal skills of psychotherapy (building rapport, emotional regulation, and navigating flexible therapy delivery) which are traditionally practiced through role-play during the mental health professional training process. Although psychotherapists were largely optimistic about the virtual patient's potential ability to bolster these skills, this utility was impacted by the extent to which the virtual patient was perceived as human-like. Our reflexive analysis additionally distilled eight aspects which influenced the psychotherapists' perception of virtual patient "humanness," and we provided associated implications for the design of human-like virtual patients. We synthesized our findings to discuss the use of human–virtual-patient collaboration to practice rapport building and comment on the challenges of prototyping novel, unpredictable AI systems within a psychotherapeutic context. Further, we stressed the importance of designing mental health AI systems which are representative of real patients and their cultural communities. As we increasingly look to AI-powered interactive technologies to augment roles traditionally played by humans, we are offered a unique opportunity to interrogate what we believe it means to be "human."

Acknowledgments

This work was supported in part by the National Science Foundation under Award Number 1915504 and by the University of Rochester Wilmot Cancer Institute.

References

[1] Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*. PMLR, 337–371.

[2] Yuna Ahn, Yilin Zhang, Yujin Park, and Joonhwan Lee. 2020. A chatbot solution to chat app problems: Envisioning a chatbot counseling system for teenage victims of online sexual exploitation. In *Extended Abstracts of the 2020 CHI Conference on human factors in computing systems*. 1–7.

[3] Mahwish Aleem, Imama Zahoor, and Mustafa Naseem. 2024. Towards Culturally Adaptive Large Language Models in Mental Health: Using ChatGPT as a Case Study. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. 240–247.

[4] Mustafa Ali. 2014. Towards a decolonial computing. (2014).

[5] Mohammad Rafayet Ali, Seyedeh Zahra Razavi, Raina Langevin, Abdullah Al Mamun, Benjamin Kane, Reza Rawassizadeh, Lenhart K. Schubert, and Ehsan Hoque. 2020. A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Scotland, UK) *(IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 8 pages. https://doi.org/10.1145/3383652.3423900

[6] Mohammad Rafayet Ali, Taylan Sen, Benjamin Kane, Shagun Bose, Thomas M Carroll, Ronald Epstein, Lenhart Schubert, and Ehsan Hoque. 2023. Novel Computational Linguistic Measures, Dialogue System and the Development of SOPHIE: Standardized Online Patient for Healthcare Interaction Education. *IEEE Trans. Affect. Comput.* 14, 1 (jan 2023), 223–235. https://doi.org/10.1109/TAFFC.2021.3054717

[7] Syed Mustafa Ali. 2016. A brief introduction to decolonial computing. *XRDS: Crossroads, The ACM Magazine for Students* 22, 4 (2016), 16–21.

[8] Adriana Alvarado Garcia, Juan F Maestre, Manuhuia Barcham, Marilyn Iriarte, Marisol Wong-Villacres, Oscar A Lemus, Palak Dudani, Pedro Reynolds-Cuéllar, Ruotong Wang, and Teresa Cerratto Pargman. 2021. Decolonial pathways: Our manifesto for a decolonizing agenda in hci research and design. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–9.

[9] DSMTF American Psychiatric Association, American Psychiatric Association, et al. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*. Vol. 5. American psychiatric association Washington, DC.

[10] Tal August, Nigini Oliveira, Chenhao Tan, Noah Smith, and Katharina Reinecke. 2018. Framing effects: Choice of slogans used to advertise online experiments can boost recruitment and lead to sample biases. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–19.

[11] Julia Ayache, Andrew M Connor, Stefan Marks, Alexander Sumich, and Nadja Heym. 2022. Humanness lies in unpredictability: Role of Theory of Mind on anthropomorphism in human-computer interactions. In *Proceedings of the 10th International Conference on Human-Agent Interaction*. 306–308.

[12] Federico Barnabe, Cristiano Busco, Pål I Davidsen, Maurizio Lambri, and Gianfranco Zatta. 2013. The strategic micro-firm: a role play in management training for dynamic businesses. *Journal of Workplace Learning* 25, 5 (2013), 328–342.

[13] Amine Benamara, Jean-Claude Martin, Elise Prigent, Laurence Chaby, Mohamed Chetouani, Jean Zagdoun, Hélène Vanderstichel, Sébastien Dacunha, and Brian Ravenet. 2022. Copalz: A computational model of pathological appraisal biases for an interactive virtual alzheimer patient. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 72–81.

[14] Karthik S Bhat, Mohit Jain, and Neha Kumar. 2021. Infrastructuring telehealth in (in) formal patient-doctor contexts. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.

[15] Nanyi Bi and Janet Yi-Ching Huang. 2023. I create, therefore I agree: Exploring the effect of AI anthropomorphism on human decision-making. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 241–244.

[16] Jonathan I Bisson, Neil P Roberts, Martin Andrew, Rosalind Cooper, and Catrin Lewis. 2013. Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane database of systematic reviews* 12 (2013).

[17] Ann Blandford, Jo Gibbs, Nikki Newhouse, Olga Perski, Aneesha Singh, and Elizabeth Murray. 2018. Seven lessons for interdisciplinary research on interactive digital health interventions. *Digital health* 4 (2018), 2055207618770325.

[18] René Börner, Jürgen Moormann, and Minhong Wang. 2012. Staff training for business process improvement: The benefit of role-plays in the case of KreditSim. *Journal of Workplace Learning* 24, 3 (2012), 200–225.

[19] Kim Bosman, Tibor Bosse, and Daniel Formolo. 2019. Virtual agents for professional social skills training: An overview of the state-of-the-art. In *Intelligent Technologies for Interactive Entertainment: 10th EAI International Conference, INTETAIN 2018, Guimarães, Portugal, November 21-23, 2018, Proceedings 10*. Springer, 75–84.

[20] Claus Bossen and Kathleen H Pine. 2023. Batman and robin in healthcare knowledge work: Human-AI collaboration by clinical documentation integrity specialists. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–29.

[21] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[22] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
[23] Virginia Braun and Victoria Clarke. 2023. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International journal of transgender health* 24, 1 (2023), 1–6.
[24] V. Braun, V. Clarke, N. Hayfield, and G Terry. 2019. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*. Springer, 843–860.
[25] Thomas Breideband, Jeffrey Bush, Chelsea Chandler, Michael Chang, Rachel Dickler, Peter Foltz, Ananya Ganesh, Rachel Lieber, William R Penuel, Jason G Reitman, et al. 2023. The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 376–380.
[26] Eleanor R Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuścińska, Suzanne Currie, J Marc Overhage, Erika S Poole, and Jofish Kaye. 2023. Healthcare AI Treatment Decision Support: Design Principles to Enhance Clinician Adoption and Trust. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
[27] Yu-Han Chang, Rajiv Maheswaran, Tomer Levinboim, and Vasudev Rajan. 2011. Learning and evaluating human-like NPC behaviors in dynamic games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 7. 8–13.
[28] Tiffany Chen, Cassandra Lee, Jessica R Mindel, Neska Elhaouij, and Rosalind Picard. 2023. Closer Worlds: Using Generative AI to Facilitate Intimate Conversations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
[29] Janghee Cho and Emilee Rader. 2020. The role of conversational grounding in supporting symbiosis between people and digital assistants. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.
[30] Frederik Roland Christiansen, Linus Nørgaard Hollensberg, Niko Bach Jensen, Kristian Julsgaard, Kristian Nyborg Jespersen, and Ivan Nikolov. 2024. Exploring Presence in Interactions with LLM-Driven NPCs: A Comparative Study of Speech Recognition and Dialogue Options. In *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology*. 1–11.
[31] C Donald Combs and P Ford Combs. 2019. Emerging roles of virtual patients in the age of AI. *AMA journal of ethics* 21, 2 (2019).
[32] Sarah C Cook, Ann C Schwartz, and Nadine J Kaslow. 2017. Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics* 14 (2017), 537–545.
[33] Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F Jung, Nicola Dell, Deborah Estrin, and James A Landay. 2024. The illusion of empathy? notes on displays of emotion in human-computer interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[34] Fabio de Almeida and Sónia Rafael. 2024. Bias by Default.: Neocolonial Visual Vocabularies in AI Image Generating Design Practices.. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.
[35] Munmun De Choudhury, Sachin R Pendse, and Neha Kumar. 2023. Benefits and harms of large language models in digital mental health. *arXiv preprint arXiv:2311.14693* (2023).
[36] Martin Johannes Dechant, Robin Welsch, Julian Frommel, and Regan L Mandryk. 2022. (Don't) stand by me: How trait psychopathy and NPC emotion influence player perceptions, verbal responses, and movement behaviours in a gaming task. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.
[37] Alice Delbosc, Magalie Ochs, Nicolas Sabouret, Brian Ravenet, and Stéphane Ayache. 2023. Towards the generation of synchronized and believable non-verbal facial behaviors of a talking virtual agent. In *Companion Publication of the 25th International Conference on Multimodal Interaction*. 228–237.
[38] Tiffany D Do, Ryan P McMahan, and Pamela J Wisniewski. 2022. A new uncanny valley? The effects of speech fidelity and human listener gender on social perceptions of a virtual-human speaker. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–11.
[39] Charlotte Dubosc, Geoffrey Gorisse, Olivier Christmann, Sylvain Fleury, Killian Poinsot, and Simon Richir. 2021. Impact of avatar facial anthropomorphism on body ownership, attractiveness and social presence in collaborative tasks in immersive virtual environments. *Computers & Graphics* 101 (2021), 82–92.
[40] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864.
[41] Edna B. Foa, Elizabeth Ann Hembree, Barbara Olasov Rothbaum, and Sheila A. M. Rauch. 2019. *Prolonged exposure therapy for PTSD: emotional processing of traumatic experiences: therapist guide* (second edition ed.). Oxford University Press, New York. OCLC: on1111296019.
[42] Yue Fu, Rebecca Michelson, Yifan Lin, Lynn K Nguyen, Tala June Tayebi, and Alexis Hiniker. 2022. Social Emotional Learning with Conversational Agents: Reviewing Current Designs and Probing Parents' Ideas for Future Ones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–23.

[43] Anthony Gelis, Sophie Cervello, Romain Rey, Guy Llorca, Philippe Lambert, Nicolas Franck, Arnaud Dupeyron, Marion Delpont, and Benjamin Rolland. 2020. Peer role-play for training communication skills in medical students: a systematic review. *Simulation in Healthcare* 15, 2 (2020), 106–111.

[44] Narayan Ghiotti, David Clulow, Serene Cheon, Kevin Cui, and Hyo Kang. 2023. Prototyping Kodi: Defining Design Requirements to Develop a Virtual Chat-bot for Autistic Children and Their Caregivers. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 126–131.

[45] Veslemøy Guise, Mary Chambers, and Maritta Välimäki. 2012. What can virtual patient simulation offer mental health nursing education? *Journal of psychiatric and mental health nursing* 19, 5 (2012), 410–418.

[46] Yuexing Hao, Zeyu Liu, Robert N Riter, and Saleh Kalantari. 2024. Advancing Patient-Centered Shared Decision-Making with AI Systems for Older Adult Cancer Patients. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[47] Donna Haraway. 2016. 'Situated Knowledges: the Science Question in Feminism and the Privilege of Partial Perspective'. In *Space, gender, knowledge: Feminist readings*. Routledge, 53–72.

[48] Christina N Harrington and Lisa Egede. 2023. Trust, Comfort and Relatability: Understanding Black Older Adults' Perceptions of Chatbot Design for Health Information Seeking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[49] Masum Hasan, Cengiz Ozel, Sammy Potter, and Ehsan Hoque. 2023. SAPIEN: affective virtual agents powered by large language models. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–3.

[50] Allyson I Hauptman, Wen Duan, and Nathan J Mcneese. 2022. The components of trust for collaborating with ai colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 72–75.

[51] Siu Loon Hoe and Tamsin Greulich-Smith. 2022. Using role play to develop an empathetic mindset in executive education. *Industrial and Commercial Training* 54, 1 (2022), 145–151.

[52] Matthias Hoppe, Beat Rossmy, Daniel Peter Neumann, Stephan Streuber, Albrecht Schmidt, and Tonja-Katrin Machulla. 2020. A human touch: Social touch increases the perceived human-likeness of agents in virtual reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–11.

[53] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: my automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 697–706. https://doi.org/10.1145/2493432.2493502

[54] Huei-Tse Hou. 2012. Analyzing the learning process of an online role-playing discussion activity. *Journal of Educational Technology & Society* 15, 1 (2012), 211–222.

[55] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. 2024. Navigating llm ethics: Advancements, challenges, and future directions. *arXiv preprint arXiv:2406.18841* (2024).

[56] Daeho Lee Jieon Lee and Jae gil Lee. 2024. Influence of Rapport and Social Presence with an AI Psychotherapy Chatbot on Users' Self-Disclosure. *International Journal of Human–Computer Interaction* 40, 7 (2024), 1620–1631. https://doi.org/10.1080/10447318.2022.2146227

[57] Patrick Kenny, Thomas D Parsons, Jonathan Gratch, Anton Leuski, and Albert A Rizzo. 2007. Virtual patients for clinical therapist skills training. In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007*. Springer, 197–210.

[58] Dajung Kim, Niko Vegt, Valentijn Visch, and Marina Bos-De Vos. 2024. How Much Decision Power Should (A) I Have?: Investigating Patients' Preferences Towards AI Autonomy in Healthcare Decision Making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[59] Hanseob Kim, Bin Han, Jieun Kim, Muhammad Firdaus Syawaludin Lubis, Gerard Jounghyun Kim, and Jae-In Hwang. 2024. Engaged and Affective Virtual Agents: Their Impact on Social Presence, Trustworthiness, and Decision-Making in the Group Discussion. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[60] Jennifer G Kim, Taewan Kim, Sung-In Kim, So-youn Jang, Eun Bin Lee, Heejung Yoo, Kyungsik Han, and Hwajung Hong. 2022. The workplace playbook VR: exploring the design space of virtual reality to foster understanding of and support for autistic people. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[61] Michal Kolomaznik, Vladimir Petrik, Michal Slama, and Vojtech Jurik. 2024. The role of socio-emotional attributes in enhancing human-AI collaboration. *Frontiers in Psychology* 15 (2024), 1369957.

[62] Hadas Kotek, Rikker Dockum, and David Sun. 2023. Gender bias and stereotypes in large language models. In *Proceedings of the ACM collective intelligence conference*. 12–24.

[63] Claire Lane and Stephen Rollnick. 2007. The use of simulated patients and role-play in communication skills training: a review of the literature to August 2005. *Patient education and counseling* 67, 1-2 (2007), 13–20.

[64] Matthew J Leach. 2005. Rapport: A key to treatment success. *Complementary therapies in clinical practice* 11, 4 (2005), 262–265.

[65] Tae Kyoung Lee, Kevin Crowston, Carsten Østerlund, and Grant Miller. 2017. Recruiting messages matter: message strategies to attract citizen scientists. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 227–230.

[66] Michael Leyer, Ann-Kathrin Hirzel, and Juergen Moormann. 2018. Achieving sustainable behavioral changes of daily work practices: The effect of role plays on learning process-oriented behavior. *Business Process Management Journal* 24, 4 (2018), 1050–1068.

[67] Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2021. Towards understanding and mitigating social biases in language models. In *International Conference on Machine Learning*. PMLR, 6565–6576.

[68] Mengqi Liao and S Shyam Sundar. 2021. How should AI systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-AI interaction. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.

[69] Q Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.

[70] Zihan Liu, Han Li, Anfan Chen, Renwen Zhang, and Yi-Chieh Lee. 2024. Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. *arXiv preprint arXiv:2402.16039* (2024).

[71] Poornima Madhavan and Douglas A Wiegmann. 2007. Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science* 8, 4 (2007), 277–301.

[72] Sarah Mandl, Maria Kobert, Maximilian Bretschneider, Frank Asbrock, Bertolt Meyer, Anja Strobel, and Thomas Süße. 2023. Exploring key categories of social perception and moral responsibility of AI-based agents at work: Findings from a case study in an industrial setting. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[73] Rakesh K Maurya. 2024. A qualitative content analysis of ChatGPT's client simulation role-play for practising counselling skills. *Counselling and Psychotherapy Research* 24, 2 (2024), 614–630.

[74] Jingbo Meng, Minjin Rheu, Yue Zhang, Yue Dai, and Wei Peng. 2023. Mediated social support for distress reduction: AI Chatbots vs. Human. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–25.

[75] Stephanie Milani, Arthur Juliani, Ida Momennejad, Raluca Georgescu, Jaroslaw Rzepecki, Alison Shaw, Gavin Costello, Fei Fang, Sam Devlin, and Katja Hofmann. 2023. Navigates like me: Understanding how people evaluate human-like AI in video games. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[76] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.

[77] Sharon Mozgai, Albert Rizzo, and Arno Hartholt. 2023. Human-Centered Design for a Virtual Human led mHealth Intervention for Suicide Prevention. *Human Systems Engineering and Design (IHSED 2023): Future Trends and Applications* 112, 112 (2023).

[78] Hendrik Müller, Aaron Sedley, and Elizabeth Ferrall-Nunge. 2014. Survey Research in HCI. In *Ways of Knowing in HCI*. Springer.

[79] Moin Nadeem, Anna Bethke, and Siva Reddy. 2020. StereoSet: Measuring stereotypical bias in pretrained language models. *arXiv preprint arXiv:2004.09456* (2020).

[80] Taichi Nakamura, Un Kai, and Yuki Tachikawa. 2014. Requirements engineering education using expert system and role-play training. In *2014 IEEE international conference on teaching, assessment and learning for engineering (TALE)*. IEEE, 375–382.

[81] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[82] Chinasa T Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It cannot do all of my work": community health worker perceptions of AI-enabled mobile health applications in rural India. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–20.

[83] Teresa K O'Leary, Michael K Paasche-Orlow, and Timothy Bickmore. 2024. Something I Can Lean On': A Qualitative Evaluation of a Virtual Palliative Care Counselor for Patients with Life-Limiting Illnesses. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[84] Katherine E Ottman, Brandon A Kohrt, Gloria A Pedersen, and Alison Schafer. 2020. Use of role plays to assess therapist competency and its association with client outcomes in psychological interventions: A scoping review and competency research agenda. *Behaviour Research and Therapy* 130 (2020), 103531.

[85] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[86] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. 2019. Identifying challenges and opportunities in human-AI collaboration in healthcare. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*. 506–510.

[87] Sachin R Pendse, Daniel Nkemelu, Nicola J Bidwell, Sushrut Jadhav, Soumitra Pathare, Munmun De Choudhury, and Neha Kumar. 2022. From treatment to healing: envisioning a decolonial digital mental health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.

[88] Daniel Pillis, Pat Pataranutaporn, Pattie Maes, and Misha Sra. 2024. AI Comes Out of the Closet: Using AI-Generated Virtual Characters to Help Individuals Practice LGBTQIA+ Advocacy. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 686–698.

[89] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–21.

[90] Sheila A Rauch, Afsoon Eftekhari, and Josef I Ruzek. 2012. Review of exposure therapy: a gold standard for PTSD treatment. *J Rehabil Res Dev* 49, 5 (2012), 679–687.

[91] Jungwook Rhim, Minji Kwak, Yeaeun Gong, and Gahgene Gweon. 2022. Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior* 126 (2022), 107034.

[92] Albert Rizzo, Patrick Kenny, and Thomas D Parsons. 2011. Intelligent virtual patients for training clinical skills. *JVRB-Journal of Virtual Reality and Broadcasting* 8, 3 (2011).

[93] Albert Rizzo, Scherer Scherer, David DeVault, Jonathan Gratch, Ronald Artstein, Arno Hartholt, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, et al. 2016. Detection and computational analysis of psychological signals using a virtual human interviewing agent. *Journal of Pain Management* 9, 3 (2016), 311–321.

[94] Albert A Rizzo, Gale M Lucas, Jonathan Gratch, Giota Stratou, Louis-Philippe Morency, Kenneth Chavez, Russ Shilling, and Stefan Scherer. 2016. Automatic Behavior Analysis During a Clinical Interview with a Virtual Human.. In *MMVR*. 316–322.

[95] Solrun Brenk Rønning and Stål Bjørkly. 2019. The use of clinical role-play and reflection in learning therapeutic communication skills in mental health education: an integrative review. *Advances in medical education and practice* (2019), 415–425.

[96] Barbara Olasov Rothbaum and Ann C Schwartz. 2002. Exposure therapy for posttraumatic stress disorder. *American journal of psychotherapy* 56, 1 (2002), 59–75.

[97] Ofir Sadka and Alissa Antle. 2020. Interactive technologies for emotion-regulation training: Opportunities and challenges. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–12.

[98] Pedro Sanches, Axel Janson, Pavel Karpashevich, Camille Nadal, Chengcheng Qu, Claudia Daudén Roquet, Muhammad Umair, Charles Windlin, Gavin Doherty, Kristina Höök, et al. 2019. HCI and Affective Health: Taking stock of a decade of studies and charting future research directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–17.

[99] Shruti Sannon, Brett Stoll, Dominic DiFranzo, Malte Jung, and Natalya N Bazarova. 2018. How personification and interactivity influence stress-related disclosures to conversational agents. In *companion of the 2018 ACM conference on computer supported cooperative work and social computing*. 285–288.

[100] Ava Elizabeth Scott, Daniel Neumann, Jasmin Niess, and Paweł W Woźniak. 2023. Do you mind? User perceptions of machine consciousness. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–19.

[101] Katie Seaborn, Peter Pennefather, Norihisa Miyake, and Mihoko Otake-Matsuura. 2021. Crossing the Tepper line: An emerging ontology for describing the dynamic sociality of embodied AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

[102] Anna-Maria Seeger and Armin Heinzl. 2018. Human versus machine: Contingency factors of anthropomorphism as a trust-inducing design strategy for conversational agents. In *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017*. Springer, 129–139.

[103] Irving Seidman. 2006. *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers college press.

[104] Mark Sendak, Madeleine Clare Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. 2020. " The human body is a black box" supporting clinical decision-making with deep learning. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 99–109.

[105] William Seymour and Max Van Kleek. 2021. Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–16.

[106] Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* 623, 7987 (2023), 493–498.

[107] Shefaly Shorey, Emily NK Ang, Esperanza Debby Ng, John Yap, Lydia ST Lau, Chee Kong Chui, and Yiong Huak Chan. 2023. Evaluation of a theory-based virtual counseling application in nursing education. *CIN: Computers, Informatics, Nursing* (2023), 10–1097.

[108] Petr Slovák, Nikki Theofanopoulou, Alessia Cecchet, Peter Cottrell, Ferran Altarriba Bertran, Ella Dagan, Julian Childs, and Katherine Isbister. 2018. " I just let him cry... Designing Socio-Technical Interventions in Families to Prevent Mental Health Disorders. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–34.

[109] Petr Slovák, Anja Thieme, David Murphy, Paul Tennent, Patrick Olivier, and Geraldine Fitzpatrick. 2015. On becoming a counsellor: Challenges and opportunities to support interpersonal skills training. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1336–1347.

[110] Everett S Spain, Yasmine L Konheim-Kalkstein, Ryan G Erbe, Corrine N Wilsey, and Stacey F Rosenberg. 2023. We Don't Do that Here: Using Role Playing and Character Battle Drills to Develop Upstander Behavior at West Point. *Journal of Character and Leadership Development* 10, 2 (2023), 1–13.

[111] Frank W Stahnisch and Marja Verhoef. 2012. The flexner report of 1910 and its impact on complementary and alternative medicine and psychiatry in north america in the 20th century. *Evidence-Based Complementary and Alternative Medicine* 2012, 1 (2012), 647896.

[112] Jan-Philipp Stein and Peter Ohler. 2017. Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160 (2017), 43–50.

[113] Yvonne Steinert. 1993. Twelve tips for using role-plays in clinical teaching. *Medical Teacher* 15, 4 (1993), 283–291.

[114] Claire Jane Stewart and Aiesha Ba Mashmous. 2023. Training for psychiatric assessments using virtual simulation. *The Journal of Mental Health Training, Education and Practice* 18, 3 (2023), 261–273.

[115] Isabel Straw. 2020. The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future. *Artificial intelligence in medicine* 110 (2020), 101965.

[116] Thomas Talbot and Albert "Skip" Rizzo. 2019. Virtual human standardized patients for clinical training. *Virtual reality for psychological and neurocognitive interventions* (2019), 387–405.

[117] Marcus Cheng Chye Tan, Stefanie Yen Leng Chye, and Kylin Shu Min Teng. 2022. "In the shoes of another": immersive technology for social and emotional learning. *Education and Information Technologies* 27, 6 (2022), 8165–8188.

[118] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2017. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PloS one* 12, 8 (2017), e0182151.

[119] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.

[120] Vincent B Van Hasselt, Stephen J Romano, and Gregory M Vecchi. 2008. Role playing: Applications in hostage and crisis negotiation skills training. *Behavior modification* 32, 2 (2008), 248–263.

[121] Gary R VandenBos. 2007. *APA dictionary of psychology.* American Psychological Association.

[122] Himanshu Verma, Jakub Mlynar, Roger Schaer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquoz, and Adrien Depeursinge. 2023. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[123] Janet Vertesi. 2014. Seamful spaces: Heterogeneous infrastructures in interaction. *Science, Technology, & Human Values* 39, 2 (2014), 264–284.

[124] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI doctor" in rural clinics: challenges in AI-powered clinical decision support system deployment. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.

[125] Isaac Wang and Jaime Ruiz. 2021. Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction* 37, 17 (2021), 1648–1673.

[126] Zhiyuan Wang, Varun Reddy, Karen Ingersoll, Tabor Flickinger, and Laura E Barnes. 2024. Rapport Matters: Enhancing HIV mHealth Communication through Linguistic Analysis and Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[127] Lauren Wilcox, Robin Brewer, and Fernando Diaz. 2023. AI Consent Futures: A Case Study on Voice Data Collection with Clinicians. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30.

[128] Jedrek Wosik, Marat Fudim, Blake Cameron, Ziad F Gellad, Alex Cho, Donna Phinney, Simon Curtis, Matthew Roman, Eric G Poon, Jeffrey Ferranti, et al. 2020. Telehealth transformation: COVID-19 and the rise of virtual care. *Journal of the American Medical Informatics Association* 27, 6 (2020), 957–962.

[129] Qin Wu. 2024. Designing Interactive Technology to Support Children with Autism. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[130] Zihan Yan, Yaohong Xiang, and Yun Huang. 2024. Social Life Simulation for Non-Cognitive Skills Learning. *arXiv preprint arXiv:2405.00273* (2024).

[131] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[132] Michael Yin and Robert Xiao. 2024. Press A or Wave: User Expectations for NPC Interactions and Nonverbal Behaviour in Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 8, CHI PLAY (2024), 1–25.

[133] Dong Whi Yoo, Hayoung Woo, Sachin R Pendse, Nathaniel Young Lu, Michael L Birnbaum, Gregory D Abowd, and Munmun De Choudhury. 2024. Missed Opportunities for Human-Centered AI Research: Understanding Stakeholder Collaboration in Mental Health AI Research. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–24.

[134] Christopher You, Rashi Ghosh, Andrew Maxim, Jacob Stuart, Eric Cooks, and Benjamin Lok. 2022. How does a virtual human earn your trust? guidelines to improve willingness to self-disclose to intelligent virtual agents. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. 1–8.

[135] Nima Zargham, Dmitry Alexandrovsky, Thomas Mildner, Robert Porzel, and Rainer Malaka. 2023. " Let's Face It": Investigating User Preferences for Virtual Humanoid Home Assistants. In *Proceedings of the 11th International Conference on Human-Agent Interaction*. 246–256.

[136] Xiao Zhan, Noura Abdi, William Seymour, and Jose Such. 2024. Healthcare Voice AI Assistants: Factors Influencing Trust and Intention to Use. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–37.

[137] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. "An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

[138] Shuo Zhou and Timothy Bickmore. 2022. A virtual counselor for breast cancer genetic counseling: adaptive pedagogy leads to greater knowledge gain. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

Received July 2024; revised December 2024; accepted March 2025