

## Credibility and Appropriateness in AI-Based Simulation as Empathy Training Tools (Short paper)

**Ronen Grinberg**  
Technion – Israel Institute of  
Technology  
[ronen.g@campus.technion.ac.il](mailto:ronen.g@campus.technion.ac.il)

**Dan Waisman**  
Technion – Israel Institute of  
Technology  
[dwaisman@technion.ac.il](mailto:dwaisman@technion.ac.il)

**Rinat B. Rosenberg-Kima**  
Technion – Israel Institute of  
Technology  
[rinatros@technion.ac.il](mailto:rinatros@technion.ac.il)

### אמינות והתאמת סימולציה מבוססת בינה מלאכותית כבלי להכשרת תקשורת אמפתית (מאמר קצר)

**רינת ב. רוזנברג-קימה**  
הטכניון – מכון טכנולוגי לישראל  
[rinatros@technion.ac.il](mailto:rinatros@technion.ac.il)

**דן וייסמן**  
הטכניון – מכון טכנולוגי לישראל  
[dwaisman@technion.ac.il](mailto:dwaisman@technion.ac.il)

**רון גרינברג**  
הטכניון – מכון טכנולוגי לישראל  
[ronen.g@campus.technion.ac.il](mailto:ronen.g@campus.technion.ac.il)

#### Abstract

Empathic communication is a central competency for educational rehabilitation counselors, yet existing training methods often struggle to provide the socially grounded practice needed for developing such skills. Advances in generative artificial intelligence offer new opportunities for scalable simulation-based training. Guided by Vygotsky's Sociocultural Theory and the Media Equation and Computers Are Social Actors (CASA) frameworks, this study examines a chatbot-based system designed to support empathy training through a Virtual Participant (VP) simulating a young adult with a disability. Using a qualitative case study approach, eight senior social workers from a rehabilitation organization evaluated two VP-practitioner interaction transcripts. Findings regarding the VP revealed six credibility challenges: verbal credibility, emotional and behavioral credibility, credibility in expressing resistance and initiative, inconsistencies between verbal utterances and described behavior, absence of physical presence, and absence of shared relational history. These issues limited the VP's perceived authenticity and reduced its effectiveness as a simulated partner for practicing empathic dialogue. Overall, the findings highlight both the potential and limitations of AI-mediated empathy training. While generative agents hold great potential for practice, meaningful development of empathic communication appears to require human facilitation to contextualize, refine, and carefully prompt AI-generated interactions.

**Keywords:** Simulation-based learning, Simulated Participant, the Media Equation, GenAI, AIED.

המחברים רוצים לציין באופן מיוחד את רונן גרינברג ז"ל שעבד באופן מסור על הנושא הנחקר במסגרת הדוקטורט שלו על הכשרה לשיח אמפתי בעזרת AI של סטודנטים לרפואה בטכניון ובמסגרת עמותת כוונים. רונן היה איש חינוך אהוב ודגול, בעבודתו ניהל את מחלקת מחקר, פיתוח והדרכה בעמותת כוונים, המכשירה צעירים עם מגבלה לחיים עצמאיים בקהילה, היה אדם כותב ויצירתי, איש משפחה וחברים. רונן האמין ביכולתה של הבינה לעזור לבני האדם להיות אנושיים יותר. רונן נהרג בתאונת דרכים ב-18/11/25.

## Introduction

In modern rehabilitation settings, educational rehabilitation counselors play a critical role in fostering the personal, emotional, and social development of individuals with disabilities going through complex transitions (Hartley & Tarvydas, 2022). Because of their direct interpersonal support for vulnerable individuals, the ability to communicate with empathy becomes a crucial professional skill rather than simply a desirable personal trait.

Advances in educational technology, especially the integration of AI and conversational agents, have created new opportunities for communication training in social professions. AI-powered chatbots have been successfully used to support learning across various fields, offering learners safe, repetitive practice environments with immediate, personalized feedback (Liaw et al., 2023). These systems can simulate emotionally detailed interactions, making them particularly effective for fostering empathy in professionals (Levkovich et al., 2025; Pira et al., 2025).

Vygotsky's Sociocultural Theory (Vygotsky, 1978) suggests that higher cognitive functions, such as professional communication, develop through social interaction with a More Knowledgeable Other (MKO). But can a digital agent serve as a social entity? the Media Equation Theory posits that people automatically and unconsciously treat computers, media, and artificial entities as if they were real social beings (Reeves & Nass, 1996) and the Computers Are Social Actors (CASA) framework (Nass & Moon, 2000) further explains that users apply social norms, expectations, and interpersonal scripts to computers.

Rooted in this theoretical framework, we introduce a chatbot-based system designed to train educational rehabilitation counselors in empathetic communication. This research aims to evaluate the quality and limitations of the VP as perceived by eight senior social workers. The following question guided this qualitative evaluation: How do social workers evaluate the credibility and appropriateness of the Virtual Participant (VP)?

## Method

### Participants

Eight senior social workers from Kivunim (<https://www.kvn.org.il/>), a nonprofit organization that runs preparatory programs for independent living, participated in this study. The social workers were the system's main users and had already worked with it in practice. The participants signed a consent form approved by the university ethics committee.

### Materials

The study used the Cesura.ai platform (<https://cesura.ai/>). The system's conversational and clinical reasoning capabilities were powered by Google's Gemini 2.5 Pro large language model. Two senior social workers, one specializing in cerebral palsy (CP) and the other in autism spectrum disorder (ASD), answered the agent-builder's prompts and supplied example dialogue to define the personas for two distinct VPs: a young adult with CP and a young adult with ASD. During the interaction, the

VP delivers its verbal responses as text, accompanied by a description of its non-verbal behaviors (e.g., position, tone of voice, eye movements) enclosed in brackets. For example: "Fine, as usual, nothing special happened. [Lifts her head slightly, glances for a moment toward the counselor, and then returns to her phone]".

### Data Collection and Analysis

Data were collected in a single one-hour focus group with eight senior social workers from Kivunim, seven of them had gained hands-on experience with the tool prior to the session. The focus group followed a multi-stage process. The social workers received one of two transcripts of interactions between a VP and a practitioner and were asked to assess individually the VP's credibility. A full-group discussion on the credibility and appropriateness of *both* VPs followed. The group discussion was recorded and transcribed verbatim for analysis, and the transcripts annotated with the social workers' individual notes were also collected. The data were analyzed using thematic analysis (Braun & Clarke, 2006). The process included close reading of the transcripts, identifying and grouping initial codes, and developing themes that reflected recurring insights.

### Findings

Five themes emerged from the thematic analysis with respect to the research question ("How do social workers evaluate the credibility of the Virtual Participant (VP)?") (see **Table 1**): (1) *Verbal credibility* - Some VPs' responses were judged by the social workers as incongruent with expected linguistic and cognitive profiles, described as "*too complicated*" or overly articulated, while others were perceived as vague or stereotypical, (2) *Emotional and behavioral credibility* - Emotional and behavioral cues were also experienced by the social workers as exaggerated or artificial, (3) *Credibility in representing resistance and initiative* - The participants noted that the portrayal of resistance often appeared linguistically rigid rather than behaviorally grounded, (4) *Inconsistent verbal utterances and described behavior* - A further source of difficulty was the inconsistency between the VP's verbal utterances and the accompanying behavioral descriptions, (5) *Absence of physical presence* - A recurring theme was the absence of physical presence, which they described as a central aspect of their work with young adults. In their professional practice, subtle embodied cues provide crucial information about a participant's emotional state, and (6) *Absence of shared history* - The social workers also emphasized the difficulties posed by the VP's lack of personal history. In practice, they build a deep familiarity with each participant's personal history, routines, and idiosyncrasies, knowledge that grounds their interpretations and guides empathetic responses. The VP, however, could not "remember" previous encounters, refer to earlier conversations, or reflect a personalized daily schedule. Without this continuity, the VP felt detached from the lived context of the preparatory programs and therefore less authentic as a relational partner.

**Table 1.** Main themes concerning the credibility and appropriateness of the Virtual Participant (VP)

Theme	Example Quotes
1. Verbal and Content Credibility	<ul style="list-style-type: none"> <li>• "The character's responses are too general."</li> <li>• "It's also like an autistic trait that isn't always true, that people think they only respond to their areas of interest."</li> <li>• "During the conversation, the character always stays on topic, doesn't lose focus and scatter to other topics like in reality."</li> </ul>
2. Emotional and Behavioral Credibility	<ul style="list-style-type: none"> <li>• "The really extreme reactions stood out. The direct gaze, sort of piercing and all, like you imagine it, but it doesn't really happen that way."</li> <li>• "I felt it was too extreme, like very, very extreme reactions."</li> <li>• "That binary that was there, like, between softening and rigid, softening and rigid, there was something too dichotomous about it."</li> </ul>
3. Credibility in representing resistance and initiative	<ul style="list-style-type: none"> <li>• "I don't need any incentive to go to the Lego store. I go there when I want to" – marked as a non-credible sentence.</li> </ul>
4. Inconsistent verbal utterance and described behavior	<ul style="list-style-type: none"> <li>• "The character says: 'Just. It doesn't interest me' and it's described as: [Shifts gaze back to the phone, ignores the question], the interpretation that the character ignoring the question is incorrect."</li> <li>• "The resistance here is very linguistic."</li> </ul>
5. Absence of physical presence	<ul style="list-style-type: none"> <li>• "You can, really, without listening, just by looking at him, you know he's anxious right now, you know he's tense, you know he's getting annoyed with you, but you know he's with you right now."</li> <li>• "It feels to me that the difficulty in the conversation is sitting with the person in the room and feeling him."</li> </ul>
6. Absence of shared history	<ul style="list-style-type: none"> <li>• "You learn each person, you know that if he folds his fingers, he's stressed, and this is how he talks to you when he's calm."</li> <li>• "Is there no option to create a personal history in the prompt? Like an ongoing conversation... We're talking about the personal plan, and he remembers."</li> <li>• "Maybe [we need] to add another distinction between a participant who is new, just arrived, just got settled, and a participant who has been in the program for a year, a year and a half"</li> </ul>

## Discussion

The findings point to a mixed picture: on the one hand, participants experienced the system as a structured, engaging space for practicing empathic dialogue; on the other hand, they identified clear limitations regarding the credibility and appropriateness of the VP.

From the perspective of the Media Equation and CASA frameworks, the VP was expected to function as a believable social actor that could support sociocultural learning processes (Nass & Moon, 2000; Reeves & Nass, 1996). However, the social workers' critiques indicate only a partial realization of this promise. Rather than a specific, coherent individual, the VP often appeared as a statistical "average". This gap is consistent with concerns raised by Kirkeby-Hinrup & Stenseke (2025), who describe how LLM-based language can approximate human expression while still leaving users with a sense of unease and uncertainty about the "mind" behind the text.

While the VP can generate rich material for practice, the social workers saw the human trainer as the primary MKO who must contextualize AI outputs, restore nuance, and transform transcripts into meaningful learning experiences. This suggests that, in line with sociocultural perspectives, AI in this domain is best positioned as a mediating tool within a hybrid pedagogical model.

While this study highlighted the limitations of the current simulations, some of these limitations may be addressed through more rigorous prompt engineering. A key question moving forward is which limitations are amenable to improved prompting, and which are intrinsic to present-day LLMs' architecture.

## References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Gómez-León, M. I. (2022). Development of empathy through social emotional artificial intelligence. *Papeles Del Psicólogo, 43*(3), 218–224.
- Hartley, M. T., & Tarvydas, V. M. (2022). *The professional practice of rehabilitation counseling*. Springer Publishing Company.
- Kirkeby-Hinrup, A., & Stenseke, J. (2025). The psychology of LLM interactions: The uncanny valley and other minds. *Journal of Psychology and AI, 1*(1), 2457627.
- Komasawa, N., & Yokohira, M. (2023). Simulation-based education in the artificial intelligence era. *Cureus, 15*(6).
- Levkovich, I., Haber, Y., Levi-Belz, Y., & Elyoseph, Z. (2025). A step toward the future? Evaluating GenAI QPR simulation training for mental health gatekeepers. *Frontiers in Medicine, 12*, 1599900.
- Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., & Chua, W. L. (2023). Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study. *Nurse Education Today, 122*, 105718.
- Pira, G. L., Ruini, C., Vescovelli, F., Baños, R., & Ventura, S. (2025). Could Empathy Be Taught? The Role of Advanced Technologies to Foster Empathy in Medical Students and Healthcare Professionals: A Systematic Review. *Journal of Medical Systems, 49*(1), 6. <https://doi.org/10.1007/s10916-025-02144-9>
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK, 10*(10), 19-36.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.
- Zeng, Y., Brown, C., & Rounsevell, M. (2025). Too Human to Model: The Uncanny Valley of LLMs in Social Simulation—When Generative Language Agents Misalign with Modelling Principles. *arXiv Preprint arXiv:2507.06310*.