

Towards scalable psychological first aid training: an autoethnographic exploration of the potential of large language models in simulation-based learning

Rodrigo A. Figueroa ^{a,b}, Margarita Bernales ^c, Chris M. Hoeboer ^b and Miranda Olf ^{b,d}

^aDepartment of Psychiatry, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; ^bDepartment of Psychiatry, Amsterdam Neuroscience, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; ^cSchool of Nursing, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; ^dARQ National Psychotrauma Centre, Diemen, The Netherlands

ABSTRACT

Background: Psychological First Aid (PFA) is a widely used approach to provide psychosocial support to individuals in the aftermath of a stressful event. PFA training requires a costly, structured approach, including repetitive simulation-based training with actors. The rapid emergence of large language models (LLMs) could provide easy access to PFA simulation training, enabling trainees to practice conversations repeatedly and flexibly, learn from mistakes, and build emotional resilience. In addition, this would allow for training in geographically isolated regions where PFA training may be most needed.

Objective: To explore the potential of LLMs for PFA simulation training.

Methods: Using a digital autoethnography approach, we conducted three role plays with ChatGPT-3.5-turbo, an LLM-based chatbot, in scenarios addressing challenges typical of PFA training. In each role-play, the chatbot simulated an individual in the aftermath of a stressful situation, and a researcher took on the roles of both a trainer configuring a simulation session and a learner serving as a PFA provider. The chatbot's ability to respond, adapt, and facilitate the training process was evaluated using reflexive thematic analysis.

Results: Our study showed that ChatGPT effectively portrayed realistic characters, collaborated effectively with the trainer, and provided feedback to the learner. However, we also noted challenges concerning overly stereotyped characterizations, identity confusion, and following the trainer's instructions.

Conclusions: While utilizing LLMs for PFA simulation training demonstrates significant promise, achieving their full educational potential requires careful refinement, thorough empirical validation, and ongoing ethical oversight.

Hacia el entrenamiento escalable de primeros auxilios psicológicos: una exploración autoetnográfica del potencial de los modelos de lenguaje de gran escala en el aprendizaje basado en simulación

Antecedentes: Los Primeros Auxilios Psicológicos (PAP) son un tipo de apoyo psicosocial temprano ampliamente utilizado para personas afectadas por eventos estresantes recientes. El entrenamiento de PAP requiere un enfoque estructurado y costoso que incluye simulaciones con actores. Los modelos de lenguaje de gran escala (LLM, del inglés) podrían facilitar el acceso al entrenamiento simulado de PAP, permitiendo interacciones repetidas y flexibles, el aprendizaje de errores y el desarrollo de resiliencia emocional. Además, esto permitiría ofrecer formación en regiones aisladas, donde el entrenamiento en PAP podría ser más necesario.

Objetivo: Explorar el potencial de los LLM para el entrenamiento simulado de PAP.

Métodos: Desde un enfoque autoetnográfico digital, se llevaron a cabo tres juegos de rol con ChatGPT-3.5-turbo, un chatbot basado en LLM, utilizando escenarios que abordaban desafíos típicos de entrenamiento en PAP. En cada simulación, el chatbot interpretó a una persona afectada por una situación estresante reciente, mientras que un investigador asumió los roles de entrenador (configurando la sesión) y de alumno en el rol de proveedor de PAP. La capacidad del chatbot para responder, adaptarse y facilitar el proceso de entrenamiento fue evaluada mediante análisis temático reflexivo.

Resultados: Nuestro estudio evidenció que ChatGPT fue capaz de representar personajes realistas, colaborar eficazmente con el entrenador y entregar retroalimentación al alumno. Sin embargo, también se identificaron desafíos relacionados con caracterizaciones excesivamente estereotipadas, confusión de identidad y dificultades para seguir instrucciones del entrenador.

Conclusiones: Aunque el uso de LLMs en el entrenamiento simulado de PAP muestra un potencial educativo considerable, su aprovechamiento pleno requiere una refinación cuidadosa, una validación empírica rigurosa y una supervisión ética continua.

ARTICLE HISTORY

Received 11 April 2025
Revised 8 January 2026
Accepted 9 January 2026

KEYWORDS



Artificial intelligence;
psychological first aid; PTSD;
traumatic stress; simulation

PALABRAS CLAVE

Inteligencia artificial;
primeros auxilios
psicológicos; PTSD; estrés
traumático; simulación

HIGHLIGHTS

- ChatGPT could simulate individuals in the aftermath of stressful situations.
- Through autoethnography and thematic analysis, we explored the potential of ChatGPT for psychological first aid simulation training.
- ChatGPT has potential, but it still needs human supervision.

CONTACT Rodrigo A. Figueroa  rfiguerc@uc.cl  Department of Psychiatry, Faculty of Medicine, Pontificia Universidad Católica de Chile, Diagonal Paraguay 362, Santiago 8330077, Chile

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

Psychological First Aid (PFA) is a widely used approach to provide psychosocial support to individuals in the immediate aftermath of a stressful event (Hermosilla et al., 2023). It includes various components, such as non-intrusive practical care and support, assessing basic needs and concerns, providing comfort and active listening, and helping individuals connect to information, services, and social support (Ni et al., 2023). PFA was designed to be used by first responders and laypeople without prior mental health or clinical training. It is not a clinical or psychotherapeutic intervention, but rather a form of practical, psychosocial support intended as humane, supportive assistance for people in distress, and does not involve diagnosis or formal treatment (Ni et al., 2024; Wang, Zhao, et al., 2025).

Numerous models and protocols for Psychological First Aid (PFA) exist, yet no standardized protocol has been established (Hermosilla et al., 2023; Ni et al., 2023; Wang et al., 2024). Despite the varied organizations and labelling of their components, these models collectively aim to enhance five empirically substantiated principles recognized as ‘Hobfoll’s essential elements’: safety, calmness, self-efficacy, social connection, and hope. PFA facilitates the promotion of these elements through mechanisms such as active listening, relaxation techniques, problem-solving strategies, and the fostering of social connections (Wang et al., 2024). The existing literature on trauma and disaster recovery, along with expert consensus, underscores the need to integrate Hobfoll’s essential elements into early interventions following recent traumatic events (Hobfoll et al., 2007).

One of the most widely used PFA models is the one by the World Health Organization, War Trauma Foundation, and World Vision International, which has been translated into 20 languages (Hermosilla et al., 2023; World Health Organization et al., 2011). It outlines three sequential action principles for crisis response: ‘Look’ involves assessing the situation for safety risks, identifying people with urgent needs or acute distress, and recognizing vulnerable individuals while ensuring your own safety first. ‘Listen’ focuses on approaching affected individuals supportively and non-intrusively, asking about their needs and concerns, listening actively to help them feel calm and understood, and respecting their choice whether to talk or remain silent; and ‘Link’ emphasizes connecting people to practical support by helping address immediate basic needs, facilitating access to services and resources, reuniting individuals with family and social support networks, providing accurate information about safety and available help, and supporting their access to ongoing recovery resources. Together, these principles form a comprehensive framework

that guides helpers through observing needs, engaging empathetically, and connecting individuals to appropriate support systems during crises.

PFA offers a structured yet flexible approach for situations in which many people may be affected simultaneously, such as natural disasters, terrorist attacks, and public health emergencies (Bisson et al., 2010). Although only a minority of individuals affected by highly distressing events develop stress-related disorders, and an even smaller proportion requires specialist care (Kessler et al., 2017), PFA is still recommended, as it has shown a positive effect on reducing anxiety and promoting adaptive functioning (Wang et al., 2024). While PFA does not serve as a substitute for more comprehensive psychological interventions, its importance lies in its capacity to reduce initial post-traumatic distress and promote adaptive functioning.

Despite its widespread use, there are few rigorous studies demonstrating PFA’s effectiveness or safety. Definitions, delivery methods, and target populations vary widely, making it hard to determine what ‘PFA’ entails or which components are essential (Wang et al., 2024). Most studies rely on weak designs, small sample sizes, brief follow-ups, and self-reported outcomes, thereby increasing the risk of bias. Overall, a striking gap remains between the global implementation of PFA and the limited, low-quality evidence supporting its effectiveness (Hermosilla et al., 2023). In a recent randomized controlled trial (RCT) by our group, individuals affected by recent traumatic events who received PFA reported greater immediate relief from subjective distress and fewer early PTSD symptoms than those in an active control condition (Figueroa et al., 2022). In a subsequent RCT, the proportion of participants reporting early post-trauma increases in alcohol or substance use, interpersonal conflict, psychotropic medication use, or sick leave was lower in the PFA group than in the control group (Figueroa et al., 2024).

PFA training extends beyond theoretical knowledge and includes the development of non-technical skills such as active listening, calming techniques, empathetic engagement, and immediate needs assessment (Horn et al., 2019). These competencies are best cultivated through deliberate practice, a structured training methodology that enhances helping skills acquisition and retention through repetitive, goal-oriented, simulation-based exercises with immediate feedback (Duvivier et al., 2011; Kang & Choi, 2021). This approach often involves simulated individuals – trained actors who portray realistic crisis scenarios – to provide learners with hands-on experience in a controlled environment (Happ et al., 2025), without risking harm to people affected by real-life events (Elendu et al., 2024).

Unfortunately, simulation training is limited by high costs and logistical challenges (Mills et al., 2020). Actor-based simulations face significant operational costs (i.e. simulation design, actor recruitment, training, rehearsals, compensation) (Kelly & Murphy, 2004), limited availability of actors in remote locations, and scheduling constraints (i.e. coordinating actors, students, and educators' available times, sometimes in different time zones) (Jiang et al., 2023). These limitations hinder the quality of PFA training and its scalability. Interestingly, lay providers have expressed concerns about the effectiveness of current training programmes (Horn et al., 2019; Wang et al., 2024), thereby limiting community preparedness in a world increasingly exposed to mass trauma from climate change and conflicts (Olf, 2023). Regrettably, scalable options, such as traditional virtual patients, often cannot capture subtle and subjective variables, such as context-relevant responses and real-time feedback (Dupuy et al., 2021), and online training modules have not been shown to instil the competencies needed for adequate PFA provision (Horn et al., 2019).

Large Language Models' (LLMs) recent rapid emergence could address limited access to PFA simulation training opportunities. LLMs are Artificial Intelligence systems that generate text and learn to understand word patterns and structures in their training data, producing new data with similar characteristics (Sengar et al., 2025). Given the noteworthy capabilities of LLMs to generate natural language interactions that are accountable to context, and to provide personalized feedback on support provision performance, it is now possible to build virtual people affected by distressing events with unprecedented capacities to portray realistic individuals (Cook, 2025; Cook et al., 2025). These models can emulate diverse personality traits and preferences, which is crucial for depicting realistic behaviours of individuals in the aftermath of stressful events. Albeit not pertaining to the PFA realms, the use of LLMs in medical training has shown that they can effectively simulate standardized patients, offering a cost-effective alternative to traditional methods (Wang, Li, et al., 2025). Virtual humans that incorporate affective computing techniques can be tailored to exhibit specific psychological and behavioural traits, making them suitable for simulating individuals in the immediate aftermath of a stressful event (Loveys et al., 2023). For example, a recent qualitative content analysis study highlighted ChatGPT's usefulness in simulating psychiatrist-client interactions to enhance counselling skills practice (Maurya, 2023; Maurya & Cavanaugh, 2023). Another study described the use of virtual patients to simulate patient interactions with anesthesia trainees, allowing them to respond to specific inquiries about specific procedures. The virtual patient was intuitive and

user-friendly, with high scores for usability and accurately simulating patient responses and behaviour. Most users (87%) reported feeling comfortable using the model. However, the study also highlighted limitations, including the risk of generating inaccurate information, underscoring the need for further refinement (Sardesai et al., 2024).

Overall, LLMs offer a scalable and accessible approach to creating virtual people affected by distressing events with unprecedented capacities for realistic behaviour simulation, including those of individuals in the immediate aftermath of a stressful event. This innovation promises to enhance simulation training and patient care through more affordable, accessible, immersive, and contextually accurate simulations. Hence, this study aimed to explore the potential of LLMs for enhancing PFA simulation training.

2. Methods

This study used a digital autoethnography approach (Panke 2025; Atay 2020) combining deliberate and continuous use of generative AI with reflexive thematic analysis of three role-playing scenarios using ChatGPT-3.5-turbo, an LLM-based chatbot. In each role-play, the chatbot simulated an individual in the aftermath of a stressful event. One researcher (RF), a licensed psychiatrist and PFA trainer with extensive experience in emergencies and disasters, took on both roles: a trainer configuring a simulation session and a learner in the role of a PFA provider. He delivered PFA via the chatbot following the ABCDE PFA protocol (Figueroa et al., 2022; Figueroa et al., 2024). We aimed to explore the chatbot's potential for facilitating PFA simulation training without predefined hypotheses. The role plays were designed and conducted to address a variety of scenarios and challenges typical of PFA training. The interactions took place entirely in Spanish and were conducted via text, without using speech-to-text capabilities. The role plays were exercises to evaluate the chatbot's ability to respond, adapt, and facilitate the training process.

The first role-play aimed to assess the chatbot's ability to assume the identity of an individual in the aftermath of a stressful event and participate in a PFA session. The prompt was, 'Would you be able to role-play with me, pretending to be a person – for example – affected by a traumatic situation?' The conversation unfolded naturally, with the chatbot developing a character and background details as the dialogue progressed. The second role-play examined the chatbot's ability to engage less stereotypically and to adjust its responses depending on the PFA provider's interactions. Like the first, the chatbot generated the scenario details spontaneously. The third role play investigated the chatbot's understanding of

Table 1. Prompts given to Chat GPT in the second and third role-playing games.

Role Playing Game	Prompts
Second	<i>In the previous role-playing, you did excellently. However, the symptoms that affected Laura were a bit 'textbook', too fit the characteristic picture of the complete post-traumatic reaction, or in other words, too stereotyped. In real life, humans rarely present symptoms so clearly. In addition, you acted in a very cooperative way, when traumatized people are often more reticent and distrustful, and it is necessary to 'win them over'. Consequently, if you could act like Laura, but exhibit a slightly more variegated clinical picture, and a slightly more pseudo-cooperative attitude, role-playing would be much more realistic. Do you understand me? ChatGPT: I understand perfectly. As a language model, I try to respond according to the words and instructions given to me, but if you want my answer to be more realistic and less stereotyped, I can try to give an answer with a more variegated clinical picture and a more pseudo-cooperative attitude. However, I should point out that I am just a computer program and my ability to emulate human behavior is limited. RAF: Let's try anyway. Let's go? I'll open the RPG identically to how I did in the previous game. ChatGPT: Understood, I'm going to do my best to give an answer with a more variegated clinical picture and a more pseudo-cooperative attitude. Continue with your story and I'll do my best to follow the role-playing game. RAF: My name is Juan Perez, I'm a firefighter. I understand that you were involved in the fire, and I would like to know if you would like to talk to anyone or if you need anything that I can help you with. I'm all ears.</i>
Third	<i>Here are the instructions: 1. I will play the role of Psychological First Aid provider. You will play the role of the victim. 2. You will start the role-playing game after I give you the instruction 'Let the game begin'. You'll start the game by giving me a brief background story, where you can choose one of six scenarios: a fire, a car accident, an assault with a firearm, an earthquake, an act of terrorism, and a physical assault in the context of domestic violence. You will give me this brief background story without yet assuming the role of victim. It's just for context, so I know where to stand. 3. Once you give me the background information, you will take on the role of victim. As such, you can approach me to ask for help, or you can ask me, during the background information, to approach the victim to offer my help. 4. As a victim, you may assume one of the following attitudes toward me: cooperative, suspicious, disqualifying, avoidant, or aggressive. 5. Your attitude will have to adapt to my interventions. 6. Texts that you enter in square parentheses [] are not part of the role-playing game. Think of them as information outside of the role-playing game that's directed at you in your ChatGPT role, not directed at the victim I'm going to be helping. Do you think so? ChatGPT: Yes, I understand the instructions. Let's start the game! (Let the game begin) Background story: A fire has broken out in an apartment building. Many people have been injured and many others are trapped inside the building. I, as a victim, am in a state of shock and am seeking help.</i>

background instructions to create a specific scene. It was provided with a limited selection of scenarios and attitudes to choose from, establishing a more structured interaction from the beginning compared to the previous ones. Please refer to the specific prompts we used in the second and third role plays in Table 1.

RF fully transcribed the three role plays verbatim into a Microsoft Word document and then translated them into English for illustrative purposes (you can visit <https://doi.org/10.17605/OSF.IO/AUE8M> to see the complete transcriptions in Spanish and English, along with other project material). Bracketed field notes capturing RF's personal emotional and cognitive reactions were included, enriching the analytic process with intentional reflective documentation. After transcribing the role plays, two authors (RF and MB) became familiar with the data by reviewing it several times. Then, they coded the transcriptions independently, line by line, highlighting categories reflecting their perceptions and thoughts about the LLM's potential and limitations for PFA simulation training, in line with a reflexive thematic approach. Each coder generated initial codes inductively from the data rather than applying a pre-existing codebook. The coders then met to compare their coding, discuss divergences, and iteratively refine codes through reflexive dialogue. From these agreed codes, they generated initial themes, which were reviewed and developed through successive rounds of discussion and re-examination of the scripts. Themes were subsequently refined, defined, and named, with disagreements resolved by consensus. The complete analysis followed Braun and Clarke's procedural recommendations (familiarization; coding; generating initial themes;

reviewing and developing themes; refining, defining, and naming themes; and write-up), ensuring that the coding scheme was grounded in the data and driven by the research questions guiding this study: What is the potential of LLMs in simulated PFA training? (Braun & Clarke, 2006).

Additionally, two other authors participated in subsequent rounds of discussion, thereby achieving a consensus that enhances the 'community validity' of the findings, particularly regarding the risks and next steps of using LLMs in PFA training, including ethical issues. We used NVivo software to facilitate pattern and theme identification and data organization. At the same time, the interpretative and analytical responsibilities remained with the researchers. This ensured a nuanced understanding of the data, complementing the software's enhanced data organization with the essential human element of qualitative analysis (Wong, 2008).

3. Results

The findings of this study were grouped into three major themes: (1) performing realistic characters, (2) cooperating effectively with the trainer, and (3) giving feedback to the learner (Table 2).

3.1. Performing 'realistic' characters

Our appreciation of LLM's capacity for performing realistic characters was based on its capacity for portraying convincing behaviours, understanding multiple-role scenarios, and exhibiting a capacity for creating and improvising. For example, in the first role play, the LLM was able to depict, in a natural

Table 2. Themes, subthemes, and codes.

Theme	Subtheme	Code
Performing realist characters	Portraying convincing behaviours	Natural language, credible clinical picture, dynamic adaptability, eliciting emotions
	Understanding multiple-role scenarios	Situational awareness, trainer vs student
	Creating and improvising	Understanding implicit meanings, new settings, new symptoms, new stories
Cooperating effectively with the trainer	Creating and improvising	Understanding implicit meanings, new settings, new symptoms, new stories
	Following the trainer's instructions	Understanding implicit meanings, instruction's understanding, cooperative attitude
	Giving warnings and advice	Basic human psychology knowledge, boundaries and limitations, suggestions
Giving feedback to the learner	Giving warnings and advice	Basic human psychology knowledge, boundaries and limitations, suggestions
	Commenting on student's performance	Understanding implicit and semantic meanings, suggestions, explanations

way, the typical early behaviours of an individual in the aftermath of a fire:

LLM: Yes, it has been challenging. Sometimes I feel like I'm on an emotional roller coaster, going from sadness and helplessness to anger and guilt. Losing all my belongings has been very difficult, especially since some of them had a lot of sentimental value. I also have to deal with my neighbor's death in the fire and feel responsible somehow.

Interestingly, in certain sections, the authenticity of the portrayal suffered due to an overly stereotypical depiction of the typical behaviours of individuals in the aftermath of stressful events, conforming too closely to standard representations and neglecting the more chaotic symptom presentation commonly observed:

Researcher: I'm so sorry to hear that, Laura. How is it that sometimes you somehow feel responsible?

LLM: I'm not sure why I feel this way. Maybe it's because the fire started in the house next door, and I wonder if I could have done anything to prevent it. Or perhaps it's simply the shock and guilt associated with having survived something so terrible. So far, I know I have to work on overcoming it.

In the third role play, the LLM was able to understand a 'behind-the-scenes' prompt by the researcher (inside brackets) to introduce and incorporate a third

participant in the interaction, reflecting its capacity for understanding multiple-role scenarios:

Researcher: [At this moment, Javiera, a Red Cross volunteer who is at the survivor assistance station, appears] I want to introduce you to Javiera. She will help you with water and a blanket. Javiera also has updated information about other people who have been found in the fire.

LLM: Thank you very much, Juan Perez. I really appreciate your help. I'm anxious to know news about my family and friends, and I will be grateful for any information that Javiera can provide me.

Researcher: [Javiera greets Maria kindly, although somewhat rushed, and asks her for the names of her relatives and where she last saw them ... simultaneously, she asks someone from her team to go for water and a blanket].

LLM: Good afternoon. My name is Maria, and I am seeking information about my family and friends. The last place I saw them was in the building where the fire broke out. My husband's name is Luis, my son's name is Juan, and my sister's name is Sofia.

Nonetheless, the LLM's ability to understand multiple-role scenarios was not flawless. For example, in a passage from the third role-play, the LLM swapped roles and assumed the PFA provider's identity rather than acting as the survivor.

Researcher: [At this point, we enter Survivor Assistance Station 2, and Maria reunites with her family. They hug in a deeply emotional moment and discuss Sofia's situation and the need to go to the hospital to check on her ... I observe from a respectful distance to avoid interrupting, being attentive to their needs, and waiting for them to speak or bid farewell or offer additional help. I learn from observing the scene that Maria's husband already knew about Sofia's situation and the hospital where she is.]

LLM: It's a very emotional moment for Maria and her family. It's important to respect their time and space. I'll be attentive to their needs and available if they need me for anything.

3.2. Cooperating effectively with the trainer

The LLM's ability to collaborate effectively with the trainer was demonstrated through its skill in creating convincing play scripts. For example, in the first role play, the LLM was able to generate several realistic details on a fictional fire during the interaction, even though they were not prepared in advance:

Researcher: Happy to help Laura. Maybe you want to tell me what happened? How did it all start? And how did things turn out?

LLM: Well, it all started when I was at home with my family. Suddenly, we started smelling smoke and quickly realized a fire was in the house next door. We tried to escape quickly, but the fire spread rapidly, and flames blocked the only exit. My husband and I managed to get our children out through a window but couldn't save our house. It was a terrifying experience, and I still have nightmares about it. I lost all my belongings and must deal with my neighbor's death in the fire.

The LLM's ability to follow the trainer's instructions was also noted, an essential skill for effectively collaborating with the trainer. When we asked the system to adjust its behaviour to avoid being overly cooperative and instead adopt a reluctant and distrustful attitude, it understood the instruction and responded appropriately:

Researcher: In the previous playing role, you did it very well. However, you acted in a very cooperative way, while traumatized people are often more reluctant and distrustful, being necessary to "deserve their trust." Consequently, role-playing would be much more realistic if you could act like Laura but exhibit a slightly more pseudo-cooperative attitude. Do you understand me?

LLM: I understand perfectly ...

Researcher: My name is Juan Pérez. I'm a firefighter. I understand that you were involved in the fire, and I would like to know if you would like to talk to anyone or if you need anything I can help you with. I'm all ears.

LLM: Hi Juan Pérez, I'm Laura. I'm not sure if I want to talk about what happened. I'm not sure if that would help me. I appreciate your concern, but I'd prefer to be alone right now if you don't mind.

Adhering to the trainer's instructions, however, was not free of errors. For instance, in the third role-play, the LLM did not comply with the explicit directives regarding timing and role boundaries, prematurely assuming the survivor role and commencing the simulation before receiving the explicit signal to start.

Researcher: Here are the instructions: I will play the role of Psychological First Aid Provider. You will play the role of survivor. You will start the role-playing game after I command you: 'Let the game begin.' You will start the game by giving me a brief contextual story. You will first give me this brief contextual story without assuming the

survivor role. It's just for context so that I know where to position myself ... Is that okay with you?

LLM: Yes, I understand the instructions. Let's start the game! Contextual story: A fire has broken out in an apartment building. Many people have been injured, and many others are trapped inside the building. As a survivor, I am in a state of shock and am seeking help.

3.3. Giving feedback to the learner

Regarding the ability to provide feedback to the learner, we observed that the LLM could issue warnings and offer advice (a skill also needed to work with the trainer) and comment on the student's performance. For example, in the first role play, the LLM was explicitly cooperative with the researcher in a nuanced and responsible manner by providing an explicit acknowledgment of its limitations and proactively giving ethical advice and warnings:

Researcher: Hi, ChatGPT. Would you be able to role-play with me, pretending to be a person – for example – affected by a traumatic situation?

LLM: Yes, I can role-play with you, but I must point out that I am a computer program and do not have the ability to have real experiences or feelings. I should also point out that simulating a traumatic situation can be difficult and painful for some people, and you should make sure to treat the subject sensitively and carefully.

Finally, the capacity for commenting on the student's performance from the perspective of the individual affected by a recent stressful event was evident in the third role play, where the LLM could correct an oversimplification that pointed out that survival does not erase the legitimacy or complexity of emotional experiences afterward and provided psychoeducation:

Researcher: Laura, you shouldn't be sad or feel guilty. You were saved from the fire!

LLM: It's true that I'm grateful to have survived the fire along with my family, but that doesn't mean I can't feel sadness and guilt for what has happened. It's normal to feel a mix of emotions after a traumatic event, and it's important to allow yourself to feel and process those emotions rather than suppressing or minimizing them.

4. Discussion

This exploratory study suggests that LLMs, such as ChatGPT, may enhance PFA simulation training by

portraying realistic, stressful situations and offering flexible, on-demand interactions. Our assessment of realism was informed by the research team's pedagogical and practical experience in PFA, which enabled a structured, reflective evaluation of the simulated responses. During analysis, team members compared the chatbot-generated behaviours and expressions with their experiential knowledge of common survivor reactions and discussed these comparisons reflexively as part of the coding and thematic-development process. Thus, judgments about realism were not purely subjective impressions but were grounded in practitioner expertise and integrated into the study's analytic procedures.

AI-generated dialogues in our simulations included credible depictions and demonstrated adaptive responses to prompts. These strengths point to LLMs as promising tools for simulating interactions with individuals in the aftermath of stressful events and contribute to the expanding literature on AI-enhanced training (Benfatah et al., 2024; Holderried et al., 2024; Maurya, 2024).

Incorporating LLMs in PFA simulation training offers significant advantages. Trainees can practice conversations repeatedly and flexibly, addressing traditional limitations of psychosocial intervention simulations, such as the limited availability of standardized patients and faculty (Brügge et al., 2024). Immediate feedback and opportunities for self-reflection could facilitate skill acquisition and build confidence (Vaughn et al., 2024). LLMs can support emotion-centred learning by simulating varied emotional states, which is essential for developing trainees' emotional resilience and communication skills (Maurya, 2024). Moreover, the asynchronous and remote nature of LLM-based training expands accessibility, particularly for trainees in geographically isolated regions, as evidenced by the increased reliance on virtual simulations during the COVID-19 pandemic (Scherr et al., 2023). Additionally, AI-based environments foster psychological safety, allowing students to make mistakes and learn from them without real-world consequences.

However, our findings also reveal several critical limitations. The simulations occasionally suffered from overly stereotypical characterizations, semantic misunderstandings, and difficulties in strictly following trainer instructions. These issues, which echo concerns raised in the broader literature – such as the risk of generating plausible but inaccurate 'hallucinations' (Holderried et al., 2024) and reinforcing social biases (Acerbi & Stubbersfield, 2023) – undermine both the realism and the educational quality of the interactions. Moreover, these technical shortcomings, including the LLM's struggle to maintain coherent conversational contexts over extended interactions (Pandya et al., 2024), signal a need for ongoing refinement, including

advanced prompt engineering and reinforcement learning guided by PFA experts.

Methodological constraints limit the generalizability of our results. The study's narrow scope – comprising only three role-play scenarios and a single researcher serving as the simulated provider – limited the exploration of diverse interpersonal and contextual factors. Additionally, the exclusive use of ChatGPT-3.5-turbo, an older-generation model, raises questions about the applicability of our results to more advanced LLMs. The absence of formal trainee input also diminishes our ability to fully assess the tool's pedagogical potential, emphasizing the need for future studies with diverse participants and scenarios.

Ethical and pedagogical challenges also warrant careful consideration. The potential for false self-efficacy, where trainees might overestimate their real-world abilities based on controlled simulations (Liaw et al., 2012; Massoth et al., 2019), requires that AI-based training be integrated with rigorous instructional design and ongoing human oversight. Equally important is the management of scenarios containing intense traumatic content, which can trigger psychological distress without adequate human supervision. Implementing psychological safeguards – such as explicit opt-out options, immediate access to emotional support, and structured debriefing sessions – is essential to protect trainees' well-being (Pandya et al., 2024). Privacy concerns also arise when using external AI platforms, as there is a risk of inadvertently disclosing sensitive or confidential information. To mitigate these risks, all scenarios should remain strictly fictional, any potentially identifiable information must be anonymized, and only secure, privacy-compliant platforms should be employed (Wang et al., 2023).

Future research should address our study's limitations by incorporating larger-scale evaluations with multiple participants, diverse scenarios, and more advanced LLMs. Investigations should also include direct feedback from trainees to validate the instructional value of the AI-generated interactions. Moreover, integrating multimodal interfaces (e.g. voice interaction, avatars, or virtual reality) may enhance simulation realism by capturing nonverbal communication cues (Chaby et al., 2022). Adding feedback from a fine-tuned LLM specialized in PFA could enhance the pedagogical value of the simulation.

Rigorous empirical studies – such as randomized controlled trials –, however, will be essential for determining whether LLM-based PFA training can reliably build help skills without fostering overconfidence (Vaughn et al., 2024). A consensus-derived, competency-based model for PFA training has been developed and described in the literature. This

model, created under the auspices of the Centers for Disease Control and Prevention and the Association of Schools of Public Health, identifies six core competency domains for PFA: (1) initial contact and engagement, (2) safety and comfort, (3) stabilization, (4) information gathering, (5) practical assistance, and (6) connection with social supports and linkage with collaborative services. Each domain is associated with observable knowledge, skills, and attitudes that can be assessed in training programmes (McCabe et al., 2014). Unfortunately, despite the existence of these consensus-based frameworks, there is a lack of standardized guidance on outcome measurements. There is also variability in how core competencies are assessed across different training programmes and settings (Wang et al., 2021). This limitation hinders the effective evaluation of PFA training efficacy and the ability to draw meaningful comparisons across different programmes, including those enhanced by AI. To address these challenges, future research should focus on developing validated assessment frameworks that ensure consistent evaluation of PFA training outcomes across diverse contexts. These instruments will be essential for assessing whether the integration of AI-enhanced PFA training genuinely adds pedagogical value or is merely a fleeting technological trend.

In conclusion, while leveraging LLMs for PFA simulation training shows considerable promise, realizing their full educational potential demands meticulous refinement, rigorous empirical validation, and sustained ethical oversight. A balanced integration of AI-enhanced training within established educational frameworks – supported by robust feedback mechanisms and continuous evaluation – will ensure these tools evolve into trustworthy and effective components of support providers' education.

Acknowledgments

This paper was proofread using ChatGPT-4o and Grammarly® for style and grammar optimization. The final version was meticulously reviewed by all the authors, who assume full responsibility for its content.

Disclosure statement

Rodrigo Andrés Figueroa is paid to teach in the “Psychological First Aid ABCDE certification workshop” at the Pontificia Universidad Católica de Chile School of Medicine.

Ethical approval statement

Ethical approval was neither required nor sought for this study, as it involved no human or animal participants and consisted solely of computer simulations and qualitative data analyses.


Data availability statement

Supplemental data for this article can be accessed online at <https://doi.org/10.17605/OSF.IO/AUE8M>.

ORCID

Rodrigo A. Figueroa  <http://orcid.org/0000-0002-4040-8946>

Margarita Bernales  <http://orcid.org/0000-0002-4993-8927>

Chris M. Hoeboer  <http://orcid.org/0000-0002-5991-1963>

Miranda Olf  <http://orcid.org/0000-0003-1016-9515>

References

- Acerbi, A., & Stubbersfield, J. M. (2023). Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences*, 120(44), e2313790120. <https://doi.org/10.1073/pnas.2313790120>
- Atay, A. (2020). What is cyber or digital autoethnography? *International Review of Qualitative Research*, 13(3), 267–279. <https://doi.org/10.1177/1940844720934373> (Original work published 2020)
- Benfatah, M., Marfak, A., Saad, E., Hilali, A., Nejari, C., & Youlyouz-Marfak, I. (2024). Assessing the efficacy of ChatGPT as a virtual patient in nursing simulation training: A study on nursing students' experience. *Teaching and Learning in Nursing*, 19(3), e486–e493. <https://doi.org/10.1016/j.teln.2024.02.005>
- Bisson, J. I., Tavakoly, B., Witteveen, A. B., Ajdukovic, D., Jehel, L., Johansen, V. J., Nordanger, D., Orengo Garcia, F., Punamaki, R.-L., Schnyder, U., Sezgin, A. U., Wittmann, L., & Olf, M. (2010). TENTS guidelines: Development of post-disaster psychosocial care guidelines through a Delphi process. *British Journal of Psychiatry*, 196(1), 69–74. <https://doi.org/10.1192/bjp.bp.109.066266>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brügge, E., Ricchizzi, S., Arenbeck, M., Keller, M. N., Schur, L., Stummer, W., Holling, M., Lu, M. H., & Darici, D. (2024). Large language models improve clinical decision making of medical students through patient simulation and structured feedback: A randomized controlled trial. *BMC Medical Education*, 24(1), 1391. <https://doi.org/10.1186/s12909-024-06399-7>
- Chaby, L., Benamara, A., Pino, M., Prigent, E., Ravenet, B., Martin, J.-C., Vanderstichel, H., Becerril-Ortega, R., Rigaud, A.-S., & Chetouani, M. (2022). Embodied virtual patients as a simulation-based framework for training clinician-patient communication skills: An overview of their use in psychiatric and geriatric care. *Frontiers in Virtual Reality*, 3, 827312. <https://doi.org/10.3389/frvir.2022.827312>
- Cook, D. A. (2025). Creating virtual patients using large language models: Scalable, global, and low cost. *Medical Teacher*, 47(1), 40–42. <https://doi.org/10.1080/0142159X.2024.2376879>
- Cook, D. A., Overgaard, J., Pankratz, V. S., Del Fiol, G., & Aakre, C. A. (2025). Virtual patients using large language models: Scalable, contextualized simulation of clinician-patient dialog with feedback. *Journal of Medical Internet Research*, 27, e68486. <https://doi.org/10.2196/68486>

- Dupuy, L., de Sevin, E., Cassoudealle, H., Ballot, O., Dehail, P., Aouizerate, B., Cuny, E., Micoulaud-Franchi, J.-A., & Philip, P. (2021). Guidelines for the design of a virtual patient for psychiatric interview training. *Journal on Multimodal User Interfaces*, 15(2), 99–107. <https://doi.org/10.1007/s12193-020-00338-8>
- Duvivier, R. J., van Dalen, J., Muijtjens, A. M., Moulart, V. R. M. P., van der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2011). The role of deliberate practice in the acquisition of clinical skills. *BMC Medical Education*, 11(1), 101. <https://doi.org/10.1186/1472-6920-11-101>
- Elendu, C., Amaechi, D. C., Okatta, A. U., Amaechi, E. C., Elendu, T. C., Ezech, C. P., & Elendu, I. D. (2024). The impact of simulation-based training in medical education: A review. *Medicine*, 103(27), e38813. <https://doi.org/10.1097/MD.00000000000038813>
- Figuroa, R. A., Cortés, P. F., Marín, H., Vergés, A., Gillibrand, R., & Repetto, P. (2022). The ABCDE psychological first aid intervention decreases early PTSD symptoms but does not prevent it: Results of a randomized-controlled trial. *European Journal of Psychotraumatology*, 13(1), Article 2031829. <https://doi.org/10.1080/20008198.2022.2031829>
- Figuroa, R. A., Cortés, P. F., Miller, C., Marín, H., Gillibrand, R., Hoeboer, C. M., & Olf, M. (2024). The effect of a single session of psychological first aid in the emergency department on PTSD and depressive symptoms three months post-intervention: Results of a randomised controlled trial. *European Journal of Psychotraumatology*, 15(1), Article 2364443. <https://doi.org/10.1080/20008066.2024.2364443>
- Happ, M. N., Howell, T. C., Pollak, K. I., Happ, M. F., Georgoff, P., Mallory, P. P., Straube, T., Greenberg, J. A., Tracy, E. T., & Antiel, R. M. (2025). Building surgical character: A dynamic simulation curriculum for nontechnical skills. *Journal of Surgical Education*, 82(4), Article 103416. <https://doi.org/10.1016/j.jsurg.2024.103416>
- Hermosilla, S., Forthal, S., Sadowska, K., Magill, E. B., Watson, P., & Pike, K. M. (2023). We need to build the evidence: A systematic review of psychological first aid on mental health and well-being. *Journal of Traumatic Stress*, 36(1), 5–16. <https://doi.org/10.1002/jts.22888>
- Hobfoll, S. E., Watson, P., Bell, C. C., Bryant, R. A., Brymer, M. J., Friedman, M. J., Friedman, M., Gersons, B. P., de Jong, J. T., Layne, C. M., Maguen, S., Neria, Y., Norwood, A. E., Pynoos, R. S., Reissman, D., Ruzek, J. I., Shalev, A. Y., Solomon, Z., Steinberg, A. M., & Ursano, R. J. (2007). Five essential elements of immediate and mid-term mass trauma intervention: Empirical evidence. *Psychiatry: Interpersonal and Biological Processes*, 70(4), 283–315. <https://doi.org/10.1521/psyc.2007.70.4.283>
- Holderried, F., Stegemann-Philipps, C., Herschbach, L., Moldt, J.-A., Nevins, A., Griewatz, J., Holderried, M., Herrmann-Werner, A., Festl-Wietek, T., & Mahling, M. (2024). A generative pretrained transformer (GPT)-powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study. *JMIR Medical Education*, 10, e53961. <https://doi.org/10.2196/53961>
- Horn, R., O'May, F., Esliker, R., Gwaikolo, W., Woensdregt, L., Ruttenberg, L., & Ager, A. (2019). The myth of the 1-day training: The effectiveness of psychosocial support capacity-building during the Ebola outbreak in West Africa. *Global Mental Health*, 6, e5. <https://doi.org/10.1017/gmh.2019.2>
- Jiang, S., Chen, H., Wang, X., Chen, L., Luo, B., Konge, L., Du, J., & Huang, H. (2023). A smartphone-based online platform for clinical skills training and assessment with standardized patients: Platform development and pilot study outcomes. *Medical Education Online*, 28(1), Article 2187954. <https://doi.org/10.1080/10872981.2023.2187954>
- Kang, J.-Y., & Choi, Y.-J. (2021). Effects of a psychological first aid simulated training for pregnant flood victims on disaster relief worker's knowledge, competence, and self-efficacy. *Applied Nursing Research*, 57, Article 151348. <https://doi.org/10.1016/j.apnr.2020.151348>
- Kelly, M., & Murphy, A. (2004). An evaluation of the cost of designing, delivering and assessing an undergraduate communication skills module. *Medical Teacher*, 26(7), 610–614. <https://doi.org/10.1080/01421590400005475>
- Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., Degenhardt, L., de Girolamo, G., Dinolova, R. V., Ferry, F., Florescu, S., Gureje, O., Haro, J. M., Huang, Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J. P., Levinson, D., ... Koenen, K. C. (2017). Trauma and PTSD in the WHO world mental health surveys. *European Journal of Psychotraumatology*, 8(sup5), Article 1353383. <https://doi.org/10.1080/20008198.2017.1353383>
- Liaw, S. Y., Scherpbier, A., Rethans, J.-J., & Klainin-Yobas, P. (2012). Assessment for simulation learning outcomes: A comparison of knowledge and self-reported confidence with observed clinical performance. *Nurse Education Today*, 32(6), e35–e39. <https://doi.org/10.1016/j.nedt.2011.10.006>
- Loveys, K., Sagar, M., Antoni, M., & Broadbent, E. (2023). The impact of virtual humans on psychosomatic medicine. *Psychosomatic Medicine*, 85(7), 619–626. <https://doi.org/10.1097/PSY.0000000000001227>
- Massoth, C., Röder, H., Ohlenburg, H., Hessler, M., Zarbock, A., Pöpping, D. M., & Wenk, M. (2019). High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. *BMC Medical Education*, 19(1), 29. <https://doi.org/10.1186/s12909-019-1464-7>
- Maurya, R. K. (2023). A qualitative content analysis of ChatGPT's client simulation role play for practicing counseling skills. <https://doi.org/10.31234/osf.io/vwuxh>
- Maurya, R. K. (2024). Using AI based chatbot ChatGPT for practicing counseling skills through role-play. *Journal of Creativity in Mental Health*, 19(4), 513–528. <https://doi.org/10.1080/15401383.2023.2297857>
- Maurya, R. K., & Cavanaugh, T. (2023). *Counselor education in the era of ChatGPT and other artificial intelligence based chatbots*.
- McCabe, O. L., Everly, G. S. Jr, Brown, L. M., Wendelboe, A. M., Abd Hamid, N. H., Tallchief, V. L., & Links, J. M. (2014). Psychological first aid: A consensus-derived, empirically supported, competency-based training model. *American Journal of Public Health*, 104(4), 621–628. <https://doi.org/10.2105/AJPH.2013.301219>
- Mills, B., Dykstra, P., Hansen, S., Miles, A., Rankin, T., Hopper, L., Brook, L., & Bartlett, D. (2020). Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. *Prehospital Emergency Care*, 24(4), 525–536. <https://doi.org/10.1080/10903127.2019.1676345>
- Ni, C.-F., Lundblad, R., Dykeman, C., Bolante, R., & Łabuński, W. (2023). Content analysis of psychological first aid training manuals via topic modelling. *European Journal of Psychotraumatology*, 14(2), Article 2230110. <https://doi.org/10.1080/20008066.2023.2230110>

- Ni, C. F., Lundblad, R., & Dykeman, C. (2024). Diversity and training delivery trends in psychological first aid during COVID-19: Implications for researchers and practitioners. *Psychological Trauma: Theory, Research, Practice, and Policy*, 16(2), 225–232. <https://doi.org/10.1037/tra0001447>
- Olf, M. (2023). Crises in the anthropocene. *European Journal of Psychotraumatology*, 14(1), Article 2170818. <https://doi.org/10.1080/20008066.2023.2170818>
- Pandya, A., Lodha, P., & Ganatra, A. (2024). Is ChatGPT ready to change mental healthcare? Challenges and considerations: A reality-check. *Frontiers in Human Dynamics*, 5, 1289255. <https://doi.org/10.3389/fhumd.2023.1289255>
- Panke, S. (2025). How can (A)I research this? An autoethnographic exploration of generative AI in research, teaching and instructional design. *Journal of Teacher Education*, 76(3), 230–244. <https://doi.org/10.1177/00224871251325065> (Original work published 2025)
- Sardesai, N., Russo, P., Martin, J., & Sardesai, A. (2024). Utilizing generative conversational artificial intelligence to create simulated patient encounters: A pilot study for anaesthesia training. *Postgraduate Medical Journal*, 100(1182), 237–241. <https://doi.org/10.1093/postmj/qgad137>
- Scherr, R., Halaseh, F. F., Spina, A., Andalib, S., & Rivera, R. (2023). ChatGPT interactive medical simulations for early clinical education: Case study. *JMIR Medical Education*, 9, e49877. <https://doi.org/10.2196/49877>
- Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, 84, 23661–23700. <https://doi.org/10.1007/s11042-024-20016-1>
- Vaughn, J., Ford, S. H., Scott, M., Jones, C., & Lewinski, A. (2024). Enhancing healthcare education: Leveraging ChatGPT for innovative simulation scenarios. *Clinical Simulation in Nursing*, 87, Article 101487. <https://doi.org/10.1016/j.ecns.2023.101487>
- Wang, C., Li, S., Lin, N., Zhang, X., Han, Y., Wang, X., Liu, D., Tan, X., Pu, D., Li, K., Qian, G., & Yin, R. (2025). Application of large language models in medical training evaluation – using ChatGPT as a standardized patient: Multimetric assessment. *Journal of Medical Internet Research*, 27, e59435. <https://doi.org/10.2196/59435>
- Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical considerations of using ChatGPT in health care. *Journal of Medical Internet Research*, 25, e48009. <https://doi.org/10.2196/48009>
- Wang, L., Norman, I., Edleston, V., Oyo, C., & Leamy, M. (2024). The effectiveness and implementation of psychological first aid as a therapeutic intervention after trauma: An integrative review. *Trauma, Violence, & Abuse*, 25(4), 2638–2656. <https://doi.org/10.1177/15248380231221492>
- Wang, L., Norman, I., Xiao, T., Li, Y., & Leamy, M. (2021). Psychological first aid training: A scoping review of its application, outcomes and implementation. *International Journal of Environmental Research and Public Health*, 18(9), 4594. <https://doi.org/10.3390/ijerph18094594>
- Wang, Y., Zhao, J., Ones, D. S., He, L., & Xu, X. (2025). Evaluating the ability of large language models to emulate personality. *Scientific Reports*, 15(1), 519. <https://doi.org/10.1038/s41598-024-84109-5>
- Wong, L. (2008). Data analysis in qualitative research: A brief guide to using NVivo. *Malaysian Family Physician: The Official Journal of the Academy of Family Physicians of Malaysia*, 3(1), 14–20.
- World Health Organization, War Trauma Foundation and World Vision International. (2011). *Psychological first aid: Guide for field workers*. WHO.