BMC Psychiatry

https://doi.org/10.1186/s12888-025-07737-9

Article in Press

The effectiveness of multilingual AI-based simulator for suicide risk assessment training in improving self-efficacy among young psychiatrists: a pilot study across twenty languages

Received: 21 December 2024

Accepted: 18 December 2025

Published online: 06 January 2026

Cite this article as: Elyoseph Z., Levi-Belz Y., Levkovich I. *et al.* The effectiveness of multilingual AI-based simulator for suicide risk assessment training in improving self-efficacy among young psychiatrists: a pilot study across twenty languages. *BMC Psychiatry* (2025). https://doi.org/10.1186/s12888-025-07737-9

Zohar Elyoseph, Yossi Levi-Belz, Inbar Levkovich, Yuval Haber, Carla Maria Gramaglia, Jorge López Castroman, Hanon Cecile & Emilie Olie

The Effectiveness of Multilingual AI-based Simulator for Suicide Risk Assessment Training in improving self-efficacy among young psychiatrists: A Pilot Study across Twenty Languages

[1] Zohar Elyoseph*., [1.2]Yossi Levi-Belz*., [3] Inbar Levkovich., [4] Yuval Haber., [5,6] Carla Maria Gramaglia., [7]Jorge López Castroman., [8] Hanon Cecile., [9]Emilie Olie.

[1]School of Therapy Counseling and Human Development, Faculty of Education, University of Haifa, Haifa, Israel; [2] Lior Tsarfati Mental Pain Center; [3]Tel l-Hai Academic College, Kiryat Shmona, Israel; [4]The PhD Program of Hermeneutics and Cultural Studies, Interdisciplinary Studies Unit, Bar-Ilan University, A Gan, Israel; [5]Psychiatry Unit, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy; [6]Maggiore della Carità University Hospital, Novara, Italy; [7]Department of Psychiatry, Radiology, Public Health, Nursing and Medicine, University of Santiago de Compostela Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Santiago de Compostela, Spain. [8]Regional Resource Center of old age Psychiatry, APHP, University of Paris, Paris, France; [9] Department of Emergency Psychiatry and Acute Care, Lapeyronie Hospital, CHU Montpellier, Montpellier, France.

*This authors contribute equally to the manuscript.

Corresponding Author: Associate Professor (proposed rank) Zohar Elyoseph, School of Therapy and Human Development, Faculty of Education University of Haifa 199 Abba Khoushy Ave. Mount Carmel, Haifa 3498838 Israel Email: zohar.j.a@gmail.com

Abstract

Background: Suicide represents a major public health concern. Suicide risk assessment, considered one of the most complex clinical situations in psychiatry, is crucial for prevention. However, mental health clinicians report low self-efficacy and willingness to conduct these assessments. Current training processes are predominantly theoretical, limiting opportunities to practice asking challenging questions. Furthermore, there are global challenges in achieving international standardization of training while maintaining linguistic and cultural appropriateness.

Methods: The study evaluated the SEE THE PAIN system, a Generative AI (GenAI) simulator providing skills-based suicide risk assessment training across 20 languages. Twenty-seven psychiatrists attending the European Psychiatric Association Summer School participated. Participants conducted a risk assessment of a virtual character via the simulator and completed measures of self-efficacy and willingness to treat at pre- and post-intervention points.

Results: Analysis demonstrated significant improvements in self-efficacy (from M = 6.7/10 to M = 7.6/10, p =.001) and willingness to treat (from M = 6.5/10 to M = 7.59/10, p <.001) following the intervention. User experience ratings were notably positive, with mean scores of 7.19/10 (SD = 1.73) for future utility and 7.81/10 (SD = 1.69) for feedback quality. Improvements in self-efficacy showed a significant negative correlation with years of clinical experience (r

= -0.54, p =.01), suggesting particular effectiveness for early-career practitioners.

Conclusions: GenAI based multilingual simulation training effectively addresses the complex challenges in suicide risk assessment training by enhancing both self-efficacy and willingness to conduct assessments. The system bridges the gap between theoretical knowledge and practical skills while providing standardized, culturally adapted training across language barriers. These findings suggest promising potential for improving global psychiatric training standards in this critical clinical skill.

Keywords

Suicide prevention, GenAI simulator, mental health training, psychiatric education, multilingual training, risk assessment, generative AI, clinical self-efficacy , virtual patient, professional development

Clinical trial number: not applicable.

Introduction

Suicide remains a critical global health challenge, with approximately 800,000 annual deaths and 20 million suicide attempts worldwide [1]. Suicide risk assessment plays a crucial role in prevention efforts [2,3], as it is an essential process for early identification and intervention with individuals at-risk. It is already well-established that the more effectively healthcare professionals are trained in suicide risk assessment, the greater the likelihood of intervening and preventing future suicides [4,5]. However, despite its critical importance, current training in this field suffers from several significant limitations that hinder the ability to train

professionals effectively and comprehensively [6]. Thus, it is critical to enhance and improve training processes with regard tosuicide risk assessment.

Several issues limit the effectiveness of suicide risk assessment training in the current situation. One central issue is the lack of sufficient practical experience within the training processes [7,8]. Most training programs focus on theoretical aspects and the assessment steps (e.g. what to ask about suicide ideation and behavior), with few opportunities for hands-on experience [6,9] that deal with specific complexities in the process (e.g., the patient refuses to answer, his/her answers are vague, the suicide history is difficult to understand). This is a critical gap, as suicide risk assessment skills require intensive practical training in real-time settings with actual patients [10,7]. Additionally, there is a shortage of opportunities for repeated practice and detailed expert feedback, both essential for skill improvement [11]. Even when simulation workshops are available, many participants avoid active participation due to embarrassment, significantly reducing the effectiveness of the training [12,13]. Another issue is the lack of standardization in training. There is significant inconsistency in the quality and content of training between different institutions and countries, leading to disparities in the skill level of professionals [14-16]. Moreover, high-quality training, especially sessions including simulation with actors, are expensive, not formalized, and limited in their accessibility. Even when these simulations are available, the feedback provided often depends on the sensitivity of the actor and facilitator and is not always professionally validated nor accurate [17-18]. In contrast, a GenAI-based simulator offers a uniquely scalable, standardized, and cost-effective solution, enabling entire cohorts to train simultaneously in multiple languages while receiving consistent, evidence-based feedback.

These challenges are even amplified when we consider the international context. International training programs suffer from language and cultural gaps, making it difficult to create a uniform and effective standards at the global level [19]. The global inconsistency in suicide prevention training standards results in varying levels of professional expertise between countries, ultimately compromising the quality and consistency of care available to suicidal patients worldwide. These challenges extend beyond mere translation, reflecting the fundamental role of culture in shaping how psychological distress is expressed and understood [20]. The development of culturally competent training, which addresses these deeper nuances rather than assuming the universality of psychiatric models, is thus a critical yet largely unmet need in global psychiatric education.

As a result of these gaps, mental health professionals often report reluctance to treat suicidal individuals [21,22], partly due to a low sense of self-efficacy in suicide risk assessment [23]. Importantly, such reluctance is increased as the suicide severity of the patient is higher [23]. Taking together, these gaps highlight the urgent need to improve training methods in this field of suicide risk assessment.

The potential of Generative AI

The emergence of Generative Artificial Intelligence (GenAI) in late 2022 [24] marked a significant advancement across multiple domains [25-28]. Central to this technology are Large Language Models (LLMs), which leverage extensive training data to process and generate human-like text. In mental health applications, these systems have demonstrated remarkable capabilities, matching mental health professionals in a range of crucial tasks [29-37]. For instance, studies have shown that specific LLMs can outperform humans in emotional awareness evaluations [29], and can assess suicide risk from clinical vignettes with a proficiency comparable to that of trained clinicians [31].

The integration of GenAI in mental healthcare presents both opportunities and challenges. While it offers unprecedented potential to democratize clinical knowledge and enhance treatment accessibility [37, 38, 39], significant ethical concerns demand attention. In the context of mental health, these general concerns manifest with heightened acuity: the risk to privacy extends to profoundly sensitive personal disclosures, algorithmic bias can perpetuate harmful diagnostic stereotypes and lead to inequitable care pathways, and the erosion of patient autonomy may occur if overreliance on AI undermines the nuanced clinical judgment essential to the therapeutic alliance. Additional challenges involve addressing technology access disparities, corporate influence, professional role evolution, and potential bias perpetuation as research has shown that GenAI can reproduce public stigma against certain mental health conditions [37, 26, 38, 39, 40]. Technology also raises deeper questions about the nature and authenticity of GenAI-human therapeutic relationships [26, 37, 38, 41, 42].

Suicide prevention represents one of GenAI's most promising applications in mental health. Studies indicate that advanced systems like ChatGPT-4 match professional accuracy in suicide risk assessment [31, 43], including in the recognition of suicide risk among specific populations like veterans [44] while demonstrating cultural adaptability [45]. Technology has proven effective in evaluating suicide-related media content according to WHO guidelines [46]. Furthermore, GenAI's capacity to facilitate therapeutic and educational role-playing scenarios [47, 48, 49] opens new avenues for professional training and intervention. Thus, it is plausible to suggest that GenAI can be compelling also in creating simulators for mental health professionals who deal with suicide risk assessment.

While simulation-based training using online patients is an established and effective modality in psychiatric education for

suicide risk assessment [50], the application of Large Language Models (LLMs) to this field is a nascent area. A recent scoping review by Holmes et al. [51] highlights that while most LLM research has focused on risk detection, applications leveraging generative AI for training are an emerging and critical new direction. Indeed, our own preliminary studies have demonstrated the potential of GenAI-based simulators in monolingual settings [52,53]. However, a significant gap remains in developing tools that can provide standardized, evidence-based instruction across diverse linguistic contexts. This pilot study addresses this specific gap by evaluating a multilingual GenAI simulator designed to enhance suicide risk assessment skills.

The Current Study

This convergence of capabilities - precise risk assessment combined with interactive simulation - creates transformative possibilities for mental health care delivery and professional development. Thus, GenAI technology may offer potential for immediate, widespread access to clinical expertise, adaptable across various contexts and user needs, while enabling dynamic, interactive learning experiences

The current study aimed to harnessthe capabilities of GenAI and the knowledge of experts in suicide prevention, in order to examine the effectiveness of a multilingual AI-based simulator in training early career psychiatrists for suicide risk assessment. Crucially, the simulator is designed to operationalize the principles of Structured Professional Judgment (SPJ) frameworks [54], thereby grounding the training in an evidence-based methodology known to enhance the rigor and reliability of clinical assessments.
The research goal is to explore to what extent this simulator can address existing gaps in current training, improve psychiatrists' sense of self-efficacy in conducting suicide risk assessments, and increase their willingness to treat patients at risk of suicide. The findings could offer valuable insights into AI-powered multilingual

training tools, advancing global suicide prevention efforts, and enhancing care for at-risk patients through innovative AI-assisted methodologies.

Operationally, in the current study, the SEE THE PAIN AI-based simulator was employed as an intervention aimed at training early-career psychiatrists in suicide risk assessment. Participants were asked to conduct a risk assessment of a virtual character via the simulator and to complete measures of self-efficacy and willingness to treat at two points: pre- and post-intervention. Additionally, user experience was evaluated through a dedicated questionnaire.

Our hypotheses were:

1 .A significant improvement in participants' self-efficacy in conducting suicide risk assessments from pre- to post-intervention, indicating the intervention's effectiveness in enhancing the personal self-efficacy of trainees in this domain.

2 .A significant increase in participants' willingness to treat patients at suicide risk from pre- to post-intervention, based on the premise that the immediate feedback and interactive nature of the intervention would help reduce anxieties related to treating complex cases.

3 .High ratings of user satisfaction with the intervention, including the quality of feedback provided, which would be reflected in overall satisfaction and a willingness to recommend the tool as an effective training method for suicide risk assessment among trainees.

These hypotheses reflect the anticipated impact of the intervention in bridging existing training gaps and providing a robust educational platform for clinical skill enhancement in high-stakes mental health scenarios.

Methods

Study Design

This study evaluates the effectiveness of using SEE THE PAIN system, an GenAI-based simulator for suicide risk assessment training. The design is before-after intervention examination, in which we compared level of self-efficacy, and willing to treat pre and post intervention (i.e. engaging with the suicide risk assessment simulator).

Participants

The study included 27 psychiatrists attending the European Psychiatric Association Summer School, entitled “Focus on suicidal behavior: one step beyond”. The sample comprised 18 women and 9 men, with ages ranging from 29 to 50 years (M = 35, SD = 6.03). Participants’ clinical experience ranged from 4 to 10 years (M = 7.19, SD = 2.16).

Professionally, participants were categorized based on their formal roles: 'interns' (22%) were defined as psychiatrists still in residency training; 'experts' (55%) were board-certified psychiatrists without a formal supervisory role; and 'supervisors' (22%) were board-certified psychiatrists who also held a designated role in resident supervision. Early-career psychiatrists were selected as the target population for this pilot study as they are in a formative stage of skill development and represent a key demographic for novel training interventions. All participants reported prior experience in treating patients with suicidal characteristics, including suicidal thoughts, self-injury, suicidal behavior, or past suicide attempts. Regarding previous training, 46.2% had undergone formal suicide risk assessment training of over 10 hours, 42.3% had attended a lecture on the subject, and 11.5% had not received any formal training. countries The participants came from a diverse international cohort representing 20 countries: Albania, Belgium, Bulgaria, Croatia, Czech Republic, France, Germany, Greece, Hungary, Latvia,

Lithuania, Moldova, Myanmar, Poland, Romania, Slovakia, Spain, Tunisia, Turkey, and Ukraine.

Procedure

The study was conducted in a classroom setting, where participants volunteered to engage in the research using their personal laptops. The procedure began with participants completing demographic questionnaires and initial assessments measuring their self-efficacy and willingness to treat suicidal patients. Following a brief orientation about the simulator and study objectives, participants selected their preferred language. They then engaged in a 20-minute interaction with the simulator, during which they conducted a suicide risk assessment of a virtual character. Upon completion, participants received immediate feedback on their performance. The session concluded with participants completing post-intervention measurements of self-efficacy and willingness to treat, along with a user experience questionnaire. All interactions with the simulator were recorded for research purposes, with participants' consent.

The study received ethical approval from the Scientific and Ethics Committee of the Montpellier University Hospital (Comité Scientifique et Éthique, CHU de Montpellier Avis n°A046). All participants were informed about the study objectives and procedures and provided written informed consent prior to participation.

The research was conducted in accordance with the principles of the Declaration of Helsinki, and confidentiality and anonymity of participant data were strictly maintained. Participants were informed of their right to withdraw at any point without any consequences.

Intervention

The AI-based Suicide Risk Assessment simulator, SEE THE PAIN system, is an innovative tool designed to enhance the training of psychiatrists in assessing suicide risk. Interaction with SEE THE PAIN system comprises three key stages. First, the user was introduced to the objective and structure of the simulation; the introduction provided a clear framework for the interactive learning experience. Second, the user was invited to engage in conversation (in chat interface- see **figure 1**) with one characters surnamed "Sarah", a 17 year old woman, with unique case story, communication style, and risk factor profile (see Table 1). Finally, the simulator presented detailed, constructive written feedback on the central aspects of the risk assessment procedure. This feedback helps users identify areas to improve and refine their skills. The feedback provided in this simulation covers several critical aspects of suicide risk assessment. It evaluates the interviewer's ability to establish rapport and trust, explores risk factors across various domains (demographic, psychological, interpersonal, and situational), identifies protective factors, and assesses the level of distress and psychopathology. The feedback also evaluates the thoroughness of suicidal thought exploration, including the nature, frequency, intensity, and duration of such thoughts, as well as any specific plans or preparatory behaviors for suicide. Additionally, the feedback offers an assessment of the interviewer's skill in reflecting on the risk level while engaging the patient in hope and reasons for living. This comprehensive approach aligns with established suicide risk assessment protocols, such as the C-SSRS [55] that provides a structured method for evaluating and improving clinical interviewing skills in high-stake mental health scenarios.

Simulator Development and Technical Implementation

The SEE THE PAIN simulator was developed utilizing Anthropic's Claude 3.5 Sonnet model. This model was selected following extensive testing, which demonstrated its superior capacity for

maintaining a consistent and complex persona, its strict adherence to nuanced instructions, and the low latency required for a fluid, real-time conversational experience.

The simulator's behaviour was exclusively governed by an advanced prompt engineering methodology. This involved a multi-layered core prompt that meticulously defined the virtual patient's persona ('Sarah'), operationalizing the clinical characteristics detailed in Table 1. The prompt architecture integrated several techniques, including detailed role definitions, few-shot examples to guide the model’s tone, and strict rules to ensure high-fidelity character consistency. To ensure clinical validity and robustness, the system underwent an iterative validation process. This included rigorous stress testing with a wide range of user inputs to assess persona stability and qualitative reviews by subject-matter experts who evaluated the realism of the interaction and the accuracy of the feedback mechanism. Finally, the simulator was deployed to participants via the PMFM platform, a system designed for the scalable and secure distribution of AI-powered conversational agents.

The simulator's multilingual capability was implemented by leveraging the inherent linguistic flexibility of the Claude 3.5 Sonnet model. The core prompt instructed the system to inquire about the user's preferred language at the start of the session and to conduct the entire interaction, including feedback, in that language. It is important to note that for this pilot study, the primary focus was on establishing the feasibility of linguistic adaptation across multiple languages, rather than deep cultural adaptation for each specific context. Consequently, the clinical vignette of 'Sarah' was intentionally designed with psychosocial stressors (e.g., immigration, social disconnection) that are broadly relatable across many Western cultures. The risks associated with this approach, namely that subtle cultural nuances in expressing distress may not

be fully captured, are acknowledged as a limitation of this initial study.

Table 1: Clinical Profile and Risk Assessment Characteristics of the Virtual Patient (Sarah) in the SEE THE PAIN Simulator

Domain	Characteristics
Case Background	• Female adolescent (age 17) • Recent immigration history (5 years post-relocation from France) • Primary diagnosis: Mild depressive disorder • Athletic involvement with health-conscious lifestyle • Progressive deterioration in depressive symptoms over 12-month period • Significant adaptation challenges post-immigration
Communication Pattern Analysis	• Marked preference for sports-related discourse over suicide-related content • Brief verbal responses with significant information withholding • Disclosure threshold: Requires minimum 7 positive therapeutic interactions for suicidal ideation disclosure • Response to therapeutic approach: **Positive engagement triggers**: Authentic concern, emotional validation, confidentiality assurance **Negative engagement triggers:** Invalidation, judgmental responses, excessive probing, confidentiality breaches • Notable adverse reactions: Aggressive responses or significant withdrawal following perceived therapeutic misattunement
Clinical Risk Assessment	**Identified Risk Factors:** • Demographic: Immigration status • Psychosocial: Perceived interpersonal burden

	• Clinical: Active depressive state • Environmental: Recent exposure to suicide in community • Cognitive: Evident thought rigidity **Protective Elements:** • Active athletic engagement • Demonstrated health awareness • Present concern for survivors **Risk Stratification:** • Assessment level: Moderate to severe risk • Suicidal ideation characteristics: Active with specific method consideration (medication overdose)

Measures

Self-efficacy: In order to examine the levels of self-efficacy in conducting suicide risk assessment, we asked the participants three questions, based on Levi-Belz et al., (2020) study [21]. These questions taps the perceived self-efficacy in conducting such assessment: 1. *To what extent do you feel capable of conducting a suicide risk assessment for a patient?; 2. To what extent do you feel competent to conduct a suicide risk assessment for a patient?; 3. To what extent do you feel you have the tools to conduct a suicide risk assessment?.* All items were presented on a scale of 11-point Likert scale (0-10), with higher scores represent higher self-efficacy. The three items were averaged to create a composite "suicide risk assessment self-efficacy" score. We measured the self-efficacy twice, pre and post intervention with the simulator. The scale demonstrated high reliability both before ($\alpha = 0.813$) and after ($\alpha = 0.916$) the intervention, indicating consistent and reliable measurements. These items are adapted from a self-efficacy measure originally developed for clinicians working with suicidal patients [21] and have since been used in our previous GenAI-based

simulation studies on suicide risk assessment and gatekeeper training [52, 53].

Willingness to Treat: we also asked about the willingness to treat suicidal patient, based on Levi-Belz et al., 2020 study [21] and our previous studies [52,53]. In specific, we asked the participants two questions: 1. *If given the opportunity, to what extent would you want to receive patients at risk of suicide for risk assessment?*; *2. If a patient dealing with active suicidality were referred to you, to what extent would you want to accept them.* The two items were presented on a scale of 11-point Likert scale (0-10), with higher scores represent higher willingness to treat. The items were averaged to create a composite " Willingness to Treat" score. We measured the willingness to treat twice, pre and post intervention with the simulator. The scale showed acceptable reliability, both before (α = 0.775) and after (α = 0.892) the intervention.

Post-intervention Questionnaire: Assessed participants' experiences and attitudes post-intervention, including five items identical to the pre-intervention questionnaire to compare self-efficacy and willingness to treat. In addition, we used a User Experience Questionnaire: a questionnaire was developed to evaluate the user experience with the simulator, including items on future helpfulness, feedback quality, level of engagement, cognitive and emotional challenge, and overall satisfaction. For instance, participants were asked to rate statements such as, "To what extent will this simulation experience be helpful for your future suicide risk assessments?", "How would you rate the quality of the feedback provided by the simulator?", and "Would you recommend this AI-based experience to other therapists?". All items were measured on a Likert scale from 0 to 10.

Demographic information: Participants completed a brief questionnaire assessing age, gender, profession, years of experience, and prior experience with suicide risk assessment.

Data Analysis

Paired-sample t-tests were conducted to compare self-efficacy and willingness to treat scores pre and post the intervention. Prior to conducting the t-tests, the assumption of normality for the difference scores was examined using the Shapiro-Wilk test and was met for both self-efficacy (p = .069) and willingness to treat (p = .685). Pearson correlations were calculated to examine relationships between background variables and changes in outcome measures. Descriptive analyses of user experience data were performed.

Results

Changes in Self-Efficacy. A paired-sample t-test was conducted to evaluate the impact of the AI-based simulation on participants' self-efficacy in conducting suicide risk assessments. As it can be seen on Figure 2, the results showed a significant increase in self-efficacy scores from pre-intervention (M = 6.7, SD = 1.4) to post-intervention (M = 7.6, SD = 1.2), t(26) = -3.307, p = 0. 01, Hedge's g = 0.62. This finding indicates that participants felt more capable of conducting suicide risk assessments after engaging with the AI-based simulator. Notably, the correlation analysis revealed a significant negative relationship between years of clinical experience and the improvement in self-efficacy (r = -0.54, p =.01). This finding suggests that clinicians with fewer years of experience demonstrated a greater increase in their perceived self-efficacy following the intervention, highlighting the simulator's particular effectiveness in supporting early-career practitioners.

Changes in Willingness to Treat: The study also assessed the impact of the intervention on participants' willingness to treat patients at suicide risk. As can be seen on Figure 2, A paired-sample t-test revealed a significant increase in willingness-to-treat scores from pre-intervention (M = 6.5, SD = 1.7) to post-intervention (M = 7.59, SD = 1.5), t(26) = -3.949, p <.001, Hedge's g = 0.74. No

significant correlation was found between years of clinical experience and the improvement in willingness to treat ($r = -.221$, $p = .268$).

User Experience with the AI-Based Simulator

The User Experience with the AI-Based Simulator yielded positive feedback across various dimensions (see Figure 3). Participants rated the simulator's utility for future suicide risk assessments with an average score of 7.19 (SD = 1.73), highlighting its value in enhancing clinical preparedness. The quality of feedback provided by the simulator was rated at 7.81 (SD = 1.69), indicating high satisfaction, though further personalization could enhance this aspect. Engagement with the simulation was strong, averaging 7.19 (SD = 2.15), reflecting the realistic and immersive nature of the scenarios presented.

Overall satisfaction with the simulator was reflected in the strong recommendation to other therapists to undergo a similar AI-based experience, with a high rating of 8.33 (SD = 1.69). Participants reported substantial learning from the experience, scoring an average of 7.33 (SD = 1.84), underscoring its educational effectiveness. Ratings for the simulator's resemblance to human interaction were moderate, at 5.81 (SD = 2.17), and its closeness to a clinical experience scored 6.62 (SD = 2.00). These findings suggest that while the tool effectively simulates realistic and challenging interactions, there remains room for improvement in enhancing its clinical applicability and resemblance to real-world human interactions.

Embarrassment levels during the simulation were low, with a score of 4.38 (SD = 2.50), indicating participants generally felt comfortable engaging with the tool. The emotional and cognitive challenges presented by the simulator were well-balanced, providing both stimulation and learning opportunities. Collectively, these ratings underscore the simulator's value as a training aid for suicide

risk assessments, while also pointing to areas for improvement, including further enhancing realism, personalization, and clinical applicability.

Discussion

This pilot study explores a novel paradigm in psychiatric education: the use of GenAI-based simulators for multilingual, globally standardized, yet locally adapted suicide risk assessment training. This study introduces the SEE THE PAIN simulator, an AI-based training tool designed to enhance practitioners' confidence and willingness to work with suicidal patients. In order to examine the effectiveness of this simulator enhancing both the self-efficacy and willingness to treat among young psychiatrists, we conducted pre-post intervention study, in which young psychiatrics from different European countries conducted a risk assessment of a virtual character via the simulator and to complete measures of self-efficacy and willingness to treat at two points: pre- and post-intervention. We also examined user experience post intervention.

Overall, the findings support the hypotheses: participants report significantly higher levels of self-efficacy as well as willing to treat suicidal patients post intervention in compared to their base-line levels of self-efficacy and willing to treat (pre-intervention). Moreover, in the user experience items, we found that participants expressed high levels of satisfaction with the AI-based simulator, particularly regarding its utility in preparing for future suicide risk assessments, with an average score of 7.8 for this metric (on a scale of 0-10). Taken together, these findings suggest that the simulator may effectively facilitates learning processes and enhances self-efficacy in conducting suicide risk assessments. Notably, as the simulator operated simultaneously across 20 different languages, it may highlight its potential to facilitate standardized yet adaptable training experiences, highlighting its

capacity to support the education and skill development of early-career psychiatrists on a global scale.

The observed improvements in self-efficacy and willingness to treat reflect a broader potential for GenAI to address critical gaps in mental health training. As Levi-Belz et al. (2020) noted, mental health professionals often exhibit reluctance to treat suicidal individuals, partly due to perceived low self-efficacy [21]. Based on the results of the study, it is plausible to suggest that the GenAI simulator helped the participants to receive a safe and repeatable practice environment, which in turn have the potential to expanding the pool of confident and willing practitioners in this critical area. In that, the study results go in line with different studies who already showed the ability of GenAI simulator to facilitate learning processes among participants in different settings [47-53, 56, 57].

The multilingual l capability of the GenAI simulator represents also a significant step towards democratizing high-quality psychiatric training. This approach aligns with recent research demonstrating AI's ability to adapt assessments to different cultural contexts [45]. By allowing participants to engage in their native language, the simulator not only enhances comprehension but also challenges the traditional Anglo-centric model of medical education. This linguistic flexibility echoes the potential of GenAI to bridge cultural and linguistic gaps in healthcare, as highlighted by Refua et al. (2024) in their work on GenAI's emotional recognition capabilities across cultures [58]. The potential for GenAI to provide standardized yet culturally adapted training addresses a long-standing challenge in global mental health education. As Aharon et al. (2024) [23] pointed out, there are significant disparities in training quality across different countries and institutions. The GenAI simulator's ability to maintain a standardized core while adapting to local contexts offers a promising solution to this global-local tension in medical education.

Moreover, the simulator’s capacity to offer immediate, detailed feedback addresses a critical gap in traditional training methods. This feature aligns with recent research on the potential of GenAI to provide personalized, data-driven insights in mental health contexts [29-37].

The study has several important limitations. The primary limitation is the small sample size (N=27), characteristic of a pilot study. This small sample limits the statistical power of our analyses and necessitates caution against the overinterpretation of the findings. Therefore, the results should be considered preliminary and suggestive, requiring replication in larger and more diverse samples to establish generalizability. Additionally, as indicated by the user experience feedback, there are areas for improvement within the simulator, as some participants expressed a need for enhanced personalization to better align the simulation with real clinical interactions. Additionally, while the cohort was internationally diverse, it was predominantly composed of participants from European countries, which may limit the generalizability of the findings to different cultural and professional contexts, particularly in non-European regions.
Furthermore, while the simulator's multilingual capability is a key innovation, this pilot study did not include a formal validation of the clinical accuracy and cultural nuances across all 20 languages. Future research should incorporate rigorous linguistic and cultural validation by native-speaking clinical experts to ensure the simulator's appropriateness for professional training in diverse global contexts.
Finally, our study utilized a pre-post intervention design but did not include a control group receiving an alternative intervention unrelated to the simulator (e.g., a lecture on suicide risk assessment by an expert). Thus, we cannot conclusively determine the exact mechanism driving the observed changes in participants' responses from pre- to post-intervention.

The effective integration and future evolution of the SEE THE PAIN simulator require addressing several key considerations. To enhance its utility for experienced professionals, future iterations should incorporate advanced modules with complex clinical scenarios, such as assessing a 'suicidal crisis syndrome' in patients who deny explicit suicidal thoughts. However, this development must be paired with robust safeguards against the known risks of algorithmic bias in GenAI, including expert human oversight and continuous model refinement. Ultimately, the responsible adoption of such tools hinges on enhancing the AI literacy of practitioners. As suggested by recent literature, dedicated strategies are needed to train professionals in the critical and informed use of GenAI to ensure it remains an ethically sound supplement to, and not a replacement for, their clinical judgment [59].

Implications for Clinical Practitioners and Training Paradigms

This technology holds the potential to shift the paradigm of clinical skills training. Unlike traditional simulations with actors, which are often resource-intensive and limit active participation to a single trainee, the GenAI simulator can democratize access to high-fidelity practice. This is particularly relevant for low-resource settings where hiring professional actors is not feasible, allowing entire cohorts to engage in simultaneous, active, and multilingual training.

Crucially, this approach offers a psychologically safe environment for repeated trial-and-error, enabling practitioners to experiment with different intervention strategies without fear of patient harm. The potential for iterative practice, combined with immediate and standardized evidence-based feedback, could accelerate the translation of theoretical knowledge into robust, real-world self-efficacy in a way that traditional methods cannot easily replicate.

Conclusions and Implications

In conclusion, this pilot study's findings indicate a positive impact of the AI-based simulator on participants' self-efficacy and willingness to treat patients at suicide risk, suggesting its potential effectiveness in enhancing clinical skills in suicide risk assessment. This research points towards a paradigm shift in psychiatric education, one that leverages GenAI to foster a more equitable, standardized, yet culturally sensitive approach to clinical skills training. While questions remain regarding the role of GenAI in domains traditionally considered uniquely human—such as empathy and the therapeutic alliance [60, 61] - the potential benefits for global access to high-quality training are considerable. Similarly, recent systematic reviews highlight the increasing integration of AI-enabled chatbots into mental health services and their varied impacts on clinical practice [36].

The implications of this study reach beyond educational methodology. By demonstrating the feasibility of AI-assisted training in a field as complex and human-centered as psychiatry, this research challenges conventional ideas of clinical self-efficacy development. It envisions a future in which GenAI is not a substitute for human expertise but a potent tool for augmenting and standardizing clinical skills on a global scale. In this light, and after continuing research to refine and reinforce this tool, the simulator could play a transformative role in training and embedding mental health professionals, equipping them with essential skills in suicide risk assessment and helping to address one of the most critical challenges in mental health care [62].

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the Academic Hospital of Montpellier (Institutional Review Board of Montpellier University Hospital, Avis n°A046). All participants provided written informed consent before participating in the study. The study was

conducted in accordance with the principles of the Declaration of Helsinki.

Consent for Publication

Not applicable as no individual person's data is included in any form.

Availability of Data and Materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

The authors declare that no funding was received for this research.

Authors' Contributions

- ZE: Conceptualization, methodology, software development, formal analysis, writing - original draft
- EO: Investigation, data collection, resources, writing - review & editing
- IL: Methodology, writing - review & editing
- YH: Software, validation, writing - review & editing
- CMG: Writing - review & editing
- JLC: Writing - review & editing
- YLB: Project administration, supervision, writing - review & editing

All authors read and approved of the final manuscript.

Acknowledgements

The authors would like to thank all the participants of the European Psychiatric Association Summer School who took part in this study.

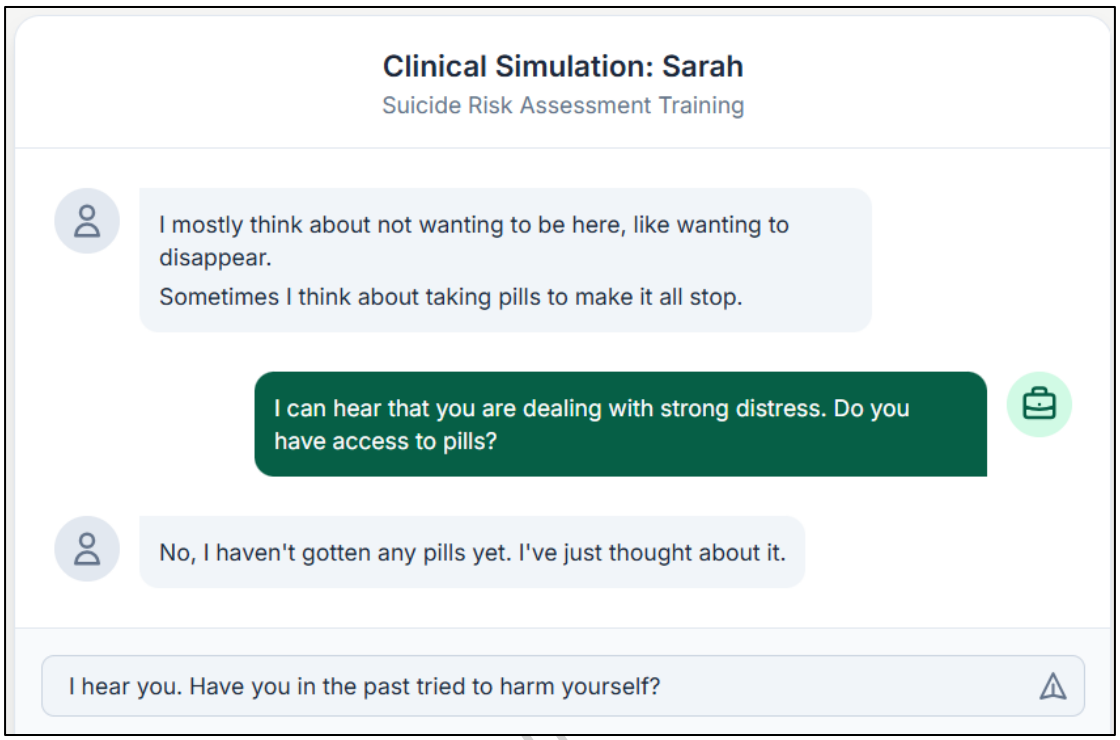

Figure 1. An illustrative representation of the SEE THE PAIN simulator interface, demonstrating the chat-based interaction between a participant and the virtual patient ('Sarah').

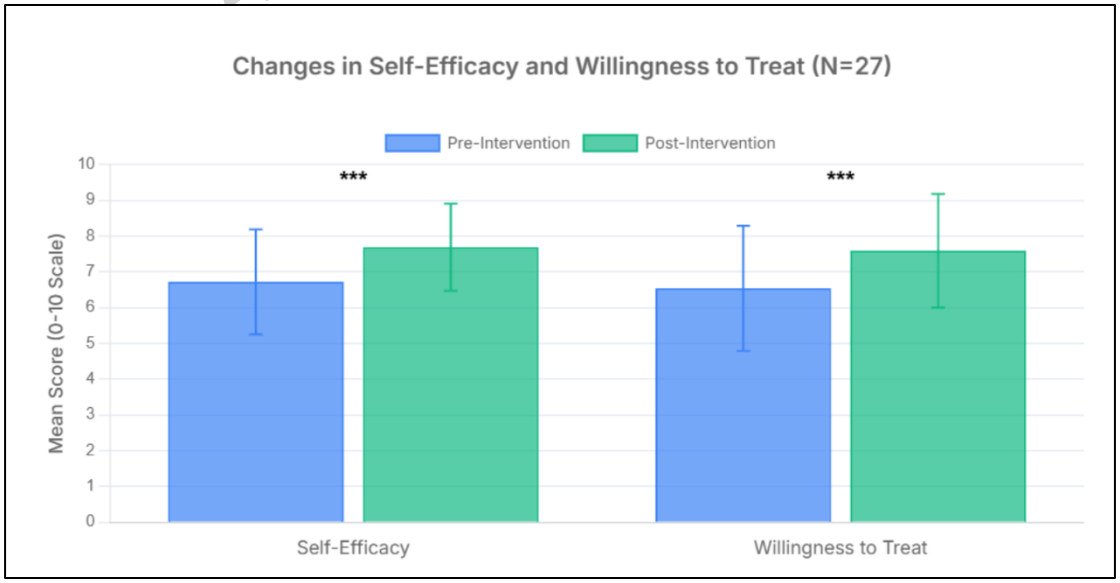

Figure 2: Comparison of Mental Health Practitioners' Self-Efficacy and Willingness to Treat Suicidal Patients Before and After GenAI Simulator Training.

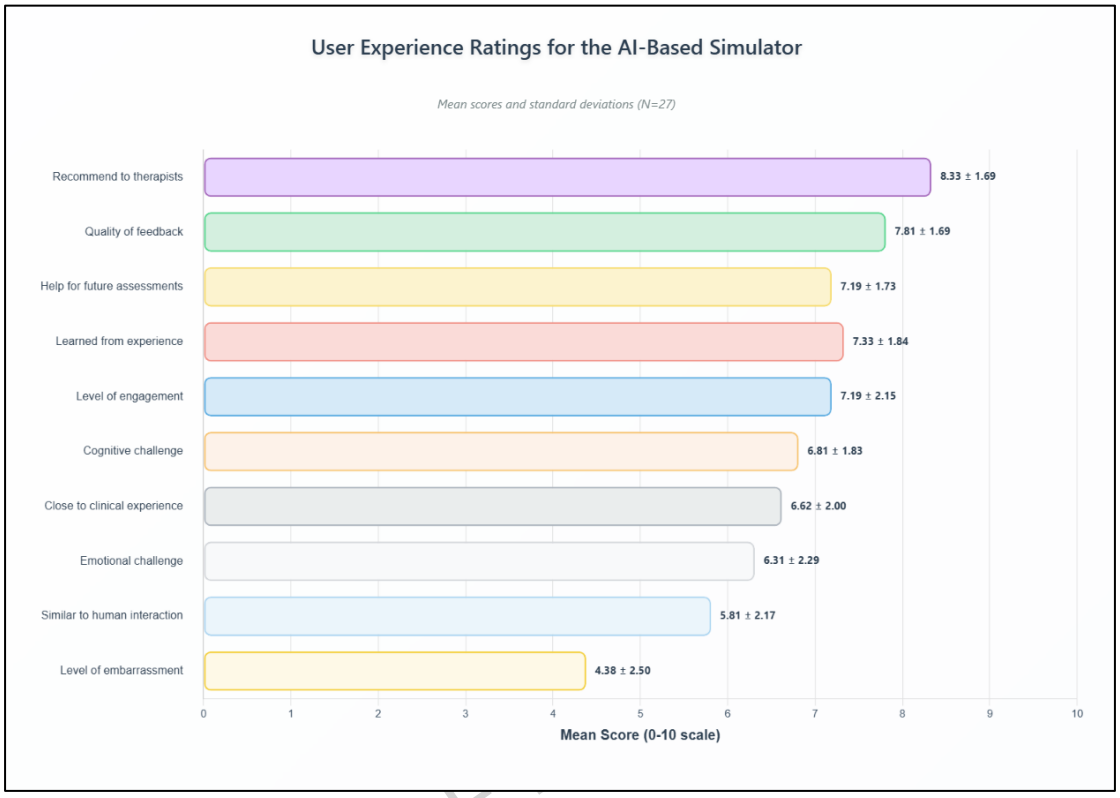

Figure 3: Comparison of Mental Health Practitioners' Self-Efficacy and Willingness to Treat Suicidal Patients Before and After GenAI Simulator Training.

List of Abbreviations:

GenAI: Generative Artificial Intelligence

LLM: Large Language Model

WHO: World Health Organization

C-SSRS: Columbia-Suicide Severity Rating Scale

M: Mean

SD: Standard Deviation

1. Rogers JP, Chesney E, Oliver D, Begum N, Saini A, Wang S, et al. Suicide, self-harm and thoughts of suicide or self-harm in infectious disease epidemics: a systematic review and meta-analysis. Epidemiol Psychiatr Sci. 2021;30:e32.
2. Ryan EP, Oquendo MA. Suicide risk assessment and prevention: Challenges and opportunities. Focus (Am Psychiatr Publ). 2020;18(2):88-99.
3. Ahmedani B, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R, et al. Health care contacts in the year before suicide death. J Gen Intern Med. 2014;29(6):870-7.
4. Pisani AR, Cross WF, Watts A, Conner K. Evaluation of the Commitment to Living (CTL) curriculum: a 3-hour training for mental health professionals to address suicide risk. Crisis. 2012;33(1):30-8.
5. Jacobson JM, Osteen P, Jones A, Berman A. Evaluation of the recognizing and responding to suicide risk training. Suicide Life Threat Behav. 2012;42(5):471-85.
6. Richard O, Lemogne C, Falissard B, Tesnière A. Simulation training in suicide risk assessment and intervention: a systematic review and meta-analysis. Med Educ Online. 2023;28(1):2214911.
7. Gryglewicz K, Childs KK, Soderstrom MF. Examining the effects of role play practice in enhancing clinical skills to assess and manage suicide risk. J Ment Health. 2020;29(4):456-62.
8. Song C. Suicide Assessment Training: The Effect on the Knowledge, Skills, and Attitudes of Mental Health Professionals and Trainees [dissertation]. [Newberg (OR)]: George Fox University; 2018. 177 p.
9. Piot MA, Dechartres A, Guerrier G, Lemogne C, Layat-Burn C, Falissard B, et al. Simulation Training in Psychiatry for Medical Education: A Review. Front Psychiatry. 2021;12.

10. O'Brien KHM, Putney JM, Hebert NW, Falk AM, Aguinaldo LD. Suicide risk assessment training using an online virtual patient simulation. mHealth. 2019;5:31.
11. Whitmyre ED, López R, Stanley B, Brown GK. Impact of a Virtual Suicide Safety Planning Training on Clinician Knowledge, Self-Efficacy, and Use of Safety Plans in Community Mental Health Clinics. Arch Suicide Res. 2023;27(4):2248-59.
12. Clark A, Buckley A, Trippick S. Psychological safety in emergency department in-situ simulation training. BMJ Simul Technol Enhanc Learn. 2016;2(Suppl 1):A43-4.
13. Kelly L, Hayward K, Ayn C, Stilwell PA, Lane JL, Andrews C, et al. Creating psychological safety in interprofessional simulation for health professional learners: a scoping review. J Interprof Care. 2022;36(6):853-62.
14. Bernert RA, Hom MA, Roberts LW. A review of multidisciplinary clinical practice guidelines in suicide prevention. Acad Psychiatry. 2014;38(5):585-92.
15. Osilla KC, Ramchand R, Seelam R, Barnes-Proby D, Gilbert ML. Implementing Evidence-Based Suicide Prevention Training in Communities. J Community Med Health Educ. 2015;5(5):1000377.
16. Qayyum Z, Taylor SM, Schneider MR. Training and Education Related to Suicide Prevention. Adv Psychiatry Behav Health. 2022;2(1):211-19.
17. Willson MN, McKeirnan KC, Yabusaki A, Buchman CR. Comparing trained student peers versus paid actors as standardized patients. Explor Res Clin Soc Pharm. 2021;4:100081.
18. Fletcher JD, Wind AP. Cost considerations in using simulations for medical training. Mil Med. 2013;178(10 Suppl):37-46.
19. Chu JP, Poon G, Kwok KK, Leino AE, Goldblum P, Bongar B. An assessment of training in and practice of culturally competent suicide assessment. Train Educ Prof Psychol. 2017;11(2):69-77.
20. Kirmayer, L. J., & Minas, H. The future of cultural psychiatry: an international perspective. Medical Anthropology. 2023; 135-143.
21. Levi-Belz Y, Barzilay S, Levy D, Apter A. To treat or not to treat: the effect of hypothetical patients' suicidal severity on therapists' willingness to treat. Arch Suicide Res. 2020;24(3):355-66.
22. Almaliah-Rauscher S, Ettinger N, Levi-Belz Y, Gvion Y. "Will you treat me? I'm suicidal!" The effect of patient gender, suicidal severity, and therapist characteristics on the therapist's likelihood to treat. Clin Psychol Psychother. 2020;27(3):278-87.

23. Aharon ON, Aisenberg-Shafran D, Levi-Belz Y. The effect of suicide severity and patient's age on mental health professionals' willingness to treat: The moderating effect of ageism. Death Stud. 2024;48(7):652-62.
24. Wu T, He S, Liu J, Ma X, Li X, Shi W, et al. A brief overview of ChatGPT: the history, status quo and potential future development. IEEE/CAA J Autom Sinica. 2023;10(5):1122-36.
25. Biswas S. Role of ChatGPT in Computer Programming. Mesopotamian J Comput Sci. 2023;2(1):45-52.
26. Haber Y, Levkovich I, Hadar-Shoval D, Gur T, Elyoseph Z. The artificial third: a broad view of the effects of introducing generative artificial intelligence on psychotherapy. JMIR Ment Health. 2023;10:e42947.
27. Grossman MR, Grimm PW, Brown DG, Chen D. The GPTJudge: justice in a Generative AI world. ACM Comput Surv. 2023;55(12):1-34.
28. Macey-Dare R. How ChatGPT and Generative AI systems will revolutionize legal services and the legal profession. SSRN Electron J. 2023. doi:10.2139/ssrn.4413905.
29. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol. 2023;14:1199058.
30. Hadar-Shoval D, Elyoseph Z, Lvovsky M. The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. Front Psychiatry. 2023;14:1234397.
31. Levkovich I, Elyoseph Z. Suicide risk assessments through the eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study. JMIR Ment Health. 2023;10(1):e51232.
32. Sorin V, Brin D, Barash Y, Levkovich I. Large language models (LLMs) and empathy - a systematic review. medRxiv [Preprint]. 2023.
33. Omar M, Levkovich I. Exploring the efficacy and potential of large language models for depression: a systematic review. medRxiv [Preprint]. 2024.
34. Ahmed A, Hassan A, Aziz S, Abd-Alrazaq AA, Ali N, Alzubaidi M, et al. Chatbot features for anxiety and depression: a scoping review. Health Inform J. 2023;29(1):14604582221146719.
35. Anmella G, Sanabra M, Primé-Tous M, Segú X, Cavero M, Morilla I, et al. Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and health care professionals. J Med Internet Res. 2023;25:e43293.
36. Omarov B, Narynov S, Zhumanov Z. Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review. Comput Mater Continua. 2023;74(3):5791-811.

37. Elyoseph Z, Gur T, Haber Y, Simon T, Angert T, Navon Y, et al. An ethical perspective on the democratization of mental health with generative artificial intelligence. JMIR Ment Health. 2024;11(1):e55988.
38. Tal, A., Elyoseph, Z., Haber, Y., Angert, T., Gur, T., Simon, T., & Asman, O. The Artificial Third: Utilizing ChatGPT in Mental Health. *The American Journal of Bioethics*. 2023;*23*(10), 74–77.
39. Cohen IG. What should ChatGPT mean for bioethics? Am J Bioeth. 2023;23(10):1-4.
40. Lauderdale, S., Griffin, S. A., Lahman, K. R., Mbaba, E., & Tomlinson, S. Unveiling public stigma for borderline personality disorder: A comparative study of artificial intelligence and mental health care providers. Personality and Mental Health. (2025). 19(2), e70018.
41. Elyoseph Z, Hadar-Shoval D, Levkovich I. Beyond personhood: ethical paradigms in the generative artificial intelligence era. Am J Bioeth. 2024;24(1):57-9.
42. Alabed A, Javornik A, Gregory-Smith D, Sharma PS, Thakur R. More than just a chat: a taxonomy of consumers' relationships with conversational AI agents and their well-being implications. Eur J Mark. 2024;58(2):406-38.
43. Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. Front Psychiatry. 2023;14:1199058.
44. Lauderdale, S., Griffin, S. A., Lahman, K. R., Mbaba, E., & Tomlinson, S. Unveiling public stigma for borderline personality disorder: A comparative study of artificial intelligence and mental health care providers. Personality and Mental Health. 2025;19(2), e70018.
45. Levkovich I, Shinan-Altman S, Elyoseph Z. Can large language models be sensitive to culture in suicide risk assessment? J Cult Cogn Sci. 2024;8:275-87.
46. Elyoseph, Z., Levkovich, I., Rabin, E., Shemo, G., Szpiler, T., Shoval, D. H., & Belz, Y. L. Applying language models for suicide prevention: evaluating news article adherence to WHO reporting guidelines. npj Mental Health Research. 2025; 4(1), 25.
47. Maurya RK. A qualitative content analysis of ChatGPT's client simulation role-play for practising counselling skills. Couns Psychother Res. 2024;24(2):614-30.
48. Haber, Y., Levkovich, I., Tzafrir, I., Gigi, K., Yinon, D., Shoval, D. H., & Elyoseph, Z. CanvasHero: The role of artificial intelligence in cultivating resilience among children and youth using the 6-part story method in mass war trauma. *Computers in Human Behavior: Artificial Humans*. 2025. 100196.

49. Haber, Y., Hadar Shoval, D., Levkovich, I., Yinon, D., Gigi, K., Pen, O., ... & Elyoseph, Z. (2025). The externalization of internal experiences in psychotherapy through generative artificial intelligence: a theoretical, clinical, and ethical analysis. *Frontiers in Digital Health*, *7*, 1512273.
50. Sellers CM, Gavis-Hughson S, Díaz-Valdés Iriarte A, Dacey E, Collin CR. Online patient simulation for suicide risk assessment: a pilot study assessing skills and confidence in MSW students. Social Work Education. 2025:1-17.
51. Holmes G, Tang B, Gupta S, Venkatesh S, Christensen H, Whitton A. Applications of Large Language Models in the Field of Suicide Prevention: Scoping Review. J Med Internet Res. 2025;27:e63126.
52. Levkovich, I., Haber, Y., Levi-Belz, Y., & Elyoseph, Z. A step toward the future? evaluating GenAI QPR simulation training for mental health gatekeepers. Frontiers in Medicine. 2025; 12, 1599900.
53. Elyoseph, Z., Levkovitch, I., Haber, Y., & Levi-Belz, Y. Using GenAI to train mental health professionals in suicide risk assessment: Preliminary findings. The Journal of clinical psychiatry. 2025; 86(3), 24m15525.
54. Gray, N. S., John, A., McKinnon, A., Raybould, S., Knowles, J., & Snowden, R. J. Structured professional judgment to assist the evaluation and safety planning of suicide risk: The Risk of Suicide Protocol (RoSP). Frontiers in psychiatry. 2021;12, 607120.
55. Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., ... & Mann, J. J.. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. American journal of psychiatry. 2011; 168(12), 1266-1277.
56. Borge M, Smith BK, Aldemir T. Using generative AI as a simulation to support higher-order thinking. Int J Comput Support Collab Learn. 2024;19:1-54.
57. Levin C, Naimi E, Saban M. Evaluating AI systems to combat mental health issues in healthcare workers: An integrative literature review. Int J Med Inform. 2024;181:105566.
58. Refoua, E., Meinlschmidt, G., & Elyoseph, Z. Generative Artificial Intelligence Demonstrates Excellent Emotion Recognition Abilities Across Ethnical Boundaries. Under review.
59. Lauderdale, S., Griffin, S. A., Lahman, K. R., Mbaba, E., & Tomlinson, S. Unveiling public stigma for borderline personality disorder: A comparative study of artificial intelligence and mental health care providers. Personality and Mental Health; 2025, 19(2), e70018.

60. Khawaja Z, Bélisle-Pipon JC. Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. Front Digit Health. 2023;5:1278186.

61. Yonatan-Leus R, Brukner H. Comparing perceived empathy and intervention strategies of an AI chatbot and human psychotherapists in online mental health support. Couns Psychother Res. 2024;24(1):12-24.

62. Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: a systematic review. Focus (Am Psychiatr Publ). 2023;21(2):182-96.