

Original Paper

Evaluation of an AI-Based Chatbot Providing Real-Time Feedback in Communication Training for Mental Health Care Professionals: Proof-of-Concept Observational Study

Lea Herschbach¹, MSc; Teresa Festl-Wietek¹, Dr rer nat; Christian Stegemann-Philipps¹, Dr rer nat; Alessandra Sonanini¹, MD; Bernd Herrmann¹; Rebecca Erschens², Dr rer nat; Anne Herrmann-Werner¹, MME, Prof Dr Med; Friederike Holderried¹, MME, MD

¹TIME - Tuebingen Institute for Medical Education, University of Tuebingen, Tuebingen, Germany

²Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Tuebingen, Germany

Corresponding Author:

Lea Herschbach, MSc

TIME - Tuebingen Institute for Medical Education

University of Tuebingen

Elfriede-Aulhorn-Strasse 10

Tuebingen, 72076

Germany

Phone: 49 70712973715

Email: lea.herschbach@med.uni-tuebingen.de

Abstract

Background: Effective communication is essential in medical practice, especially in dealing with the increasing number of patients presenting with mental health conditions. Feedback plays a crucial role in improving communication skills but is often difficult to implement. Artificial Intelligence (AI)-based tools offer promising support for practicing and improving communication techniques among physicians.

Objective: This proof-of-concept study investigated the capability of an AI-based chatbot that provided physicians with real-time feedback to train communication techniques. Specifically, it addressed three research questions: (1) How accurate is the AI-generated feedback, and how is it perceived by participants? (2) Does the use of an AI-based chatbot and AI-generated feedback lead to an increased frequency in the use of specific communication techniques? (3) Do physicians perceive the training with the AI-based chatbot and AI-generated feedback as beneficial for their daily clinical practice?

Methods: The study used a proof-of-concept design using 56 anonymized chat transcripts between physicians and simulated patients. Participating physicians received automated feedback generated by the AI-based chatbot, which was designed to assess and encourage the use of specific communication techniques. Feedback accuracy and user perception were evaluated, and changes in communication behavior were assessed. Finally, participants completed a postintervention questionnaire to evaluate perceived benefits for clinical practice.

Results: The AI-generated feedback was found to be very accurate with 85.38% of the real-time feedback being partially or even totally correct. In addition, most of the participants considered the feedback accurate, with 83.64% (n=55) agreeing or totally agreeing. Furthermore, 87.27% (n=55) of the participants agreed or totally agreed to the fact that the feedback was helpful. It also appeared to support the repeated use of recommended communication techniques. Furthermore, most of the participants agreed that practicing with an AI-based chatbot helped them practice and apply new communication techniques in clinical interactions, improved their physician-patient communication, and helped them recognize mental health conditions better in everyday medical practice.

Conclusions: The results highlight the potential of AI-supported training to address key communication challenges in medicine, particularly in the context of mental health care. The combination of real-time practice and immediate feedback may foster sustained improvements in physician-patient interactions. Integrating such tools into medical education could offer a valuable complement to traditional training methods. Future research may aim to refine AI models to improve their reliability and investigate the long-term effects and objective measures.

KEYWORDS

artificial intelligence; clinical competence; communication; feedback; medical education; physician-patient relations; simulated patients

Introduction

One of the core competencies of medical professionals is communication, with up to half of a physician's working time devoted to patient interactions [1,2]. Effective communication between physicians and patients can positively impact patient comprehension, cooperation, and trust, and therefore has a significant impact on the success of treatment [2,3]. Despite this, key communication skills are sometimes insufficiently represented in medical curricula [4,5]. Standardized patients offer realistic practice opportunities but are time and resource intensive, limiting scalability [6-8]. By using artificial intelligence (AI) to simulate patients, physicians can practice communication more flexibly and on a larger scale. Previous studies have demonstrated that AI-based simulations provide realistic patient interactions, offer subsequent feedback on history-taking, and generate mostly plausible responses [9,10]. AI is a valuable tool in medical education, offering accurate and automated assessments to enhance learning and skill development and improve patient outcomes [11-13]. However, effective integration requires maintaining students' independent and critical thinking, ensuring outputs are evidence-based and ethically sound. Accuracy, bias, privacy, and overreliance on technology must also be considered [12,14].

Feedback is essential to improve communication skills and for effective learning in clinical practice, but the type of feedback and the manner in which it is administered can have different effects [15,16]. In the context of communication, it is often challenging and costly to obtain timely and individualized feedback. Barriers such as inadequate training of supervisors, unfavorable learning environments, and lack of time make the feedback process more difficult [17]. AI models can be used to recognize and categorize communication techniques such as motivational interviewing (MI). MI is a patient-centered counseling approach that aims to enhance intrinsic motivation for behavior change through empathy, reflective listening, and guided exploration of ambivalence [18]. In addition to the categorization of communication techniques, these AI models can also deliver immediate, tailored feedback, supporting skill refinement during and after training without resource intensive supervision [19-21]. This feedback has been shown to assist participants in the refinement of their communication techniques during and following the training period [19,22].

Effective communication is crucial in the care of patients with mental health conditions, as these encounters often require heightened sensitivity, empathy, and adaptability [23-25]. The prevalence of mental illnesses is increasing worldwide [26,27], with up to 30% of outpatient and inpatient medical care allocated to the treatment of mental illnesses in Germany [2]. General practitioners (GPs) are typically the first point of contact for mental health problems [28,29]; yet, limited training and the

high demands of diagnosis and treatment make the care of these patients particularly challenging [30-33].

Most studies on AI-generated feedback provide posttask rather than real-time feedback, despite a growing demand for dynamic systems that adapt feedback focus and complexity in real time [34]. One reason for this is technical complexity, since multiple distinct large language model (LLM) calls and a control flow to accommodate for the parallelized LLM responses are required. While machine learning-based real-time feedback has shown high usability and improved knowledge elaboration, interaction, and group performance [35,36], its application in the context of LLMs and communication remains largely untested.

In summary, prior research has shown the importance of communication in medicine and highlighted training gaps, particularly in interactions with patients with mental health disorders. However, there are few studies that examine the potential of real-time AI feedback to enhance communication techniques. The benefits of simulated patients and AI for communication skills have already been demonstrated but these have mostly focused on single techniques (like MI) and provided feedback after the interaction. The accuracy of AI-generated feedback for communication techniques beyond MI, its effectiveness in real-time application, and its long-term impact on skill retention have not been systematically studied. Addressing these gaps is essential to optimize AI-supported communication training and improve clinical outcomes, which is why we aim to examine the following questions in this study: (1) How accurate is the AI-generated, real-time feedback on communication techniques and how is it perceived by the participants? (2) Does the use of an AI-based chatbot and AI-generated feedback lead to an increased frequency in the use of specific communication techniques? (3) Do physicians perceive the training with the AI-based chatbot and AI-generated feedback as beneficial for their daily clinical practice?

The objective of this study is to test the hypothesis that an AI-based chatbot providing real-time feedback enhances physicians' use of targeted communication techniques. Specifically, the study examines the accuracy, user perception, and impact of AI-generated real-time feedback on communication skills. The intervention is grounded in established frameworks of empathic and motivational communication, including Roger's therapeutic stance, Linehan's validation, and MI, which guide both the design of AI feedback and its application in clinical training.

Methods

Study Outline

This study was designed as a proof of concept to evaluate the feasibility and acceptability of real-time AI-generated feedback in communication training. Data collection took place between

February and June 2025 as part of a training class in mental health for physicians at the University Hospital of Tübingen, Germany. Participants were divided into 2 groups to counterbalance case order, with group 1 starting with case A and group 2 with case B. After a thematic introduction and instructions on the communication techniques (see Table 1), participants conducted a 20-minute chat with an AI-simulated patient. They used provided laptops and received real-time feedback on the communication techniques used. This was followed by an assessment of the patient (diagnosis and recommended actions) and receiving overall feedback on communication technique application. The feedback summarized the techniques used, presented absolute and relative usage frequencies, highlighted underused techniques, and presented examples of their correct application. In addition, the completeness of the medical history was evaluated. Participants

then rated whether this feedback was perceived as accurate and helpful, using a 5-point Likert scale from “totally disagree” to “totally agree.” The procedure was then repeated with a second AI-simulated patient chat, resulting in 2 chats per participant. After the interaction with the AI, a short survey was conducted, which collected demographic data and assessed the subjective perception of the tool’s authenticity on the same 5-point Likert scale described above. Furthermore, the following 4 questions were asked specifically: “I have the impression that practicing with an AI-supported chatbot can help me...” (1) “...practice new communication techniques,” (2) “...apply communication techniques in everyday clinical practice,” (3) “...better recognize a depression or anxiety disorder in everyday clinical practice,” and (4) “...improve physician-patient communication in everyday clinical practice.”

Table 1. Overview of communication techniques used with examples from the chats (translated to English).

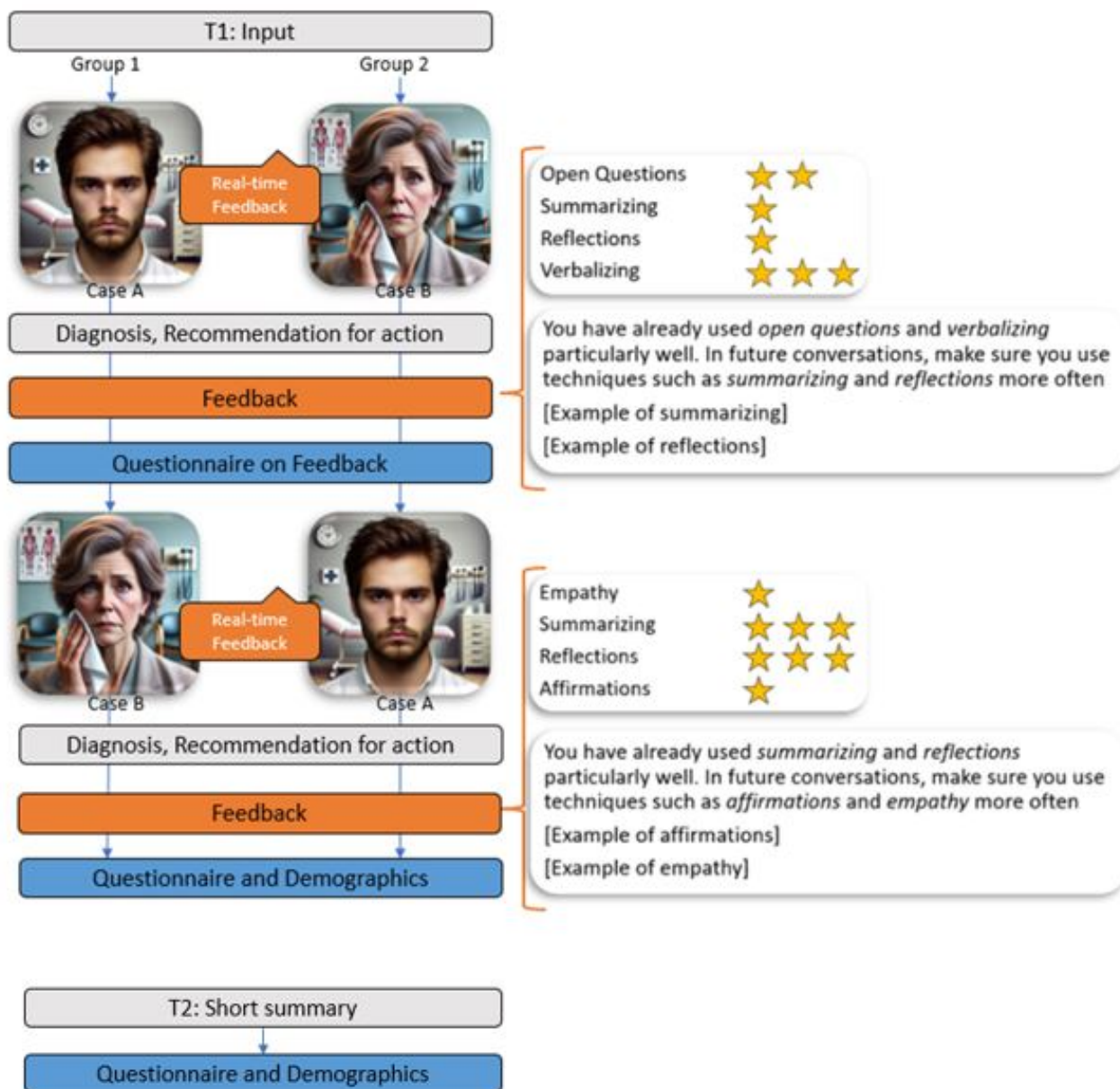
Theory and technique	Explanation	Example
Roger’s therapeutic stance		
Empathy	This refers to the therapist’s ability to understand the client’s experience without getting lost in those feelings.	GP ^a : “It’s always really unpleasant, this staying awake and digging and not being able to fall asleep. I understand that.”
Motivational interviewing		
Open questions	Open questions invite others to tell “their story” in their own words without directing them in any particular direction.	GP: “What brings you to me?”
Reflections	Rephrasing a statement to capture the implicit meaning and feeling of a patient’s statement.	P ^b : “My studies and my relationship suffer because I am constantly tired and irritable. This naturally leads to conflicts.” GP: “There is a lot of pressure on you to live up to yourself and your relationship.”
Summarizing	Summarizing what has been said at a later stage, especially at transition points. Often end with an invitation or question.	GP: “I would like to summarize once again: you have been suffering from sleep problems, a loss of concentration and reduced energy in everyday life for half a year, is that correct?”
Validation according to Linehan		
Verbalizing	Emotional reactions that are not expressed but are implied or conclusive from the behavior are validated.	GP: “I can see that you are very anxious; is that right?”
Validating in terms of past experiences	Emotional reactions are related to the medical history or biographical or biological aspects and normalized as appropriate.	GP: “After such a bad experience, it is very understandable to be afraid of another attack.”
Validating in terms of the current situation	Emotional reactions are related to the current situation and normalized as appropriate.	P: “But I’m still worried that it could be something serious” *looks anxious*. GP: “I understand that. IT’s really scary when you feel your heart racing and you start sweating.”
Radical genuineness	The patient is seen as a partner on an eye-level relationship, targeting self-opening.	GP: “Yes, I can understand that, I also prefer not to be in crowded places, too much hustle and bustle.”
Cheerleading	Believing in the patient is emphasized, positive characteristics of the patient are related to therapy.	GP: “Very good. I’m pleased that you want to get on board [with therapy].”

^aGP: general practitioner.

^bP: patient.

Participants answered the questions using the same 5-point Likert scale from “totally disagree” to “totally agree.” Additionally, a preliminary diagnosis had to be formulated. Electronic consent for anonymous data storage and analysis was obtained. Since data collection took place during a class and all participants were supposed to receive the same learning

opportunities, a control group that did not practice with the tool was not included. Due to the lack of a control group and the small sample size, causal conclusions are limited. Nevertheless, the findings provide essential initial insights into the topic. A detailed flowchart illustrating the study procedure is presented in Figure 1.

Figure 1. Outline of the experimental setup and the steps of data collection. T1: February 7, 2025; T2: June 27, 2025.

To further explore the long-term implications of training with AI-simulated patients, a follow-up survey was administered to the same participants 5 months later. They answered the same 4 questions mentioned above, with their responses this time reflecting their experiences during the preceding months.

Material

AI-Simulated Patients

We used a GPT-4 based chatbot with a custom interface, which allows us to enter individual patient vignettes, constantly manipulate chat history, and use elaborate prompting. It enabled students to submit written questions to a virtual patient, receive responses, and save the complete chat history for further analysis. This application is described in detail in a previous studies [10,37].

In this study, 2 patient cases were developed by experts. The first case depicts a 28-year-old man with depression; the second case portrays a 46-year-old woman with agoraphobia and panic

disorder. The cases are based on the *DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition])* [38] diagnostic criteria and were selected based on the prevalence of the disorders. The cases were integrated into the chatbot and comprehensively tested.

Communication Techniques

A literature review identified validated techniques for conducting empathic dialogue. Three well-established frameworks were selected, Roger's therapeutic stance [39], validation according to Linehan [40] and MI [18], each of which provides a validated framework for core communication techniques. These frameworks guided the design of the AI-based training and the feedback provided to participants. Items not applicable to text-based chats (eg, "Unconditional positive appreciation" and "Listening and observing") were not used. Similar or overlapping techniques (eg, "Congruence" and "Radical Genuineness," "Affirmations" and "Cheerleading," and "Reflections" from MI and validation according to Linehan)

were merged. The selected techniques are summarized in [Table 1](#). To prepare a training dataset, 2 experts independently evaluated 7 chats conducted for training purposes that assess which communication techniques were used and when. We did this to check how accurate we can expect AI to be when even human experts do not always agree. Initial coding yielded moderate agreement ($\kappa=0.426$, $P<.001$) due to the complexity of the material and slight differences in interpretation. After discussion and refinement, coding was repeated, resulting in good agreement ($\kappa=0.689$, $P<.001$). This evaluated training dataset was then integrated into the AI tool so that the feedback rules could be trained.

Feedback

The study incorporated various effective feedback types [41,42]. Formative feedback was provided immediately within the chat flow, specifying which communication technique had been applied, thereby confirming its use and encouraging further application. Summative feedback was visualized as a star-based graph showing the frequency of each technique after the chat. Future-oriented feedback highlighted the 2 least frequently applied techniques, illustrating them with examples to support their increased use in future conversations. A visual example of the feedback presented can be found in [Multimedia Appendix 1](#).

Training the AI

The training dataset was integrated into the application to enable AI-generated feedback during data collection. An initial F_1 -score of 0.38 revealed frequent false positives, mainly due to the low frequency of certain techniques in the training data. The F_1 -score is the harmonic mean of precision and recall [43]. To improve accuracy, the definitions of the communication techniques [44] were refined and supplemented with positive and negative examples in the prompt. This raised the F_1 -score to 0.58, which was deemed sufficient. In this training context, a slightly higher rate of false positives (ie, techniques detected that were not present) was considered didactical acceptable, even beneficial, because it encourages participants to critically engage with their communication and reflect on technique use.

Prompt Development

The basic prompt structure for the chats was already available from the AI-based patient model that was used [10]. For the real-time feedback version gpt-4o-2024-08-06 was used ([Multimedia Appendix 2](#)). Prompts for generating feedback on communication techniques were refined in a manual, iterative process, continuously evaluating error rates for all items and reviewing cases where the AI's feedback differed from our hand annotated data. Most of the major improvements resulted either from providing examples or from changing the specific wording. Wording was iteratively changed to make the prompts more precise and to make distinctions between the categories clearer. The exact prompts can be found in [Multimedia Appendix 3](#). To implement the real-time feedback, multiple, distinct LLM-calls were necessary to avoid overloading one single prompt and to be able to choose appropriate models for each task.

Data Analysis

All data were collected anonymously through the AI application. Participants' ratings of feedback helpfulness and accuracy, as well as communication performance measures, were automatically recorded. After data collection, 56 chats from 28 participants were exported and included in the analysis, with 1 participant completing the chats but not all of the questionnaires. In the follow-up survey, 26 physicians of the same cohort participated. Chats were segmented into question-answer pairs (QAPs), with each participant's question paired with the subsequent AI-generated response. They then were aggregated into a Microsoft Excel 2408 spreadsheet. The questionnaires were analyzed descriptively regarding demographics and performance. Statistical analyses were conducted in R (version 4.4.1; R Core Team, 2024) and SPSS (version 30; IBM Corp, 2023). Statistical significance was set at $P<.05$. Means, SDs, percentages, Pearson correlations, and F_1 -scores were calculated. To examine the effectiveness of AI-generated feedback in promoting the use of specific techniques, the frequency of technique usage in chat 1 was compared to usage counts in chat 2 by calculating Poisson regressions. For each participant, the 2 techniques explicitly mentioned in the feedback from the first chat, including examples, were selected for analysis. Figures were created using Excel 2408. As the evaluation required specialized domain knowledge and consistency, one expert reviewed all chats and rated the AI-generated feedback as (partially) correct or incorrect. Given that all feedback originated from the AI system, the evaluation was not blinded to this fact.

Ethical Considerations

The study was approved by the Ethics Committee of the Faculty of Medicine at University Hospital Tübingen (597/2024BO2). All participants provided informed consent. Privacy and confidentiality were safeguarded through deidentification and secure data management practices. Data were processed anonymously and no conclusions could be drawn about participants. Participation in the study was voluntary and no form of compensation was provided.

Results

Demographics

In the first data collection, 18 participants identified themselves as female and 9 as male. The average age was 38.2 (SD 11.1) years. Ten participants worked in a hospital, 15 were in private practice, 2 stated other working fields. Thirteen participants were in the field of general medicine, 4 were in internal medicine, 3 were in gynecology, 2 were in neurology, 1 was in pediatrics, and 4 was in other fields. On average, they had 10.30 (SD 8.55) years of professional experience. A total of 56 chats were evaluated. On average, a chat consists of 19.19 (SD 6.46) QAPs. A total of 1108 QAPs were included in the evaluation. In the follow-up survey, 18 participants identified themselves as female and 8 as male. The average age was 40.88 (SD 8.19) years, the professional experience was 11.15 (SD 8.80) years. Sixteen participants were in private practice, 9 worked in a hospital, and 1 stated "other setting." In total, 14 participants were in the field of general medicine, 3 were in gynecology,

and 2 each were in neurology and in pediatrics and internal medicine. Three participants stated being in other fields.

Correctness of the Diagnosis

Of the 28 participants, 17 (60.71%) correctly diagnosed both patients, 8 (28.57%) correctly diagnosed one, and 3 (10.71%) did not diagnose any patient correctly.

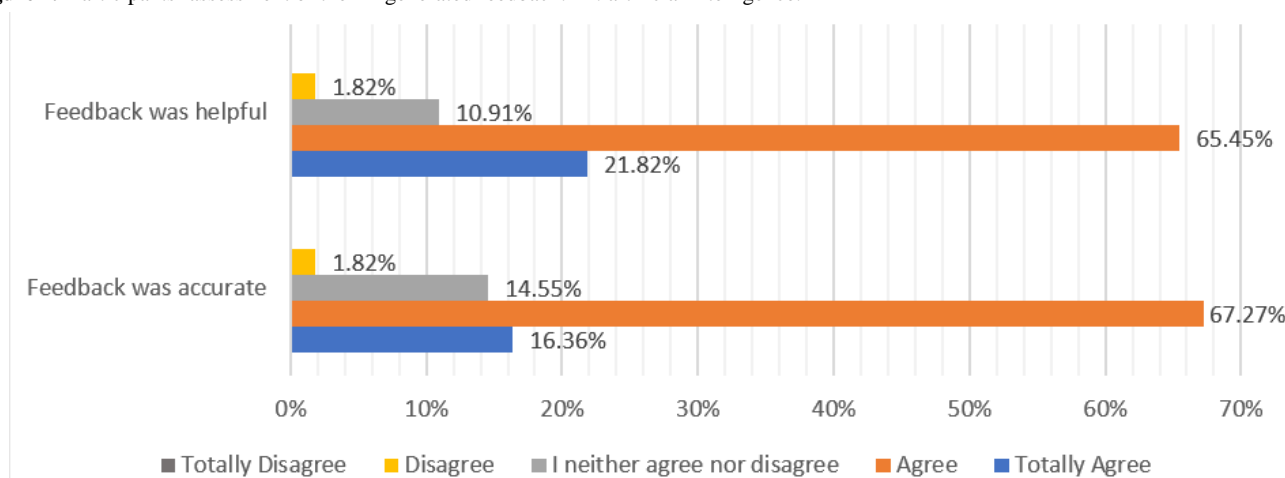
Accuracy of Feedback

Expert evaluation of the AI-generated feedback demonstrated high overall accuracy with 876 out of 1108 (79.06%) QAPs identified completely correct. Seventy (6.32%) QAPs were identified as partially correct, meaning at least one of several feedbacked techniques was correct. A total of 81 (7.31%) QAPs were categorized as false positive and 70 (6.32%) QAPs as false negative. In 11 QAPs (1.00%), the techniques mentioned in the

feedback were both false positives and false negatives (Multimedia Appendix 4). The AI rating was assessed as false positive in 4 out of 10 (40.00%) cases for “including past experiences,” in 13 out of 41 (31.71%) cases for “including current situation” and in 88 out of 277 (31.18%) cases for “open questions.” False negatives occurred most frequently for “including past experiences” (6/10, 60.00% cases), “including current situation” (17/41, 41.46% cases), and “empathy” (30/79, 37.97% cases).

These findings align with the participants’ perceptions. Across all 55 responses (28 participants completing 2 rounds each), 83.64% (n=55) agreed or strongly agreed that the feedback was accurate. Similarly, 87.27% (n=55) agreed or totally agreed to the fact that the feedback was helpful, with just one disagreement (1.79%, n=55; Figure 2).

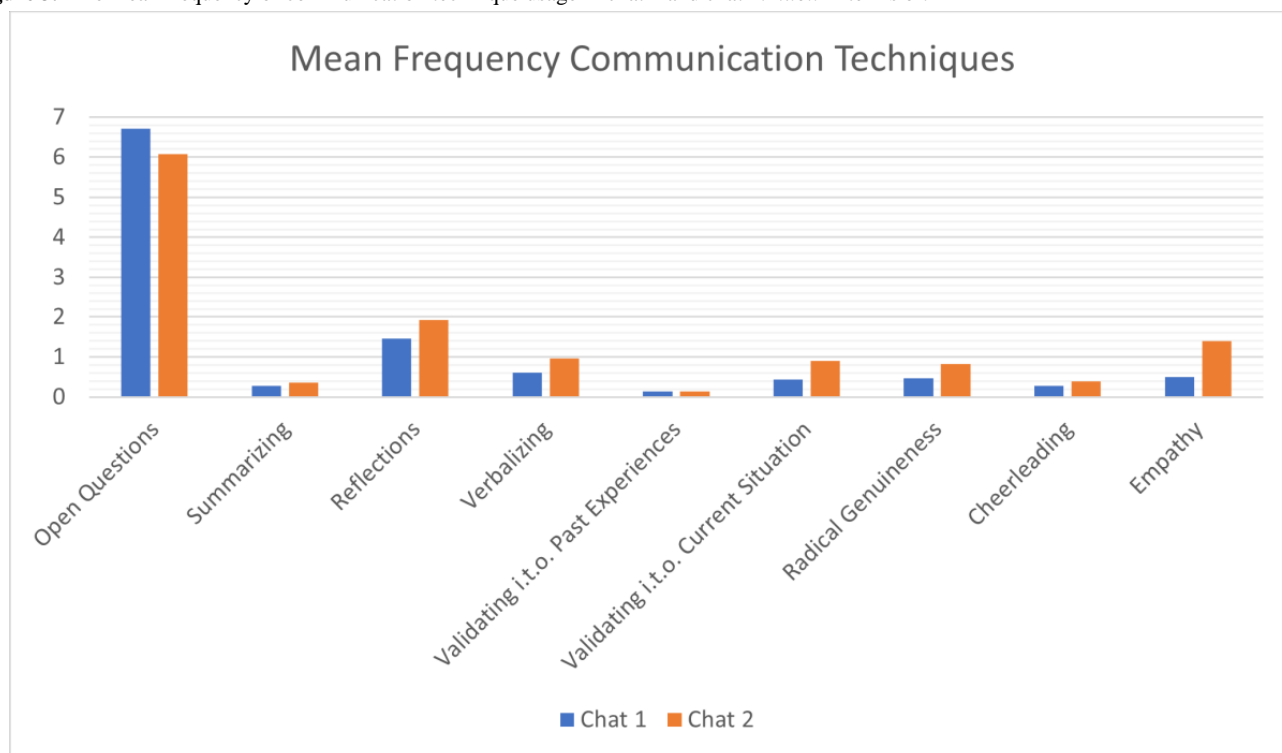
Figure 2. Participants’ assessment of the AI-generated feedback. AI: artificial intelligence.



The Use of Specific Communication Techniques

A Poisson regression was conducted to model the number of technique occurrences in the second chat as a function of prior feedback. Model criticism showed that the Poisson regression is properly modeling the data. The analysis revealed a significant

increase in technique usage after receiving AI-generated feedback ($\beta=1.365$, SE 0.323; $z=4.221$; $P<.001$). The estimated mean count of techniques (λ) was 0.21 in chat 1 and 0.84 for chat 2. These findings suggest that AI-generated feedback effectively reinforces the use of recommended techniques in subsequent conversations (Figure 3).

Figure 3. The mean frequency of communication technique usage in chat 1 and chat 2. i.t.o.: in terms of.

Subjective Benefits in Clinical Practice

In the next step, we examined the extent to which the physicians perceived practicing with the AI-generated patients as subjectively beneficial for their daily clinical work. Overall, 85.19% (n=27) of the participants agreed or totally agreed to the statement that practicing with an AI-supported chatbot can help them to practice new communication techniques and to apply them in everyday clinical practice (85.19%, n=27 agree or totally agree). Furthermore, 81.48% (n=27) agreed or totally agreed that practicing could help them to improve their physician-patient communication in their daily clinical work. Furthermore, the majority indicated that it may help them to

better recognize depression and anxiety disorders in everyday clinical practice, with 59.26% (n=27) agreeing or totally agreeing. The exact agreement rates are shown in Table 2. These results were replicated in the follow-up survey 5 months later. Again, the majority of the 26 participants stated that, according to their subjective perception, practicing with the AI-supported chatbot enabled them to practice new communication techniques, apply them in everyday clinical practice and improve their physician-patient communication. They also reported that, in their own perception, they were better able to recognize depression and anxiety in their clinical practice after the training (Table 2).

Table 2. The participants’ assessments of the benefits of the application in everyday clinical practice at measurement time point 1 (February 2025, N=27) and measurement time point 2 (June 2025, N=26).

Response to “I have the impression that practicing with an AI-generated Chatbot can...” and time point	Totally agree, n (%)	Agree, n (%)	I neither agree nor disagree, n (%)	Disagree, n (%)	Totally disagree, n (%)
...help me practice new communication techniques.					
T1	3 (11.11)	20 (74.07)	3 (11.11)	1 (3.70)	0 (0)
T2	0 (0)	16 (61.54)	7 (26.92)	3 (11.54)	0 (0)
...help me apply communication techniques in everyday clinical practice.					
T1	5 (18.52)	18 (66.67)	3 (11.11)	1 (3.70)	0 (0)
T2	2 (7.69)	14 (53.85)	9 (34.62)	1 (3.85)	0 (0)
...help me better recognize a depression or anxiety disorder in everyday clinical practice.					
T1	1 (3.70)	15 (55.56)	7 (25.93)	3 (11.11)	1 (3.70)
T2	1 (3.85)	14 (53.85)	8 (30.77)	3 (11.54)	0 (0)
...help me improve physician-patient communication in everyday clinical practice.					
T1	2 (7.41)	20 (74.07)	3 (11.11)	2 (7.41)	0 (0)
T2	0 (0)	12 (46.15)	13 (0.50)	1 (3.85)	0 (0)

Authenticity

Most of the participants (n=27, 62.96%) agreed that the AI-simulated patients appeared authentic, and an additional 11.11% (n=27) strongly agreed. Only 7.41% (n=27) disagreed; no one strongly disagreed. Pearson correlation analysis revealed a strong positive relationship between perceived authenticity and overall agreement with the chatbot’s usefulness ($r_{28}=0.696$, $P<.001$), suggesting that higher perceived authenticity is associated with greater perceived benefit.

Discussion

Principal Findings

Our findings indicate that AI-generated real-time feedback is mostly accurate and is also perceived as such by the participants. This finding demonstrates the capability of AI technologies for generating prompt feedback on communication techniques. Furthermore, training with the AI-based chatbot with real-time feedback led to an increased frequency in the application of targeted communication techniques, demonstrating its potential to enhance practical communication skills. Finally, the participating physicians reported that the training had a beneficial impact on their daily clinical practice. This suggests that feedback provided by AI may positively influence health care professionals' perceived competence and confidence.

Feedback Accuracy

Our results show that AI-generated feedback is mostly accurate, indicating very satisfactory results for a new application of AI. Most participants also found the feedback to be accurate and helpful, supporting the use of AI for real-time evaluation of communication techniques. This aligns with the findings of Imel et al [19], who found that participants generally perceived AI-generated feedback as accurate and helpful for their clinical performance. Furthermore, this confirms prior findings on the effectiveness of AI being comparable to generic experts [45]

and extends them by demonstrating that the AI is capable of detecting not only motivational interviewing techniques [19-21] but also communication techniques from Rogers and Linehan. This is particularly noteworthy as these techniques extend beyond the complexity and scope of MI, offering the potential to substantially enhance communication skills of general practitioners in everyday practice. This finding has practical relevance, as reliable AI feedback enables objective assessment of communication skills and enhances communicative competence in clinical settings. Unlike previous work in educational settings, where ChatGPT feedback was evaluated without iterative refinement [46], our study incorporated a process in which AI-generated feedback was modified and refined based on experts’ rating. In line with previous recommendations [47], we trained our AI models on context-specific data to enhance the quality and usability of the feedback.

However, performance varied across techniques, likely due to the low frequency of some techniques (eg, “including past experiences”) in both training and chat data, limiting the AI’s ability to acquire them effectively. A certain error rate with false positive and false negative errors remained. These misclassifications are a common issue [46,48] and may limit training validity and warrant further refinement of the feedback algorithms.

Increase in the Use of Communication Techniques

A key aim of the study was to assess whether AI-generated feedback encourages more frequent use of communication techniques. Our results show a significant increase between the first and second chat, suggesting that AI-based training may support the correct application of these techniques. These findings are in line with those of Cook et al [49], who reported substantial benefits of virtual patients compared with no intervention. These findings are in line with a recent review [50] on the affordances of AI-based tools, indicating that learners primarily develop their communication skills through



prompt feedback and tailored recommendations. Our findings may extend this work by suggesting that real-time AI feedback could serve as a practical approach to enhance communication training. In addition, a recent study by García-Torres et al [51] addressed implementation challenges identified in previous research, such as expanding the application of AI to new communication techniques and clinical domains [51]. While an increase in technique use was observed, causality cannot be confirmed. Further research is needed to evaluate the specific impact of AI feedback in this context.

Perceived Benefits in Clinical Practice

In addition to quantitative improvements, the majority of the participants reported that AI-based training helped them practice and apply new communication techniques and enhanced their self-perceived ability to recognize depression and anxiety in real clinical interactions. The follow-up survey confirmed these impressions and highlighted perceived long-term benefits. These findings support previous suggestions indicating that AI-based training is not only effective in skill acquisition but also translates into tangible benefits in real-world medical practice [50]. This aligns with the findings of Imel et al [19], who reported that participants expressed a high likelihood of applying the received feedback in practice, underlining the potential of such feedback to foster behavioral change. However, the perceived benefit of AI-based training was found to be influenced by the perceived authenticity of the chatbot. This aligns with prior research, indicating that the effectiveness of virtual training tools depends on their realism and user engagement [51,52]. The importance of developing AI training systems that can simulate real-world interactions as accurately as possible is emphasized by this.

Implications for Medical Training

This study highlights the potential of AI-powered training to meet the increasing demand for mental health care [53,54] by equipping GPs with key communication techniques. Such tools provide scalable, accessible, and consistent learning opportunities and therefore may represent a valuable addition to future training programs. Compared to traditional training methods, AI-based chatbot training offers several distinct advantages, such as a wide range of application, scalable and flexible access, and allowing learners to practice repeatedly and at their own pace [55]. The real-time feedback is particularly valuable in medical education as it allows learners to immediately correct and refine their skills [56,57], supplementing traditional instruction with individualized, continuous training. Furthermore, AI-generated feedback can be provided immediately, consistently, and without requiring trained human observers [15,17]. While text-based AI cannot fully replicate the complexity, emotional dynamics, or unpredictability of real patient interactions, its availability, repeatability, and standardization make it a valuable supplement for communication training. It can help to maximize learning benefits while mitigating risks such as model bias or overreliance. AI feedback should complement, not replace, expert supervision and be integrated into structured training programs with clear guidance on its interpretation and application [51,52]. In order to ensure the successful integration

of technology into the educational environment, educators must adapt curricula and assessment strategies to the evolving technological landscape [58].

Limitations and Future Research

Despite these promising results, this study presents certain limitations. A primary limitation concerns arises from its exclusive focus on short-term skill acquisition and subjective benefits. Consequently, the results primarily reflect perceived rather than objective or long-term improvements. Additionally, the study incorporated a rather small sample size of 28 physicians from a single site, and participation was voluntary. This may limit the generalizability of the findings and may have introduced self-selection bias, as participants could have been more motivated or interested than nonparticipants. Another relevant issue is the absence of a control group, limiting the ability to draw causal conclusions regarding the impact of AI-generated feedback on the correct use of communication techniques. Finally, the moderate κ values for agreement between the 2 raters of the communication techniques in the pretest likely reflect the complexity of the techniques, whose application can vary even among experts, making uniform assessment difficult. This inherent variability has to be considered when interpreting the results and examined more closely in future studies. Potential risks that also have to be considered especially in the field of medicine include privacy issues, overreliance on automated feedback, model bias, and occasional inaccuracies [59]. It is only if AI feedback is entirely reliable that it can be considered a suitable replacement for a human supervisor. Ethical aspects such as transparency, accountability, and data privacy must also be taken into account.

Future research may benefit from refining AI models to improve feedback reliability and accuracy. Additionally, incorporating longitudinal designs may help to investigate long-term retention of communication skills. Objective outcome measures, such as assessments by patients or recordings of real consultations, could provide a framework for further evaluation of the impact on patient outcomes. Expanding studies to larger, more diverse, multicenter samples may help confirm and generalize these findings. To strengthen causal inference, future research could consider comparisons with a nonintervention control group and randomized controlled trials. This can help clarify the causal contribution of AI-generated feedback to learning new communication techniques.

Conclusion

In conclusion, this proof-of-concept study highlights the potential of AI-generated feedback in enhancing physicians' communication techniques. The study addressed the research questions by demonstrating that the AI-generated feedback was mostly reliable and was rated as accurate and helpful by the participants. In addition, it has been shown that the AI-supported training resulted in an increased application of specific communication techniques, and that participants reported benefits for their daily clinical practice. These findings suggest that AI-based feedback may represent a valuable addition to medical communications training while also highlighting areas for further research and refinement.

Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy [60], the following tasks were delegated to GAI tools under full human supervision: Literature search and systematization, Adapting and adjusting emotional ton, and Translation and Identification of limitations. The GAI tool used was ChatGPT-5 (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. This declaration was submitted by LH. Additionally, after using this tool, the authors thoroughly reviewed, revised, and edited the content and take full responsibility for the final version of the manuscript.

Funding

The study itself received no specific funding. We acknowledge support from the Open Access Publishing Fund of the University of Tübingen.

Data Availability

The data that support the findings of this study are available from the corresponding author (LH) upon reasonable request.

Authors' Contributions

Conceptualization: FH and LH

Data curation: AS and RE

Formal analysis: LH and TFW

Resources: AHW and TFW

Software: CSP and BH

Supervision: AHW and TFW

Visualization: LH

Writing—original draft: FH and LH

Writing—review and editing: LH, TFW, CSP, AS, BH, RE, AHW, and FH

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exemplary feedback on communication techniques used.

[\[DOCX File , 87 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Real-time feedback prompts.

[\[DOCX File , 16 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Prompts.

[\[DOCX File , 27 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Review and evaluation of the AI-generated real-time feedback by experts. AI: artificial intelligence.

[\[PNG File , 73 KB-Multimedia Appendix 4\]](#)

References

1. Yaqoob Mohammed Al Jabri F, Kvist T, Azimirad M, Turunen H. A systematic review of healthcare professionals' core competency instruments. *Nurs Health Sci*. Mar 26, 2021;23(1):87-102. [doi: [10.1111/nhs.12804](https://doi.org/10.1111/nhs.12804)] [Medline: [33386675](https://pubmed.ncbi.nlm.nih.gov/33386675/)]
2. Fritzsche K, Schweickhardt A. Ärztliche Gesprächsführung. In: Fritzsche K, Wirsching M, editors. *Basiswissen Psychosomatische Medizin und Psychotherapie*. Berlin. Springer; 2020:57-69.
3. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Educ Couns*. Mar 2009;74(3):295-301. [doi: [10.1016/j.pec.2008.11.015](https://doi.org/10.1016/j.pec.2008.11.015)] [Medline: [19150199](https://pubmed.ncbi.nlm.nih.gov/19150199/)]

4. Venktaramana V, Loh EKY, Wong CJW, Yeo JW, Teo AYT, Chiam CSY, et al. A systematic scoping review of communication skills training in medical schools between 2000 and 2020. *Med Teach*. Sep 2022;44(9):997-1006. [doi: [10.1080/0142159X.2022.2054693](https://doi.org/10.1080/0142159X.2022.2054693)] [Medline: [35653622](https://pubmed.ncbi.nlm.nih.gov/35653622/)]
5. Winfield S, Hyland D. A literature review of whether communication skills specific to psychiatry are being taught to medical undergraduates around the world. *BJPsych open*. Jun 18, 2021;7(S1):S301-S302. [doi: [10.1192/bjo.2021.797](https://doi.org/10.1192/bjo.2021.797)]
6. Beullens J, Rethans JJ, Goedhuys J, Buntinx F. The use of standardized patients in research in general practice. *Fam Pract*. Feb 01, 1997;14(1):58-62. [doi: [10.1093/fampra/14.1.58](https://doi.org/10.1093/fampra/14.1.58)] [Medline: [9061346](https://pubmed.ncbi.nlm.nih.gov/9061346/)]
7. MacLean S, Kelly M, Geddes F, Della P. Use of simulated patients to develop communication skills in nursing education: An integrative review. *Nurse Educ Today*. Jan 2017;48:90-98. [doi: [10.1016/j.nedt.2016.09.018](https://doi.org/10.1016/j.nedt.2016.09.018)] [Medline: [27741440](https://pubmed.ncbi.nlm.nih.gov/27741440/)]
8. Quail M, Brundage SB, Spitalnick J, Allen PJ, Beilby J. Student self-reported communication skills, knowledge and confidence across standardised patient, virtual and traditional clinical learning environments. *BMC Med Educ*. Feb 27, 2016;16(1):73. [FREE Full text] [doi: [10.1186/s12909-016-0577-5](https://doi.org/10.1186/s12909-016-0577-5)] [Medline: [26919838](https://pubmed.ncbi.nlm.nih.gov/26919838/)]
9. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, Festl-Wietek T, Holderried M, Eickhoff C, et al. A language model-powered simulated patient with automated feedback for history taking: Prospective study. *JMIR Med Educ*. Aug 16, 2024;10:e59213. [FREE Full text] [doi: [10.2196/59213](https://doi.org/10.2196/59213)] [Medline: [39150749](https://pubmed.ncbi.nlm.nih.gov/39150749/)]
10. Holderried F, Stegemann-Philipps C, Herschbach L, Moldt J, Nevins A, Griewatz J, et al. A Generative Pretrained Transformer (GPT)-powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study. *JMIR Med Educ*. Jan 16, 2024;10:e53961. [FREE Full text] [doi: [10.2196/53961](https://doi.org/10.2196/53961)] [Medline: [38227363](https://pubmed.ncbi.nlm.nih.gov/38227363/)]
11. Nagi F, Salih R, Alzubaidi M, Shah H, Alam T, Shah Z, et al. Applications of artificial intelligence (AI) in medical education: A scoping review. In: *Studies in Health Technology and Informatics*. Volume 305: Healthcare Transformation with Informatics and Artificial Intelligence. Amsterdam. IOS Press; 2023:648-651.
12. Zhang W, Cai M, Lee HJ, Evans R, Zhu C, Ming C. AI in medical education: Global situation, effects and challenges. *Educ Inf Technol*. Jul 10, 2023;29(4):4611-4633. [doi: [10.1007/s10639-023-12009-8](https://doi.org/10.1007/s10639-023-12009-8)]
13. Narayanan S, Ramakrishnan R, Durairaj E, Das A. Artificial intelligence revolutionizing the field of medical education. *Cureus*. Nov 2023;15(11):e49604. [FREE Full text] [doi: [10.7759/cureus.49604](https://doi.org/10.7759/cureus.49604)] [Medline: [38161821](https://pubmed.ncbi.nlm.nih.gov/38161821/)]
14. Ghorashi N, Ismail A, Ghosh P, Sidawy A, Javan R. AI-powered chatbots in medical education: Potential applications and implications. *Cureus*. Aug 2023;15(8):e43271. [FREE Full text] [doi: [10.7759/cureus.43271](https://doi.org/10.7759/cureus.43271)] [Medline: [37692629](https://pubmed.ncbi.nlm.nih.gov/37692629/)]
15. Hattie J, Timperley H. The Power of Feedback. *Rev Educ Res*. Mar 01, 2007;77(1):81-112. [doi: [10.3102/003465430298487](https://doi.org/10.3102/003465430298487)]
16. Herrmann-Werner A, Loda T, Erschens R, Schneider P, Junne F, Gilligan C, et al. Face yourself! - learning progress and shame in different approaches of video feedback: A comparative study. *BMC Med Educ*. Mar 27, 2019;19(1):88. [FREE Full text] [doi: [10.1186/s12909-019-1519-9](https://doi.org/10.1186/s12909-019-1519-9)] [Medline: [30917820](https://pubmed.ncbi.nlm.nih.gov/30917820/)]
17. Clynes MP, Raftery SE. Feedback: an essential element of student learning in clinical practice. *Nurse Educ Pract*. Nov 2008;8(6):405-411. [doi: [10.1016/j.nepr.2008.02.003](https://doi.org/10.1016/j.nepr.2008.02.003)] [Medline: [18372216](https://pubmed.ncbi.nlm.nih.gov/18372216/)]
18. Rollnick S, Miller WR. What is motivational interviewing? *Behav Cogn Psychother*. Jun 16, 2009;23(4):325-334. [doi: [10.1017/S135246580001643X](https://doi.org/10.1017/S135246580001643X)]
19. Imel ZE, Pace BT, Soma CS, Tanana M, Hirsch T, Gibson J, et al. Design feasibility of an automated, machine-learning based feedback system for motivational interviewing. *Psychotherapy (Chic)*. Jun 2019;56(2):318-328. [doi: [10.1037/pst0000221](https://doi.org/10.1037/pst0000221)] [Medline: [30958018](https://pubmed.ncbi.nlm.nih.gov/30958018/)]
20. Pellemans M, Salmi S, Mérelle S, Janssen W, van der Mei R. Automated behavioral coding to enhance the effectiveness of motivational interviewing in a chat-based suicide prevention helpline: Secondary analysis of a clinical trial. *J Med Internet Res*. Aug 01, 2024;26:e53562. [FREE Full text] [doi: [10.2196/53562](https://doi.org/10.2196/53562)] [Medline: [39088244](https://pubmed.ncbi.nlm.nih.gov/39088244/)]
21. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Sri Kumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. *J Subst Abuse Treat*. Jun 2016;65:43-50. [FREE Full text] [doi: [10.1016/j.jsat.2016.01.006](https://doi.org/10.1016/j.jsat.2016.01.006)] [Medline: [26944234](https://pubmed.ncbi.nlm.nih.gov/26944234/)]
22. Tanana MJ, Soma CS, Sri Kumar V, Atkins DC, Imel ZE. Development and evaluation of ClientBot: Patient-like conversational agent to train basic counseling skills. *J Med Internet Res*. Jul 15, 2019;21(7):e12529. [FREE Full text] [doi: [10.2196/12529](https://doi.org/10.2196/12529)] [Medline: [31309929](https://pubmed.ncbi.nlm.nih.gov/31309929/)]
23. Grattan S, Davies K, Weise J, Burns N, Murray R, Weldon J, et al. Development of a framework of the skills and attributes needed by mental health professionals to provide optimal clinical care to people experiencing complex psychosis: A Delphi consensus study. *Aust N Z J Psychiatry*. Jan 2025;59(1):60-73. [doi: [10.1177/00048674241289032](https://doi.org/10.1177/00048674241289032)] [Medline: [39460570](https://pubmed.ncbi.nlm.nih.gov/39460570/)]
24. Papageorgiou A, Loke YK, Fromage M. Communication skills training for mental health professionals working with people with severe mental illness. *Cochrane Database Syst Rev*. Jun 13, 2017;6(6):CD010006. [FREE Full text] [doi: [10.1002/14651858.CD010006.pub2](https://doi.org/10.1002/14651858.CD010006.pub2)] [Medline: [28613384](https://pubmed.ncbi.nlm.nih.gov/28613384/)]
25. Priebe S, Dimic S, Wildgrube C, Jankovic J, Cushing A, McCabe R. Good communication in psychiatry--a conceptual review. *Eur Psychiatry*. Oct 16, 2011;26(7):403-407. [doi: [10.1016/j.eurpsy.2010.07.010](https://doi.org/10.1016/j.eurpsy.2010.07.010)] [Medline: [21571504](https://pubmed.ncbi.nlm.nih.gov/21571504/)]
26. GBD Results. Institute for Health Metrics and Evaluation. 2023. URL: <https://vizhub.healthdata.org/gbd-results> [accessed 2024-08-18]

27. Vos T. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. Oct 17, 2020;396(10262):1204-1222. [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)] [Medline: [33069326](https://pubmed.ncbi.nlm.nih.gov/33069326/)]
28. de Vries U, Schüßler G, Petermann F. Psychotherapie in der Psychosomatik – Trends in Diagnostik und Therapie. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*. Sep 2012;60(4):301-307. [doi: [10.1024/1661-4747/a000130](https://doi.org/10.1024/1661-4747/a000130)]
29. Larisch A, Heuft G, Engbrink S, Brähler E, Herzog W, Kruse J. [Treatment of mental and psychosomatic complaints - how the general population in Germany gets treatment, and what it expects and knows about treatment alternatives]. *Z Psychosom Med Psychother*. May 01, 2013;59(2):153-169. [doi: [10.13109/zptm.2013.59.2.153](https://doi.org/10.13109/zptm.2013.59.2.153)] [Medline: [23775554](https://pubmed.ncbi.nlm.nih.gov/23775554/)]
30. Bradley CP. GPs and mental health. *Eur J Gen Pract*. 2008;14(1):2-3. [FREE Full text] [doi: [10.1080/13814780802136925](https://doi.org/10.1080/13814780802136925)] [Medline: [18464165](https://pubmed.ncbi.nlm.nih.gov/18464165/)]
31. Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *J Affect Disord*. Apr 2011;130(1-2):26-36. [doi: [10.1016/j.jad.2010.07.028](https://doi.org/10.1016/j.jad.2010.07.028)] [Medline: [20708274](https://pubmed.ncbi.nlm.nih.gov/20708274/)]
32. Turki M, Babbah T, Abid W, Ellouze S, Ouali R, Charfi R, et al. Mental disorders management in general practice. *Eur Psychiatr*. Aug 13, 2021;64(S1):S397-S398. [doi: [10.1192/j.eurpsy.2021.1065](https://doi.org/10.1192/j.eurpsy.2021.1065)]
33. Fleury M, Imboua A, Aubé D, Farand L, Lambert Y. General practitioners' management of mental disorders: a rewarding practice with considerable obstacles. *BMC Fam Pract*. Mar 16, 2012;13(1):19. [FREE Full text] [doi: [10.1186/1471-2296-13-19](https://doi.org/10.1186/1471-2296-13-19)] [Medline: [22423592](https://pubmed.ncbi.nlm.nih.gov/22423592/)]
34. Ba S, Yang L, Yan Z, Looi CK, Gašević D. Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: A systematic literature review. *Comps Educ Open*. Dec 2025;9:100284. [doi: [10.1016/j.caeo.2025.100284](https://doi.org/10.1016/j.caeo.2025.100284)]
35. Biswas SS. Role of Chat GPT in public health. *Ann Biomed Eng*. May 15, 2023;51(5):868-869. [doi: [10.1007/s10439-023-03172-7](https://doi.org/10.1007/s10439-023-03172-7)] [Medline: [36920578](https://pubmed.ncbi.nlm.nih.gov/36920578/)]
36. Zheng L, Niu J, Zhong L. Effects of a learning analytics - based real - time feedback approach on knowledge elaboration, knowledge convergence, interactive relationships and group performance in CSCL. *Brit J Educational Tech*. Aug 13, 2021;53(1):130-149. [doi: [10.1111/bjet.13156](https://doi.org/10.1111/bjet.13156)]
37. Holderried F, Sonanini A, Philipps A, Stegemann-Philipps C, Herschbach L, Festl-Wietek T, et al. Using AI to train future clinicians in depression assessment: a feasibility study. *JMI Preprints*. Preprint posted online November 4, 2025. [doi: [10.2196/preprints.87102](https://doi.org/10.2196/preprints.87102)]
38. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Washington, DC. American Psychiatric Association Publishing; 2013.
39. Rogers CR. The necessary and sufficient conditions of therapeutic personality change. *J Consult Psychol*. Apr 1957;21(2):95-103. [doi: [10.1037/h0045357](https://doi.org/10.1037/h0045357)] [Medline: [13416422](https://pubmed.ncbi.nlm.nih.gov/13416422/)]
40. Linehan M. Validation and psychotherapy. In: Bohart AC, Greenberg LS, editors. *Empathy reconsidered: New directions in psychotherapy*. Washington, DC. American Psychological Association; 1997:353-392.
41. Gnepp J, Klayman J, Williamson I, Barlas S. The future of feedback: Motivating performance improvement through future-focused feedback. *PLoS One*. 2020;15(6):e0234444. [FREE Full text] [doi: [10.1371/journal.pone.0234444](https://doi.org/10.1371/journal.pone.0234444)] [Medline: [32559254](https://pubmed.ncbi.nlm.nih.gov/32559254/)]
42. Hardavella G, Aamli-Gagnat A, Saad N, Rousalova I, Sreter KB. How to give and receive feedback effectively. *Breathe (Sheff)*. Dec 30, 2017;13(4):327-333. [FREE Full text] [doi: [10.1183/20734735.009917](https://doi.org/10.1183/20734735.009917)] [Medline: [29209427](https://pubmed.ncbi.nlm.nih.gov/29209427/)]
43. Rijsbergen RC. Information retrieval: theory and practice. Newcastle University. URL: <http://homepages.cs.ncl.ac.uk/brian.randell/Seminars/146.pdf> [accessed 2025-07-31]
44. Höschel S. *Therapie-Basics Dialektisch-Behaviorale Therapie (DBT): Mit E-Book inside und Arbeitsmaterial*. Weinheim. Julius Beltz GmbH & Co. KG; 2023.
45. Nissen L, Rother JF, Heinemann M, Reimer LM, Jonas S, Raupach T. A randomised cross-over trial assessing the impact of AI-generated individual feedback on written online assignments for medical students. *Med Teach*. Sep 2025;47(9):1544-1550. [doi: [10.1080/0142159X.2025.2451870](https://doi.org/10.1080/0142159X.2025.2451870)] [Medline: [39831699](https://pubmed.ncbi.nlm.nih.gov/39831699/)]
46. Steiss J, Tate T, Graham S, Cruz J, Hebert M, Wang J, et al. Comparing the quality of human and ChatGPT feedback of students' writing. *Learn Instr*. Jun 2024;91:101894. [doi: [10.1016/j.learninstruc.2024.101894](https://doi.org/10.1016/j.learninstruc.2024.101894)]
47. Er E, Akçapınar G, Bayazit A, Noroozi O, Banihashem SK. Assessing student perceptions and use of instructor versus AI-generated feedback. *Brit J Educational Tech*. Dec 27, 2024;56(3):1074-1091. [doi: [10.1111/bjet.13558](https://doi.org/10.1111/bjet.13558)]
48. Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. *Breast*. Apr 2020;50:49-55. [FREE Full text] [doi: [10.1016/j.breast.2020.01.008](https://doi.org/10.1016/j.breast.2020.01.008)] [Medline: [32007704](https://pubmed.ncbi.nlm.nih.gov/32007704/)]
49. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Acad Med*. 2010;85(10):1589-1602. [doi: [10.1097/acm.0b013e3181edfe13](https://doi.org/10.1097/acm.0b013e3181edfe13)]
50. Celik I, Gedrimiene E, Siklander S, Muukkonen H. The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education. *Australas J Educ Technol*. Jun 01, 2024;40(3):19-39. [doi: [10.14742/AJET.9069](https://doi.org/10.14742/AJET.9069)]
51. García-Torres D, Vicente Ripoll MA, Fernández Peris C, Mira Solves JJ. Enhancing clinical reasoning with virtual patients: A hybrid systematic review combining human reviewers and ChatGPT. *Healthcare (Basel)*. Nov 11, 2024;12(22):2241. [FREE Full text] [doi: [10.3390/healthcare12222241](https://doi.org/10.3390/healthcare12222241)] [Medline: [39595439](https://pubmed.ncbi.nlm.nih.gov/39595439/)]

52. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: A systematic review. *Med Educ*. Sep 22, 2020;54(9):786-795. [doi: [10.1111/medu.14152](https://doi.org/10.1111/medu.14152)] [Medline: [32162355](https://pubmed.ncbi.nlm.nih.gov/32162355/)]
53. Bruckner TA, Scheffler RM, Shen G, Yoon J, Chisholm D, Morris J, et al. The mental health workforce gap in low- and middle-income countries: a needs-based approach. *Bull World Health Organ*. Mar 01, 2011;89(3):184-194. [FREE Full text] [doi: [10.2471/BLT.10.082784](https://doi.org/10.2471/BLT.10.082784)] [Medline: [21379414](https://pubmed.ncbi.nlm.nih.gov/21379414/)]
54. Butryn T, Bryant L, Marchionni C, Sholevar F. The shortage of psychiatrists and other mental health providers. Causes, current state, and potential solutions. *Int J Acad Med*. 2017;3(1):5. [doi: [10.4103/IJAM.IJAM_49_17](https://doi.org/10.4103/IJAM.IJAM_49_17)]
55. Kononowicz AA, Woodham LA, Edelbring S, Stathakourou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: Systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res*. Jul 02, 2019;21(7):e14676. [FREE Full text] [doi: [10.2196/14676](https://doi.org/10.2196/14676)] [Medline: [31267981](https://pubmed.ncbi.nlm.nih.gov/31267981/)]
56. Badyal DK, Bala S, Singh T, Gulrez G. Impact of immediate feedback on the learning of medical students in pharmacology. *J Adv Med Educ Prof*. Jan 2019;7(1):1-6. [FREE Full text] [doi: [10.30476/JAMP.2019.41036](https://doi.org/10.30476/JAMP.2019.41036)] [Medline: [30697542](https://pubmed.ncbi.nlm.nih.gov/30697542/)]
57. Garner MS, Gusberg RJ, Kim AW. The positive effect of immediate feedback on medical student education during the surgical clerkship. *J Surg Educ*. May 2014;71(3):391-397. [doi: [10.1016/j.jsurg.2013.10.009](https://doi.org/10.1016/j.jsurg.2013.10.009)] [Medline: [24797856](https://pubmed.ncbi.nlm.nih.gov/24797856/)]
58. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ*. Mar 14, 2024;17(5):926-931. [doi: [10.1002/ase.2270](https://doi.org/10.1002/ase.2270)] [Medline: [36916887](https://pubmed.ncbi.nlm.nih.gov/36916887/)]
59. Heston T, Khun C. Prompt engineering in medical education. *IME*. Aug 31, 2023;2(3):198-205. [doi: [10.3390/ime2030019](https://doi.org/10.3390/ime2030019)]
60. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Account Res*. Aug 08, 2025;1-27. [doi: [10.1080/08989621.2025.2544331](https://doi.org/10.1080/08989621.2025.2544331)] [Medline: [40781729](https://pubmed.ncbi.nlm.nih.gov/40781729/)]

Abbreviations

AI: artificial intelligence

DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)

GP: general practitioner

LLM: large language model

MI: motivational interviewing

QAP: question-answer pair

Edited by J Sarvestan; submitted 22.Aug.2025; peer-reviewed by AC de Oliveira, W Wei; comments to author 24.Sep.2025; accepted 06.Nov.2025; published 28.Nov.2025

Please cite as:

Herschbach L, Festl-Wietek T, Stegemann-Philipps C, Sonanini A, Herrmann B, Erschens R, Herrmann-Werner A, Holderried F. Evaluation of an AI-Based Chatbot Providing Real-Time Feedback in Communication Training for Mental Health Care Professionals: Proof-of-Concept Observational Study. *J Med Internet Res* 2025;27:e82818

URL: <https://www.jmir.org/2025/1/e82818>

doi: [10.2196/82818](https://doi.org/10.2196/82818)

PMID:

©Lea Herschbach, Teresa Festl-Wietek, Christian Stegemann-Philipps, Alessandra Sonanini, Bernd Herrmann, Rebecca Erschens, Anne Herrmann-Werner, Friederike Holderried. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 28.Nov.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.