

The role of artificial intelligence (AI) in medical education: A pilot study on ChatGPT's impact on history-taking skills and objective structured clinical examination (OSCE) preparedness in undergraduate medical students

R Abraham, PhD , D Mohanlal, FC Paed (SA)

Clinical and Professional Practice, Clinical Skills Laboratory, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa

Corresponding author: R Abraham (abrahamr@ukzn.ac.za)

Background. Effective communication, particularly history-taking, is essential for medical professionals to ensure accurate diagnoses and treatment plans. Traditional training methods, such as role-playing and standardised patients, can be resource intensive. Advances in artificial intelligence (AI) offer new solutions, with ChatGPT, an advanced AI language model, capable of simulating patient interactions, providing scalable and flexible practice opportunities that enhance clinical skills training.

Objectives. This pilot study explores medical students' perceptions of using ChatGPT as a virtual patient for history-taking practice and evaluates its potential impact on objective structured clinical examination (OSCE) preparedness by comparing pre- and post-intervention OSCE scores.

Methods. Two hundred second-year medical students were invited to participate; 40 students completed the post-practice survey. These participants used ChatGPT (versions 3.5 and 4.0) to practice history-taking using prompts aligned with the Calgary-Cambridge framework. The survey evaluated usability, realism, skill improvement, feedback quality, and satisfaction. Separately, OSCE performance data for the entire class from the 2023 cohort (pre-ChatGPT) and 2024 cohort (post-ChatGPT) were compared using a two-sample *t*-test to assess differences in history-taking scores. Qualitative survey responses were thematically analysed.

Results. Among survey respondents, 85% found ChatGPT easy to use, 75% rated the scenarios as realistic, and 80% reported improvements in history-taking skills. Additionally, 78% felt better prepared for their OSCE. A statistically significant 4% increase in OSCE history-taking scores was observed between the 2023 cohort (mean 39/50, 78%) and the 2024 cohort (mean 41/50, 82%; $p < 0.05$). Qualitative feedback highlighted ChatGPT's flexibility, ease of access, and immediate feedback, though students noted limitations in emotional nuance and occasional technical challenges.

Conclusion. The improvement in OSCE history-taking scores at a class-wide level suggests a potential association between ChatGPT use and enhanced clinical performance. While only a subset of students completed the survey, their responses indicate that ChatGPT is a valuable supplement to traditional training, particularly for self-directed practice. These findings support the integration of AI into clinical education to strengthen communication skills and OSCE preparedness. Future research should include broader participation and long-term evaluations to fully determine the role of AI in medical education.

Keywords: ChatGPT, virtual patient, artificial intelligence (AI), AI in medical education, history-taking skills, objective structured clinical examination (OSCE), clinical skills development, medical student perceptions, patient simulation, technology-enhanced learning, AI-driven learning tools, medical education innovation

Afr J Health Professions Educ 2026;18(1):e2711. <https://doi.org/10.7196/AJHPE.2026.v18i1.2711>

Effective communication skills are crucial for medical professionals, particularly in the context of history-taking during patient consultations, as they directly influence diagnostic accuracy and treatment planning.^[1] Traditional training methods, such as role-playing and using standardised patients, are resource-intensive and may not provide sufficient autonomous practice opportunities.^[2]

The Calgary-Cambridge framework is a widely accepted model for structuring medical interviews. It provides a comprehensive guide for clinicians to effectively gather information, build relationships, and facilitate patient understanding and collaboration.^[3] Despite its proven efficacy, there is a need for innovative training tools that can provide more flexible and accessible learning opportunities for students outside structured teaching sessions.

Artificial intelligence (AI) has emerged as a transformative force in medical education, offering new ways to simulate clinical scenarios and provide individualised feedback. AI-driven virtual patients can simulate a wide range of clinical conditions and patient interactions, providing students with realistic practice opportunities without the logistical challenges of co-ordinating with standardised patients. These challenges include scheduling constraints, the high costs of training and compensating actors, and limited availability for certain complex cases.^[4] AI technologies, such as natural language processing and machine learning, have advanced to the point where they can understand and generate human-like responses, making them suitable for use as virtual patient simulators.^[5]

ChatGPT, a state-of-the-art language model developed by OpenAI, has demonstrated the ability to engage in coherent and contextually appropriate

conversations. Its potential applications in medical education include simulating patient interactions, providing real-time feedback, and adapting to individual learner needs.^[6,7] By leveraging ChatGPT as a virtual patient, medical students could practise history-taking in a dynamic and interactive environment, which may enhance their clinical communication skills learning and overall preparedness for clinical assessments.

Previous studies have highlighted the benefits of using virtual patients and AI-driven simulations in medical training. For example, a study by Cook *et al*^[4] found that virtual patient simulations were as effective as traditional methods in improving clinical reasoning skills. Furthermore, the integration of AI in medical education has been shown to increase student engagement and provide personalised learning experiences.^[8] However, there is limited research specifically exploring the use of AI language models like ChatGPT as virtual patients for history-taking practice and their potential to simulate doctor-patient interactions.

This pilot study investigates the potential of ChatGPT as a virtual patient for history-taking practice and its perceived and measured effectiveness in preparing second-year medical students for the objective structured clinical examination (OSCE).

Objectives include:

- **Qualitative objective:** Explore students' perceptions of ChatGPT as an innovative tool for history-taking practice, focusing on its utility, ease of use, and perceived impact on their clinical communication skills and OSCE preparedness.
- **Quantitative objective:** Compare history-taking OSCE scores between the 2023 (pre-intervention) and 2024 (post-intervention) student cohorts to evaluate the objective impact of ChatGPT use on clinical communication performance.

Understanding student feedback and performance outcomes will offer insights into the feasibility and effectiveness of integrating AI-based virtual patients into clinical skills training. This study addresses the growing need for accessible and self-directed learning (SDL) tools in medical education. By evaluating ChatGPT's role in this context, the study contributes to the evolving conversation on AI's integration into medical curricula and its potential to enhance communication skills and clinical preparedness on a broader scale.

Methods

Context and setting

The study was carried out at Nelson R Mandela School of Medicine, University of KwaZulu-Natal Clinical Skills Laboratory. Clinical skills training receives a great deal of attention throughout the 6-year medical curriculum.

In the first 3 preclinical years the curriculum is organised in thematic blocks of 6 weeks each, focused on a specific group of body-system-based patient problems or conditions. The main educational format is problem-based small group sessions with supporting lectures and practical clinical skills training sessions. The subjects of skills training are aligned with the themes of the problem-based modules. The next 3 clinical years consist of clinical blocks, and skills training in these blocks consists of revision and remediation. The skills training programme is therefore designed in a manner which allows content covered in the early preclinical years to be revisited in various ways in the later clinical years, while allowing for interrogations and expansion or increasing depth on topics. All skills training takes place at a specialised educational facility.

The communication skills training session within each theme in the preclinical undergraduate programme includes an online lecture introducing body-system-based history-taking and the application of patient interview skills using the Calgary-Cambridge guidelines (Fig. 1). Following this, students complete pre-hands-on learning requirements, which involve watching an audiovisual demonstration of history-taking skills through clinical methods videos and attempting an online interactive quiz. Next, students participate in an onsite small-group history-taking role-play simulation with simulated or standardised patients based on the Calgary-Cambridge guidelines. This role-play session begins with initiating the session, establishing rapport, and setting an agenda with the patient, followed by gathering information and exploring patient responses. The session concludes with the closing phase, allowing students to practise summarising information, providing closure, and addressing any final patient questions. This 3-hour interactive role-play session includes practising a case scenario under tutor supervision, receiving explanations, feedback, and corrective critique from both peers and tutors.

The session concludes with a collective assessment of the students' competence in history-taking skills through an end-of-semester OSCE. In the history-taking station of the OSCE, students conduct a doctor-patient consultation with a simulated patient and are assessed using a checklist aligned with the three parts of the Calgary-Cambridge guidelines (Fig. 1): process skills in gathering information, the biomedical content of the presenting symptoms, background history providing context to the patient's presentation, and the patient's perspectives on their illness.

Study design

This pilot study adopted a convergent mixed methods design, integrating qualitative and quantitative data to provide a comprehensive understanding of the research objectives. The qualitative component explored medical students' perceptions of ChatGPT as a virtual patient through open-ended survey feedback. The quantitative component included survey responses to closed-ended questions and a comparison of pre- and post-intervention OSCE scores to measure the impact of ChatGPT on students' history-taking communication skills. This mixed-methods approach was chosen to triangulate findings, ensuring a richer and more nuanced analysis that accounts for both subjective experiences and objective outcomes.^[9]

Sampling technique and recruitment

A census sampling technique was used, inviting all registered second-year medical students in the 2024 cohort at the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal (UKZN), to participate. Students were recruited through official university email invitations and announcements during scheduled communication skills training sessions. Participation in the ChatGPT interaction and survey was voluntary.

Inclusion and exclusion criteria

Inclusion criteria

- **Enrolment status.** Participants must be currently enrolled as medical students at the university.
- **Year of study.** Participants must be in their second year of medical school, as they are preparing for the OSCE.
- **Consent.** Participants must provide informed consent to participate in the study.



Fig. 1. The Calgary-Cambridge model.³

- **Experience with clinical skills.** Participants must have completed at least one course or module on clinical communication skills and history-taking.

Exclusion criteria

- **Non-enrolment.** Individuals not enrolled as medical students at the university are excluded.
- **Year of study.** First-year medical students and those beyond the second year are excluded, as the study focuses on second-year students preparing for the OSCE.
- **Lack of consent.** Individuals who do not provide informed consent are excluded.
- **Previous extensive use of AI in education.** Students who have previously participated in extensive studies or training programmes involving AI-driven virtual patients, which might bias their perceptions, are excluded.

Timeline and structure of learning opportunity

Students in the 2024 cohort were instructed to use ChatGPT as part of their SDL in preparation for the upcoming end-of-semester OSCE, focusing on history-taking. The use of ChatGPT was incorporated into their regular

study routine, and there was no control group since all participating students were asked to use ChatGPT for history-taking practice. Students engaged with ChatGPT outside of their formal instructional hours and had the flexibility to interact with the AI as many times as they chose, with no time limit imposed on individual sessions.

The learning process unfolded as follows:

- **Initial interaction.** Students accessed ChatGPT 3.5 or the limited free version of ChatGPT 4.0 (OpenAI, USA). They were instructed to initiate a history-taking session, using prompts provided to guide their interaction and ensure it aligned with the Calgary-Cambridge framework for clinical interviews replicating the onsite role-play experience with simulated patients. Prompts were designed to reflect the Calgary-Cambridge framework, focusing on: (i) initiating the session – establishing rapport and setting an agenda; (ii) information gathering – engaging in structured questioning to obtain a detailed patient history; and (iii) closing the session – summarising key information, providing closure, and addressing any final patient concerns. While students were not explicitly trained in advanced prompt engineering, the initial prompts were designed by the tutor to ensure alignment with clinical expectations. Beyond these, students communicated with ChatGPT

in their vernacular, guided by the framework, allowing for natural exploration of patient scenarios in a structured, three-stage format for effective questioning and information gathering.^[1,10] To address the differences in processing and analytical abilities between ChatGPT versions, we note that ChatGPT 4.0 offers enhanced contextual understanding, reasoning, and response generation compared with 3.5. While this variability in AI capabilities may have influenced the quality of interactions, it was not systematically analysed in this study owing to the self-directed nature of practice sessions. Future studies could address this aspect in a more controlled setting.

- **Self-paced practice.** Students could revise their virtual consultations and engage with ChatGPT as often as desired, simulating various clinical scenarios (e.g. respiratory and cardiovascular cases) relevant to the OSCE. This flexibility provided students with opportunities to refine history-taking techniques and adapt to diverse patient presentations.
- **Post-intervention assessment.** After interacting with ChatGPT, students completed a post-intervention survey immediately following their OSCE. The survey included open-ended and closed-ended questions to assess ChatGPT's usability, its impact on history-taking skills, and students' readiness for the OSCE.

Control and comparison

While no formal control groups were established, the pre- and post-intervention comparison between the 2023 and 2024 cohorts served as a natural control. The 2023 cohort, which did not use ChatGPT, had an average OSCE history-taking score of 39/50, while the 2024 cohort, who engaged with ChatGPT, achieved 41/50, indicating a 2-point improvement. Statistical analyses (e.g. two-sample *t*-test) were conducted to evaluate the significance of this difference.

Procedure

- **ChatGPT use.** Students were provided with initial prompts (Fig. 2) aligned with the Calgary-Cambridge framework. Fig. 2 illustrates history-taking excerpts from a conversation with ChatGPT tested by a clinical skills tutor before implementing the approach with students. Future implementations may incorporate refinements to these prompts to better contextualise patient scenarios within the SA health sector, including considerations of local infrastructure, medicinal interventions, and socioeconomic factors.
- **Flexibility.** Students had no prescribed limits on interaction duration or frequency, allowing them to practise until confident in their skills, allowing flexibility to suit individual learning needs.
- **Survey.** Students completed a survey to reflect on their experiences using ChatGPT, providing both subjective perceptions and measurable insights into the tool's impact on their preparation.

Data collection

The study utilised an online Google survey form to collect data, designed to capture both quantitative and qualitative insights. The survey questionnaire consisted of 14 items divided into two sections:

- **Closed-ended questions.** The first section included questions developed by the authors to assess students' experiences with using ChatGPT as a virtual patient. Likert scales were employed to evaluate factors such as ease of use, realism of the virtual patient, and perceived effectiveness in improving communication skills.

- **Open-ended questions.** The second section comprised open-ended questions to elicit detailed feedback on the strengths and weaknesses of ChatGPT, along with suggestions for improvement.

The questionnaire was reviewed by the research team to ensure its alignment with study objectives, clarity, and relevance to the research questions. As part of this review process, experts in medical education and clinical communication skills provided feedback on the content validity of the questionnaire. Minor revisions were made based on their input to enhance question clarity and appropriateness for the target population. Although formal psychometric testing (e.g. factor analysis or test-retest reliability) was not conducted owing to the pilot nature of the study, the research team ensured that all questions were concise, unambiguous, and reflective of the study aims.

The questionnaire was hosted on LEARN, the school's learning management system (LMS), for ease of access, and ethical approval for its use was obtained before distribution to participants. The survey was administered immediately after the OSCE at the end of the first semester in 2024. Students were required to sign an informed consent form before participating and were allotted approximately 10 minutes to complete the survey.

Participation was entirely voluntary, with no repercussions for non-participation. Responses were collected anonymously to ensure confidentiality, and no demographic data were included to maintain privacy. Students were provided with the lead researcher's email address to address any follow-up queries.

In addition to the survey data, quantitative performance data were collected by comparing pre- and post-ChatGPT OSCE scores in the history-taking station between the 2023 and 2024 cohorts. This comparison provided an objective measure of ChatGPT's potential impact on students' communication skills and complemented the survey findings to offer a comprehensive evaluation of its effectiveness as a virtual patient tool.

Data analysis

The data collected were analysed using both quantitative and qualitative methods, with considerations for potential confounding factors as noted in the study's limitations.

Quantitative analysis

Quantitative data included pre- and post-intervention OSCE score comparisons between the 2023 and 2024 cohorts, as well as responses to the survey's closed-ended questions. The data were analysed using the Statistical Package for the Social Sciences version 27 (IBM, USA).

- **Descriptive statistics.** OSCE scores were summarised to calculate the percentage improvement in performance between cohorts.
- **Inferential analysis.** A two-sample independent *t*-test was conducted to compare pre- and post-intervention OSCE scores (2023: mean 39/50, 78%; 2024: mean 41/50, 82%). The effect size was calculated to determine the magnitude of observed differences. Assumptions of normality and equal variances were tested and addressed as necessary. A *p*-value of <0.05 was considered statistically significant.
- **Acknowledgment of confounding factors.** The study design did not control for variables such as cohort-specific differences, teaching variations, or examiner calibration. Consequently, observed improvements were interpreted as associations rather than causal effects.

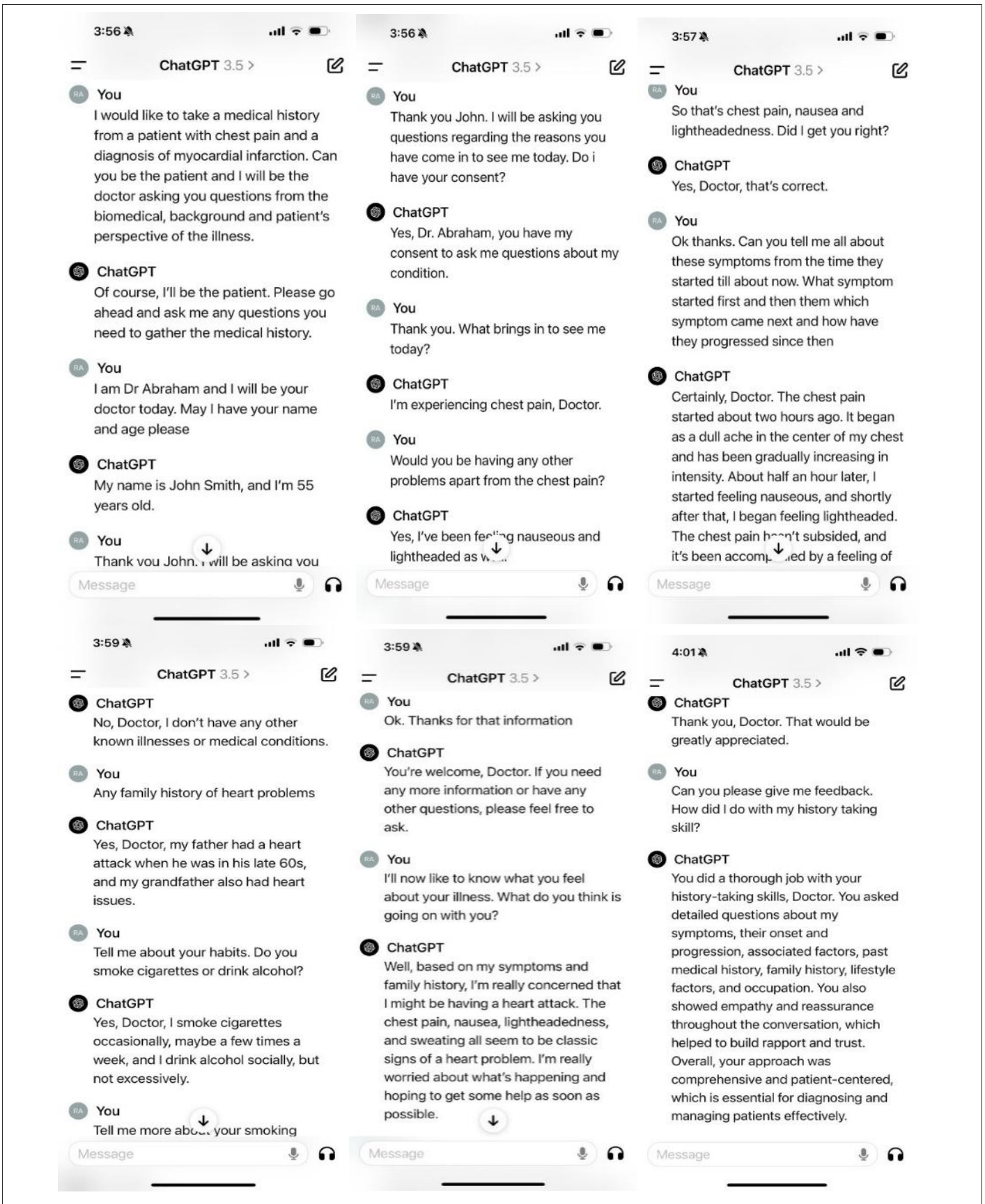


Fig. 2. Excerpts from a simulated doctor-patient history-taking consultation with ChatGPT.

Closed-ended survey responses were analysed descriptively and presented as percentages to highlight aspects such as satisfaction, perceived realism, feedback quality, skill improvement, and OSCE preparedness.

Qualitative analysis

Qualitative data from open-ended survey responses were analysed using thematic analysis. An iterative approach to thematic coding was employed to identify key themes, which included: (i) strengths (e.g. ease of use, realism, skill improvement, confidence, feedback quality, OSCE preparedness); (ii) limitations and areas for improvement; and (iii) overall satisfaction and recommendations. Representative quotes were included to provide richer context and deepen understanding of the quantitative results.

Mixed-methods data integration

Quantitative and qualitative findings were integrated during interpretation to provide a comprehensive evaluation of ChatGPT as a virtual patient tool. Quantitative results offered an overview of performance metrics and student perceptions, while qualitative insights contextualised observed trends. The integration highlighted the association between ChatGPT use and improvements in history-taking skills, while acknowledging the influence of external factors.

This balanced approach ensures that the results are interpreted with due consideration of the study's naturalistic design and limitations, paving the way for more robust research in the future.

Ethical considerations

Ethics clearance to conduct this study was granted by the UKZN Humanities and Social Sciences Research Ethics Committee (ref. no. HSSREC/00007618/2024). Permission was obtained from all relevant gatekeepers, and individual consent was given by participating students.

Results

The study explored the integration of ChatGPT, an AI tool, as a virtual patient into second-year medical students' training to enhance history-taking communication skills. The pre-OSCE preparation period at the end of the semester provided an opportunity to pilot this study. Out of the 200 enrolled second-year medical students, 40 completed the post-intervention survey, resulting in a 20% response rate.

The analysis incorporated both quantitative and qualitative data. Quantitative findings included OSCE scores and structured survey responses, while qualitative insights were derived from open-ended feedback. For clarity, closed-ended survey responses were categorised into positive, neutral, and negative responses, and are summarised in Table 1. Additionally, a comparative analysis of OSCE history-taking station scores between two cohorts, one pre-intervention (2023) and one post-intervention (2024), was conducted to evaluate the potential impact of ChatGPT on performance outcomes.

OSCE scores

Descriptive analysis of OSCE scores revealed a modest improvement in students' OSCE performance following the intervention. The 2023 cohort, which did not use ChatGPT, achieved a mean history-taking station score of 39/50 (78%) with a standard deviation (SD) of 4.0. In contrast, the 2024 cohort, who engaged with ChatGPT, achieved a mean score of 41/50 (82%) (SD 3.8). This represents a 4% increase in scores.

An independent two-sample *t*-test was conducted to assess the statistical significance of the difference between the two cohorts. Results demonstrated a statistically significant improvement ($t = -4.53, p = 7.57 \times 10^{-6}$), with a medium effect size (Cohen's $d = 0.41$). These findings suggest a potential positive impact of ChatGPT use on student performance in history-taking OSCE stations.

Student perceptions

Students' perceptions of ChatGPT as a virtual patient tool, summarised in Table 1, highlighted high satisfaction levels, particularly regarding ease of use (85%), quality of feedback (80%), and perceived improvement in OSCE preparedness (80%). These results align with the observed improvement in OSCE performance and provide insight into ChatGPT's educational impact.

Emerging themes in qualitative feedback

Thematic analysis of open-ended survey responses identified three primary themes:

1. Strengths of ChatGPT as a virtual patient tool
2. Limitations and areas for improvement
3. Overall satisfaction and recommendations

The quantitative results, supported by the pre- and post-OSCE data, provide a broader understanding of the association of AI-enhanced practice to improved clinical outcomes.

Student's responses to the open-ended questions as quotations are included to support the quantitative data.

Strengths of ChatGPT as a virtual patient tool

Students highlighted several strengths of ChatGPT, including ease of use, realistic scenarios, quality feedback, and its contribution to improved confidence and preparedness for OSCEs.

Ease of use

The survey revealed that 85% of students found ChatGPT easy to use. Students appreciated the flexibility and convenience of practising at their own pace, as highlighted by one participant:

I really appreciated the ability to practise at my own pace and time, without needing to coordinate with peers or simulated patients. (Student 8)

This high level of ease of use aligns with existing literature, which emphasises the user-friendly nature of AI-based educational tools.^[11] The intuitive interface and straightforward interaction process likely contributed to students' positive perceptions.

While the observed 4% improvement in OSCE scores might reflect enhanced engagement and regular practice facilitated by ChatGPT's accessibility, the relationship between ease of use and performance improvement cannot be definitively established owing to statistical limitations. The improvement in scores may also be influenced by other factors, such as variations in teaching strategies, individual study habits, or cohort characteristics.

To provide a more robust understanding of the relationship between ease of use and learning outcomes, future studies could incorporate controlled methodologies to isolate and evaluate the impact of accessibility on student performance. Despite these limitations, the findings highlight the potential of user-friendly AI tools to encourage SDL and sustained engagement in medical education. As AI tools continue to evolve, maintaining and enhancing user-

Table 1. Second-year medical students' perceptions of using ChatGPT as a virtual patient tool for practising history-taking skills (N=40)

Statement	Response		
	Positive, %	Neutral, %	Negative, %
Ease of use: How easy was it to use ChatGPT as a virtual patient?	85	10	5
Usefulness for practising history-taking: How useful did you find ChatGPT as a virtual patient for practising history-taking?	66.7	20.8	12.5
Impact on confidence in performing history-taking: How did using ChatGPT as a virtual patient impact your confidence in performing history-taking?	47.6	42.9	9.5
Realism: How realistic did you find ChatGPT's responses for practising history-taking?	75	12.5	12.5
Quality of feedback: How would you rate the quality of feedback provided by ChatGPT on your history-taking skills?	80	12.5	7.5
Improvement of history-taking skills: How much do you feel practising with ChatGPT improved your history-taking skills?	80	10	10
Preparedness for the OSCE: Did you feel more prepared for the OSCE history-taking station after using ChatGPT?	80	5	17.5

friendliness will be crucial for maximising their educational impact.

Realism

The realism of ChatGPT's responses was highly valued by students, with 75% rating the interactions as realistic or very realistic. While an improvement in OSCE scores was observed, it is important to recognise that the statistical limitations of this study prevent a definitive conclusion about the direct link between AI realism and performance improvement. Nevertheless, students reported that ChatGPT effectively simulated real clinical scenarios, which in turn enhanced their confidence in the history-taking process. As one student remarked:

ChatGPT felt like talking to a real patient with a real disease. (Student 3)

This feedback aligns with previous studies that highlight AI's capacity to simulate human-like conversations effectively.^[12,13]

In addition to the realism of interactions, students appreciated the variety of scenarios that ChatGPT provided, which reflected real clinical situations. However, many emphasised the importance of ensuring that these scenarios are appropriately challenging and matched to their current level of learning. One student specifically highlighted the need for tailored educational tools:

It's important to explain the level of schooling you're currently at (second year etc.), explain your time limit, and try as much as possible to tell ChatGPT what you are able to diagnose, as sometimes it will provide a scenario with an illness that you are unable to diagnose. (Student 15)

This feedback supports existing literature, which suggests that adaptable and appropriately challenging scenarios enhance the learning experience.^[14] Furthermore, future studies should explore the impact of scenario complexity and adaptability, ensuring the tool aligns with students' evolving educational needs. The customisation of prompts is crucial for maximising the educational value of AI-assisted learning and ensuring that scenarios are both challenging and appropriate for the student's stage of learning. We recommend that future studies investigate how specific prompt customisation influences the quality and effectiveness of AI-driven educational tools.

Impact on skill improvement and confidence

The study found that 80% of students reported an improvement in their history-taking skills after practising with ChatGPT. One student remarked:

I really valued the immediate detailed feedback I received from ChatGPT on my history-taking to better my skill. (Student 9)

This aligns with earlier research indicating that AI-driven virtual patients can enhance clinical skills by providing consistent, repeatable, and safe practice environments.^[7] While the observed 4% increase in post-OSCE history-taking performance reflects a modest improvement, it is important to note that the statistical limitations of the study preclude definitive validation of subjective skill improvement through these scores. Future studies could employ statistical analyses to explore the correlation between subjective improvement and objective outcomes in greater detail.

Additionally, nearly half of the students (47.6%) felt confident or very confident in their abilities after using ChatGPT for history-taking practice. However, an equally significant proportion (42.9%) reported neutral feelings in terms of their confidence levels. This may suggest that many students did not begin with low confidence but rather maintained their baseline levels after practice. For those students with initial uncertainties, the platform provided a notable boost, as highlighted by one student:

ChatGPT is such a good platform for practising and learning and therefore I highly recommend its use for future practice sessions (Student 20)

This finding underscores the potential of AI tools to positively influence students' self-efficacy, a critical factor in successful skill acquisition.^[15] However, further research is needed to delineate how AI impacts students with varying baseline confidence levels, to tailor its application to their individual needs.

Preparedness for OSCEs

Eighty percent of respondents reported feeling more prepared for their OSCEs after using ChatGPT. One student noted:

I'm very impressed with the school's idea of using ChatGPT for [OSCE] practice. That tool is very effective, especially when using its voice feature. (Student 7)

This increased sense of preparedness is reflected in the improved OSCE scores, suggesting that AI-driven practice can effectively supplement traditional training methods. While students appreciated the ability to practise autonomously and at their convenience, which likely contributed to their enhanced performance during the high-stakes OSCE, it is important to

acknowledge that the frequency and depth of practice sessions likely varied among individuals. Iterative practice may explain performance differences, as those who engaged in more frequent or prolonged sessions could have experienced greater skill refinement. Future studies could explore patterns of tool use to evaluate how sustained and consistent practice impacts performance outcomes.

This suggests that AI tools offer valuable opportunities for students to refine their skills and build confidence, further enhancing their readiness for assessments.^[18]

Apart from the realistic scenarios, students highlighted the flexibility ChatGPT offered for OSCE preparation. One student remarked:

I really appreciated the ability to practise at my own pace and time without needing to coordinate with peers or simulated patients. (Student 8)

This flexibility aligns with the shift toward learner-centred education, which prioritises adaptable tools to accommodate diverse schedules and learning styles.^[4] Collecting data on the frequency and consistency of practice sessions in future research would provide a deeper understanding of how such flexibility can influence learning outcomes and ensure equitable access to skill-building opportunities.

Limitations and areas for improvement

Despite the overall positive reception, several areas for improvement were identified. Students highlighted issues such as the lack of emotional nuance in some responses, the need for structured prompts and technical limitations of the free version of ChatGPT.

Emotional nuance

A few students felt that ChatGPT's responses sometimes lacked emotional nuance, which made it less effective for practising empathetic communication. One student observed:

At times, I felt it didn't show much emotion, so I wasn't sure if the issue was with how I phrased the questions. Maybe learning to ask the right questions would help. (Student 4)

This feedback highlights an opportunity to explore how tailored prompts, or initial responses might improve AI's ability to simulate emotional nuance. For example, guiding ChatGPT with prompts to 'respond empathetically' or 'simulate a patient in distress' could better align its responses with the emotional context required for effective communication practice.

These concerns align with broader literature on the limitations of AI in healthcare education, particularly regarding emotional intelligence and adaptability.^[19] While AI offers valuable support for skill-building, the human element in medical training remains irreplaceable, especially for teaching empathy and emotional attunement. This underscores the need for AI tools to complement, rather than replace, traditional clinical education methods, particularly in areas where emotional nuance is critical. Future studies could explore how specific prompt adjustments or scenario designs can bridge this gap, fostering more nuanced interactions.

Structured prompts

Students suggested that tutors develop structured prompts to guide them in initiating consultations with ChatGPT. This further highlights the need for more sophisticated prompts and improving the AI's ability to simulate human emotions accurately. One student mentioned:

It would [further] be helpful if we could be provided with multiple opening scenario prompts for ChatGPT to act upon as the patient. These scenarios of what the patient's presenting complaint is can then be uploaded to Learn24 so that we may copy and paste them to ChatGPT. (Student 14)

This feedback presents an excellent opportunity to create a centralised 'prompt bank', which can house diverse scenarios with clear objectives and desired outcomes. Such a resource can standardise interactions and maximise the educational value of ChatGPT. Structured scenario designs, including a variety of presenting complaints and clinical contexts, could be uploaded to learning platforms like Learn24, offering students a ready-to-use tool for practice.

Providing predesigned prompts aligns with existing literature, which underscores the importance of structured learning environments in improving educational outcomes.^[20] Additionally, training students on how to craft their own effective prompts could further enhance the quality of interactions and foster critical thinking skills.^[21] Future studies could evaluate the impact of such structured resources on student learning and engagement, ensuring that AI tools are optimally utilised in medical education.

Technical limitations

Several technical issues were noted, such as time constraints when using the voice interface with ChatGPT's free version, as mentioned by one student:

The free version of ChatGPT doesn't allow you to use the voice feature more than once. (Student 19)

Additionally, response delays posed a challenge when practising for the OSCE, which typically lasts 8 minutes. One student remarked:

The only issue with using ChatGPT for your history-taking is the time it takes to respond. It is not possible to complete the history-taking within 8 minutes. (Student 16)

To address these challenges, a hybrid approach was suggested. One student proposed using ChatGPT to generate a patient case scenario, which could then be role-played with a peer for OSCE preparation:

... Therefore, it would be advisable to ask ChatGPT to generate a script for the patient based on my initial interaction with ChatGPT. You can then ask a friend to play the role of the patient... (Student 16)

This student further highlighted ChatGPT's potential for objective assessment, suggesting that it could generate a rubric for peer evaluation:

You may even ask ChatGPT to generate a rubric so your friend can award marks to you based on your performance. This would give you a general idea of what you need to work on and what can be improved. (Student 16)

These suggestions highlight potential avenues for optimising ChatGPT's use in medical training. Addressing technical limitations, such as by adopting the paid version or improving the infrastructure, could significantly enhance the user experience. These challenges are not unique to ChatGPT but are common across many AI applications and should be prioritised for resolution.^[22]

While the observed improvement in OSCE scores provides preliminary evidence of ChatGPT's potential value, it is important to acknowledge the study's limitations, including the lack of statistical analyses to establish causation. ChatGPT's integration into medical education represents a

promising shift in educational strategies,^[6] offering capabilities ranging from personalised learning to collaborative skill development. Future studies could investigate the direct impact of resolving these technical issues on both learning outcomes and student satisfaction.

Overall satisfaction and recommendations

Despite some challenges, the overall satisfaction with ChatGPT was high, with many students recommending its continued use. A representative quote highlighting satisfaction states:

With ChatGPT, the scenarios and virtual patients had more realistic conditions and were approachable. In the OSCE the patient had too many symptoms and conditions that we could not even provide a clear summary at the end. (Student 1)

This perception may reflect students' understanding of clinical scenarios or the inherent complexity of history-taking in high-stakes assessments such as OSCEs. It also raises the question of whether the complexity of OSCE cases aligns optimally with students' training levels. While ChatGPT may offer a simplified structure for scenarios, this suggests that its alignment with assessment methods could be strengthened to ensure constructive alignment. Addressing this by calibrating virtual patient scenarios to match the OSCE's complexity may enhance the tool's educational effectiveness and better prepare students for assessments.

The positive reception aligns with studies that have reported high satisfaction rates among students using virtual patients and AI-driven learning aids.^[4]

Further, non-first-language English speakers particularly appreciated ChatGPT for practising English fluency. One student remarked:

I'm very impressed with the school's idea of using ChatGPT for practice. The tool is very effective especially when using its voice feature, it doesn't just help with the cases but also help exercise English fluency especially for me whom English is not my home language. Also, there you can learn to be fast in taking history. I really recommend further use of this tool. (Student 7)

This aspect underscores an additional benefit of AI tools in medical education: enhancing language proficiency. Practising with ChatGPT not only helps improve clinical skills but also aids in developing communication skills, which are crucial for non-first-language English speakers in an English-dominated medical field. As indicated in the study by Li *et al.*,^[23] the use of AI tools, including ChatGPT, has shown potential in improving the writing and language skills of non-native English-speaking students, suggesting that similar benefits can extend to verbal communication skills, such as those needed in clinical interactions. The dual benefit of clinical and language skill enhancement makes ChatGPT an even more valuable educational tool.

Discussion

The integration of ChatGPT as a virtual patient in clinical communication skills training demonstrated promising outcomes in enhancing second-year medical students' history-taking skills and OSCE readiness. Using a mixed-methods approach, this study evaluated ChatGPT's effectiveness in improving skill performance, perceived learning, and engagement. The findings highlight the strengths of AI-assisted learning, identify challenges, and suggest practical strategies for curriculum integration.

Impact on OSCE performance

The statistically significant improvement in history-taking OSCE scores from 2023 to 2024 ($p=7.57 \times 10^{-6}$; Cohen's $d=0.41$) suggests a moderate educational effect attributable to ChatGPT's integration. While causality cannot be definitively established in this observational design owing to potential confounding factors such as cohort differences and examiner variability, the alignment between quantitative results and qualitative feedback strengthens the case for a connection between AI-assisted practice and improved communication skills performance. These findings are consistent with prior research indicating that virtual patients can foster clinical reasoning and communication skill development.^[4,20] Nonetheless, further controlled studies are needed to isolate ChatGPT's specific effects and to assess long-term outcomes.

Student engagement and perceptions

Student feedback supported the quantitative findings, with high satisfaction levels for ChatGPT's usability, feedback quality, and contribution to OSCE preparedness. These results align with literature highlighting user-friendly AI tools as enablers of self-directed and sustained engagement in medical education.^[11] ChatGPT's realism was particularly valued, as it provided authentic scenarios fostering confidence and skill improvement.

However, mixed confidence levels were noted, with nearly half of the students expressing neutral feelings. This suggests that while ChatGPT benefited students with lower baseline confidence, those with higher initial confidence experienced fewer gains. Personalised learning approaches could address these diverse needs.

Strengths of ChatGPT as a virtual patient

Students valued ChatGPT's ability to simulate realistic clinical encounters and deliver detailed, consistent feedback. These features align with prior research demonstrating AI's capacity to create safe, repeatable environments for clinical skills practice.^[7] Its flexibility and accessibility also cater to diverse learning needs, particularly in resource-constrained settings, where standardised patients and faculty time are limited.^[4]

Limitations and recommendations for improvement

Despite its strengths, the study identified several limitations. ChatGPT's lack of emotional nuance was a recurring concern, as empathy is vital for effective patient-provider communication. Tailored prompts simulating specific emotional scenarios could enhance this aspect. Additionally, technical issues such as response delays and limitations of the free version posed challenges for time-sensitive OSCE preparation. Institutional access to premium versions and hybrid approaches combining AI with peer role-playing may mitigate these issues.

Comparison with existing literature

The findings of this study align with the broader body of research on the use of virtual patients and AI in medical education. It illustrates how AI chatbots as a virtual patient simulation can be integrated into students' clinical skills learning for their personalised, interactive and immersive learning experience. Previous studies have demonstrated the effectiveness of virtual patients in improving clinical reasoning.^[4,20] This study adds to the evidence by illustrating that ChatGPT, a more recent AI tool, can provide similar benefits in enhancing history-taking communication skills. Also, rather than merely consuming information passively from ChatGPT,

Table 2. Integration of ChatGPT into the medical curriculum

Integration area	Description	Key benefits	Challenges addressed
Supplement to traditional methods	Use ChatGPT alongside standardised patients (SPs) and role-playing in clinical skills training	Provides additional practice opportunities beyond formal settings	Resource limitations of SPs and time constraints for faculty
Early introduction in pre-clinical years	Integrate ChatGPT during early clinical skills courses (years 1-2) for history-taking practice	Builds foundational skills before interacting with real or simulated patients	Develops communication skills early on, before real patient encounters
Scenario-based learning modules	Design structured modules with varying complexity, integrating ChatGPT to practice progressive history-taking tasks.	Systematic skill development aligned with curricular goals.	Ensures students engage with diverse patient cases for well-rounded training.
Targeted support for non-first language English speakers	Provide additional ChatGPT sessions for students needing to improve medical English and communication skills	Improves fluency and communication, especially in medical contexts	Addresses language barriers in patient interactions
Formative assessment and feedback	Use ChatGPT for formative assessments, providing students with real-time feedback on their performance	Allows immediate self-correction and reflection, improving learning outcomes	Fills gaps where timely feedback is lacking in traditional methods
Practice for OSCE preparation	Incorporate ChatGPT in OSCE prep courses, simulating mock OSCE stations focused on history-taking exercises.	Enhances students' familiarity with OSCE structure, expectations, and pacing.	Overcomes limited access to standardised patients or faculty time for practice
Focus on emotional communication	Develop ChatGPT cases focusing on empathetic communication, simulating emotionally charged patient scenarios	Encourages practice in empathy and patient-centred care	Addresses the lack of emotional nuance in AI-based learning
Institutional access to advanced AI tools	Provide institutional access to premium versions of ChatGPT to overcome technical limitations like time restrictions and lag	Improves overall user experience and maximises the effectiveness of AI for educational purposes	Overcomes technical limitations found in the free version of ChatGPT
Faculty training on AI integration	Train faculty on how to effectively incorporate AI tools like ChatGPT into their teaching methods and track student progress	Ensures the seamless integration of AI tools in line with learning objectives	Addresses the lack of familiarity or expertise with AI-based learning tools
Continuing feedback and evaluation	Collect regular student feedback on ChatGPT use and adjust implementation strategies accordingly	Ensures that the integration of AI tools evolves with student needs and curriculum changes	Helps adapt to changing educational requirements and technology updates
Peer-based collaborative learning	Facilitate group-based practice where students use ChatGPT together, followed by discussions of patient cases and feedback	Encourages collaboration, peer feedback, and critical thinking	Promotes teamwork and shared learning while maintaining individual practice

students actively engage in patient-provider communication, responding to patient queries, delivering diagnoses, and discussing treatment options. These interactions help students develop critical communication skills essential for effective patient care, supporting a holistic approach to clinical reasoning and problem-solving.^[24]

Additionally, this study's emphasis on the practical use of ChatGPT in OSCE preparation highlights its potential to meet specific educational needs, an area that has been less explored in existing literature.^[19,25] The findings suggest that integrating ChatGPT into the medical curriculum can further enhance history-taking skills while optimising the learning experience. This can be achieved by addressing the identified challenges, while capitalising on its strengths to improve student learning and OSCE readiness.

Recommendations for curriculum integration

Based on this study's findings, Table 2 outlines strategies for integrating ChatGPT into medical curricula, such as using it for formative assessment, supporting non-first-language English speakers, and incorporating scenario-based learning. Training faculty on AI integration and collecting continuing feedback will ensure the tool's effective implementation and alignment with curricular goals.

Study limitations and future directions

While this study presents promising results with ChatGPT's potential to complement traditional medical training by fostering SDL, building confidence, and improving communication skills, it is important to acknowledge several limitations.

First, only 40 out of 200 second-year medical students responded to the survey, resulting in a 20% response rate. This relatively low response rate may introduce bias, as participants could differ in motivation, learning styles, or prior experience with AI tools compared with non-participants. Future studies should focus on strategies to improve student engagement, such as offering incentives or embedding the tool within mandatory course components, to achieve higher participation rates and more representative samples.

The study conducted with a small and homogenous sample limits the generalisability of the results. Future research should aim to include a larger, more diverse cohort of students from different academic years and varied educational backgrounds to provide a more comprehensive understanding of ChatGPT's effectiveness in enhancing history-taking skills. Additionally, the study was conducted over a limited period, which may not fully capture the long-term benefits and challenges of using ChatGPT for history-taking practice. Longitudinal studies are recommended to assess the sustained impact on students' skill development and confidence over time.

The use of the free version of ChatGPT also posed technical limitations, such as time restrictions and response lag, which may have impacted the overall user experience and the perceived effectiveness of the tool. Institutions should consider investing in the paid version of ChatGPT or similar AI tools to provide uninterrupted and more responsive interactions, improving the learning experience.

It is important to note the limitations of using AI tools like ChatGPT. ChatGPT may face issues with contextual vetting, training biases towards Western models, and occasional lapses in ensuring the accuracy and depth of content. Although ChatGPT was found to be helpful in many aspects of communication and history-taking, it remains crucial to balance the advantages of AI with these limitations. For example, ChatGPT's inability to fully replicate emotional nuance may hinder practice in empathetic communication, an essential part of patient interactions. This limitation underscores the need for the development of more sophisticated prompts to elicit emotional depth from AI-based tools. As AI tools evolve, integrating emotional intelligence features and addressing technical barriers will be critical to maximising their educational impact. Collaboration between educators, developers, and learners will ensure that AI-driven tools remain relevant and effective in meeting the dynamic needs of medical education.

Moreover, while this study highlighted the benefits of ChatGPT, it is important to emphasise that AI cannot replace human expertise in clinical teaching. ChatGPT should be viewed as a supplementary tool rather than a substitute for faculty-led training and interaction with real patients. Faculty training on AI integration and continued evaluation based on student feedback are critical to ensuring that these tools are used effectively and complement traditional learning methods.

Finally, the results suggest that integrating ChatGPT into medical education can enhance history-taking communication skills, as evidenced by the statistically significant improvement in OSCE scores and positive student perceptions. However, the study also acknowledges the influence of potential confounders, including cohort-specific differences and teaching variations, which could not be controlled for in this pilot study. Future studies should employ controlled designs to isolate ChatGPT's impact more effectively and validate its role in clinical skill development. Despite these limitations, this study highlights the promising potential of AI tools like ChatGPT in enhancing medical education and preparing students for clinical practice.

Conclusion

This study demonstrates that ChatGPT is a valuable tool for enhancing history-taking skills in medical education. Most students found it easy to use, realistic, and helpful for improving their skills and preparing for OSCEs. The ability to receive immediate, detailed feedback and practise independently at their own pace was particularly appreciated. Qualitative feedback from non-first-language English speakers indicated that ChatGPT was helpful for practising language fluency, highlighting a potential additional advantage of AI in medical training. However, this was not quantitatively assessed in the study and warrants further investigation. The study also identified areas for improvement, such as addressing technical limitations and enhancing the emotional depth of ChatGPT's responses.

The integration of pre- and post-intervention OSCE data further strengthens the findings, providing quantitative evidence of ChatGPT's positive impact on student performance. Specifically, the two-point increase in history-taking OSCE scores equivalent to a 4% improvement indicates

that ChatGPT may contribute to skill enhancement in this area. While the improvement is modest, it is noteworthy given the short duration of the intervention and the self-directed nature of its use. Positive student feedback further supports the potential of AI-driven tools to enhance clinical skills through accessible and flexible learning opportunities.

The incorporation of AI into medical education marks a significant advancement, offering innovative solutions to traditional training challenges. By addressing the noted limitations and implementing suggested improvements, ChatGPT and similar AI tools could become even more effective in preparing medical students for clinical practice. As technology evolves, the potential for AI to revolutionise medical education and enhance patient care outcomes remains promising. This study contributes to the growing body of literature on AI in medical training, offering valuable insights to guide future research and development in this field. Continued exploration and refinement of AI tools will be crucial in realising their full potential in medical education and beyond. ChatGPT stands out as a transformative tool with the potential to reshape medical pedagogy, driving meaningful curriculum reform to meet the evolving demands of modern healthcare practice.

Declaration. None.

Acknowledgements. The authors would like to thank students who participated in the study. This publication was made possible by the Department of Higher Education and Training (DHET) through the University Capacity Development Programme, University of KwaZulu-Natal, South Africa.

Author contributions. The authors complied with the international committee of medical journal editor's rules of authorship and participated fully in this research. RA contributed to the conception, design, data collection, analysis, interpretation of data and drafting the main manuscript text. DM contributed to the conception, design, data analysis and interpretation. Both authors critically reviewed the manuscript and approved it for publication.

Funding. This research was funded by the University Capacity Development Programme. The funding body was not involved in the study design, data collection, analysis or interpretation or the writing of this manuscript. The views expressed in this report are those of the authors and do not necessarily reflect those of the University Capacity Development Programme.

Conflicts of interest. None.

1. Kurtz SM, Silverman JD, Draper J. Teaching and Learning Communication Skills in Medicine. Boca Raton: CRC Press, 2016.
2. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc* 2007;2(2):115-125. <https://doi.org/10.1097/SIH.0b013e3180315539>
3. Silverman J, Kurtz SM, Draper J. Skills for Communicating with Patients. Boca Raton: CRC Press, 2013.
4. Cook DA, Erwin PJ, Triola MM. Computerised virtual patients in health professions education: A systematic review and meta-analysis. *Acad Med* 2010;85(10):1589-1602. <https://doi.org/10.1097/ACM.0b013e3181edfe13>
5. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 2018;22(5):1589-1604. <https://doi.org/10.1109/JBHI.2017.2767063>
6. Ahmed Y. Utilisation of ChatGPT in medical education: Applications and implications for curriculum enhancement. *Acta Inform Med* 2023;31(4):300-305. <https://doi.org/10.5455/aim.2023.31.300-305>
7. Plackett R, Kassianos AP, Mylan S, et al. The effectiveness of using virtual patient educational tools to improve medical students' clinical reasoning skills: A systematic review. *BMC Med Educ* 2022;22:365. <https://doi.org/10.1186/s12909-022-03410-x>
8. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med* 2018;93(8):1107-1109. <https://doi.org/10.1097/ACM.0000000000002044>
9. Creswell JW. Research Design: Qualitative, Quantitative and Mixed Method Approaches, 4th ed. London: Sage, 2014.
10. Kurtz SM, Silverman JD, Benson J, Draper J. Marrying content and process in clinical method teaching: Enhancing the Calgary-Cambridge guides. *Acad Med* 2003;78(8):802-809. <https://doi.org/10.1097/00001888-200308000-00011>
11. Narayanan S, Ramakrishnan R, Durairaj E, et al. Artificial intelligence revolutionizing the field of medical education. *Cureus* 2023;15(11): e49604. <https://doi.org/10.7759/cureus.49604>
12. Cameron RS, Yong GP. The evolution and role of simulation in medical education. *Anesthesia Patient Safety Foundation* 2023. <https://www.apsf.org/article/the-evolution-and-role-of-simulation-in-medical-education/> (accessed 22 July 2024).

13. Bienstock J, Heuer A. A review on the evolution of simulation-based training to help build a safer future. *Medicine (Baltimore)* 2022;101(25):e29503. <https://doi.org/10.1097/MD.00000000000029503>
14. Kruger D. Adaptive learning technology to enhance self-directed learning. In: Olivier J, ed. *Self-directed multimodal learning in higher education 2020*; (NWU Self-Directed Learning Series Volume 5) (pp. 93–116). AOSIS. <https://doi.org/10.4102/aosis.2020.BK210.03>
15. Kononowicz AA, Woodham LA, Edelbring S, et al. Virtual patient simulations in health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration. *J Med Internet Res* 2019; 21(7):L14676. <https://doi.org/10.2196/14676>
16. Boud D, Molloy E. *Feedback in Higher and Professional Education: Understanding It and Doing It Well*. Abingdon-on-Thames: Routledge, 2013.
17. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ Psychol* 2011;46(4):197-221.
18. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels*. Oakland, Calif.: Berrett-Koehler Publishers, 2006.
19. Tsang R. Practical applications of ChatGPT in undergraduate medical education. *J Med Educ Curricular Dev* 2023;10:23821205231178449 <https://doi.org/10.1177/23821205231178449>
20. McCoy L. *Virtual Patient Simulations for Medical Education: Increasing Clinical Reasoning Skills through Deliberate Practice*. Doctoral dissertation, Arizona State University, 2014. <https://core.ac.uk/download/pdf/79572495.pdf>
21. Walter Y. Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *Int J Educ Techno Higher Educ* 2024;21:15. <https://doi.org/10.1186/s41239-024-00448-3>
22. Zhu E, Lilienthal A, Shluzas LA, Masiello I, Zary N. Design of mobile augmented reality in health care education: A theory-driven framework. *JMIR Med Educ* 2015;1(2):e10. <https://doi.org/10.2196/mededu.4443>
23. Li J, Zong H, Wu E, et al. Exploring the potential of artificial intelligence to enhance the writing of English academic papers by non-native English-speaking medical students - the educational application of ChatGPT. *BMC Med Educ* 2024; 24:736. <https://doi.org/10.1186/s12909-024-05738-y>
24. Elnahal SM, Blackford A, Smith K, et al. Identifying predictive factors for incident reports in patients receiving radiation therapy. *Int J Radiat Oncol Biol Phys* 2016;94(5):993-999. <https://doi.org/10.1016/j.ijrobp.2015.11.041>
25. Consorti F, Mancuso R, Nocioni M, Piccolo A. Efficacy of virtual patients in medical education: A meta-analysis of randomised studies. *Computers Educ* 2012;59(3):1001-1008.

Received 11 October 2024. Accepted 19 January 2026.