AI-assisted Simulated Patient for Clinical Communication Skills Training

Contents

Executive summary

Key project outcomes

This project successfully developed a proof-of-concept AI-assisted platform for practicing clinical communication skills, focused on sleep medicine scenarios. Among 15 postgraduate students invited to test the tool, 7 completed the feedback form, with a mean usability score of 75% on the Chatbot Usability Questionnaire (CUQ). Users found the tool engaging, realistic, and valuable for practicing clinical history-taking, especially appreciating the replay function and low-pressure environment. Qualitative feedback highlighted the tool's potential for enhancing reflective practice, while also identifying areas for improvement, such as clearer setup instructions, expanded scenario variety, and more interactive features. Importantly, students demonstrated strong openness to integrating AI into their learning, recognising its potential to support skill development. The project produced a functioning platform with a case library (n = 2), and initial feedback system, laying a strong foundation for future expansion and refinement, particularly toward more dynamic simulations and the inclusion of complex clinical data. Unfortunately, integration with Canvas, one of our initial intended outcomes, was not possible within the timeframe of this project.

Key learning for the team around using AI for developing teaching and learning

Developing this AI-assisted tool provided the team with valuable insights into both the potential and limitations of using large language models (LLMs) for educational purposes. A key learning was the importance of **balancing realism with educational structure**—while the model often responded appropriately and naturally, it sometimes provided too much unsolicited information, reducing opportunities for students to practice eliciting details through questioning. It also became clear that LLMs are inherently polite and cooperative, making it challenging to simulate more complex or emotionally charged patient behaviours, such as frustration or resistance. Designing effective case studies required careful scripting and anticipation of user input to present the model from veering off-topic, though it generally managed unanticipated questions well. The feedback function also posed several challenges, given that it was dependent on the conversation transcript, which was not always accurate. On a positive note, we observed strong student enthusiasm for using AI in learning, which reinforced its value as a formative tool. Overall, the project emphasised the need for thoughtful scenario design, pedagogically aligned feedback system, and the importance of scaffolding AI

tools with clear instructions and user support. With careful planning, AI offers exciting opportunities to enhance experiential learning, accessibility, and student engagement across diverse clinical education settings.

Project introduction

Background and context

Communication skills, which are essential for safe and effective practice, often cannot be fully developed through textbook learning alone. Traditional methods, such as observed simulations, may not offer a psychologically safe space for all learners—particularly those with learning difficulties, neurodevelopmental disorders, or mental health challenges, who may find the pressure of real-time observation overwhelming. Furthermore, students studying remotely often lack access to regular in-person simulation opportunities, limiting their ability to practice essential skills. This project aims to address these challenges by offering accessible, supportive, and forward-thinking opportunities for communication skills development.

Objectives

The project aims to develop and evaluate an AI-assisted platform that allows healthcare students to practice clinical communication skills through realistic patient simulations with automated feedback. Specifically, it will create AI-assisted simulated patients and history-taking scenarios, provide structured OSCE-style feedback, support repeated practice, and test integration with Canvas. Expected outcomes include a functional simulation interface, a case library, implementation documentation, and preliminary evaluation data, with the potential for academic dissemination.

Scope

This project focuses on **clinical history-taking skills** in sleep medicine cases (insomnia and hypersomnia), with feedback centred on clinical **communication** rather than diagnostic accuracy or treatment. Designed to **supplement**, not replace, in-person simulations and case study discussion seminars, this proof-of concept tool has potential for future expansion to other medical areas/clinical skills, as well as deployment for research purposes (e.g. participant recruitment and assessment training).

Tools and technologies

Initial testing was done using ChatGPT 4.0 -Educational license. Later the project was deployed as a standalone application. Initially deployed using Azure VMs with a Gradio framework, leveraging Twilio as our turn server and FastRTC to interface with OpenAI's real-time API. OAuth2-Proxy provided SSO authentication as a reverse proxy. The app was transitioned to Azure

App Service for containerized hosting with built-in authentication. We replaced Gradio with Streamlit and switched from Twilio to LiveKit's hosted server solution. This iteration also introduced PostgreSQL database hosting on Azure. The LiveKit Python client handled communication between OpenAI's real-time API and the LiveKit server.

Collaboration

This project is the result of a merge of three distinct groups from two separate submissions. As such, the project combined a wide range of input from the field of medical education (sleep medicine, general medicine and pharmacology), spanning both UG and PGT students as the target audience, within online and in-person teaching modalities. Students, clinicians, educators, and members of the CTL and AI/ML teams contributed expertise and feedback throughout the process.

Project outcomes and findings

Evaluation results

15 PGT students were invited to test the platform and complete a feedback form that included the Chabot Usability Questionnaire (CUQ) and open-ended questions probing tool use experience and suggestions for improvement. Of these, 9 engaged with the tool and 7 completed the feedback form (see results below). Students had an average of 2 conversations with the AI-simulated patient (range 1 - 6), with each conversation lasting about 7 min on average (range 2 - 11 min). Overall, users found the AI simulation to be a valuable and realistic tool for practicing clinical communication, praising its replay features and immersive design, while recommending clearer setup guidance, technical improvements, and enhanced interactivity to support more effective and engaging use.

Quantitative data

Mean scores on the CUQ revealed a usability score of 75% (n = 7). Individual scores are depicted in the table and figure below.

Table 1. Individual scores on the Chabot Usability Questionnaire (CUQ)

Participant	CUQ score (%)
#1	61
#2	84
#3	92
#4	86
#5	70
#6	75
#7	59
Mean	75

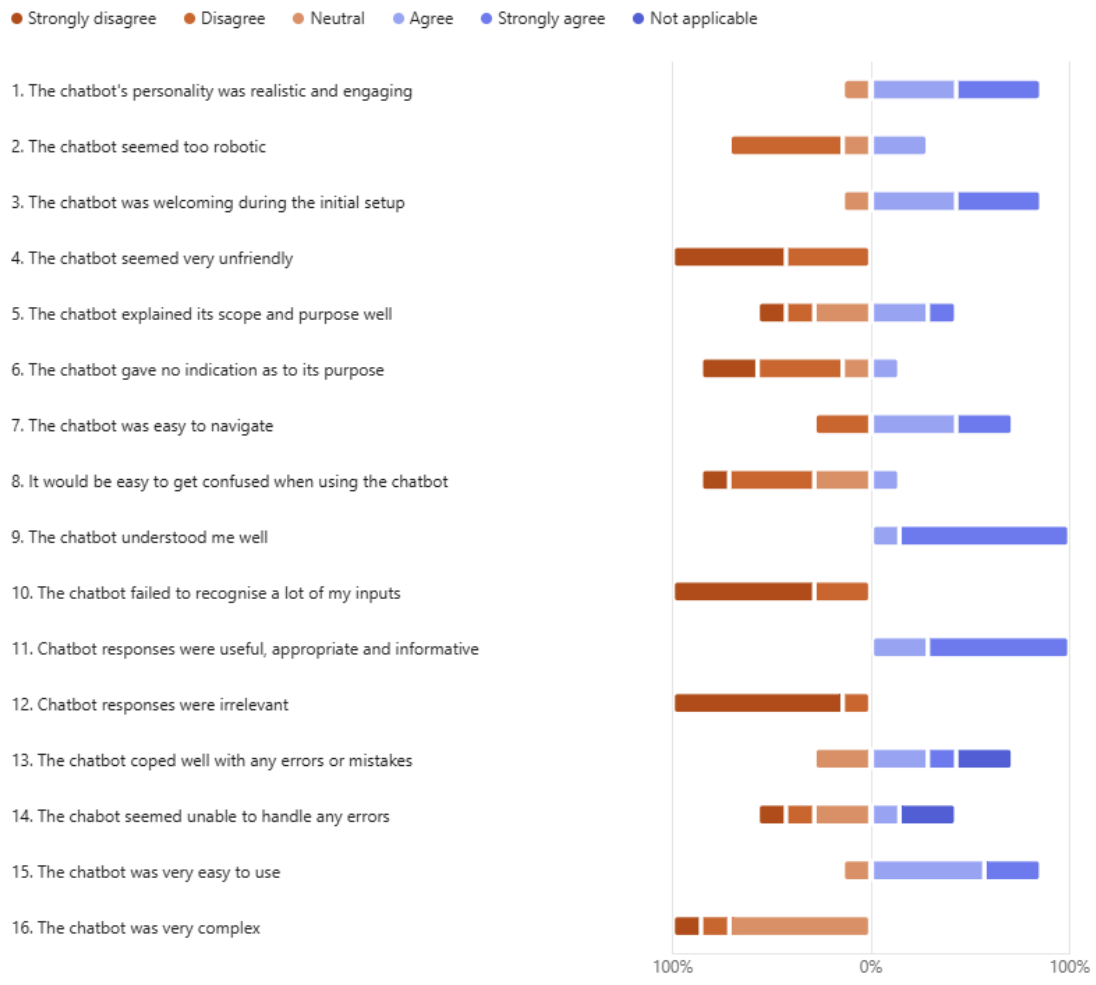

Figure 1: Pooled individual item responses on the Chabot Usability Questionnaire (CUQ)

Qualitative insights

A summary of the written feedback is provided in the table below.

Table 2. Summary of qualitative user feedback

What users liked	Suggestions for improvement
1. Engaging, realistic, and valuable for practicing clinical history-taking. 2. Ability to **replay and repeat scenarios** supported reflection and improvement. 3. Simulated patient responses were **lifelike and appropriate**, with natural inflections. 4. Allowed users to **think at their own pace** and practice in a low-pressure environment. 5. Useful for **practicing initial assessments.** 6. Described as an **excellent learning resource.**	1. Provide **clearer setup instructions**, including how to navigate, switch scenarios, and manage voice settings. 2. Address **technical issues** (e.g. voice distortion with Sage, browser access problems on Mac/Chrome). 3. Add a **summary of patient responses** to support more targeted follow-up questions. 4. Include **examples of ideal questioning** to help users model best practices. 5. Introduce **off-topic or irrelevant details** to increase realism. 6. Encourage the **simulated patient to ask questions**, making interactions more dynamic. 7. Expand the tool with **more varied scenarios** for broader practice.

Lessons learned

Challenges

Clinical/Theoretical

One of the key challenges was calibrating the AI model's responses to encourage active questioning without offering too much unsolicited information—a behaviour that, while occasionally realistic, limited opportunities for students to practice structured clinical enquiry. Replicating a broad range of patient personalities, particularly more emotionally charged or resistant ones (e.g. expressing frustration or anger), also proved difficult, as large language models are inherently optimised for courteous and cooperative dialogue. Developing each case study required considerable effort to

anticipate a wide array of possible questions, thereby minimising off-script improvisation; although the model generally handled such situations well, this added significant design complexity. Another challenge was selecting an appropriate feedback framework—one that aligned with intended learning outcomes and marking criteria, but avoided overwhelming users with too many action points.

Technological

The application frameworks were chosen based on team familiarity and ability to support rapid development and deployment. Both eventually presented constraints that restrict development. The FastRTC implementation became progressively unstable, eventually causing the application to hang indefinitely when users attempted to initiate conversations.

The transcription accuracy provided by OpenAI's real-time API proved inconsistent, compromising the quality of conversation records and subsequently, the quality of the automatic feedback provided.

Key takeaways

Insights

Students are open to incorporating AI into their learning, demonstrating engagement and enthusiasm towards the use of emerging technologies in clinical education.

Framework and service selection has long-term implications. Integrated solutions like LiveKit demonstrated superior stability, suggesting that, purpose-built systems offer better reliability for real-time features.

Recommendations for future projects

Future development should focus on creating modular libraries of case studies and patient personalities that can be mixed and matched to provide varied and personalised learning experiences. Enhancing interactivity by encouraging the simulated patient to actively ask questions and seek clarification would further improve realism, as would incorporating video alongside audio. Allowing the patient to present and discuss more complex clinical data, such as actigraphy charts or test results, would better reflect real-world consultations.

From a technical perspective, future projects should prioritise production-grade frameworks designed for scalability and complex user interactions rather than prototyping tools. Improving the accuracy of the transcription would improve the automated feedback function.

Advice for teams

Although this proof-of-concept tool has received good feedback from users, its effectiveness in objectively improving clinical communication skills has not been formally assessed as part of this project. Therefore, we do not currently recommend its use in summative assessments such as OSCEs. Instead, the tool is best suited for formative use, supporting skills development and reflective learning.

We suggest testing tools, services, and frameworks under production conditions before architectural commitment. Anticipate moving beyond prototyping frameworks to production-grade solutions quickly.

Appendices

1. Prompts & marking rubric
2. Case studies (Insomnia and Hypersomnia, .doc)
3. Guidance provided to users
4. Annotated, anonymised transcript of a conversation between the AI-Simulated Patient and a PGT student, plus automated feedback on communication skills.
5. Chatbot usability questionnaire & scoring guidelines (.pdf)
6. Feedback form questions and results (.xlsx).