The Digital Shift: Assessing ChatGPT's Capability as a New Age Standardized Patient.

Joseph Cross, Tarron Kayalackakom, Raymond E Robinson, Andrea Vaughans, Roopa Sebastian, Ricardo Hood, Courtney Lewis, Sumanth Devaraju, Prasanna Honnavar, Sheetal Naik, Jillwin Joseph, Nikilish Anand, Abdalla Mohammed, Asjah Johnson, Eliran Cohen, Teniola Adeniji, Aisling Nnenna Nnaji, Julia George

Submitted to: JMIR Medical Education
on: June 20, 2024

Table of Contents

The Digital Shift: Assessing ChatGPT’s Capability as a New Age Standardized Patient.

Joseph Cross[1] PhD; Tarron Kayalackakom[2] MD; Raymond E Robinson[3] MPH, MBA, MD; Andrea Vaughans[2] MD; Roopa Sebastian[2] MSc, PhD; Ricardo Hood[2] MD; Courtney Lewis[2] MD; Sumanth Devaraju[2] MD; Prasanna Honnavar[2] PhD; Sheetal Naik[2] PhD; Jillwin Joseph[2] PhD; Nikilish Anand[4] MD; Abdalla Mohammed[1] BSc; Asjah Johnson[1] BSc; Eliran Cohen[1] BSc; Teniola Adeniji[1] BSc; Aisling Nnenna Nnaji[1] BSc; Julia George[1]

[1]Xavier University School of Medicine Orangestad AW
[2]College of Medicine American University of Antigua St Johns AG
[3]College of Graduate Studies American University of Antigua St Johns AG
[4]School of Medicine University of Texas Edinburg US

Corresponding Author:
Joseph Cross PhD
Xavier University School of Medicine
Aruba Campus
23 Santa Helenastraat
Orangestad
AW

Abstract

Background: Standardized patients (SPs) have been crucial in medical education, offering realistic patient interactions to students. Despite their benefits, SP training is resource-intensive, and access can be limited. Advances in artificial intelligence, particularly with large language models like ChatGPT, present new opportunities for virtual SPs, potentially addressing these limitations.

Objective: To assess medical students' perceptions and experiences of using ChatGPT as a standardized patient (SP) and to evaluate ChatGPT’s effectiveness in performing as a virtual SP in a medical school setting.

Methods: This qualitative study, approved by the AUA Institutional Review Board (IRB), involved eleven medical student volunteers (5 females, 6 males, aged 20-32) from the American University of Antigua (AUA) College of Medicine. Students were observed during a live role-play, interacting with ChatGPT as an SP using a predetermined prompt. A structured 15-question survey was administered before and after the interaction. Thematic analysis was conducted on the transcribed and coded responses, with inductive category formation.

Results: Thematic analysis identified key themes pre-interaction including: technology limitations (e.g., prompt engineering difficulties), learning efficacy (e.g., potential for personalized learning, reduced interview stress), verisimilitude (e.g., absence of visual cues), and trust (e.g., concerns about AI accuracy). Post-interaction, students noted improvements in prompt engineering, some alignment issues (e.g., limited responses on sensitive topics), maintained learning efficacy (e.g., convenience, repetition), and continued verisimilitude challenges (e.g., lack of empathy and non-verbal cues). No significant trust issues were reported post-interaction. Despite some limitations, students found ChatGPT a valuable supplement to traditional SPs, enhancing practice flexibility and diagnostic skills.

Conclusions: ChatGPT can effectively augment traditional SPs in medical education, offering accessible, flexible practice opportunities. However, it cannot fully replace human SPs due to limitations in verisimilitude and prompt engineering challenges. Integrating prompt engineering into medical curricula and continuous advancements in artificial intelligence (AI) are recommended to enhance the utility of virtual SPs.

(JMIR Preprints 20/06/2024:63353)
DOI: https://doi.org/10.2196/preprints.63353

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Article

Title: The Digital Shift: Assessing ChatGPT's Capability as a New Age Standardized Patient.

Joseph Cross[1], PhD; Tarron Kayalackakom[2], MD; Raymond E Robinson[3], MD; Sumanth Devaraju[2], MD; Andrea Vaughans[2], MD; Ricardo Hood[2], MD; Prasanna Honnavar[2], PhD; Sheetal Naik[2], PhD; Roopa Sebastian[2], PhD; Courtney Lewis[2], MD; Nikhilesh Anand[2], MD; Jillwin Joseph[2], PhD; Abdalla Mohammed[1], BSc; Asjah Johnson[1], BSc; Eliran Cohen[1], BSc; Teniola Adeniji[1], BSc; Aisling Nnenna Nnaji[1], BSc; Julia George[1]

[1]School of Medicine, Xavier University, Orangestad, Aruba, West Indies

[2]College of Medicine, American University of Antigua, Coolidge, Antigua, West Indies

[3]College of Graduate Studies, American University of Antigua, Coolidge, Antigua, West Indies

Corresponding Author:

Joseph Cross, PhD
School of Medicine
Xavier University
Orangestad
Aruba
West Indies
Phone: + 297 748 6196
Email: crossphd@xusom.com

Keywords: medical education; standardized patient; artificial intelligence; AI; ChatGPT; OSCE; virtual patient; medical student

Abstract

Background: Standardized patients (SPs) have been crucial in medical education, offering realistic patient interactions to students. Despite their benefits, SP training is resource-intensive, and access can be limited. Advances in artificial intelligence, particularly with large language models like ChatGPT, present new opportunities for virtual SPs, potentially addressing these limitations.

Objective: To assess medical students' perceptions and experiences of using ChatGPT as a standardized patient (SP) and to evaluate ChatGPT's effectiveness in performing as a virtual SP in a medical school setting.

Methods: This qualitative study, approved by the AUA Institutional Review Board (IRB), involved eleven medical student volunteers (5 females, 6 males, aged 20-32) from the American University of Antigua (AUA) College of Medicine. Students were observed during a live role-play, interacting with ChatGPT as an SP using a predetermined prompt. A structured 15-question survey was administered before and after the interaction. Thematic analysis was conducted on the transcribed and coded responses, with inductive category formation.

Results: Thematic analysis identified key themes pre-interaction including: technology limitations (e.g., prompt engineering difficulties), learning efficacy (e.g., potential for personalized learning, reduced interview stress), verisimilitude (e.g., absence of visual cues), and trust (e.g., concerns about AI accuracy). Post-interaction, students noted improvements in prompt engineering, some alignment issues (e.g., limited responses on sensitive topics), maintained learning efficacy (e.g., convenience, repetition), and continued verisimilitude challenges (e.g., lack of empathy and non-verbal cues). No significant trust issues were reported post-interaction. Despite some limitations, students found ChatGPT a valuable supplement to traditional SPs, enhancing practice flexibility and diagnostic skills.

Conclusions: ChatGPT can effectively augment traditional SPs in medical education, offering accessible, flexible practice opportunities. However, it cannot fully replace human SPs due to limitations in verisimilitude and prompt engineering challenges. Integrating prompt engineering into medical curricula and continuous advancements in artificial intelligence (AI) are recommended to

enhance the utility of virtual SPs.

Introduction: SPs have been a cornerstone of medical education since the 1960s [1]. They offer students an immersive, real-world experience. Cleland *et al*. found them superior for teaching consultation skills compared to traditional methods and Flanagan and Cummings found medical students who had trained with SPs had increased confidence and competency compared to students trained using other modalities [1,2].

Despite the demonstrated utility of SPs in medical education, there are often inadequate opportunities for students to interact with SPs. Training of SPs is time-consuming and costly [1]. With a large cohort of undergraduate students, sessions with SPs are often limited in number, and in some cases, not available at all.

In the 21st century, virtual SPs have emerged. These are computer programs that simulate specific illnesses and respond to learner inputs [3]. They have become invaluable tools in both teaching and assessment. However, their development also requires significant resources, making it challenging for institutions without robust educational technology support departments [4].

As the field of AI has advanced, the potential for its application in medical education has expanded. Large Language Models (LLMs), such as ChatGPT (OpenAI, San Francisco CA), have revolutionized natural language processing. These sophisticated neural networks, trained on vast amounts of internet data, are adept at predicting subsequent words in a sequence [5]. ChatGPT, a chatbot based on the GPT-3.5 model, has an enormous 175 billion parameters and displays a remarkable capacity for understanding and reasoning, bordering on human-like proficiency [6]. Since its introduction in November 2022, sectors spanning from history to entertainment have rapidly adopted the LLM [7].

This advancement in AI has led to the development of virtual standardized patient chatbots. A number of major educational material suppliers and specialized companies are offering chatbot SPs, based on LLMs capable of natural language interactions, for students to practice clinical skills. One example is Oscer, which can present over 200 virtual patient conditions and boasts above 90% accuracy in symptomology [8]. Similarly, the University of Texas Medical Branch makes use of an AI agent termed Virti, which they use to conduct virtual Observed Structured Clinical Examinations (OSCEs) with medical students [9].

However, for this new generation of virtual patients there is again considerable time and resources required for the company or institution to develop the program and train the LLM on specific datasets and student access can be limited by cost and/or locality [4].

The debut of ChatGPT sparked inquiries into its potential as a standardized patient. Liu et al. crafted ten medical histories with ChatGPT, which were then vetted by experienced physicians. Their results highlight ChatGPT’s promise in clinical education, though some responses may come across as robotic [10]. Suarez et.al. gathered dental student’s feedback after interacting with an AI chatbot. The majority found the experience valuable, especially those who made a correct diagnosis. This underscores the potential of integrating AI into health sciences training [11].

Weidener et al. emphasized the growing consensus on incorporating AI into medical education. Their study indicated the importance of both practical and technological skills for leveraging AI in medicine [12]. Similarly, Jowsey et al have recommended adoption of AI into medical education as a way of preparing future physicians for the reality of modern practice [13].

We were aware that SPs at our school, AUA, were in limited supply and we had received a number of comments to the effect that SPs were effective, but students would like greater access to them. Some students in fact had no access during their course, depending on which cohort they were part of.

One of our study's aims was to assess medical student’s perceptions and experience of the use of AI in medicine and to gain and understanding of their perceptions of AI before and after interaction with ChatGPT performing the role of an SP. A second aim was to assess, by observation, whether ChatGPT could perform adequately in the role of a virtual SP in a medical school. Our investigation

therefore focused on the following research questions:
1. How do students perceive the effectiveness of ChatGPT compared to traditional SPs in medical training scenarios?
2. To what extent can ChatGPT function effectively as a virtual SP in medical education?
To inform our research, we conducted a qualitative study with student volunteers, surveying them before and after interacting with ChatGPT. Using a few-shot approach (i.e. limited pretraining of the LLM about the expected output), we defined ChatGPT's role, setting clear expectations, such as not revealing the simulated disease.
Our findings suggest that ChatGPT can complement traditional SPs, especially when access to SPs is limited, but cannot fully substitute in this role.
Methods: This study was given expedited approval by the AUA IRB. Eleven medical student volunteers enrolled in the MD course at AUA were recruited via a campus-wide email. Students were 5 female and 6 males, ages ranging from 20-32 years, comprising students from both 1st and 2nd year of the basic sciences course section of the MD program. Participants were explicitly informed that their involvement in the research was completely voluntary. They were also assured that their responses would remain confidential and anonymous and all participants signed informed consent agreements.
Students were given access to ChatGPT version 4 accounts, the most recent available at the time of the study. Students were observed during a live role-play, in which a student inputted a predetermined prompt, provided by the study authors, into ChatGPT. The prompt directed the LLM to present as a patient with a neurological condition (Figure 1.).
Figure 1. Prompt used in ChatGPT role play.
The student, in the role of physician, then interviewed the ChatGPT and attempted to make a differential diagnosis (Figure 2.).
Figure 2. Screenshot example of ChatGPT SP interaction.
The role play was conducted verbally, as a voice control extension added to the ChatGPT accounts allowed natural language conversation between the student and the LLM [14]. A structured questionnaire consisting of 15 open-ended questions was administered before and after interaction with ChatGPT in the role of an SP. Students were asked about specific elements of their interaction and interviews were conducted in person by faculty team members. (Multimedia Appendix 1.).
Thematic Analysis:
The results of the students' group work were recorded, transcribed, and coded by three different authors (JC, TK, RR). Following discussions in regular meetings, findings were summarized, and a category system consisting of main and subcategories, according to Mayring's qualitative content analysis, was agreed upon [15].
Selected text passages were used as quotations to illustrate each category. Inductive category formation was performed to reduce the content of the material to its essentials (bottom-up process).

Results:

The thematic analysis of student feedback prior to interaction with ChatGPT as an SP identified several key themes and subthemes (Table 1.). Under the theme of technology limitations, students noted challenges with prompt engineering, such as difficulty in asking effective questions, because the AI could not roleplay a physical examination. In terms of learning efficacy, students mentioned the potential for personalized learning materials, grammatical assistance, and the ability for repeated practice without the constraints of limited SP availability. Additionally, some students highlighted the potential for increased convenience, as they could practice as often and whenever they wanted. A potential reduction in SP interview stress was also seen as a benefit of increase virtual practice. However, under the theme of verisimilitude, students expressed concerns about the absence of visual cues and rapport, which are important in real patient interactions. Finally, trust issues were raised regarding the accuracy of the LLMs output.

Table 1. Thematic analysis of student feedback pre-interaction with ChatGPT SP.

Themes and Subthemes	Representative Quotations
Theme 1. Technology Limitations	
Prompt engineering	“The challenges might be just asking the right questions, because it's an AI you can't ask them to do physical examinations”
Theme 2. Learning Efficacy	
Personalized Learning materials	“Triple checking work and not only getting the right answer, but getting explanations for the right answer and then why the wrong answer is wrong”
Grammatical assistance	“It would be helpful because English is not my first language”
Repetition	“There's usually 10 medical students to one patient, and sometimes you're fighting over each other to get the interview, so this allows us to get more repetitions”
Depth of medical knowledge	“The sky's the limit with regards to what we can practice”
Interview stress/anxiety	“It will kind of be a bit more stress free because you know you're talking to a computer rather than an actual patient”
Convenience	“Be able to practice it as much as I want, as often as I want and any time I want”
Theme 3. Verisimilitude	
Absence of visual cues	“You have to figure out ways to ask the question without the visual cues”
Absence of rapport/empathy	“Building the communication and the relationship with your patient is important”
Theme 4. Trust	
Inaccurate output	“One incident was in the small group activity, where it gave us the wrong answer”

Following interaction with ChatGPT, the thematic analysis of student feedback revealed some changes in perceptions (Table 2.). While technology limitations were still noted, students mentioned they had learnt to improve the output from ChatGPT by tailoring prompts. They also reported alignment issues, such as ChatGPT not providing information on sensitive topics like patient sexual history. Learning efficacy remained a significant theme, with students appreciating the convenience and repetition benefits. They found the ability to practice history taking without stress and receive feedback useful for skill development. However, verisimilitude issues persisted as a theme, with students noting the absence of visual and tonal cues, and the lack of rapport and empathy, all of which impacted the effectiveness of the patient interview and the ability to make a diagnosis. Some students experienced information overload, feeling that ChatGPT provided more information than a real patient would.

Table 2. Thematic analysis of student feedback post-interaction with ChatGPT SP.

Themes and Subthemes	Representative Quotations
Theme 1. Technology Limitations	
Prompt engineering	“You could put in the prompt that you want to tailor the responses you want to get back”
Alignment	“When I asked like about sexual history, they were not able to give information”
Theme 2. Learning Efficacy	
Convenience	“Having ChatGPT to practice history whenever we want, I think that's the improvement”
Repetition	“You are able to have a lot more repetitions than you are in lab”
Interview stress/anxiety	“Since it's a computer, it's not real. I had less anxiety”
Feedback	“I can ask “hey, how did you think I did?””
Skills development	“It highlighted the importance of on-the-spot thinking and memory recall in a medical scenario”
Overall Enhanced learning	“It's going to make you sharper. You know, you're probably going to be ahead of your peers, you're going to be able to answer a patient in a better, more detailed manner. Give them a better treatment or care plan”
Theme 3. Verisimilitude	
Absence of visual cues	“For the standardized patient you physically see them. You can see if they're in pain, they don't have to explain where they are in pain”
Absence of tonal cues	“ChatGPT had the same tone, even if it was saying something sad”
Absence of rapport/empathy	“It takes away the personal connection between the doctor and the patient”
Information Overload	“It felt like it was offering more information than a regular patient would”

Discussion:

Principle Findings: This study investigated the use of ChatGPT as an SP by qualitative analysis of students' responses to a questionnaire pre-interaction and post-interaction with ChatGPT performing the role of SP. Students overall were positive about the potential for ChatGPT to augment their clinical skills via this method, echoing the results of other studies [16–18]. Thematic analysis provided insights into student perceptions. Major themes identified were technology limitations, learning efficacy and verisimilitude. Our results suggest that the current version of ChatGPT (ChatGPT version 4 at time of this study) can function effectively as an augmentation to traditional SPs, but cannot fully substitute for SPs.

The technological limitations of LLMs in the context of SP exercises were both anticipated and confirmed in our study. The subtheme of prompt engineering was particularly important. Students were aware of the importance of correctly worded prompts before the exercise, and post-interaction interviews indicate students were using a trial-and-error approach to prompt engineering to improve the output. The importance of prompt engineering has been noted in previous studies [16,19,20]. Our results suggest instruction in prompt engineering should form part of medical curricula which are incorporating virtual SPs in future and should also be part of a more general integration of AI into medical curricula.

The post-interaction interviews also revealed an additional subtheme of alignment. Alignment refers to the problem of ensuring that AI acts in accordance with human intentions and human values [21]. Students noted that the LLM did not provide a response when asking about a patient's sexual history, a standard question in any medical consultation. Ensuring ChatGPT does not output material which could be considered offensive under societal norms is a component of alignment [22]. However, our results demonstrate an “alignment tax”, in that the model becomes less useful due to constraints imposed by the alignment. The development of LLMs designed specifically for medical education may overcome this issue [23].

Learning efficacy was also a major theme identified in this study. Important subthemes in this category were repetition and convenience. Students noted the benefits of having access to ChatGPT for practice at any time or place and having virtually unlimited ability to repeat the exercises. As mentioned earlier, access to SPs is limited in many medical schools [10].The ability to augment this shortfall with a virtual SP may be a positive option for many medical students and medical schools.

Interestingly, some students expressed that they experienced considerable anxiety as much as a day before they were scheduled to interact with an SP, even though they were aware that the SP was not a real patient. The ability to practice with an LLM, like ChatGPT, was seen as beneficial, because students could develop questioning techniques to a point where even during the session with a real SP they could still perform well.

Some differences between pre-interaction and post-interaction in terms of subthemes were evident under the major theme of learning efficacy. Before the exercise students were focused more on anticipated or previous experiences of the use of LLMs for personalized learning materials, for example, developing mnemonics, practice questions or flashcards. This reflects the experience of other medical students [24]. Responses following the exercise were focused on diagnostic patient interaction skills. This is to be expected as students now had actual experience of ChatGPT in this role and knew this was to be the focus of our study.

Verisimilitude was a major theme in both pre-interaction and post-interaction responses. All students mentioned this as a limiting factor. Absence of facial cues, changes in tone, or body language and an inability to develop rapport were all seen as drawbacks of the virtual SP. Some students also mentioned this impacted their role as physician. For example, a student physician leaning in to the patient to show interest, or other types of body language, was redundant in the exercise. Other studies have also highlighted that the output from ChatGPT cannot replicate the true stimuli a physician relies on in a patient visit [20,25,26]. We note virtual patients are developing rapidly, so issues with verisimilitude may be overcome in future, although it may take some time before

ChatGPT, specifically, is able to incorporate a visual or physical layer.

Trust as a theme was evident in pre-interview responses, but had disappeared in post-interview responses. We note that our faculty team, consisting of clinicians and PhD-qualified members, did not notice any “hallucinations” in output, despite multiple repetitions of the exercise. Yanagita et al. recently found high quality illness scripts, used for improving medical student’s clinical reasoning, could be generated by ChatGPT with relatively few errors [27]. Magalhaes et al. also found a majority of students trusted ChatGPT’s output [17].

Nevertheless, even a single error in ChatGPT output, given multiple healthcare providers may receive the same output, could affect many patients. It is therefore imperative that the veracity of AI output be thoroughly tested before it is fully integrated into healthcare and medical education settings.

Other subthemes for learning efficacy evident post-interaction were feedback and information overload. Our prompt included a direction for ChatGPT to provide feedback on how students could improve their performance. We note it was necessary to revise the prompt several times during the study, as initially it only provided positive feedback, which did not help in identifying areas for improvement. Responses under the information overload subtheme suggested students found the LLM tended to provide more information in regard to a given question than perhaps a real patient or SP would. This presumably related to the depth of medical knowledge of the LLM, but should be considered in further iterations of this exercise. It may be possible to refine the prompt to reduce this effect.

Overall, in terms of its performance as a virtual SP, most students were positive about ChatGPT's potential. Post-interviews reflected this sentiment, and our observations confirmed its capability in this role. Minimal errors were detected in its outputs, and although there was some response latency (gap of a few seconds), it was negligible. The overall interaction felt natural, and well-engineered prompts enhanced its utility by allowing for tasks like providing laboratory reports and feedback.

Study Limitations: This study had a small sample size, making it difficult to conduct any statistical analysis. However, we note this was a qualitative exploratory study only, intended to establish if ChatGPT could perform the role of SP and to gain a qualitative view of student user’s perceptions. Future studies will involve a much larger sample size, allowing for quantitative, as well as qualitative, data to be obtained. A second potential limitation was ascertainment bias. It is possible students interested in technology in general, and ChatGPT in particular, were drawn to volunteer for the study, leading to skewed results. Indeed, many of the students appeared to be “early adopters” of technology judging from interview comments. More careful vetting and a larger sample size may help to overcome this issue in future studies.

Conclusions: This study found ChatGPT to be an effective supplement, though not a full replacement, to traditional SPs. Students and faculty appreciated its potential, noting benefits like flexible practice times, reduced stress, and improved diagnostic skills. Some shortcomings were noted, including the need for effective prompt engineering and the lack of non-verbal cues affecting realism. Despite these challenges, its reliability and convenience make it a valuable training tool.

Overall, ChatGPT offers a significant adjunct to traditional SPs, providing accessible, flexible practice opportunities for medical students. The study underscores the importance of integrating prompt engineering into medical curricula and refining AI interactions for balanced information delivery. Continuous advancements in virtual patient technology and AI capabilities, including improved verbal and auditory flow, are expected to further enhance ChatGPT’s utility in medical education. Future studies are planned with a larger sample size and utilizing the recently released ChatGPT version 4o, which has enhanced auditory and verbal capacities.

Acknowledgements: We thank Richard Millis, PhD, for support and guidance during this study.
Authors Contributions: JC conceived the original idea and study design. JC, TK, RR, SD, AV, RH, PH, SN, RS, CL, NA and JJ refined the study design and conducted the study activity. TK, RS and RR conducted study participant interviews. AM, AJ, EC, TA, ANN and JG data organization, input and analysis. JC, TK and RR coding and theme construction. JC, TK and RR writing original draft of paper. All authors contributed to the review and editing of the paper and approved the submitted version

Conflicts of Interest: None declared.

References

1. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42 1. Med Teach. Informa Healthcare; 2009. p. 477–486. PMID:19811162
2. Flanagan OL, Cummings KM. Standardized Patients in Medical Education: A Review of the Literature. Cureus 2023 Jul 17; doi: 10.7759/cureus.42027
3. Phanudulkitti C, Puengrung S, Meepong R, Vanderboll K, Farris KB, Vordenberg SE. A systematic review on the use of virtual patient and computer-based simulation for experiential pharmacy education. Exploratory Research in Clinical and Social Pharmacy. Elsevier Inc.; 2023. doi: 10.1016/j.rcsop.2023.100316
4. Huang G, Reynolds R, Candler C. Virtual Patient Simulation at U.S. and Canadian Medical Schools. Academic Medicine 2007 May;82(5):446–451. doi: 10.1097/ACM.0b013e31803e8a0a
5. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S, Nori H, Palangi H, Ribeiro MT, Zhang Y. Sparks of Artificial General Intelligence: Early experiments with GPT-4. 2023 Mar 22; Available from: http://arxiv.org/abs/2303.12712
6. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: Potential for AI-

assisted medical education using large language models. Dagan A, editor. PLOS Digital Health 2023 Feb 9;2(2):e0000198. doi: 10.1371/journal.pdig.0000198
7. Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, Aljamaan F, Malki KH, Halwani R, Al-Tawfiq JA, Temsah M-H, Al-Eyadhy A. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. Cureus 2023 Apr 8;15(4):e37281. PMID:37038381
8. Oscer ai. Oscer. 2024. URL: https://www.oscer.ai/about. [accessed 2024-06-17]
9. University of Texas Medical Branch. Virti. 2024. URL: https://www.utmb.edu/som/som-educational-affairs/office-of-clinical-education-(oce)/clerkship---pediatrics/pedi-clerkship. [accessed 2024-06-17]
10. Liu X, Wu C, Lai R, Lin H, Xu Y, Lin Y, Zhang W. ChatGPT: when the artificial intelligence meets standardized patients in clinical training. J Transl Med. BioMed Central Ltd; 2023. doi: 10.1186/s12967-023-04314-0
11. Suárez A, Adanero A, Díaz-Flores García V, Freire Y, Algar J. Using a Virtual Patient via an Artificial Intelligence Chatbot to Develop Dental Students' Diagnostic Skills. Int J Environ Res Public Health MDPI; 2022 Jul 1;19(14). PMID:35886584
12. Weidener L, Fischer M. Artificial Intelligence Teaching as Part of Medical Education: Qualitative Analysis of Expert Interviews. JMIR Med Educ JMIR Publications Inc.; 2023;9. doi: 10.2196/46428
13. Jowsey T, Stokes-Parish J, Singleton R, Todorovic M. Medical education empowered by generative artificial intelligence large language models. Trends Mol Med 2023 Sep; doi: 10.1016/j.molmed.2023.08.012
14. Aidia ApS. Voice control for ChatGPT x MIA AI. 2024. URL: https://chromewebstore.google.com/detail/voice-control-for-chatgpt/eollffkcakegifhacjnlnegohfdlidhn?hl=en&pli=1. [accessed 2024-06-17]
15. Mayring P. Qualitative Content Analysis Theoretical Foundation, Basic Procedures and Software Solution. Approaches to Qualitative Research in Mathematics Education Advances in Mathematics Education 2015. doi: https://doi.org/10.1007/978-94-017-9181-6_13
16. Kavadella A, Da Silva MAD, Kaklamanos EG, Stamatopoulos V, Giannakopoulos K. Evaluation of ChatGPT's Real-Life Implementation in Undergraduate Dental Education: Mixed Methods Study. JMIR Med Educ JMIR Publications Inc.; 2024;10(1). doi: 10.2196/51344
17. Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in Medical Informatics Education: Exploring Student Perceptions and Proposing Experiential Integration (Preprint). JMIR Med Educ JMIR Publications Inc.; 2023 Jul 27; doi: 10.2196/51151
18. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online Taylor and Francis Ltd.; 2023;28(1). PMID:36855245
19. Gray M, Baird A, Sawyer T, James J, DeBroux T, Bartlett M, Krick JA, Umoren RA. Increasing Realism and Variety of Virtual Patient Dialogue: Novel Application of the Artificial Intelligence ChatGPT to enhance a Conversational Virtual Patient for Prenatal Counseling Education (Preprint). JMIR Med Educ JMIR Publications Inc.; 2023 Jul 17; doi: 10.2196/50705
20. Webb JJ. Proof of Concept: Using ChatGPT to Teach Emergency Physicians How to Break Bad News. Cureus Springer Science and Business Media LLC; 2023 May 9; doi: 10.7759/cureus.38755
21. Open AI. Open AI. 2024. URL: https://openai.com/index/instruction-following/. [accessed 2024-06-17]
22. Open AI. Open AI Safety. 2024. URL: https://openai.com/index/our-approach-to-ai-safety/. [accessed 2024-06-17]
23. Tu T, Palepu A, Schaekermann M, Saab K, Freyberg J, Tanno R, Wang A, Li B, Amin M, Tomasev N, Azizi S, Singhal K, Cheng Y, Hou L, Webson A, Kulkarni K, Mahdavi SS, Semturs C, Gottweis J, Barral J, Chou K, Corrado GS, Matias Y, Karthikesalingam A, Natarajan V. Towards Conversational Diagnostic AI. arXiv preprint 2024 Jan 10; doi: https://doi.org/10.48550/arXiv.2401.05654

24. Hashemi-Sabet F, Leung F-H. Equity-driven MCAT prep: a ChatGPT advantage. Can Med Educ J The Association of Faculties of Medicine of Canada; 2024 Jan 15; doi: 10.36834/cmej.78169
25. Musallam E, Alhaj Ali A, Alkhafaji M. OpenAI's ChatGPT Clinical Simulation. Nurse Educ 2024 May 9; doi: 10.1097/NNE.0000000000001657
26. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The Role of Large Language Models in Medical Education: Applications and Implications. JMIR Med Educ 2023 Aug 14;9:e50945. PMID:37578830
27. Yanagita Y, Yokokawa D, Fukuzawa F, Uchida S, Uehara T, Ikusaka M. Expert assessment of ChatGPT's ability to generate illness scripts: an evaluative study. BMC Med Educ 2024 May 15;24(1):536. doi: 10.1186/s12909-024-05534-8

Abbreviations
AUA: American University of Antigua
LLM: Large Language Model
AI: Artificial Intelligence
SP: Standardized Patient
OSCE: Observed Structured Clinical Examination
IRB: Institutional Review Boar

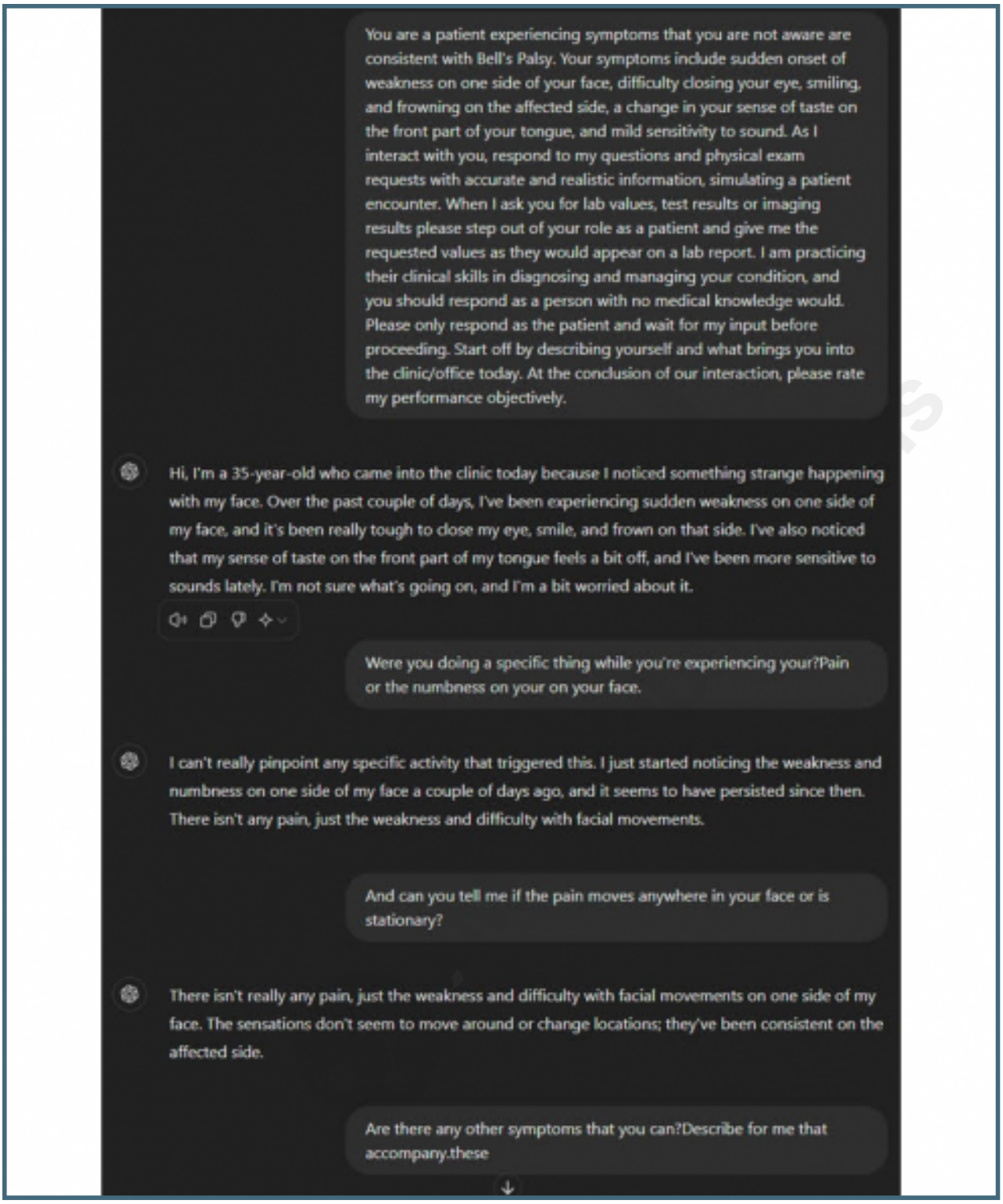
Supplementary Files

Figures

Prompt used in ChatGPT role play.

You are a patient experiencing symptoms that you are not aware are consistent with Bell's Palsy. Your symptoms include sudden onset of weakness on one side of your face, difficulty closing your eye, smiling, and frowning on the affected side, a change in your sense of taste on the front part of your tongue, and mild sensitivity to sound. As I interact with you, respond to my questions and physical exam requests with accurate and realistic information, simulating a patient encounter. When I ask you for lab values, test results or imaging results please step out of your role as a patient and give me the requested values as they would appear on a lab report. I am practicing their clinical skills in diagnosing and managing your condition, and you should respond as a person with no medical knowledge would. Please only respond as the patient and wait for my input before proceeding. Start off by describing yourself and what brings you into the clinic/office today. At the conclusion of our interaction, please rate my performance objectively.

Screenshot example of ChatGPT SP interaction.



Multimedia Appendixes

Interview questions.
URL: http://asset.jmir.pub/assets/331f0dd77c616d3456888fa535f96484.docx