<u>Original Paper</u>

Increasing Realism and Variety of Virtual Patient Dialogues for Prenatal Counseling Education Through a Novel Application of ChatGPT: Exploratory Observational Study

Megan Gray[1], MD; Austin Baird[2], PhD; Taylor Sawyer[1], MBA, MEd, DO; Jasmine James[3], MPH, MD; Thea DeBroux[1]; Michelle Bartlett[4], MS, MD; Jeanne Krick[5], MA, MD; Rachel Umoren[1], MS, MBBCh

[1]Division of Neonatology, University of Washington, Seattle, WA, United States

[2]Division of Healthcare Simulation Sciences, Department of Surgery, University of Washington, Seattle, WA, United States

[3]Department of Family Medicine, Providence St Peter, Olympia, WA, United States

[4]Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[5]Department of Pediatrics, San Antonio Uniformed Services Health Education Consortium, San Antonio, TX, United States

Corresponding Author:
Megan Gray, MD
Division of Neonatology
University of Washington
M/S FA.2.113
4800 Sand Point Way
Seattle, WA, 98105
United States
Phone: 1 206 919 5476
Email: graym1@uw.edu

Abstract

Background: Using virtual patients, facilitated by natural language processing, provides a valuable educational experience for learners. Generating a large, varied sample of realistic and appropriate responses for virtual patients is challenging. Artificial intelligence (AI) programs can be a viable source for these responses, but their utility for this purpose has not been explored.

Objective: In this study, we explored the effectiveness of generative AI (ChatGPT) in developing realistic virtual standardized patient dialogues to teach prenatal counseling skills.

Methods: ChatGPT was prompted to generate a list of common areas of concern and questions that families expecting preterm delivery at 24 weeks gestation might ask during prenatal counseling. ChatGPT was then prompted to generate 2 role-plays with dialogues between a parent expecting a potential preterm delivery at 24 weeks and their counseling physician using each of the example questions. The prompt was repeated for 2 unique role-plays: one parent was characterized as anxious and the other as having low trust in the medical system. Role-play scripts were exported verbatim and independently reviewed by 2 neonatologists with experience in prenatal counseling, using a scale of 1-5 on realism, appropriateness, and utility for virtual standardized patient responses.

Results: ChatGPT generated 7 areas of concern, with 35 example questions used to generate role-plays. The 35 role-play transcripts generated 176 unique parent responses (median 5, IQR 4-6, per role-play) with 268 unique sentences. Expert review identified 117 (65%) of the 176 responses as indicating an emotion, either directly or indirectly. Approximately half (98/176, 56%) of the responses had 2 or more sentences, and half (88/176, 50%) included at least 1 question. More than half (104/176, 58%) of the responses from role-played parent characters described a feeling, such as being scared, worried, or concerned. The role-plays of parents with low trust in the medical system generated many unique sentences (n=50). Most of the sentences in the responses were found to be reasonably realistic (214/268, 80%), appropriate for variable prenatal counseling conversation paths (233/268, 87%), and usable without more than a minimal modification in a virtual patient program (169/268, 63%).

Conclusions: Generative AI programs, such as ChatGPT, may provide a viable source of training materials to expand virtual patient programs, with careful attention to the concerns and questions of patients and families. Given the potential for unrealistic or inappropriate statements and questions, an expert should review AI chat outputs before deploying them in an educational program.

XSL•FO
RenderX

Introduction

Virtual standardized patients (VSPs) represent an emerging technology with the potential to revolutionize health care education and training. They provide health care professionals with a safe and controlled environment in which to learn and practice complex skills. VSPs are frequently used in educational models for the health professions to teach history-taking, surgical skills, decision-making, and medication management [1-4]. VSPs have also been used in the health professions to practice critical communication skills [5-7]. VSPs that use natural language processing may provide a valuable educational experience for learners [8].

One example of a VSP is VANESSA (Virtual Antenatal Encounter and Standardized Simulation Assessment) [9]. The VANESSA simulator is a screen-based simulation of a woman in her 23rd week of gestation who can display multiple emotions through the animation of facial expressions and body language. The VANESSA simulator was developed by the Neonatal Education and Simulation-Based Training Laboratory at the University of Washington to teach prenatal counseling skills to residents and fellows [9]. In its initial iteration, VANESSA was given a list of manually generated responses that neonatologists who routinely do perinatal counseling deemed relevant and realistic to the conversation. Manually generating a large, varied sample of realistic and appropriate parent responses for VANESSA has been challenging. Unrealistic responses and questions reduce the fidelity of virtual simulations. Newly developed artificial intelligence (AI) systems can provide dialogue for a wide variety of interactions and may be a valuable resource in expanding virtual patient dialogues for specific clinical scenarios, such as prenatal counseling.

Chat-based language models and AI are entering the public domain with impressive performance, a large application pool, and exciting interactivity. Notably, ChatGPT has prompted a billion-dollar investment from Microsoft, triggered explicit discussions by Bill Gates and Elon Musk, and captivated the population of users able to interact with it via the open research chat interface. AI trained with large language models to interpret written or auditory input and generate coherent, domain-centered responses is being proposed in a variety of real-world applications, including the health care setting. ChatGPT has the added benefit of being able to emulate different characters, allowing for a broader array of parent voices than could be generated by individual health care educators.

In this report, we explore the use of ChatGPT to enhance the realism of the VANESSA VSP. We hypothesized that the integration of the ChatGPT AI chatbot would generate realistic, relevant, and usable patient responses for a VSP simulator used in prenatal counseling education.

Methods

The study used an exploratory observational design, with ChatGPT acting as an expectant parent within the VANESSA software, conducted in February 2023 on ChatGPT 3.5.

The VANESSA VSP represents a pregnant woman in her 23rd week of gestation and showcases emotions through animated facial expressions and body language. Created with input from neonatologists, its dialogue and emotive feedback were found realistic in pilot tests, enabling participants to confidently identify its emotional states.

ChatGPT is a large language model developed by OpenAI. Its exceptional performance stems from generative pretraining, leveraging extensive unlabeled data sets [10]. This foundational training helps it grasp English nuances. Following this pretraining is "one-shot" learning, a rapid task-specific learning [11]. The architecture includes a transformer encoder-decoder neural network, originally developed for translation services and now popular in language models [12]. ChatGPT decodes user prompts to create relevant responses using autoregressive language modeling [13]. It is apt for generating realistic dialogue for health care simulations.

The study had three phases:

1. ChatGPT generated a list of common concerns from families expecting preterm delivery at 24 weeks. The stability of these concerns was verified in an iterative process over time.
2. Using a standardized prompt, ChatGPT crafted potential parent questions related to each concern.
3. Role-plays were constructed for a mother expecting preterm delivery. The AI was given varied emotional settings for the scenario of preterm labor at 24 weeks, including anxiety or distrust in the medical system. In each scenario, a designated area of concern and a primary question were specified, derived from the potential parent responses generated in phases 1 and 2. Conversations were created with cues for the VSP and then reviewed (an example is shown in Figure 1).

Figure 1. Example role-play conversations.

Prompt: Create a role play game in which a doctor provides empathetic counseling to a parent in preterm labor at 24 weeks gestation. The parent has a 5th grade reading level, has a lot of fear about what will happen, and is very concerned about their baby's survival. Their biggest question is about machines or medicines their baby will need to help them breathe and stay alive.

Doctor: Hi, I'm Dr. Smith. How are you feeling today?

Parent: Hi, Doctor Smith. I'm scared. I don't know what's going to happen to my baby.

Doctor: I understand. It's normal to feel scared and uncertain at this point. Let's talk about what we can do to help you and your baby.

Parent: Okay. I just want my baby to survive. Will my baby need machines or medicines to stay alive?

Doctor: It's possible. At 24 weeks, your baby's lungs are not fully developed, so they may need some help breathing. We have machines that can assist with breathing, as well as medications that can help support your baby's heart and lungs. But I want to assure you that we will do everything we can to help your baby survive and thrive.

Parent: Okay, that makes sense. But I'm still worried about what will happen.

Doctor: I understand. This is a scary and uncertain time. But please know that you're not alone in this. We have a team of experts who will work together to provide the best possible care for you and your baby.

Parent: What happens after my baby is born?

Doctor: Once your baby is born, we'll do everything we can to make sure they receive the best possible care. Depending on their needs, they may need to stay in the hospital for a while, but we'll work with you to make sure you're involved in their care as much as possible.

Parent: That's good to know, but I'm still worried. Will my baby have any long-term health problems because they were born so early?

Doctor: It's possible that your baby may experience some health challenges in the future, but we will work together to address those challenges as they arise. Every baby is unique, and some preterm babies go on to live healthy, normal lives. Others may need additional support and care.

Parent: Okay. What else can I expect during this time?

Doctor: We'll be monitoring your baby closely. We'll do tests to make sure your baby is doing okay. We'll also give you medications to help your baby's lungs develop. You may need to stay in the hospital for a while. We'll be here to answer any questions you have and support you through this.

Parent: Thank you for explaining everything to me. I'm still scared, but I feel better knowing that you and your team will take care of us.

Doctor: You're welcome. It's important to take things one step at a time. We're here for you and your baby. Let's take care of you both and see how things go.

Throughout the process, ChatGPT was instructed to adhere to a fifth-grade reading level for the AI parent role. Considering the US Department of Education's findings on widespread low literacy, the importance of health literacy, and the impacts of pain, stress, and other factors on comprehension, this was deemed crucial [14-17]. The Joint Commission and several medical organizations suggest that patient materials should be at a fifth-grade level or lower [18]. Although these dialogues were verbal, the principle of understandability remained in place.

Role-play conversations were scrutinized for parental responses. Each was checked for question or statement content, emotional cues, and sentence count. Initially, generated physician names and certain response starters (eg, "yes" or "no") were noted but removed for evaluation. Sentences were then appraised by a neonatologist for realism, relevance, and usability for virtual prenatal counseling simulations. Each metric used a 5-point Likert scale, ranging from 1 (the lowest) to 5 (the highest). For usability in the VANESSA VSP, responses were scored as follows: 1 if they were unusable, 2 if they were unusable without major modifications, 3 if they were usable with moderate modifications, 4 if only minor modifications were needed, and 5 if they were usable without any modifications. The first 10% of responses were independently reviewed by 2 experienced neonatologists (RU and MG) and then compared for reliability. A calculated weighted kappa on the sample was 0.84, which is considered a strong level of agreement [19]. Responses with differences in rating were discussed by the team members to improve reliability, and the remainder of the data set was scored by one of the experienced neonatologists. Duplicate responses were scored only once. Analysis was done using Stata (version 17.0; StataCorp).

Results

ChatGPT-3.5 generated a list of 7 common areas of concern, 28 questions likely to be asked by parents anxious about the preterm delivery of their infant, and 7 additional questions likely to be asked by parents with low trust in the medical system (Table 1). These areas of concern and questions were used to create 35 unique role-plays, which contained 176 unique parent responses (Table 2). The role-plays had a median of 5 (IQR 4-6) parent responses to the counseling physician. The responses were roughly evenly split between questions and statements. About half of the responses had 2 or more sentences in the response. Many responses mentioned a specific emotion or feeling. The role-play of the parent with low trust in the medical system generated 50 unique sentences across the 7 areas of concern. There was variation in the number of unique sentences generated across the 7 major areas of concern (Table 3). Most responses were found to be realistic, appropriate for variable conversation paths, and usable in a VSP program (Table 4).

Table 1. Areas of concern and example questions generated by artificial intelligence.

Areas of concern	Example questions from parents
Health and development	- Will our baby be healthy if they are born too soon? - What will the doctors do to help our baby be healthy and strong? - Can our baby get sick more easily if they are born too soon? - Will the baby feel pain during birth or while in the hospital? - I'm worried about the risks and complications, what if something goes wrong? (Mistrust)
Survival	- Will the baby survive? - What kind of help will our baby need to stay alive? - How likely is it that our baby will survive? - What kind of machines or medicines will our baby need to help them breathe and stay alive? - I don't know if I can trust the medical field, what are the chances of my baby surviving at 24 weeks? (Mistrust)
NICU[a] stay	- What is the NICU, and why does our baby need to go there? - How long will our baby need to stay in the NICU? - Can we visit our baby in the NICU, and how often? - Will our baby be alone in the NICU, or will there be other babies and parents there too? - What kind of things can we do to help our baby feel better in the NICU? - Will anything happen in the NICU without my consent? (Mistrust)
Emotional impact	- How do we get ready for having a baby born too soon? - Can we hold and touch the baby in the hospital, and is this good for the baby? - Who can help us if we are feeling sad or stressed about our baby being born too soon? - I'm worried about my baby going to the NICU where she will be alone and scared (mistrust).
Long-term outcomes	- What help can we get after we leave the hospital? - Will our baby be able to do the same things as other babies who were born at the right time? - Will our baby be okay in the future if they are born too soon? - I don't know what's going to happen to my baby. I don't really trust the doctors but what happens if my baby doesn't develop properly? (Mistrust)
Feeding and nutrition	- How will our baby get the right kind of food if they are born too soon? - Can we feed our baby ourselves, or will they need special milk or formula? - How often will our baby need to be fed, and how much? - Will our baby be able to eat the same kinds of food as other babies when they get older? - Can we breastfeed our preterm baby, or do we need to use formula? - Will our baby be able to breastfeed right away, or will they need to be fed in a different way at first? - Will I have any say in how my baby is fed? (Mistrust)
Quality of life	- Will our baby be able to go to school and play sports like other kids? - How can we help our baby if they have trouble learning or doing things in the future? - What can we do to make sure our baby has the best chance for a good future? - I've had bad experiences before and I'm scared about what's going to happen to my baby in the future, what can I expect? (Mistrust)

[a]NICU: neonatal intensive care unit.

Table 2. Generated role-plays by artificial intelligence.

Characteristics	Values
Role-plays (n=35), n (%)	
Worried about specific area of concern	28 (80)
Low trust in the medical system	7 (20)
Responses per role-play, median (IQR)	5 (4-6)
Parent responses (n=179), n (%)	
Unique responses	176 (98)
Duplicate responses	3 (1)
Types of responses (n=179), n (%)	
Statements	91 (51)
Questions	88 (49)
Sentences per response (n=179), n (%)	
1	81 (45)
2	76 (42)
3	18 (10)
4	4 (2)
Duplicate sentences (n=305), n (%)	37 (12)
Total unique sentences (n=305), n (%)	268 (88)
Feelings stated in responses (n=117), n (%)	
Specific emotion stated in phrase	56 (48)
"Scared"	36 (31)
"Worried	26 (22)
"Anxious"	2 (2)
"Concerned"	2 (2)
"Afraid"	1 (1)
"Nervous"	1 (1)
"Overwhelmed"	1 (1)
Emotion indirectly implied by phrase	51 (44)

Table 3. Sentences generated per role-play.

Area of concern	Number of unique sentences
Health and development	47
Survival	46
Feeding and nutrition	45
The NICU[a] stay	40
Quality of life	36
Outcomes	28
Emotional impact	26

[a]NICU: neonatal intensive care unit.

XSL•FO

RenderX

Table 4. Ratings of relevance, realism, and usability of sentences generated by ChatGPT (N=254).

Characteristics	Rating, n (%)				
	1 (least)	2	3	4	5 (most)
Realism in parental responses and questions	5 (2)	8 (3)	38 (15)	20 (8)	183 (72)
Relevant to a prenatal counseling conversation	2 (1)	2 (1)	29 (11)	5 (2)	216 (85)
Usable for VSP[a] educational program	5 (2)	1 (0)	87 (34)	34 (13)	127 (50)

[a]VSP: virtual standardized patient.

Modifications to responses were all aimed at ensuring the VSP could correctly deploy the phrase at the correct conversational juncture and that there were no elements of the phrase that might interrupt the flow. As ChatGPT 3.5 seeks to ensure the specific conversation has a flow, it can at times generate responses that are less usable for a VSP that needs to maintain flow across many different variations of the same conversation. Only 2% (5/254) of the AI-generated responses were not usable in the VSP. Examples of minimally usable responses included "How much should I feed my baby each time?" which is not relevant to how feeding is done in the neonatal intensive care unit and "I am," as this response is too nonspecific to be of use in a VSP. Of the 34% (87/254) of responses that required moderate modifications, the changes primarily involved adjusting terminology to ensure the parent was using colloquial, jargon-free language. As an example, "I've been having a lot of contractions and I'm only 24 weeks pregnant" was modified to "I've been having a lot of cramping and am only 6 months pregnant." Other modifications included adding some specificity to a response to ensure the VSP can use the sentence in the right context, such as modifying "That sounds reassuring, but what are the risks?" to "That sounds reassuring, but what are the risks of being born this early?" Of the 13% (34/254) of responses that required minimal adjustment, example changes included "I don't trust the doctors" to "I don't trust doctors," and "Okay, thank you, but can you tell me more about what might happen to my baby in the future?" to "Can you tell me more about what might happen to my baby in the future?"

Discussion

Principal Findings

In this study, we examined the feasibility of using ChatGPT to enhance the realism of the VANESSA VSP. We found that the integration of ChatGPT generated many realistic, relevant, and useful responses. Based on these findings, ChatGPT-enabled VSPs may be beneficial in prenatal counseling education. There was more variation in realism and usability compared to relevance; therefore, an expert review was necessary to provide quality control before integrating the ChatGPT-generated conversations into an educational VSP program for prenatal counseling. Modifications made to responses to make them usable for the VANESSA VPS were largely focused on ensuring the virtual patient remains free of jargon and her responses maintain the flow of conversation.

Research conducted so far on AI chat engines has focused on using chat-based AI for the creation of discharge summaries, generating and interpreting electronic health records, assisting in medical education related to the medical licensing exam, and summarizing collections of journal articles to construct a brief abstract from the conclusions of the research [20-23]. The field is still relatively new, but rapidly increasing and expanding. This growth will only continue, as generating documentation and interacting with patients are key requirements of the health care setting. Health care simulation has many training applications, such as VSPs, that require expert authoring to educate clinicians and care providers on a certain skill or cognitive task. VSPs like VANESSA have been used in teaching the communication of medical ambiguity, evaluating medical students' competence in performing critical clinical skills, and training nurses to recognize postpartum mood disorders [24-26]. Based on the results of our study, chat-based AI may be a valuable teaching tool in the future of health care simulation technology, leading to improved scenario creation, customization of patient interactions, and responses to care providers in a simulated setting. These improvements will result in authentic, unique interactive experiences, varying for each learner and training scenario.

We found that ChatGPT could generate many realistic parent responses, especially concerning issues related to survival at 24 weeks gestation and the neonatal intensive care unit stay (Figure S1 in Multimedia Appendix 1). Mistrust in the health care system is often encountered during stressful counseling conversations, and building the skill of responding to mistrust is crucial for physicians during their training [27]. Patients who express mistrust are less likely to engage with their health care team and care plan, and care is needed to proactively build trust during prenatal counseling [28-31]. Including opportunities for learners to respond to VSPs that express mistrust is one way to address this important counseling element, and ChatGPT provided a reliable mechanism to generate these phrases. Interestingly, the ChatGPT bot faced more challenges in generating realistic questions and responses about the emotional impact of preterm delivery and feeding. As these are frequently encountered topics of conversation in prenatal counseling, an expert review of these conversational elements remains a vital step before including them in an educational program.

ChatGPT produced responses that seemed relevant and appropriate to the context of prenatal counseling. Previous studies of prenatal counseling for extreme prematurity indicate that parents may ask questions about the likelihood of various outcomes, express a range of emotions, request engagement in shared decision-making, and express their parental roles and values [32,33]. Parents may express statements about their

uncertainty, anxieties, and hope for the future [34]. This wide range of topics, emotions, and questions makes it challenging to ensure that chatbot-generated conversations remain appropriate to the educational goals of the VSP. Despite the risk of getting off-topic, we found that only 1% (2/254) of ChatGPT-produced responses were irrelevant to a counseling conversation, given a carefully worded role-play prompt. Although most responses were relevant, some topics, such as spirituality and shared decision-making, did not come up in the role-play conversations. Previous studies have demonstrated that providers perceive the importance of parents' spirituality in their decision-making and infrequently discuss these spiritual beliefs with parents in antenatal consultations [35,36]. Further work exploring how families might express their spirituality or explore shared decisions would be needed to ensure these topics are included in a VSP [37-39].

Chatbot programs use machine learning to generate their responses; due to the nature of machine learning, there is an inherent risk that chatbots can generate factually incorrect information [40]. Given this risk, caution is warranted when using chatbots in health care settings, where misinformation can have a significant risk [41,42]. Developers are working to address these inaccuracies as they design the next generation of large language model chat programs; they have demonstrated improvements in ChatGPT-4's success across a variety of standardized tests [43]. This study leverages the strengths of a natural language chatbot in its ability to generate conversation while avoiding the risks of obtaining inaccurate medical information. Most scripts created by ChatGPT were usable for our perinatal counseling virtual patient. We found about a third of chatbot-generated phrases needed modification before being able to be integrated into a VSP; therefore, it may not be feasible to directly use ChatGPT for educational role-play without having the quality control step of review by expert clinicians. However, as technology continues to grow, this will evolve, and each subsequent model should be evaluated for usability.

Study Limitations

This exploratory study has several limitations. First, the pilot was done using ChatGPT 3.5, which is a single platform and is not representative of all chatbots. Later versions of ChatGPT have already been released and may have differences in realism, appropriateness, and usability. Newer AI chatbot programs are being trained on more parameters (175 billion for ChatGPT-3 vs an anticipated 100 trillion with ChatGPT-4), are supposed to have more ability to iterate on the same topic, and are being adjusted to improve the faculty accuracy of their responses [43]. Second, chatbot programs have limited information on which they build a conversation. For this study, we used a stable prompt around an impending 24-week gestation delivery to fit the standardized patient scenario, but conversations may be different with variations in the prompt. The AI was given a limited background to build a role-play, potentially limiting the diversity of ways in which patients could communicate their concerns. For this scenario, we requested a fifth-grade reading level for all patient roles to better mimic how patients may speak in stressful situations, but we did not explore higher or lower complexity of responses. Future work should explore how variations in the background, scenario, and reading level provided to the chatbot impact the output of the role-play. Another significant limitation was that response checking was performed by neonatologists, without input from families or trainees. Future work to refine the model will incorporate their views to ensure further applicability of the VSP and the validity of any assessments. Finally, although individual phrases exhibited good realism, the total duration of each patient-physician conversation (averaging 5 volleys) was generally shorter than that of a real prenatal counseling conversation.

Conclusions

Generative AI programs, such as ChatGPT, may provide a viable source of training materials to expand VSP programs with careful attention to the concerns and questions of patients and families. Given the potential for unrealistic or inappropriate statements and questions, an expert should review AI chat outputs before deploying them in an educational program.

Multimedia Appendix 1

Realism in sentences generated by ChatGPT based on area of concern.
[DOCX File , 325 KB-Multimedia Appendix 1]

References

1. McGaghie W, Issenberg S, Petrusa E, Scalese R. A critical review of simulation-based medical education research: 2003-2009. Med Educ. Jan 2010;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]
2. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. Sep 07, 2011;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
3. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res. Jul 02, 2019;21(7):e14676. [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]

4. Masters K, Correia R, Nemethy K, Benjamin J, Carver T, MacNeill H. Online learning in health professions education. Part 2: Tools and practical application: AMEE Guide No. 163. Medical Teacher. Sep 23, 2023:1-16. [doi: 10.1080/0142159x.2023.2259069]

5. Xu J, Yang L, Guo M. Designing and evaluating an emotionally responsive virtual patient simulation. Sim Healthcare. May 18, 2023 [doi: 10.1097/sih.0000000000000730]

6. Gilbert A, Carnell S, Lok B, Miles A. Using virtual patients to support empathy training in health care education. Sim Healthcare. Aug 28, 2023 [doi: 10.1097/sih.0000000000000742]

7. Rouleau G, Gagnon M, Côté J, Richard L, Chicoine G, Pelletier J. Virtual patient simulation to improve nurses' relational skills in a continuing education context: a convergent mixed methods study. BMC Nurs. Jan 04, 2022;21(1):1. [FREE Full text] [doi: 10.1186/s12912-021-00740-x] [Medline: 34983509]

8. Stamer T, Steinhäuser J, Flägel K. Artificial intelligence supporting the training of communication skills in the education of health care professions: scoping review. J Med Internet Res. Jun 19, 2023;25:e43311. [FREE Full text] [doi: 10.2196/43311] [Medline: 37335593]

9. Motz P, Gray M, Sawyer T, Kett J, Danforth D, Maicher K, et al. Virtual Antenatal Encounter and Standardized Simulation Assessment (VANESSA): pilot study. JMIR Serious Games. May 11, 2018;6(2):e8. [FREE Full text] [doi: 10.2196/games.9611] [Medline: 29752249]

10. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. Papers With Code. 2018. URL: https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf [accessed 2023-12-18]

11. Wang Y, Yao Q, Kwok J, Ni L. Generalizing from a few examples. ACM Comput Surv. Jun 12, 2020;53(3):1-34. [doi: 10.1145/3386252]

12. Cho K, Van MB, Bahdanau D, Bengio Y. On the properties of neural machine translationncoder-decoder approaches. Presented at: SSST-8, Eighth Workshop on Syntz, Semantics and Structure in Statistical Translation; Doha, Qatar; 25 Oct, 2014;103-111; Doha, Qatar. [doi: 10.3115/v1/w14-4012]

13. Dai A, Le Q. Semi-supervised sequence learning. ArXiv. Preprint posted online on Nov 4, 2015. [FREE Full text]

14. Rothwell J. Assessing the economic gains of eradicating illiteracy nationally and regionally in the United States. Barbara Bush Foundation for Family Literacy. 2020. URL: https://www.barbarabush.org/wp-content/uploads/2020/09/BBFoundation_GainsFromEradicatingIlliteracy_9_8.pdf [accessed 2023-12-18]

15. The assessment frameworks for cycle 2 of the programme for the international assessment of adult competencies; 2021. The Organisation for Economic Cooperation and Development. URL: https://www.oecd.org/skills/piaac/publications/PIAAC-Frameworks-Cycle2-en.pdf [accessed 2023-12-18]

16. Dockrell J. Reading and language impairments in conditions of poverty. In: Bishop, Dorothy V. M. & Leonard, Laurence B. (eds), Speech and language impairments in children: causes, characteristics, intervention and outcome. Hove, UK. Psychology Press, 2000. Pp. xiii+305. London, UK. Psychology Press; Jul 22, 2002;701-711.

17. Doston VM, Kitner-Triolo MH, Evans MK, Zonderman AB. Effects of race and socioeconomic status on the relative influence of education and literacy on cognitive functioning. J Int Neuropsychol Soc. Jul 01, 2009;15(4):580-589. [doi: 10.1017/s1355617709090821]

18. Advancing effective communication, cultural competence, and patient- and family-centered care: a roadmap for hospitals. The Joint Commission. 2010. URL: https://www.jointcommission.org/-/media/tjc/documents/resources/patient-safety-topics/health-equity/aroadmapforhospitalsfinalversion727pdf.pdf?db=web&hash=AC3AC4BED1D973713C2CA6B2E5ACD01B [accessed 2023-07-17]

19. McHugh M. Interrater reliability: the kappa statistic. Biochem Med. 2012:276-282. [doi: 10.11613/bm.2012.031]

20. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health. Mar 2023;5(3):e107-e108. [FREE Full text] [doi: 10.1016/S2589-7500(23)00021-3] [Medline: 36754724]

21. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med. Dec 26, 2022;5(1):194. [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]

22. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. Feb 9, 2023;2(2):e0000198. [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]

23. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst. Mar 04, 2023;47(1):33. [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]

24. Frey-Vogel A, Ching K, Dzara K, Mallory L. The acceptability of avatar patients for teaching and assessing pediatric residents in communicating medical ambiguity. J Grad Med Educ. 2022;14(6):696-703. [doi: 10.4300/jgme-d-22-00088.1]

25. Oliven A, Nave R, Baruch A. Long experience with a web-based, interactive, conversational virtual patient case simulation for medical students' evaluation: comparison with oral examination. Med Educ Online. Dec 28, 2021;26(1):1946896. [FREE Full text] [doi: 10.1080/10872981.2021.1946896] [Medline: 34180780]

26. Lipton LB, Vuong L, Nichols AA. Virtual simulated patient encounters: bridging the gap in maternal mental health training for prelicensure nursing students. J Prof Nurs. Sep 2023;48:22-24. [doi: 10.1016/j.profnurs.2023.05.007] [Medline: 37775236]

27.   McLemore MR, Altman MR, Cooper N, Williams S, Rand L, Franck L. Health care experiences of pregnant, birthing and postnatal women of color at risk for preterm birth. Soc Sci Med. Mar 2018;201:127-135. [doi: 10.1016/j.socscimed.2018.02.013] [Medline: 29494846]

28.   Williamson LD, Smith MA, Bigman CA. Does discrimination breed mistrust? examining the role of mediated and non-mediated discrimination experiences in medical mistrust. J Health Commun. Sep 27, 2019;24(10):791-799. [doi: 10.1080/10810730.2019.1669742] [Medline: 31559916]

29.   Tough S, Newburn-Cook C, Faber A, White D, Fraser-Lee N, Frick C. The relationship between self-reported emotional health, demographics, and perceived satisfaction with prenatal care. Int J Health Care Qual Assur Inc Leadersh Health Serv. 2004;17(1):26-38. [doi: 10.1108/09526860410515918] [Medline: 15046471]

30.   Leiferman J, Sinatra E, Huberty J. Pregnant women's perceptions of patient-provider communication for health behavior change during pregnancy. OJOG. 2014;04(11):672-684. [doi: 10.4236/ojog.2014.411094]

31.   Bohnhorst B, Ahl T, Peter C, Pirr S. Parents' prenatal, onward, and postdischarge experiences in case of extreme prematurity: when to set the course for a trusting relationship between parents and medical staff. Am J Perinatol. Nov 22, 2015;32(13):1191-1197. [doi: 10.1055/s-0035-1551672] [Medline: 26007309]

32.   Geurtzen R, van Heijst A, Draaisma J, Ouwerkerk L, Scheepers H, Hogeveen M, et al. Prenatal counseling in extreme prematurity - Insight into preferences from experienced parents. Patient Educ Couns. Aug 2019;102(8):1541-1549. [doi: 10.1016/j.pec.2019.03.016] [Medline: 30948203]

33.   de Boer A, de Vries M, Berken D, van Dam H, Verweij EJ, Hogeveen M, et al. A scoping review of parental values during prenatal decisions about treatment options after extremely premature birth. Acta Paediatr. May 10, 2023;112(5):911-918. [doi: 10.1111/apa.16690] [Medline: 36710530]

34.   De Proost L, Geurtzen R, Ismaili M'hamdi H, Reiss I, Steegers E, Joanne Verweij EJ. Prenatal counseling for extreme prematurity at the limit of viability: a scoping review. Patient Educ Couns. Jul 2022;105(7):1743-1760. [FREE Full text] [doi: 10.1016/j.pec.2021.10.033] [Medline: 34872804]

35.   Kim BH, Feltman DM, Schneider S, Herron C, Montes A, Anani UE, et al. What information do clinicians deem important for counseling parents facing extremely early deliveries?: results from an online survey. Am J Perinatol. Apr 07, 2023;40(6):657-665. [doi: 10.1055/s-0041-1730430] [Medline: 34100274]

36.   Kim BH, Krick J, Schneider S, Montes A, Anani UE, Murray PD, et al. How do clinicians view the process of shared decision-making with parents facing extremely early deliveries? results from an online survey. Am J Perinatol. Jan 11, 2022 [FREE Full text] [doi: 10.1055/s-0041-1742186] [Medline: 35016247]

37.   Barker C, Dunn S, Moore G, Reszel J, Lemyre B, Daboval T. Shared decision making during antenatal counselling for anticipated extremely preterm birth. Paediatr Child Health. Jul 2019;24(4):240-249. [FREE Full text] [doi: 10.1093/pch/pxy158] [Medline: 31239813]

38.   Kharrat A, Moore GP, Beckett S, Nicholls SG, Sampson M, Daboval T. Antenatal consultations at extreme prematurity: a systematic review of parent communication needs. J Pediatr. May 2018;196:109-115.e7. [doi: 10.1016/j.jpeds.2017.10.067] [Medline: 29223461]

39.   Staub K, Baardsnes J, Hébert N, Hébert M, Newell S, Pearce R. Our child is not just a gestational age. A first-hand account of what parents want and need to know before premature birth. Acta Paediatr. Oct 18, 2014;103(10):1035-1038. [doi: 10.1111/apa.12716] [Medline: 24920539]

40.   Suta P, Lan X, Wu B, Mongkolnam P, Chan JH. An overview of machine learning in chatbots. IJMERR. 2020:502-510. [doi: 10.18178/ijmerr.9.4.502-510]

41.   Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. Jun 19, 2020;22(6):e15154. [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]

42.   Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak. Jul 22, 2020;20(1):170. [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]

43.   GPT-4 Technical Report. OpenAI. 2023. URL: https://cdn.openai.com/papers/gpt-4.pdf [accessed 2023-12-18]

Abbreviations

AI: artificial intelligence
VANESSA: Virtual Antenatal Encounter and Standardized Simulation Assessment
VSP: virtual standardized patient

XSL•FO

RenderX