Generative AI versus Faculty-Facilitated Scenario-Based Simulation Design by Medical Students

Chad Taniguchi[1], Keely Myers[1], Erin Jyo[1], Benjamin Berg MD[1], Jannet Lee-Jayaram MD[1]

[1]John A. Burns School of Medicine, University of Hawaii, Honolulu, HI

Background

Interest in generative AI and its application to various disciplines, including medical education, has been exponentially growing. ChatGPT was released in 2022 and has garnered much attention due to its free public access. However, research exploring its use for design scenario-based simulations (SBS) is limited.

Rodgers' (2023) described ChatGPT's potential as a useful tool for simulationists to streamline instructional design. Yet, they underscore the crucial role of human intervention in addressing shortcomings related to errors, complexity, and formatting. A background in simulation educational design may be a prerequisite. Often when SBS design is undertaken by novice simulationists such as medical students, the process can be overwhelming and instructional design may be incomplete, especially without the guidance of experienced simulationists.

Supplementing scenario design with AI may allow medical students to explore topics that students are interested in but are not taught in the regular curriculum. The applicability of ChatGPT in aiding non-simulationists with SBS design in healthcare education has not been explored.

Objectives

To describe the instructional design process and outcomes of SBS created by medical students using ChatGPT and compare them to SBS created by medical students with expert simulation faculty guidance.

Methods

Five existing SBSs designed by medical student interest groups (SIG) with simulation faculty guidance were retrieved from simulation center archives. Scenario goals and patient synopsis were extracted by one medical student. Two medical students were blinded to all details except for the extracted scenario goals and synopsis, which was used to generate a scenario with ChatGPT v.3.5. A blank scenario design template (QR code) provided by the simulation center guided generation of the AI-generated scenario. The ChatGPT conversation tool facilitated iterative refinement of missing elements, errors, or desired modifications. Five AI-generated scenarios were produced in one session, with elapsed time recorded. The number of design elements and objectives from the scenarios were quantified. Statistical analysis on quantitative data was performed using two-tailed T-tests.

Results

On average (n=5), the ChatGPT scenarios design time was 36.6 minutes and 5.8±0.6 prompts were needed to produce the final scenario. In contrast, SBSs designed by SIGs with faculty input were created over months, and required multiple faculty-student meetings. ChatGPT produced an average of 4.0±0.3 learning objectives, compared to 3.2±0.7 when developed with faculty. ChatGPT's objectives were frequently repetitions of the initial input goals. ChatGPT fulfilled an average of 11.8±0.4 out of 18 template elements, compared to 12.8±1.5 in faculty-guided scenarios. There were no statistical differences in the number of fulfilled elements (p=0.53) or objectives (p=0.35).

Figure 1. Comparison of AI Generated and Faculty Facilitated Scenarios by Medical Students

	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Average[a]	
	AI Generated	Faculty Facilitated	AI Generated	Faculty Facilitated	AI Generated	Faculty Facilitated	AI Generated	Faculty Facilitated	AI Generated	Faculty Facilitated	AI Generated	Faculty Facilitated
Prompts[b]	5	N/A	7	N/A	6	N/A	4	N/A	7	N/A	5.8 ± 0.6	N/A
Learning Objectives[c]	3	4	4	1	4	4	5	2	4	5	4.0 ± 0.3	3.2 ± 0.7
Components Present[d]	12	15	13	11	12	14	11	8	11	16	11.8 ± 0.4	12.8 ± 1.5
Components Absent[d]	6	3	5	7	6	4	7	10	7	2	6.2 ± 0.4	5.2 ± 1.5
Time (min)[e]	54	nd	43	nd	31	nd	30	nd	25	nd	36.6 ± 5.3	nd

[a]Values are means ± SEM. Differences between the AI generated and faculty facilitated scenarios were not statistically significant ($p < 0.05$) using a two tailed student t-test.
[b]The number of inputs/prompts given to the AI software to generate a completed SBS
[c]Number of learning objectives generated by the AI software or number of learning objectives listed in the simulation design form by students
[d]Components are defined using a simulation center's simulation design form
[e]Time to complete the scenario was measured as the time of the first input to the AI sofware until a satisfactory result was achieved as determined by students. Time was not determined (nd) for the faculty facilitated scenarios.

	FACULTY-FACILITATED	AI-GENERATED
Synopsis	17 year old female brought in by mom for confusion and fever	You are the facilitator for a simulation exercise involving a team of 6 fourth-year medical students. The scenario will take place in an emergency room (ER) setting, using a high-fidelity manikin. The focus is on identifying and managing a medical emergency involving altered mental status in a realistic clinical environment.
Equipment / Supplies	Mannikin with wig	- High-fidelity manikin capable of simulating vital signs and responses. - Standard medical equipment (e.g., stethoscope, blood pressure cuff, EKG machine). - IV access supplies. - Medications (simulated).
Scenario Duration	12 minutes	9 Minutes
Debrief Location & Duration	Following all SimTiki scenarios, we will have a more extensive debrief with all participants in JABSOM MEB for approximately 30 minutes.	N/A
Scenario Goal	1. To identify pathological signs of TCA overdose based on history and physical exam. 2. To apply understanding of arrhythmia and autonomic symptoms to a new clinical scenario. 3. To learn how to manage and treat TCA overdose to prevent complications such as Vtach or Vfib. 4. To see management of intentional overdose.	1. To identify pathological signs of tricyclic antidepressant overdose based on history and physical examination 2. To apply an understanding of arrhythmia and autonomic symptoms to a new clinical scenario 3. To learn how to manage and treat tricyclic antidepressant overdose to prevent complications such as ventricular tachycardia and ventricular fibrillation 4. To see management of intentional overdose
Learning Objectives	1. Form a DDX for hypotension, AMS, and fever and determine what steps to do next to refine diagnosis. 2. Identify signs of anti-cholinergic toxicity on the physical exam 3. Apply knowledge of EKG readings to this case. 4. Learn to manage and treat TCA overdose complications - in this case, arrhythmia 5. Learn about the medical importance of intentional overdose (in the debrief as a big group)	1. To identify pathological signs of a potential overdose based on history and physical examination. 2. To apply an understanding of arrhythmia and autonomic symptoms to a new clinical scenario. 3. To learn how to manage and treat overdoses of unknown substances to prevent complications such as ventricular tachycardia and ventricular fibrillation. 4. To see the management of intentional overdose.

Table 1: Example of scenario comparison between faculty-facilitated and AI-generated scenario. Categories are based off simulation center scenario design template.

Conclusions/Discussion

The most notable difference between ChatGPT and faculty guided scenarios is substantial reduction in creation time. AI-assisted scenarios were created in minutes, while faculty guided scenarios took months to complete (based on simulation center staff estimates). Time efficiency could allow students to jumpstart the design process and time saved would support further simulation refinement under faculty guidance. However, the quality and accuracy of the ChatGPT scenarios have yet to be examined by simulation experts. Challenges experienced while using ChatGPT include the omission of requested scenario components, inadvertent removal of desired elements during the iterative process, and inconsistencies in formatting between scenarios.

Limitations of this study include utilizing the free version of ChatGPT, which does not access the internet and has no knowledge past 2022. When generating SBSs from AI, users were not blinded to details of the faculty-facilitated SBSs. Comparison was limited to the number of available SBSs.

Future Directions

Although ChatGPT is able to generate a SBS, further research evaluating the quality of the SBS is necessary. We plan to have simulation experts score the existing SBSs designed by medical SIGs and their corresponding SBSs created by ChatGPT using Hernandez's (2020) validated Simulation Scenario Evaluation Tool (SSET). AI-generated SBS will be compared to student-created SBS based on integral components for optimized student learning such as learning objectives, clinical context overview, critical actions, patient states, scenario materials/resources, and debriefing plan.

References

- Rodgers DL, Needler M, Robinson A, et al. Artificial Intelligence and the Simulationists. *Sim Healthcare*. Published online September 21, 2023. doi:10.1097/SIH.0000000000000747
- Hernandez J, Frallicciardi A, Nadir N-A, et al. BMJ Simul Technol Enhanc Learn 2020;6:344–350.doi: 10.1136/bmjstel-2019-000521

Acknowledgments

We would like to thank Kris Hara for providing us with the scenario design template and other resources associated with SimTiki.