Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240705

Feasibility and Educational Value of Clinical Cases Generated Using Large Language Models

Anna BERBENYUK[a,1] , Leigh POWELL[a] and Nabil ZARY[a]
[a] *Institute of Learning, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai Health, Dubai, UAE*
ORCiD ID: Anna Berbenyuk https://orcid.org/0000-0002-6579-3772, Leigh Powell https://orcid.org/0000-0001-6714-7371, Nabil Zary https://orcid.org/0000-0001-8999-6999

Abstract. In medical education, case-based learning (CBL) is a fundamental method for training healthcare professionals across different levels of expertise. This approach hinges on using authentic or fabricated clinical cases to bridge the gap between theoretical knowledge and its practical application. It fosters active engagement and knowledge application among learners in healthcare domains. While creating effective cases demands substantial clinical understanding and time investment, the integration of Generative Artificial Intelligence (AI) presents a promising solution to this challenge. AI can efficiently analyze extensive medical data to generate diverse and realistic clinical scenarios, continuously updating case content based on emerging medical literature and guidelines. This study explores AI-generated cases' feasibility and educational value in continuing medical education, focusing on COVID-19 scenarios tailored for the MENA region. Results indicate the potential of AI-generated cases to foster engagement and critical thinking among learners, suggesting their suitability for different levels of education. This study highlights the advantages of integrating AI into CBL and emphasizes the need for future efforts to tackle obstacles and facilitate its successful adoption.

Keywords. Artificial Intelligence (AI), Medical Education, Case-Based Learning (CBL), Clinical cases

1. Introduction

In the realm of continuing medical education, case-based learning (CBL) stands as a cornerstone method for instructing healthcare professionals at varying levels of expertise. The essence of CBL lies in its utilization of authentic or generated clinical cases to bridge the gap between theoretical knowledge and practical application, thereby fostering active engagement and knowledge application among learners in healthcare domains [1,2]. By transforming learners from passive recipients of information to active participants who critically analyze cases and apply insights to practical clinical contexts, CBL enhances clinical performance, attitudes, and collaborative skills [3].

A standard case structure used in CBL includes a short description of a patient and questions to help gather important information or ideas. Effective cases, as outlined by

[1] Corresponding Author: Anna Berbenyuk; E-mail: anna.berbenyuk@dubaihealth.ae.

the National Centre for Case Study Teaching in Science, exhibit authenticity, relatable scenarios, narrative engagement, alignment with learning objectives, educational value, curiosity stimulation, empathetic connections with characters, incorporation of patient quotes for authenticity, cultivation of decision-making skills, and broad relevance [4].

Creating cases requires extensive clinical knowledge, time, and resources. AI offers a solution by efficiently generating diverse and realistic clinical scenarios by analyzing vast medical data. Additionally, GenAI platforms can update case content based on new medical literature and guidelines, ensuring relevance [5].

AI is predominantly integrated into undergraduate-level teaching through tools such as conversational agents and supplementary diagnostic tools [6]. By leveraging AI's capabilities, educators have the potential to generate clinical scenarios that are precisely tailored to specific contexts and learning objectives. This can fundamentally transform the integration of CBL throughout all levels of medical education.

This research explores the feasibility of using AI-generated cases in continuing medical education. This study investigated two research questions: i) Can AI create clinical cases of adequate complexity and quality suitable for continuing medical education (CME)? ii) Do clinical cases constructed by AI possess substantial educational value for health professionals?

2. Methods

Two clinical cases on COVID-19 were created using OpenAI ChatGPT 4.0. Prompting strategy involved the application of a persona pattern (i.e., “senior clinical educator and clinician well-versed in high risk and long covid”) and a predefined case framework including the following sections: patient presentation, background, clinical examination, and investigations. Cases were enriched with details specific to the MENA region to tailor them to the audience (Table 1). The generated cases underwent review by subject matter experts without any modifications. Two cases were presented to working clinicians and medical educators for solving during a focused 90-minute workshop. Participants were allocated 20 minutes for discussing each case, following which they evaluated the cases using a predefined framework adapted for the CME level by the research team, as outlined in the framework proposed by Scott Kohlert et al. [7]. Assessment involved rating a set of statements on a scale ranging from 1 to 5 for each case. The statements included: "Case scenario is close to real-life clinical practice," "Case scenario prompts reflection on clinical practice," "Case presentation is challenging current practice," "Case incorporates knowledge that is grounded in evidence-based practices," "Case fosters active and engaging discussions," "Case scenario is of enough complexity for a healthcare professional," and "Case scenario is coherent and concise."

Furthermore, participants were asked to indicate which group the cases were most suitable for: medical students, medical residents, health professionals, or all of the above. Demographic data, including years of experience, specialties, involvement in teaching and CBL, as well as the use of AI for teaching, was also acquired. Survey data was collected via Microsoft Forms and subjected to statistical analysis employing descriptive and comparative techniques using statistical software.

Table 1. Prompting strategy for case generation

Prompting process
Setting the Role for ChatGPT Start by Assigning a Specific Role: "Act as a senior clinical educator well-versed in high-risk and long COVID-19."
Specifying the Output Clearly Define Desired Content: "You will help me create clinical cases for educational content relevant for working clinicians."
Outlining the Case Structure Establish Framework and Boundaries: "Let's create a versatile clinical case framework...", "Include only … sections", "Create a case in a form of a narrative".
Incorporating Specific Requirements Add Details for Customization: "Focus on creating different clinical scenarios around high-risk and long COVID for educating working clinicians. Use Arabic names for the patients and make cases as if all the patients come from the MENA region."
Requesting Supplementary Materials Ask for Additional Educational Tools: "Assume you are going to facilitate the discussions around these cases among working clinicians during a short workshop. Suggest a facilitator guide for that."

3. Results

Forty physicians and clinical educators participated in the workshop and were randomly divided into six small groups. The groups consisted of individuals with diverse levels of experience and specializations. Most participants had between 20 and 30 years of professional experience, with the majority (N=30) reporting having used case-based learning in their teaching. Half of the participants had not actively participated in case development. None of the participants reported using artificial intelligence (AI) technology as a teaching tool. We analyzed data only for those participants who evaluated both cases, resulting in a total of 36 individuals. On average, Case 1 received a score of 4.13 across all 7 domains, with mean (SD) scores for each respective domain ranging from 3.69 to 4.36. The lowest individual mean score (SD) of 3.69 (.980) was in the domain "Case presentation is challenging current practice.", while the highest (4.36 (.639)) was reported for "Case scenario prompts reflection on clinical practice". Respondents indicated that this case was most suitable for medical residents (42% of respondents). Despite utilizing the same prompting strategy, Case 2 received a lower average score (3.9). Respondents indicated that this case was most suitable for health professionals (41.7%). There was a statistically significant difference between the two cases in the following domains: "Case scenario prompts reflection on clinical practice," "Case incorporates knowledge that is grounded in evidence-based practices," "Case scenario is of enough complexity for a healthcare professional," and "Case scenario is coherent and concise" as determined by Wilcoxon's test (Table 2).

Table 2. Mean scores per statement (domain) for case 1 and case 2.

Statements	Case 1 Mean (SD)	Case 2 Mean (SD)	p- value
Case scenario is close to real-life clinical practice.	4.25 (.906)	4.17 (.737)	.567
Case scenario prompts reflection on clinical practice.	4.36 (.639)	3.97 (.845)	.012

Case presentation is challenging current practice.	3.69 (.980)	3.78 (1.072)	.501
Case incorporates knowledge that is grounded in evidence-based practices.	4.03 (.696)	3.75 (.967)	.018
Case fosters active and engaging discussions.	4.28 (.701)	4.11 (.854)	.180
Case scenario is of enough complexity for a healthcare professional.	4.17 (.811)	3.75 (.996)	.004
Case scenario is coherent and concise.	4.17 (.737)	3.81 (.856)	.009

4. Discussion

The results of this study highlight the potential role of AI in CBL. Notably, despite the prevalence of educators who employ CBL methods, none reported using AI technology for this purpose. This raises questions about the readiness of health professionals to use AI technologies and the level of trust towards them.

Using the same prompting strategy, Case 1 fostered deeper engagement and problem-solving skills than the more general Case 2, scoring well across all domains except "Case presentation challenging current practice." Respondents noted this difference, indicating suitability for learners at different levels. This emphasizes the need for an iterative AI generation process with clear objectives and highlights that AI cannot replace medical educators in case creation but rather serve as a supportive tool.

Scoring above 4 in the domain of " Case scenario is close to real-life clinical practice" in both cases underscores that AI is producing high-quality cases according to established standards[4]. High scores on domains related to prompting reflection and fostering engagement indicate that AI-generated cases can exemplify sound educational principles[8], highlighting that AI-generated cases have the potential to foster high levels of engagement and critical thinking among learners. However, the absence of AI in case creation highlights the necessity to delve into barriers hindering its adoption in medical education. These barriers may encompass perceived complexity, lack of awareness or training, and concerns regarding the authenticity and relevance of AI-generated cases.

Despite the promising potential of AI in medical education, it is essential to acknowledge the presence of selection bias in the study, particularly with the inclusion of "AI" in the title, which may have influenced participant responses and perceptions. Future research should address this bias and explore current barriers to integrating AI in case-based learning. Qualitative interviews or observational studies could provide valuable insights into educators' perspectives on AI, which could help to inform the development of tailored interventions to facilitate further adoption of AI for case-based learning.

Our study is limited by a small workshop cohort, necessitating more participants for comprehensive insights into AI-generated clinical cases. The predominance of educators in our sample may have skewed enthusiasm for AI tools, so a more diverse participant pool is crucial for representative findings. Moreover, the workshop's focus on AI likely

attracted individuals who were biased toward exploring such tools. Additionally, our evaluation framework and prompted scenarios could introduce bias, and reliance on self-reported measures may add interpretation biases. Finally, our intervention's accessibility to English speakers at a medical education symposium restricts the generalizability of our findings.

5. Conclusions

In conclusion, the findings highlight AI-generated cases' potential for fostering active learning and clinical reasoning. Further research is necessary to assess the feasibility of implementing AI in CBL on a larger scale. Exploring this area can advance the integration of innovative technologies to enhance medical education.

Acknowledgements

The MBRU institutional review board approved the study in February 2024 (submission number: MBRU IRB-2023-241). This work was funded by a Pfizer Independent Medical Education Grant (agreement number: 67504787).

References

[1] Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. Med Teach 2012;34:e421–44. https://doi.org/10.3109/0142159X.2012.680939.
[2] McLean SF. Case-Based Learning and its Application in Medical and Health-Care Fields: A Review of Worldwide Literature. J Med Educ Curric Dev 2016;3:JMECD.S20377. https://doi.org/10.4137/JMECD.S20377.
[3] James M, Baptista AMT, Barnabas D, Sadza A, Smith S, Usmani O, et al. Collaborative case-based learning with programmatic team-based assessment: a novel methodology for developing advanced skills in early-years medical students. BMC Med Educ 2022;22:81. https://doi.org/10.1186/s12909-022-03111-5.
[4] Herreid CF. What makes a good case. J Coll Sci Teach 1997;27.
[5] Rampton V, Mittelman M, Goldhahn J. Implications of artificial intelligence for medical education. Lancet Digit Health 2020;2:e111–2. https://doi.org/10.1016/S2589-7500(20)30023-6.
[6] MIR MM, MIR GM, RAINA NT, MIR SM, MIR SM, MISKEEN E, et al. Application of Artificial Intelligence in Medical Education: Current Scenario and Future Perspectives. J Adv Med Educ Prof 2023;11:133–40. https://doi.org/10.30476/JAMP.2023.98655.1803.
[7] Kohlert S, Brulotte M, Bell R, Roy J, Jalali A. A Quality Assurance Template for Revision of Case Based Learning Modules. Educ Med J 2018;10:47–56. https://doi.org/10.21315/eimj2018.10.3.5.
[8] Chickering AW, Gamson ZF. Seven Principles for Good Practice in Undergraduate Education. 1987.