

A Critical Analysis of “The Revival of Essay-Type Questions in Medical Education: Harnessing Artificial Intelligence and Machine Learning”

Sir,

Artificial intelligence (AI) has emerged as a crucial tool in recent times. It is being continuously evaluated and researched to incorporate into our lives, considering its evolving potential.¹ ChatGPT has gained much attention since it was introduced in late 2022. ChatGPT, a language model of AI, has been proposed to play important roles in medical education, including personalised assessment, quick access to content, and generating case scenarios.² We commend the authors for shedding light on the potential of ChatGPT in assessing essay-type questions in medical education.³ While the topic of this manuscript is intriguing, several aspects require discussion.

The objective of this article is to compare the assessment of human-written and machine-written essays. A correctly stated objective would be to compare the assessment of human-written essays by humans and ChatGPT, as there are no machine-written essays. The decision to use ChatGPT 3.5 (free version) instead of the more advanced ChatGPT 4 (costs only \$20 ≈ 5000 PKR) potentially limits the harnessing of AI for assessment, as ChatGPT 4 offers advanced data analysis and research capabilities that could have provided more comprehensive evaluation. The research methodology also lacks clarity about the criteria used to judge AI responses. It is uncertain what criteria were used to label AI responses “fascinating” or “thought-provoking”, as some other studies have reported caution while using ChatGPT assessment for complex medical questions because of unreliable results.⁴ An independent expert assessment of the explanations and feedback provided by the ChatGPT should have been conducted to determine their credibility. Moreover, the inclusion of plagiarism detection as one of the four prompts seems irrelevant and unnecessary for such essays as these were not research articles but medical scenarios.

Lastly, the authors agree that ChatGPT, being an AI language model, lacks critical thinking abilities, and its assessment depends on the existing knowledge. These limitations raise ethical concerns, especially in the medical field, where innovative solutions require human judgement.⁵

The study has provided valuable insights about the application of ChatGPT feedback responses for essay-type questions but there could have been improvement in the methodology in certain aspects to increase the reliability of results as stated above. We propose that further research regarding the use of AI in medical

education should be carried out in collaboration with AI experts to harness the maximum benefit of these AI tools.

COMPETING INTEREST:

The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:

UM, MAR, MAR: Literature review and manuscript writing.

REFERENCES

1. Aubignat M, Diab E. Artificial intelligence and ChatGPT between worst enemy and best friend: The two faces of a revolution and its impact on science and medical schools. *Rev Neurol (Paris)* 2023; **179(6)**:520-2. doi: 10.1016/j.neurol.2023.03.004.
2. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci* 2023; **39(2)**:605-7. doi: 10.12669/pjms.39.2.7653.
3. Shamim MS, Zaidi SJA, Rehman A. The revival of essay-type questions in medical education: Harnessing artificial intelligence and machine learning. *J Coll Phys Surg Pak* 2024; **34(5)**:595-9. doi: 10.29271/jcpsp.2024.05.595.
4. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? *Med Educ Online* 2023; **28(1)**:2181052. doi: 10.1080/10872981.2023.2181052.
5. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J* 2023; **3(1)**:e103. doi: 10.52225/narra.v3i1.103.

Umar Maqbool¹, Muhammad Ahmad Raza² and Muhammad Adnan Ramzan²

¹Department of Emergency Medicine, King Edward Medical University, Lahore, Pakistan

²Department of Internal Medicine, Quaid-e-Azam Medical College, Bahawalpur, Pakistan

Correspondence to: Dr. Umar Maqbool, Department of Emergency Medicine, King Edward Medical University, Lahore, Pakistan

E-mail: umarmaqbool597@gmail.com

Received: May 11, 2024; Revised: July 08, 2024;

Accepted: July 17, 2024

DOI: <https://doi.org/10.29271/jcpsp.2024.10.1522>

AUTHOR'S REPLY:

Sir,

Thank you for your thoughtful and detailed response to our article. We appreciate your engagement and critical analysis, which have contributed significantly to the ongoing discourse surrounding the integration of AI in medical education.

You have raised several valid points that merit further discussion. Your suggestion to frame more accurately the objective of the study by comparing the assessment of human-written essays by both humans and ChatGPT is well taken. Clarifying this distinction can indeed improve the precision of our research objectives and findings.

The version of ChatGPT used does not impact qualitative data analysis capabilities, as the core functionalities remain consistent across versions. The primary difference is that, with ChatGPT 4, users have the option to upload essays for analysis, which is not available in ChatGPT 3.5. However, for qualitative data analysis, both versions offer similar capabilities and effectiveness. The author's team had access to paid Chatgpt, and all prompt engineering and backhand Python codes were entered through paid ChatGPT.

Your concerns about the clarity of our research methodology and the criteria used to evaluate the AI responses are important. Prompt engineering was used to elicit desired responses from ChatgPT. Rubrics were developed and independent checkers evaluated the essays. We also verified the AI-generated assessment using Turnitin's AI detection tool, which showed a 0% match.

Your dismissal of the relevance of plagiarism detection in our study shows a misunderstanding of its broader applicability. Plagiarism detection is not limited to research articles but is equally important in ensuring the originality and integrity of all educational content, including medical scenarios. Medical

scenarios require clinical reasoning skills, and merely copying and pasting information is unacceptable. Therefore, all summative essays were checked routinely for plagiarism.

Finally, your concerns about AI's limitations in critical thinking and its ethical implications are well-trodden arguments. However, our study does not claim that AI can or should replace human judgement. Instead, it highlights AI as a complementary tool that, when used judiciously, can significantly enhance educational outcomes.

Your concerns about collaborating with AI experts are well-taken. The author has a dedicated digital learning team at the Digital Learning Centre of the Dow University of Health Sciences, comprising front-end and back-end developers, web and app developers, content managers, a graphic designer, a DevOps manager, and an LMS manager. All AI research is conducted in collaboration with the digital learning team.

Thank you for your valuable feedback.

Syed Jaffar Abbas Zaidi

.....
*Department of Oral Biology Digital Learning Centre,
Dow University of Health Sciences, Karachi, Pakistan*

.....
*Correspondence to: Dr. Syed Jaffar Abbas Zaidi,
Department of Oral Biology Digital Learning Centre,
Dow University of Health Sciences, Karachi, Pakistan
E-mail: jaffar.zaidi@duhs.edu.pk*

