

Performance Evaluation of ChatGPT in Forensic Medicine (Including Medical Jurisprudence & Toxicology) of Third Professional MBBS: Part 1

Joshua Ashok¹, Hariprasad V², Ashok Kumar Mohan³, Krishna Mohan Surapaneni^{4,5}

¹Professor and Head, ²Associate Professor, ³Assistant Professor, Department of Forensic Medicine, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India, ⁴Professor, Department of Biochemistry, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India, ⁵Head, Departments of Medical Education, Molecular Virology, Clinical Skills & Simulation, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India.

How to cite this article: Joshua Ashok, Hariprasad V, Ashok Kumar Mohan et. al. Performance Evaluation of ChatGPT in Forensic Medicine (Including Medical Jurisprudence & Toxicology) of Third Professional MBBS: Part 1. Indian Journal of Forensic Medicine and Toxicology 2023;17(4).

Abstract

Background: ChatGPT, the state-of-the-art innovation in the field of Artificial Intelligence (AI) is gaining massive popularity with its notable ability to generate human-like responses within the shortest period. The utility of ChatGPT is being assessed possibly in all fields now with medical education predicted to be one of the key areas to be highly benefitted by ChatGPT.

Objective: We aim to evaluate the performance of ChatGPT in Forensic Medicine of the third professional MBBS part 1.

Methods: The performance of ChatGPT was assessed using the final university examination question papers. The questions were used to converse with ChatGPT and each of the responses was recorded. The answers generated by ChatGPT were later evaluated by the expert faculties in the Department of Forensic Medicine. **Results:** All our questions regarding the subject were satisfactorily addressed by ChatGPT. The answers were concise without any pictorial representation. Before generating content related to issues like "sodomy", potential content violation messages were sent.

Conclusion: ChatGPT has performed moderately in the evaluation. Although there were invalid and incorrect responses, with additional inputs and training, these chatbot systems can be improved to assist medical educators in various subjects and can emerge as one of the innovative teaching-learning strategies in medical education. We can successfully use these technologies to improve medical education and, ultimately, patient care by leveraging AI's benefits while addressing its drawbacks

Key Words: "Artificial Intelligence"; "ChatGPT"; "Forensic medicine"; "Medical education"

Introduction

The world is astounded by the emergence of a

new technological buzz, ChatGPT that is creating history in many domains including medical education

Corresponding Author: Krishna Mohan Surapaneni, PhD., MHPE., FRSC, Departments of Biochemistry, Medical Education, Molecular Virology, Clinical Skills & Simulation, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India.

E-mail: krishnamohan.surapaneni@gmail.com

and assessment of knowledge¹. ChatGPT is rapidly gaining millions of users overtaking Google and Facebook ever since its public release on November 30, 2022². ChatGPT is a huge conversational Artificial Intelligence (AI) based language model that uses Generative Pre-Transformer 4 to generate enormous amounts of data based on deep learning algorithms³.

Recognizing its potential, ChatGPT is being tested in various domains, including medical education in which ChatGPT has become a game change after it had successfully cleared the United States Medical Licensing Exam (USMLE)⁴. Since then, numerous studies have been conducted to test its usability in medical education and scientists believe ChatGPT could be an asset in completely renovating medical education, research, and clinical management⁵. It has been employed in healthcare to give medical aid and information by responding to medical inquiries or providing differential diagnoses for common complaints and symptoms of patients⁶.

Medical education continues to advance as technology develops. But despite worries about plagiarism, misinformation or disinformation expressed by ChatGPT, there are still several methods to use this technology to promote academic standards⁷. It is important to take into account ChatGPT's problems and limits, including any moral dilemmas or negative consequences. Medical educators must stay up with the rapidly changing environment of technology and take into account the consequences of curriculum design, assessment techniques, and methods of teaching⁸.

Many studies are continuously being conducted to evaluate and appraise the knowledge of ChatGPT in various medical subjects. These assessments have diverse opinions, and more research is required

to prove the efficiency of ChatGPT in an evidence-based manner. But much research has not been done in forensic medicine. In this regard, we assessed the performance of ChatGPT in forensic medicine of Phase III Part 1 in the MBBS curriculum.

Materials and Methods:

Study Design and Setting

This observational study was conducted at the Panimalar Medical College Hospital & Research Institute, Chennai, India. We carried out the conversation using ChatGPT, an accessible interactive AI language model, recorded it, and then assessed the replies.

Ethical Consideration

Neither people nor animals are employed in this experiment. Given that it is only observational research that looks at the responses generated by an AI system, this study does not fulfil the requirements for receiving an ethical clearance.

Evaluation Material

We used the Forensic Medicine (Including Medical Jurisprudence & Toxicology) University Examination question paper which was conducted in the month of February 2023 according to the new Competency-Based Medical Education (CBME) curriculum. The paper is evaluated for a total of 40 marks and a total of 13 questions were used with one essay for 10 marks each and six notes carrying a maximum of 4 marks each. The remaining 6 short answers are allotted 1 mark each. A minimum of 50% (20 marks) is required to pass this evaluation. Table 1 shows the questions which were used to converse with ChatGPT.

Table 1: Evaluation Questions

S.No	Question	Type of question	Maximum marks (40)
1.	Define Poisoning. Discuss in detail about signs, symptoms and postmortem	Essay	10
2.	Whiplash injury	Short Notes	4
3.	Inquest	Short Notes	4
4.	Signs of recent delivery	Short Notes	4
5.	Sodomy	Short Notes	4

Continue.....

6.	Rigor mortis	Short Notes	4
7.	Impulse	Short Notes	4
8.	Kevorkian sign	Short Answer	1
9.	Run amok	Short Answer	1
10.	Locards principle	Short Answer	1
11.	Super fecundation	Short Answer	1
12.	Suis	Short Answer	1
13.	Hasse's rule	Short Answer	1

Data Collection

The questions and answers were put into ChatGPT precisely as they appeared on the exam paper. There was no instruction regarding word count, or length was provided. The answers were recorded and afterwards given to the expert team for review. We generated the responses only once.

Pattern of Evaluation

To provide a fair and unbiased result, each of the responses generated by ChatGPT was evaluated independently by two raters (Faculty from the Department of forensic medicine). A total of 40 points were awarded for the evaluation, with essays receiving a maximum score of 10, brief notes receiving a maximum score of 4, and short answers receiving a maximum score of 1. Readability, answer presentation, content validity, and visual or flowchart representation were taken into consideration while evaluating the material. The professors discussed how well ChatGPT had grasped the questions and what significant information it had missed. In light of this study, an evaluation of ChatGPT's forensic medical knowledge and interpretation skills was conducted.

Findings:

ChatGPT has generated brief and 'to the point answers' for all the questions. No pictorial representation was generated. The explanation provided for each question was satisfactory. The answers lacked specificity and technical terms which are required to deem the answers complete. Few answers deviated away from the aspects of forensic medicine. Further, the responses on the "Kevorkian

sign"; "Run amok"; "Super fecundation"; "Suis" and "Hasse rules" were totally invalid and out of context in explanation and description. All the responses generated by ChatGPT are included in the supplementary material.

Also, during the process of generating the answers, we encountered a possible content policy violation in ChatGPT. Upon conversing with ChatGPT for the generating short notes on "Sodomy", ChatGPT generated an alert message stating, "This content may violate our content policy. If you believe this to be in error, please submit your feedback - your input will aid our research in this area" as shown in Figure 1. This message indicates that the content which was requested to be generated may violate the platform's content policy. The platform has flagged this content for review and asked for feedback to help them improve their content policies emphasizing the necessity to keep in mind that content policies are in place to guarantee an online environment that is both secure and respectful for all users.

On a final note, ChatGPT has managed to "Pass" our evaluation however the generated explanations for all the questions were generalized and a few invalid responses were generated. Table 2 provides the evaluation report of the responses generated by ChatGPT with rather 1, rater 2 and mean score provided for all the answers. ChatGPT scored a total of 62.5% in the Rater 1 evaluation and 60% in the Rater 2 evaluation. By calculating the mean score, ChatGPT has secured 61.25% in the forensic medicine examination. Although it had cleared the Forensic Medicine exam, there is still room for improvements that can be made before AI-based tools like ChatGPT could be used for academic and research purposes.



Sodomy

This content may violate our [content policy](#). If you believe this to be in error, please [submit your feedback](#) — your input will aid our research in this area.



Sodomy is a term that historically referred to any sexual act that was considered "unnatural" or "sinful." In modern times, it generally refers to anal sex, but can also include other sexual acts that are deemed non-procreative or non-heterosexual.

It is important to note that the term "sodomy" is often used in a derogatory or stigmatizing way against individuals who engage in non-heterosexual sexual acts. This is inappropriate and harmful, as it perpetuates discrimination and reinforces harmful stereotypes.

It is important to remember that consensual sexual acts between adults, regardless of their gender or sexual orientation, should be respected and protected. Additionally, it is important to prioritize informed and enthusiastic consent in all sexual interactions.



Figure 1: Content policy Violation Alert on ChatGPT

Table 2: Evaluation Report

S.No	Question	Type of question	Maximum marks (40)	Rater 1	Rater 2	Mean score
1.	Define Poisoning. Discuss in detail about signs, symptoms and postmortem	Essay	10	7	6	6.5
2.	Whiplash injury	Short Notes	4	3	2.5	2.75
3.	Inquest	Short Notes	4	3	3	3
4.	Signs of recent delivery	Short Notes	4	3	3.5	3.25
5.	Sodomy	Short Notes	4	3	3	3
6.	Rigor mortis	Short Notes	4	3	2.5	2.75
7.	Impulse	Short Notes	4	2.5	3	2.75
8.	Kevorkian sign	Short Answer	1	0	0	0
9.	Run amok	Short Answer	1	0	0	0
10.	Locards principle	Short Answer	1	0.5	0.5	0.5
11.	Super fecundation	Short Answer	1	0	0	0
12.	Suis	Short Answer	1	0	0	0
13.	Hasse's rule	Short Answer	1	0	0	0
				Total MARKS: 25 (62.5%)	Total MARKS: 24 (60%)	Mean Total MARKS: 24.5 (61.25%)

Discussion

In our study, the third MBBS professional part 1 test for forensic medicine evaluated the performance of ChatGPT. The outcomes showed that ChatGPT was capable of passing the test but with moderately satisfactory performance. The study did draw attention to the fact that there is still space for development and to ChatGPT's potential effects on medical instruction. The assessment has identified gaps in ChatGPT's knowledge of particular forensic medicine principles and its capacity to give precise and thorough answers. This suggests that further modelling and fine-tuning of the model are required to improve its performance in this niche area.

A similar study analyzing the potential of ChatGPT in clinical toxicology using a simple case-based query revealed that ChatGPT was able to diagnose the case and provide acceptable explanations for the same⁹. This is consistent with the findings of our study where we have provided evidence that ChatGPT generates satisfactory responses. However, in medical practice, more than diagnosing the condition, proper history taking, clinical interpretation and effective management are key components. ChatGPT has to be proven in these domains also. Another study which compared the essay writing skills of second-year medical students in Croatia and ChatGPT in forensic medicine stated that since the control group outperformed the experimental group in most of the categories, no proof utilising ChatGPT as a writing aid enhances essay quality¹⁰.

Furthermore, insights on the potential use of ChatGPT in digital forensic investigations state that the adoption of large language models, such as ChatGPT, may undoubtedly strengthen and speed up digital forensic medical investigation procedures. But as with everything at the nexus of machine learning (AI) and computational forensics, it's crucial to validate the AI-generated content and ensure human supervision when it comes to its application rather than ever approaching an era wherein, we become excessively reliant on such technologies and forfeit our fundamental comprehension of the evidence and human abilities¹¹.

The fact that ChatGPT passed the test shows that it has made progress, but it's moderately acceptable

performance indicates that there are still certain areas where the model needs to be improved. The evaluation revealed shortcomings in ChatGPT's understanding of certain forensic medical concepts or in its ability to provide accurate and comprehensive responses. More training and development are necessary if it is to operate as effectively as possible in this particular industry. It's crucial to understand that ChatGPT and other AI models should be used in addition to conventional teaching techniques and human skills, not as a replacement for them. The greatest way to improve critical thinking, clinical judgement, and patient contact skills is through hands-on learning and mentoring from seasoned role models.

Limitations of our study included a small sample size. We used only 13 simple and direct questions to test the potential of ChatGPT. Thus, the results of this study cannot be generalized to all aspects of forensic medicine. However, to proceed with complex scenarios, its fundamental concepts and comprehension should be proven. We recommend that future research and development should concentrate on enhancing ChatGPT's functionality in the field of forensic medicine. The knowledge and precision of the model can be improved by including domain-specific training data and working with forensic medicine specialists. For future versions of ChatGPT or other AI models for medical education, continuous assessment and feedback loops are essential for identifying areas for improvement and directing growth.

Conclusions

While ChatGPT's performance in the forensic medicine examination for the third MBBS professional part 1 was satisfactory, there is considerable room for improvement. The study highlights the potential implications of AI models like ChatGPT in medical education, emphasizing the need for further refinement and domain-specific training. It is important to understand ChatGPT's implications in medical education. ChatGPT, an AI language model, can improve and enhance medical education by giving students access to new learning materials, responding to their inquiries, and helping them prepare for exams. For medical students looking for clarity on difficult forensic medicine concepts, its

capacity to produce pertinent and clear comments might be especially helpful. However, the limits of AI in medical education must be well-understood, though as misinformation and disinformation are potential threats. Instead of replacing conventional teaching techniques and human skills, ChatGPT should be considered a tool to augment them. Problem-solving, clinical judgement, and patient care are all important components of medical education that are best developed via practical experiences and support from knowledgeable mentors rather than heavy reliance on AI-based tools.

Conflicts of Interest: None

Funding: This study did not receive funding from any source or in any form.

Ethical Clearance: This study does not require ethical approval in any form

Acknowledgements: Authors would like to extend their gratitude to OpenAI, an American artificial intelligence research laboratory for providing free access to ChatGPT.

References

- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR medical education*, 9, 2023 e45312. <https://doi.org/10.2196/45312>
- ChatGPT becomes the fastest application to hit 100 million users, overtaking Facebook and Google, Accessed on 26.03.23; <https://indianexpress.com/article/technology/chatgpt-100-million-users-8423362/>
- Introducing ChatGPT- OpenAI., Accessed on 26.03.23; <https://openai.com/blog/chatgpt>
- Kung, T. H., 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. *PLOS digital health*, 2(2), 2023, e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*. 11(6):887, 2023. <https://doi.org/10.3390/healthcare11060887>
- Arif, T. B., Munaf, U., & Ul-Haque, I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise?. *Medical education online*, 28(1), 2023, 2181052. <https://doi.org/10.1080/10872981.2023.2181052>
- Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. ChatGPT - Reshaping medical education and clinical management. *Pakistan journal of medical sciences*, 39(2), 605-607, 2023. <https://doi.org/10.12669/pjms.39.2.7653>
- Lee, H. The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*, 00, 1- 6, 2023. <https://doi.org/10.1002/ase.2270>
- Sabry Abdel-Messih, M., & Kamel Boulos, M. N. ChatGPT in clinical toxicology. *JMIR Medical Education*, 9, 2023, e46876. doi: 10.2196/46876
- Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. Better by You, better than Me? ChatGPT-3 as writing assistance in students' essays, 2023.
- Scanlon, M., Nikkel, B., & Geradts, Z. Digital forensic investigation in the age of ChatGPT. *Forensic Science International: Digital Investigation*, 44, 2023. <https://doi.org/10.1016/j.fsidi.2023.301543>