

# ChatGPT's Performance in Answering Formative Pharmacology Examination Questions for Phase II Medical Undergraduate Students: A Cross-sectional Study

NETRAVATHI BASAVARAJ ANGADI<sup>1</sup>, VIDYA MAHALMANI<sup>2</sup>, BHARGAV ANNAM<sup>3</sup>

## ABSTRACT

**Introduction:** The emergence of Artificial Intelligence (AI) tools like ChatGPT has sparked interest in their potential role in medical education. The increasing volume and complexity of pharmacological knowledge pose significant challenges for undergraduate medical students. AI-based language models such as ChatGPT offer rapid access to information, explanations, and problem-solving support, potentially supplementing traditional teaching methods. However, systematic evaluation of their performance, reliability, and educational accuracy in pharmacology assessments remains limited, necessitating formal investigation.

**Aim:** To evaluate ChatGPT's performance in answering formative pharmacology examination questions for Phase II medical undergraduate students.

**Materials and Methods:** This was a cross-sectional study conducted in the Department of Pharmacology at Jawaharlal Nehru medical college in Belagavi, Karnataka, India over a period of one month (December 2024). All pharmacology questions included in the December 2024 formative undergraduate examination, covering general pharmacology, central nervous system drugs, cardiovascular drugs, and autacoids were included. The study protocol was reviewed and approved by the department scientific review committee and ethical committee approval was waived off as it did not involve human subjects. ChatGPT (version 3) was

used to answer Multiple-Choice Questions (MCQs), short answer questions, short essay questions, and long essay questions. The responses generated by ChatGPT were evaluated for accuracy, completeness, quality of explanation, and level of knowledge. Acceptability of responses was assessed through qualitative review by pharmacology subject experts. Comparison between groups was performed using Pearson's chi-square test. Descriptive statistics were applied, and results were expressed as frequencies and percentages.

**Results:** ChatGPT demonstrated high accuracy, with correct response rates of 13 (86.7%) for recall, 7 (70%) for interpretation, and 9 (81.8%) for problem-solving questions. ChatGPT demonstrated comparable performance across all three knowledge levels, with no statistically significant difference observed ( $p$ -value=0.8). The majority of explanations 29 (80.6%) were rated as good, 4 (11.1%) as needs revision and 3 (8.3%) as unacceptable. While performance was consistent across knowledge levels, challenges were noted in complex problem-solving scenarios requiring deep integration of concepts.

**Conclusion:** ChatGPT showed strong potential as an AI-assisted learning tool in pharmacology education, particularly for factual recall and conceptual understanding. However, improvements are needed in complex reasoning and nuanced explanations. This study supports further research into AI's role in enhancing medical curricula.

**Keywords:** Computer-assisted instruction, Knowledge assessment, Learning tools, Medical students

## INTRODUCTION

In the year 2022, a San Francisco based open AI launched chat generative pretrained transformer (Chat GPT) that evolved from AI Large Language Models (LLMs). It was designed to simulate human conversation in response to prompts or questions based on the context of input text [1,2]. ChatGPT is being tested in various domains, and one remarkable achievement is its clearance in the United States Medical Licensing Examination (USMLE) [3]. Recently, Kung TH et al., found that ChatGPT performed at or near the passing threshold for all three United States Medical Licensing Exams, suggesting that LLMs can assist with medical education and clinical decision-making [4]. Chatbots such as ChatGPT could provide tutoring and homework help by answering questions and providing explanations to help students understand complex concepts [5]. However, there are concerns that the use of AI software by students to write university assessments could diminish the value of the assessments and the overall quality of the university program [5].

Previous studies have demonstrated that ChatGPT can generate fluent and contextually appropriate responses to a wide range of academic queries, including examination-style questions, raising both interest and concern regarding its use in medical education [6,7]. Stokel-Walker C and Park SH et al., have highlighted the potential risks of confidently presented yet inaccurate AI-generated content, underscoring the importance of critical appraisal by learners [8,9]. While emerging literature has evaluated ChatGPT's performance in select medical domains and licensing examinations, evidence specific to pharmacology- particularly in the context of undergraduate formative assessments and varied question formats-remains limited. There is a notable gap in the literature regarding systematic evaluation of ChatGPT's accuracy, explanatory quality, and interpretative ability across multiple pharmacology subdisciplines and assessment formats. Given the complexity of pharmacology and its relevance to safe prescribing, understanding the reliability of AI-generated responses is essential before such tools can be recommended for educational use.

The present study addresses this gap by comprehensively assessing ChatGPT's performance in answering formative pharmacology examination questions for Phase II medical undergraduate students, thereby providing novel insights into its potential role, limitations, and implications as an AI-assisted learning adjunct in pharmacology education.

## MATERIALS AND METHODS

This cross-sectional study was conducted in the Department of Pharmacology at Jawaharlal Nehru Medical College, Belagavi, Karnataka, India over a one-month period (December 2024). The study protocol was reviewed and approved by the department scientific review committee and ethical committee approval was waived off as it did not involve Human subjects.

**Inclusion criteria:** All pharmacology questions from the second professional year formative (sessional) examination conducted in December 2024. Questions covered general pharmacology, drugs acting on the central nervous system, drugs acting on the cardiovascular system, and autacoids.

**Exclusion criteria:** Questions from other subjects, non-pharmacology components, and questions with ambiguous or incomplete wording were excluded from the study.

### Methodology and Parameters Studied

Chat GPT- 3.5 Turbo was used to generate responses to MCQs, short answer questions, short essay questions, and long essay questions from the examination paper. The level of cognitive domain addressed was categorised as recall, interpretation and problem-solving types and the overall acceptability of responses was assessed through qualitative expert review. Content validity of the question paper was ensured through expert review and alignment with the syllabus and learning objectives. Reliability of the questions in the question paper of the examination was supported by adherence to blueprinting principles and uniform difficulty distribution across pharmacology topics. The total marks for the exam were 100, divided as follows: 20 MCQs (1 mark each), two long essay questions (10 marks each), nine short essay questions (5 marks each), and five short answer questions (3 marks each). The study aimed to evaluate factual recall and conceptual understanding and problem solving, which are critical components in pharmacology learning. ChatGPT (version 3) was used to answer all the questions, with MCQs presented as four options, and theoretical questions typed directly into the interface. After collecting ChatGPT's responses, all answers were systematically analysed for accuracy, coherence, and alignment with the expected answers in pharmacology textbooks and peer-reviewed literature. Study focused on two aspects of ChatGPT's performance: the correctness of the responses and the acceptability of explanations provided, particularly in theoretical questions, which often require critical reasoning and conceptual clarity.

Each response was independently evaluated by pharmacology subject experts for factual accuracy, completeness, clarity of explanation, and level of knowledge. Each response generated by ChatGPT was categorised based on its level of correctness as correct and incorrect response. For MCQs, a correct answer was awarded one mark, while for theoretical questions; marks were assigned based on accuracy, completeness, and depth of explanation. Additionally, the acceptability of the explanations generated by ChatGPT was assessed through expert review by pharmacology faculty members. This qualitative assessment focused on the overall appropriateness, clarity, and educational usefulness of the responses. The responses were graded based on content as good when the explanation was accurate and comprehensive, needs revision when the response was partially correct or lacked sufficient depth, and unacceptable when the answer was incorrect or entirely irrelevant [10].

By systematically analysing both objective (MCQ) and subjective (theoretical) questions, this study provides valuable insights into the

potential role of AI in enhancing pharmacology education and supports future research into the integration of AI in medical curricula.

## STATISTICAL ANALYSIS

Data were entered and analysed using SPSS version 25.0 Descriptive statistics were applied, and results were expressed as frequencies and percentages. Comparison between the variables was analysed using Pearson's Chi-square test and p-value <0.05 was considered as statistically significant.

## RESULTS

The distribution of questions according to knowledge level is presented in [Table/Fig-1]. ChatGPT demonstrated comparable performance across all three knowledge levels, with no statistically significant difference observed (p-value=0.8) [Table/Fig-2]. Recall-based questions showed the highest accuracy, reflecting strong factual knowledge retrieval. Interpretation and problem-solving questions exhibited marginally lower accuracy; however, overall performance remained consistent across domains. 6 (37.5%) theoretical questions were incorrectly answered whereas only 1 (5%) of the MCQ were incorrectly answered by Chat GPT. The response generated by ChatGPT for a case-based question is represented in [Table/Fig-3].

Cognitive domain of questions	N (%)
Recall	15 (41.7)
Interpretation	10 (27.8)
Problem-solving	11 (30.6)

[Table/Fig-1]: Categorisation of questions based on knowledge level.

Cognitive domain of questions	Correct responses N (%)	Incorrect responses N (%)
Recall	13 (86.7)	2 (13.3)
Interpretation	7 (70)	3 (30)
Problem-solving	9 (81.8)	2 (18.2)
Type of question		
MCQ	19 (95)	1 (5)
Theoretical	10 (62.5)	6 (37.5)

[Table/Fig-2]: Comparison of ChatGPT's correct answer rate according to items' knowledge level.

ChatGPT

Give answer for the following question following the pattern of marks distribution for an Indian medical undergraduate student A 71-year-old woman who underwent hip replacement was discharged from the hospital with an appropriate post discharge therapy that included daily subcutaneous injection of regular heparin. Compare and contrast heparin with low molecular weight heparin. Mention the antagonist for heparin overdose. (4 + 1 = 5 marks)

**Comparison of Heparin and Low Molecular Weight Heparin (LMWH) (4 marks)**

Feature	Unfractionated Heparin (UFH)	Low Molecular Weight Heparin (LMWH)
Molecular Size	Large, heterogeneous	Smaller, more uniform
Mechanism of Action	Inhibits both thrombin (factor IIa) and factor Xa via antithrombin III	Primarily inhibits factor Xa, with less effect on thrombin
Bioavailability	Variable (30-70%)	Higher (about 90%)
Half-life	Short (1-2 hours)	Longer (3-6 hours)
Administration	IV or SC	SC only
Monitoring	Requires aPTT monitoring	No routine monitoring required
Reversal Agent	Protamine sulfate (fully reverses)	Protamine sulfate (partially reverses)
Risk of HIT	Higher risk	Lower risk

**Antagonist for Heparin Overdose (1 mark)**

Protamine sulfate is the antidote for heparin overdose, which binds to heparin and neutralizes its anticoagulant effect.

[Table/Fig-3]: Screenshot of ChatGPT's answer to a question item from Pharmacology examination for medical students.

The acceptability analysis revealed that 29 (80.6%) of responses were categorised as “good,” indicating clear and appropriate reasoning. A smaller proportion, required revision due to minor deficiencies in clarity or completeness, while only a few responses, were deemed unacceptable [Table/Fig-4]. These findings suggest that ChatGPT generally provides explanations suitable for undergraduate medical education, with limitations noted in handling more complex reasoning.

Acceptability level	Responses N (%)
Good	29 (80.6)
Needs revision	4 (11.1)
Unacceptable	3 (8.3)

**[Table/Fig-4]:** Acceptability of ChatGPT's explanations of question items (n=36).

## DISCUSSION

The present study evaluated the performance of ChatGPT in answering undergraduate pharmacology formative examination questions and found that the tool demonstrated better performance in structured question formats, particularly MCQs than the descriptive formats such as short and long essay questions. While ChatGPT showed satisfactory factual recall, its performance was variable in questions requiring detailed mechanistic explanations and higher-order interpretative skills.

The popularity of AI applications, particularly Chat-GPT, has experienced significant increase in recent years due to its extensive utilisation in both academic and industrial settings [11]. Concerns over ChatGPT and similar AI algorithms' accuracy and usefulness in healthcare have divided opinion.

AI can be used outside of typical computer settings and is widely accessible through inexpensive smart gadgets, making it widely available to the general public [12]. A new pretrained language model ChatGPT was launched on Nov 30, 2022. Pretrained language models have achieved striking success in Natural Language Processing (NLP), leading to a paradigm shift from supervised learning to pretraining followed by fine-tuning [13].

One of the key findings of the study was ChatGPT's overall performance in answering MCQs. The AI performed well in MCQs, providing correct answers in line with the expected responses. This was consistent with the findings of other studies [14,15] that suggest AI models like ChatGPT excel in factual recall, particularly when presented with structured, closed-ended questions. However, despite its accuracy in factual responses, ChatGPT faced challenges in providing contextually rich explanations for theoretical questions, especially in areas requiring a deeper understanding of pharmacological mechanisms. This was similar to the observations made by Goodman RS et al., who found that while AI systems excel at providing accurate, concise answers, they often struggle to explain complex concepts in a manner that meets educational standards [10]. However, in theoretical short essay or long essay questions, ChatGPT occasionally offered responses that were overly simplified or lacking in the depth required to fully explain the underlying pharmacological principles.

In particular, questions about drug actions on the cardiovascular system and central nervous system posed challenges for ChatGPT, where responses sometimes lacked the necessary linkage between pharmacodynamic and pharmacokinetic concepts, which are critical for an in-depth understanding of pharmacology. This shortcoming highlights the importance of not relying solely on AI for explaining complex scientific concepts, as the AI may fail to provide a comprehensive context or bridge gaps in understanding. These findings are in consistent with study done by Zhang Y et al., [16]. On the other hand, in more factual areas such as drug classifications or receptor binding mechanisms, ChatGPT performed well, aligning closely with textbook knowledge.

The classification of responses as “good,” “needs revision,” or “unacceptable” offered a clear view of the AI's strengths and limitations. The theoretical explanations given by ChatGPT were often deemed as “needs revision” because of their tendency to oversimplify or omit essential details. This finding aligns with research by Wang J and Fan W who suggested that AI's ability to generate responses with contextual relevance and clarity can sometimes fall short in subjects that demand deep conceptual knowledge [17]. For example, when answering questions about the mechanisms of action of drugs like beta-blockers or Selective Serotonin Reuptake Inhibitors (SSRIs), ChatGPT could describe basic mechanisms but struggled to explain the nuances of their interactions within broader systems like the sympathetic nervous system or serotonin pathways. Other recent observational studies have similarly highlighted deficiencies in AI-generated explanations for mechanistic and application-based questions [18].

However, it is important to acknowledge that ChatGPT's ability to assist in pharmacology education remains promising. Its strong performance in MCQs, a common assessment method in medical education, indicates its potential as a study tool or revision aid. Additionally, AI can be particularly beneficial in helping students practice recalling facts, testing their knowledge, and even offering explanations for some of the more straightforward pharmacological concepts. For example, ChatGPT's ability to correctly answer questions related to drug classes (e.g., antihypertensives, analgesics) could be useful in a self-directed learning environment, where students can instantly check their responses and receive immediate feedback.

This study highlights the need for careful monitoring and human oversight when using AI for more complex theoretical learning. While ChatGPT can serve as a valuable supplementary tool, it should not replace critical thinking or deep engagement with the subject matter. Human experts must continue to be involved in the learning process to ensure that AI-generated explanations are accurate, detailed, and tailored to the learners' needs. From an educational standpoint, these results support the need for curriculum adaptations that emphasise critical appraisal of AI-generated content, guided use under faculty supervision, and reinforcement of core pharmacological concepts through conventional teaching methods.

### Limitation(s)

The present study did not evaluate learning outcomes or student performance; therefore, conclusions regarding ChatGPT's effectiveness as a learning tool cannot be drawn. Instead, the findings suggest that ChatGPT may function as a supplementary information source rather than a replacement for traditional teaching.

## CONCLUSION(S)

ChatGPT demonstrated good performance in answering undergraduate pharmacology formative examination questions, particularly in structured formats such as MCQs, reflecting strong factual recall. However, its responses to descriptive questions were variable, with limitations in providing in-depth mechanistic explanations and higher-order interpretative insights. These findings suggest that while ChatGPT can serve as a useful supplementary tool for revision and self-directed learning, especially for fact-based content, it cannot replace traditional teaching or expert guidance for complex conceptual understanding. Careful faculty oversight and critical appraisal of AI-generated content are essential to ensure its effective and responsible integration into pharmacology education. Further research is needed to explore how AI tools like ChatGPT can be better integrated into the curriculum. This includes refining the AI's ability to generate more accurate, contextually relevant and detailed explanations for complex pharmacological concepts. Additionally, future studies should consider incorporating a wider range of pharmacology topics, including more specialised areas like

clinical pharmacology or toxicology, to assess the adaptability and performance of AI across the entire pharmacology curriculum.

## Acknowledgement

Authors would like to acknowledge OpenAI, an American AI research laboratory for providing free access to ChatGPT.

## REFERENCES

- [1] OpenAI. OpenAI: Models GPT-3 [Internet]. OpenAI; 2025 [cited 2025 Feb 14]. Available from: <https://beta.openai.com/docs/models>.
- [2] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. *arXiv [Preprint]*. 2020;arXiv:2005.14165.
- [3] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination? Implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. Doi: 10.2196/45312.
- [4] Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on the USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2:e0000198.
- [5] O'Connor S. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Educ Pract*. 2022;66:103537.
- [6] Garabet R, Mackey BP, Cross J, Weingarten M. ChatGPT-4 Performance on USMLE Step 1 style questions and its implications for medical education: A comparative study across systems and disciplines. *Med Sci Educ*. 2023;34(1):145-52.
- [7] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. Doi: 10.2196/45312. Erratum in: *JMIR Med Educ*. 2024;10:e57594. Doi: 10.2196/57594
- [8] Stokel-Walker C. AI bot ChatGPT writes smart essays: Should professors worry? *Nature [Internet]*. 2022 Dec 9 [cited 2025 Feb 14]. Available from: <https://doi.org/10.1038/d41586-022-04397-7>.
- [9] Park SH, Do KH, Kim S, Park JH, Lim YS. What should medical students know about artificial intelligence in medicine? *J Educ Eval Health Prof*. 2019;16:18. Doi: 10.3352/jeehp.2019.16.18.
- [10] Goodman RS, Patrinely JR, Stone CA Jr, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open*. 2023;6(10):e2336483. Doi: 10.1001/jamanetworkopen.2023.36483. PMID: 37782499; PMCID: PMC10546234.
- [11] Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *IoT CPS*. 2023;3:121-54.
- [12] Alam A. Possibilities and apprehensions in the landscape of artificial intelligence in education. In: *Proceedings of the 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*; 2021. p. 1-8.
- [13] Wang H, Li J, Wu H, Hovy E, Sun Y. Pre-trained language models and their applications. *Engineering*. 2023;25(6):51-65.
- [14] Choi W. Assessment of the capacity of ChatGPT as a self-learning tool in medical pharmacology: A study using MCQs. *BMC Med Educ*. 2023;23(1):864.
- [15] Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT knowledge evaluation in basic and clinical medical sciences: Multiple choice question examination-based performance. *Healthcare (Basel)*. 2023;11(14):2046.
- [16] Zhang Y, Ren S, Wang J, Lu J, Wu C, He M, et al. Aligning large language models with humans: A comprehensive survey of ChatGPT's aptitude in pharmacology. *Drugs*. 2025;85(2):231-54. Doi: 10.1007/s40265-024-02124-2. Epub 2024 Dec 20. PMID: 39702867; PMCID: PMC11802629.
- [17] Wang J, Fan W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*. 2025;12:621.
- [18] Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: A systematic review and meta-analysis. *J Biomed Inform*. 2024;151:104620.

### PARTICULARS OF CONTRIBUTORS:

1. Associate Professor, Department of Pharmacology, Jawaharlal Nehru Medical College, Belagavi, Karnataka, India.
2. Assistant Professor, Department of Pharmacology, Jawaharlal Nehru Medical College, Belagavi, Karnataka, India.
3. Senior Resident, Department of Pharmacology, Jawaharlal Nehru Medical College, KLE Academy of Higher Education and Research, Belagavi, Karnataka, India.

### NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:

Dr. Netravathi Basavaraj Angadi,  
Associate Professor, Department of Pharmacology, Jawaharlal Nehru Medical College, Belagavi-590010, Karnataka, India.  
E-mail: drnetra.angadi@gmail.com

### AUTHOR DECLARATION:

- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? No
- Was informed consent obtained from the subjects involved in the study? NA
- For any images presented appropriate consent has been obtained from the subjects. NA

### PLAGIARISM CHECKING METHODS: [Jain H et al.]

- Plagiarism X-checker: Dec 21, 2025
- Manual Googling: Feb 23, 2026
- iThenticate Software: Feb 27, 2026 (3%)

### ETYMOLOGY: Author Origin

EMENDATIONS: 8

Date of Submission: Dec 03, 2025

Date of Peer Review: Jan 01, 2026

Date of Acceptance: Mar 03, 2026

Date of Publishing: May 01, 2026