

# A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, and Microsoft Copilot in Solving NEET PG Medical Entrance Test

**Patil Anuradha Vishwanath<sup>1</sup>, Kale Sachin Sadanand<sup>2</sup>, Jali Nandita Vivekanand<sup>3</sup>, Bhise Manik Changoji<sup>4\*</sup>**

<sup>1</sup>Associate Professor, Department of Pathology, MGM Medical College and Hospital, Chhatrapati Sambhajnagar (Aurangabad), Maharashtra, India - 431001 (University - MGM Institute of Health Science, Navi-Mumbai), India

<sup>2</sup>Professor, Department of Pathology, MGM Medical College and Hospital, Chhatrapati Sambhajnagar (Aurangabad), Maharashtra, India - 431001 (University - MGM Institute of Health Science, Navi-Mumbai), India

<sup>3</sup>Intern, MGM Medical College and Hospital, Chhatrapati Sambhajnagar (Aurangabad), Maharashtra, India - 431001 (University - MGM Institute of Health Science, Navi-Mumbai), India

<sup>4\*</sup>Professor and Head, Department of Psychiatry, MGM Medical College and Hospital, Chhatrapati Sambhajnagar (Aurangabad), Maharashtra, India - 431001 (University - MGM Institute of Health Science, Navi-Mumbai), India.

Email: [dr.manik.bhise@gmail.com](mailto:dr.manik.bhise@gmail.com) (Corresponding Author)

**Received Date: 10/10/2025; Accepted: 22/01/2026; Published Date: 28/03/2026**

## ABSTRACT

**Background:** Artificial intelligence Chatbots (AI Chatbots) can assist medical students in preparation and cracking the different exams. Existing literature shows their accuracy varies while doing this. Present study aims to fill this gap by comparing accuracy of 3 AI Chatbots.

**Objectives:** Primary objective was to assess and compare the accuracy of ChatGPT-4, Google Gemini and Microsoft Copilot in solving the NEET PG 2023 exam. Secondary objective was to compare their accuracy for different types of questions and questions across different medical subjects.

**Methods:** All 200 questions of NEET PG 2023 exam paper were presented 'as it is' to the three AI chatbots. Accuracy was assessed as percentage of correct responses to each question. We compared their overall accuracy, and accuracy for different question types and subject taxonomy.

**Results:** Accuracy of Microsoft Copilot was 82.5%, ChatGPT was 80.5%, and Google Gemini 77.5% (Chi Square=1.6, p=0.4). Performance of three AI Chat Bots didn't differ in terms of different subjects (Chi Square=2.7, p=0.9) or types of questions (Chi Square=0.35, p=0.9).

**Conclusion:** All three AI Chatbots showed good accuracy with no significant difference in solving NEET PG exam questions. There was no difference in accuracy of the three chatbots in terms of subject taxonomy of questions or the type of question.

**Keywords:** Artificial intelligence (AI), entrance test, NEET-PG, multiple choice questions, graduate medical education, machine learning, deep learning

**How to cite this article:** Vishwanath PA, Sadanand KS, Vivekanand JN, Changoji BM. A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, and Microsoft Copilot in Solving NEET PG Medical Entrance Test. *Int J Drug Deliv Technol.* 2026;16(21s): 474-481. DOI: 10.25258/ijddt.16.21s.50

**Source of support:** Nil.

**Conflict of interest:** None

## INTRODUCTION

Postgraduate National Eligibility cum Entrance Test (NEET-PG) is conducted in India by The National Board of Examinations in Medical Sciences (NBEMS). This is

the only entrance test for seeking admission to postgraduate medical courses in the country. This exam is one of the toughest exams considering its vast syllabus, complex nature of questions, rising competition, and lack

## A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test

of quality training at low cost.<sup>1</sup> Approximately 2,00,517 students appeared for the NEET PG exam held in the March 2023. Of these, only 111871 (55.7%) qualified to appear for seat allotment.<sup>2</sup> With the advent of Artificial Intelligence (AI), its use in medical education and healthcare is being increasingly emphasized. Availability of free AI chatbots is assisting all stake holders in these fields in different ways. AI's ability to analyze these subjects and clinical case simulations could translate into real-life applications, enabling physicians to make quicker and more informed diagnoses. A recent review shows AI tools can assist in diagnosis, identify susceptibility to a disease using genomics data, can predict spectrum of phenotypes, can guide the treatment and overall care of patients. Authors further emphasized that AI tools can also be useful in providing mental health support, can reduce physician burn out and power the healthcare programs for public at large. However, these advantages come with draw backs, including errors in present AI tools due to lack of quality medical data, issues with data privacy, security and mixed preferences by patients for availing AI based services.<sup>3</sup>

Potential of AI to offer diagnostic assistance and suggest treatment as per guidelines can be used for solving clinical case vignettes that are tested in the NEET PG. Ability of AI Chatbots to score high marks in different exams has been studied. Their accuracy is well established for the United States Medical Licensing Examination, Ophthalmic Knowledge Assessment Program exam, United Kingdom Medical Licensing Assessment (UKMLA), NEET UG, and NEET-PG entrance tests.<sup>1,4,5,6,7</sup> There are relatively few studies which compared the accuracy of different AI Chatbots. Two studies found the superiority of ChatGPT-3.5 over Microsoft Copilot and Google Bard in solving cases in hematology and physiology.<sup>8,9</sup> A recent study has also observed that ChatGPT-4 outperformed other AI Chatbots in solving clinical questions in dentistry.<sup>10</sup> Another study aiming to solve Peruvian National Medical Licensing Examination revealed that ChatGPT-4 and Microsoft Bing outperformed Google Bard and Claude.<sup>10,11</sup> These studies also suggest that despite higher scores, AI Chatbots do not have 100% accuracy in solving questions of medical exams and there is some degree of variable performance amongst them. Thus, there seems to be a potential hierarchy in accuracy of AI chatbots. Medical education and health care in general needs to be precise as it deals with lives of human beings. Scope for mistakes can cost in terms of money

and lives. This Indicates a need for further research to establish any differences in the accuracy of these AI Chatbots. Present study aims to fill this gap in existing knowledge by comparing accuracy of 3 AI Chatbots in solving the NEET PG exam. Present study was conducted with the aim of assessing and comparing the accuracy of ChatGPT-4.0, Google Gemini and Microsoft Copilot in solving the NEET PG 2023 exam. Another objective was to compare their performance for different types of questions and different medical subjects.

### METHODS

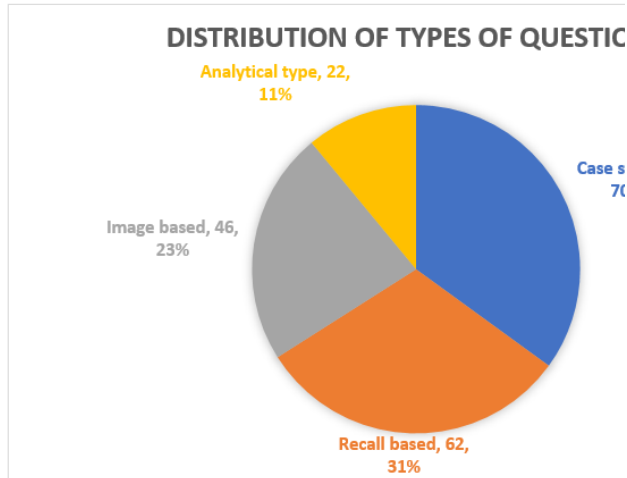
**Study design:** This is an observational cross-sectional study. We followed high-quality reporting standards of Strengthening of Reporting of Observational Studies in Epidemiology (STROBE).<sup>12</sup>

**Setting:** The study was designed by a multidisciplinary team of professors, associate professor, and intern from various medical disciplines at a medical college within a deemed to be university. The study was conducted over a three-month period, from September 2024 to November 2024.

**Sample:** The NEET PG exam is a single-day, single-session computer-based examination held at designated centers across India. The paper consists of 200 multiple-choice questions (MCQs). The NEET PG syllabus covers all preclinical, clinical, and paraclinical disciplines of the MBBS program. Candidates must select the correct, best, or most appropriate response from four answer choices for each question.<sup>13</sup> The National Board of Examinations in Medical Sciences (NBEMS), the official board, does not publicly release the question paper. Therefore, we used a question set and answers compiled by the online teaching platform PrepLadder, which gathers these based on inputs from exam takers. Although recall bias is a possibility, this was the only method available to obtain all the questions from the NEET-PG 2023 exam. Questions were categorized into four types as: pure recall-based (focused on straightforward facts), analytical (requiring critical thinking and reasoning), image-based (involving the interpretation of radiographs, histology slides, pathology specimens, and diagrams) and case study questions (clinical vignettes requiring multi-layered thinking) as shown in pie chart in fig-1. Subject wise distribution of questions was as follows: Anesthesia 3(1.5%), Radiology 4(2%), Dermatology 4(2%), Psychiatry 5 (2.5%), Orthopedics 6 (3%), ENT 6(3%), Physiology 8 (4%), Forensic medicine 8(4%),

## A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test

Ophthalmology 8(4%), Anatomy 9 (4.5%), Microbiology 10(5%), Pediatrics 10 (5%), Pharmacology 12 (6%), Biochemistry 15(7.5%), Community medicine 15 (7.5%), Pathology 16(8%), Medicine 17 (8.5%), Obstetrics and gynecology 18(9%), and Surgery 27 (13.5%).



**Figure 1: Pie chart showing different types of questions in NEET PG 2023 exam.**

**Study Tools:** Three AI chatbots were used to solve NEET PG 2023 exam questions. The ChatGPT 4.0 Plus plan was subscribed to for two months (OpenAI, Mission District, San Francisco, United States), the Microsoft Copilot Pro plan was subscribed to for two months (Microsoft Corporation, WA, US), and the Google Gemini Advanced plan was subscribed to for two months (Alphabet Inc., CA, US).<sup>15</sup> These paid versions of AI chatbots are advanced language models trained on large datasets using deep learning algorithms. They are capable of generating coherent responses and predicting text by identifying relationships between words.<sup>16</sup>

**Ethics:** As no human subjects were involved, ethical approval was not required.<sup>17</sup>

**Study procedure:** We evaluated the accuracy of three AI chatbots—ChatGPT (GPT-4.0), Gemini AI (Gemini 1.5), and Microsoft Copilot—in solving NEET PG 2023 questions. In September 2024, NEET PG 2023 questions and answers from PrepLadder were validated against standard medical textbooks, and an answer key was prepared. Each AI chatbot was presented with one question at a time, and each question was imputed only once to minimize bias when comparing results among different AI chatbots. Questions were presented in their

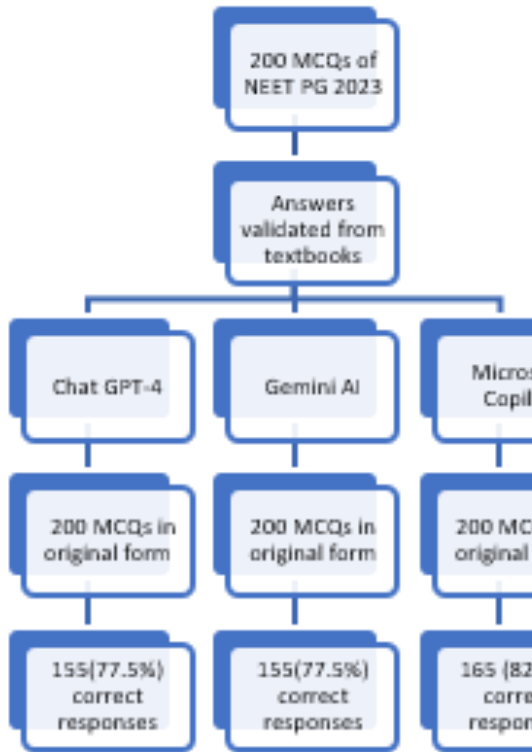
original NEET PG 2023 format, without modifications or additional hints, ensuring that responses were based solely on the input provided. To prevent memory recall bias, each question was submitted in a new conversation. All these strategies were employed to minimize the bias. Responses were marked as “Correct” or “Incorrect” using the validated answer key. Since all questions were presented to the chatbots, there was no missing data. Accuracy was measured as the percentage of correct responses out of 200 questions presented to all three chatbots. The performance of the chatbots was compared across different subjects and question types.

**Statistical Analysis:** The outputs were recorded as binomial outcomes (correct or incorrect) for each AI chatbot. Categorical variables were analyzed and presented as absolute frequencies with percentages, while continuous variables were analyzed as means. A chi-square test was conducted to determine whether the percentage of correct answers differed among the three AI chatbots. A p-value of < 0.05 was considered statistically significant. Data analysis was performed using Epi Info software version 7.2.6.0.

## RESULTS

### Study Flowchart:

**A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test**



**Figure 2: Study Flow Chart:**

The accuracy of Microsoft Copilot in solving NEET PG 2023 questions was the highest, with 165 (82.5%) correct responses. It was followed by ChatGPT-4, which achieved 161 (80.5%) correct responses, and Google Gemini, which had 155 (77.5%) correct responses. However, the difference in performance among the three chatbots was not statistically significant (Chi-square = 1.6, p = 0.4). The subject-wise performance of the chatbots is presented in Table 1.

**Table 1: Subject wise distribution of correct answers of three AI Chatbots for NEET PG 2023 Exam questions.**

Subject	Total Questions	ChatGPT	Gemini	Copilot
Anatomy	9(4.5%)	5(55.6%)	5(55.6%)	6(66.7%)
Biochemistry	14 (7%)	14 (100.0%)	13 (92.9%)	13 (92.9%)
Physiology	8 (4%)	6 (75.0%)	6 (75.0%)	6 (75.0%)
Pathology	16 (8%)	13 (81.2%)	14 (87.5%)	15 (93.8%)
Microbiology	10 (5%)	10 (100.0%)	10 (100.0%)	10 (100.0%)
Pharmacology	12 (6%)	9 (75.0%)	10 (83.3%)	10 (83.3%)
Community Medicine	15 (7.5%)	11 (73.3%)	10 (66.7%)	11 (73.3%)

Forensic Medicine and Toxicology	8 (4%)	4 (50.0%)	3 (37.5%)
Ophthalmology	8 (4%)	7 (87.5%)	7 (87.5%)
ENT	6 (3%)	5 (83.3%)	5 (83.3%)
Internal Medicine	17 (8.5%)	12 (70.6%)	12 (70.6%)
Surgery	27 (13.5%)	23 (85.2%)	22 (81.5%)
Pediatrics	10 (5%)	8 (80.0%)	8 (80.0%)
Obstetrics and Gynecology	18 (9%)	15 (83.3%)	13 (72.2%)
Radiology	4 (2%)	2 (50.0%)	2 (50.0%)
Orthopedics	6 (3%)	5 (83.3%)	4 (66.7%)
Anesthesia	3 (1.5%)	3 (100.0%)	3 (100.0%)
Dermatology	4(2%)	4 (100.0%)	3 (75.0%)
Psychiatry	5 (2.5%)	5 (100.0%)	5 (100.0%)
Total	200 (100%)	161 (80.5%)	155 (77.2%)

All three AI chatbots achieved 100% accuracy in solving questions from the Microbiology, Anesthesia, and Psychiatry subjects. However, they faced challenges in answering questions from Anatomy and radiology with average performance in Community Medicine, Forensic Medicine and Toxicology, Internal Medicine. Overall, there was no statistically significant difference in the accuracy of AI chatbots across different subjects (Table 1). On average, the AI tools demonstrated the highest accuracy in solving recall-based questions (85.5%), followed by case study questions (82.38%), and analytical questions (77.27%). The lowest accuracy was observed in interpreting image-based questions (71.03%), suggesting that these questions were relatively more difficult compared to the other three types. However, the difference in performance among the AI chatbots was not statistically significant (Table 2).

**Table 2: Distribution of correct answers of three AI Chatbots as per different types of questions in NEET PG 2023 Exam.**

Type of question	Total questions	ChatGPT	Gemini
Recall based	62(31.0%)	52(83.87%)	51(82.25%)
Case study type	70(35.0%)	60(85.7%)	56(80.0%)
Analytical type	22(11.0%)	17(77.27%)	17(77.27%)
Image based	46(23.0%)	31(67.4%)	31(67.39%)
Total	200	161	155

**DISCUSSION**  
 Microsoft Copilot demonstrated the highest accuracy (82.5%), followed by ChatGPT (80.5%) and Google Gemini (77.5%) with no significant difference in overall performance. There was also no statistically significant

## A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test

difference in accuracy across different subjects. However, all three AI chatbots exhibited lower accuracy when solving image-based questions.

Of these three Chatbots, Chat GPT had proven high accuracy in solving questions of many medical entrance exams like USMLE, Japanese medical licensing exam, United Kingdom medical licensing exam and German state medical exam.<sup>6,18-20</sup> Studies that compared the performance among these chatbots reported variable findings. A study done for the Italian entrance test for healthcare science degree reported the accuracy of Chat GPT and Microsoft copilot being better than that of Google Gemini.<sup>21</sup> However a recent study found relatively low accuracy (64.5%) of Chat GPT in solving NEET PG 2023 questions as compared to our study.<sup>1</sup> One reason for this differential accuracy might be due to difference in optimization of the algorithms and their training. Difference in their neural network architecture may also explain this difference in accuracy of the AI tools. Microsoft copilot and Chat GPT 4 use GPT (generative pre trained transformer) architecture, while Google Gemini uses LaMDA (Language model for dialogue application) in combination with web search.<sup>22</sup> These AI chatbots are periodically updated. A study reported accuracy of Chat GPT 3.5 (older version) in solving NEET PG type practice questions as only 65.3%.<sup>23</sup> Available literature also suggests inconsistency in responses to the same question imputed on different occasions by these AI Chatbots.<sup>6</sup> Performance of AI Chatbots varied among different subjects and types of questions. All three AI tools showed 100% accuracy in subjects like microbiology, Psychiatry and Anesthesia which had recall type questions and lower accuracy of 50% in Forensic medicine and Radiology subjects where questions were of image based or analytical type.<sup>24-26</sup> AI tools frequently face difficulties with image-based questions as they lack sufficient training data and have limitations in processing visual and textual information together. The inherent ambiguity in medical images like subtle patterns or overlapping features, requires a deep contextual understanding and clinical correlation that general AI systems struggle to provide at present.<sup>27,28</sup>

In India, many students are unable to get good coaching for the entrance exams. Most coaching centers are concentrated in few cities and are costlier for students belonging to low and middle-class family.<sup>29</sup> Availability of AI-Chatbots on smart phones can provide high-quality educational materials like video lectures, textbooks, interactive exercises with instant doubt resolution at little

to no cost. However exclusive reliance of these Chatbots may not be appropriate till these are made full proof. None of these AI Chatbot showed 100% accuracy in studies so far. Thus, excess reliance on these AI chatbots for NEET exam preparation, may lead students to false answers and negative marks, causing fall in their rank. Due to differential accuracy of these AI Chatbots, cross checking for answers may leave students in confusion. Over reliance on AI Chatbots may hamper development of critical thinking and problem-solving skills which are essential in medical field.<sup>30</sup> AI models need large data set for training, raising ethical concerns for data privacy, data copy rights and security. Repeated imputation of same instruction can lead to different out puts using same AI Chatbots.<sup>6,23</sup> This further limits their utility in medical education and healthcare.

Utilization of AI chatbots in healthcare can assist in diagnosis, devising management plan and prognosis for index patient. In resource constrained communities, at primary healthcare level AI Chatbots may help healthcare workers in providing accurate treatment, appropriate referral and easy access to medical knowledge. However, this can raise certain ethical issues and practical challenges. Lack of human touch, empathy, and emotional support is big hurdle in their use in patient care. Confidentiality and security of patient's data while training these AI models as well as while using them in practice can raise ethical and legal concerns.<sup>3</sup> Lack of accuracy (bias) across AI tools can lead to conflicting healthcare decisions with risk of mismanagement of treatment. Medical professionals rely on critical thinking and complex decision-making abilities. Reliance on AI can reduce the development of these abilities in medical professionals.<sup>30</sup>

Results of our study should be viewed from the limitations it has. Limited sample size, as this study tested only the NEET PG 2023 question paper, which may not fully represent the diversity and complexity of questions across multiple years. NBEMS, the official board does not release the question paper publicly. Hence, we used question set compiled by online teaching platform PrepLadder, which compiles them based on inputs from exam takers. Hence reliability of exact questions can't be ascertained. We fed the questions only once, without repeated trials to assess the consistency in their responses, as this was not the objective of study. This limits our understanding of the reliability of their performance. While stratification based on accuracy, subject taxonomy and question type provided useful data,

# A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test

the study didn't evaluate the reasoning process or quality of explanations provided by the models. Lastly, the performance of ChatGPT, Gemini and Copilot was evaluated based on their versions available at the time of study. Subsequent updates to these models may lead to improved or altered performance, limiting the generalizability of our findings.

## CONCLUSIONS

All three AI Chatbots showed good accuracy in solving the NEET PG 2023 exam questions with no significant difference. There was no difference in accuracy of the three chatbots in terms of subject taxonomy of questions or the type of question. Thus, all three AI Chatbots could potentially crack the NEET PG 2023 exam with good scores. Use of AI chatbots as mainstay educational or healthcare tools is limited by their relative inaccuracy in present form.

**Conflict of interest:** nil

**Funding-** We have funded our self for subscription plans of AI Chatbots. Chat GPT Plus plan was subscribed for two months (USD 20/ Month), Microsoft Copilot Pro subscribed for 2 months (USD 20/month) and Google Gemini 2.0 Advanced plan was subscribed for 2 months (1<sup>st</sup> month free, second month USD 23). So total self-funding was USD 103.

**Acknowledgments:** nil

## REFERENCES

1. Paul S, Govindaraj S, Jk J. ChatGPT Versus National Eligibility cum Entrance Test for Postgraduate (NEET PG). *Cureus*. 2024 Jun 24; 16(6):e63048. doi:10.7759/cureus.63048
2. Careers360 [Internet]. Careers360.com. 2025 [cited 2025 Apr 20]. Available from: <https://medicine.careers360.com/>
3. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya A, Almohareb SN, *et al*. Revolutionizing healthcare: the Role of Artificial Intelligence in Clinical Practice. *BMC Medical Education* [Internet]. 2023 Sep 22;23(1):1–15. Available from: <https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-023-04698-z>
4. 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? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education* [Internet]. 2023 Feb 8;9(9):e45312. Available from: <https://mededu.jmir.org/2023/1/e45312/PDF>
5. Teebagy S, Colwell L, Wood ER, Yaghy A, Faustina M. Improved Performance of ChatGPT-4 on the OKAP Examination: A Comparative Study with ChatGPT-3.5. *Journal of academic ophthalmology*. 2023 Jul 1;15(02):e184–7.doi:10.1055/s-0043-1774399.
6. U Hin Lai, Wu K, Hsü TC, Kai J. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. *Frontiers in Medicine*. 2023 Sep 19;10. doi:10.3389/fmed.2023.1240915
7. Farhat F, Chaudhry BM, Nadeem M, Sohail SS, Madsen DØ. Evaluating Large Language Models for the National Premedical Exam in India: Comparative Analysis of GPT-3.5, GPT-4, and Bard. *JMIR Medical Education* [Internet]. 2024 Feb 21 [cited 2024 Jun 23];10(1):e51523. Available from: <https://mededu.jmir.org/2024/1/e51523>
8. Kumari A, Kumari A, Singh A, Sanjeet Kumar Singh, Juhi A, Kumar A, *et al*. Large Language Models in Hematology Case Solving: A Comparative Study of ChatGPT-3.5, Google Bard, and Microsoft Bing. *Cureus*. 2023;15(8):e43861
9. Kumar A, Mohammed Jaffer Pinjar, Nitin Dhokane, Sorte SR, Kumari A, Mondal H. Performance of Large Language Models (ChatGPT, Bing Search, and Google Bard) in Solving Case Vignettes in Physiology. *Cureus*. 2023 Aug 4;15(8):e42972
10. Giannakopoulos K, Kavadella A, Salim AA, Stamatopoulos V, Kaklamanos EG. Evaluation of the Performance of Generative AI Large Language Models ChatGPT, Google Bard, and Microsoft Bing Chat in Supporting Evidence-Based Dentistry: Comparative Mixed Methods Study. *Journal of Medical Internet Research* [Internet]. 2023 Dec 28;25(1):e51580. Available from: <https://www.jmir.org/2023/1/e51580/>
11. Torres-Zegarra BC, Rios-Garcia W, Ñaña-Cordova AM, Arteaga-Cisneros KF, Chalco XCB, Ordoñez MAB, *et al*. Performance of ChatGPT, Bard, Claude, and Bing on the Peruvian National Licensing Medical Examination: a cross-sectional study. *Journal of Educational Evaluation for Health Professions* [Internet]. 2023;20:30. Available from: <https://pubmed.ncbi.nlm.nih.gov/37981579/>
12. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* [Internet]. 2007 Oct 18;335(7624):806–8. Available from:

**A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test**

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2034723/>
13. NBEMS [Internet]. Natboard.edu.in. 2021 [cited 2024 May 02]. Available from: <https://natboard.edu.in/viewnbeexam?exam=neetpg>
  14. Walia K. NEET PG Previous Year Question Papers of Last 7 Years with PDF [Internet]. prepladder. Kashika Walia; 2025 [cited 2024 Aug 20]. Available from: <https://www.prepladder.com/neet-pg-study-material/previous-years-question-papers/neet-pg-previous-year-question-papers-of-last-5-years>
  15. OpenAI. ChatGPT [Internet]. ChatGPT. OpenAI; Available from: <https://chatgpt.com/>
  16. Giacomo Rossetini, Cook C, Palese A, Paolo Pillastrini, Turolla A. Pros and cons of using artificial intelligence Chatbots for musculoskeletal rehabilitation management. *Journal of Orthopaedic & Sports Physical Therapy*. 2023 Dec 1;53(12):1–7.
  17. Nowell J. Guide to ethical approval. *BMJ* [Internet]. 2009 Feb 12;338(338):b450. Available from: <https://www.bmj.com/content/338/bmj.b450>
  18. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, *et al*. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. Dagan A, editor. *PLOS Digital Health* [Internet]. 2023 Feb 9;2(2):e0000198. Available from: <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000198>
  19. Takagi S, Watari T, Ayano Erabi, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study. *JMIR medical education* [Internet]. 2023 Jun 29 [cited 2023 Aug 20];9:e48002–2. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC103652915/>
  20. Ung LB, Gudera JA, Wiegand TLT, Allmendinger S, Dimitriadis K, Koerte IK. ChatGPT passes German state examination in medicine with picture questions omitted. 2023 May 30; 120(21):373-374.
  21. Giacomo Rossetini, Rodeghiero L, Corradi F, Cook C, Paolo Pillastrini, Turolla A, *et al*. Comparative accuracy of ChatGPT-4, Microsoft Copilot and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study. *BMC medical education*. 2024 Jun 26;24(1):694.
  22. Giannakopoulos K, Kavadella A, Salim AA, Stamatopoulos V, Kaklamanos EG. Evaluation of the Performance of Generative AI Large Language Models ChatGPT, Google Bard, and Microsoft Bing Chat in Supporting Evidence-Based Dentistry: Comparative Mixed Methods Study. *Journal of Medical Internet Research* [Internet]. 2023 Dec 28;25(1):e51580. Available from: <https://www.jmir.org/2023/1/e51580/>
  23. Ramnani S, Bhalla M, Bassi R. A Comparative Study of ChatGPT and BingAI in Answering the National Eligibility Entrance Test for Postgraduates (NEET-PG)-Style Practice Questions: A Cross-Sectional Analysis. *Cureus*. 2024 Dec 20; ;16(12):e76108
  24. Nader Al-Shakarchi, Haq IU. ChatGPT Performance in the UK Medical Licensing Assessment: How to Train the Next Generation? *Mayo Clin Proc Digit Health*. 2023 Jul7;1(3):309-310. doi: 10.1016/j.mcpdig.2023.06.004.
  25. Sadeq MA, Ghorab RMF, Ashry MH, Abozaid AM, Banihani HA, Salem M, *et al*. AI chatbots show promise but limitations on UK medical exam questions: a comparative performance study. *Scientific Reports*. 2024 Aug 14;14(1). doi:https://doi.org/10.1038/s41598-024-68996-2.
  26. Noda M, Ueno T, Ryota Kosu, Yuji Takaso, Mari Dias Shimada, Saito C, *et al*. Performance of GPT-4V in Answering the Japanese Otolaryngology Board Certification Examination Questions: Evaluation Study. *JMIR medical education*. 2024 Mar 28;10:e57054–4.
  27. Lin SY, Hsu YY, Ju SW, Yeh PC, Hsu WH, Kao CH. Assessing AI efficacy in medical knowledge tests: A study using Taiwan’s internal medicine exam questions from 2020 to 2023. *Digital Health*. 2024 Jan 1;10.
  28. Mumuni AN, Hasford F, Udemé NI, Dada MO, Awojoyogbe BO. A SWOT analysis of artificial intelligence in diagnostic imaging in the developing world: making a case for a paradigm shift. *Physical Sciences Reviews*. 2022 Oct 21; 9(1):443-476
  29. Team Careers360. Best NEET Coaching Institutes in India with Fees Structure -Coachings, Tutorials [Internet]. Careers360. 2022 [cited 2025 May 03]. Available from: <https://medicine.careers360.com/articles/best-neet-coaching-institutes-in-india-with-fees-structure>
  30. Sapra R. An investigation into the introduction of artificial intelligence in the medical curriculum for healthcare professionals in developing countries such as India. *International education and research journal* [Internet]. 2023 Jun 15 [cited 2025 Apr 24];9(6):79–81. Available from: <https://ierj.in/journal/index.php/ierj/article/view/2790/2919>

**A Cross-Sectional Descriptive Study of Comparative Accuracy of ChatGPT, Google Gemini, And Microsoft Copilot in Solving NEET PG Medical Entrance Test**