

# To Compare the Efficiency of ChatGPT and Bard in Medical Education: An Analysis of MCQ-Based Learning and Assessment

Sharjeel Husain<sup>1</sup>, Sabaa Shahid<sup>2</sup>, Zaid Ansari<sup>3</sup>, Tahera Ayoob<sup>4</sup>, Azhar Hussain<sup>5</sup>, Rimsha Mujahid<sup>6</sup>

## Abstract

**Objective:** This study aimed to compare the efficacy of ChatGPT and Google Bard as virtual tutors in supporting students across various levels of cognition in MCQ-based assessments in the field of Internal Medicine.

**Methods:** This cross-sectional study was conducted in the Department of Internal Medicine in collaboration with the Department of postgraduate medical education from June 2023 to October 2023. A comprehensive collection of multiple-choice questions (MCQs) covering various aspects of Internal Medicine was compiled by the research team's consensus. The items were systematically organized into chapters and further categorized based on cognitive complexity levels (C1, C2, and C3). The chosen MCQs were entered into separate sessions of both ChatGPT and Google Bard. The responses from each Artificial Intelligence platform were then compared with the corresponding answers in the designated MCQs book. Recorded responses were classified as accurate, inaccurate, or partially accurate.

**Results:** The ChatGPT exhibited an overall success rate of 64%, providing 199 correct responses out of 307 queries, of which 10 were partially correct. By contrast, Google Bard achieved an overall success rate of 58.95%, yielding 181 correct responses out of 307 queries, where 16 were partially correct. When stratified by cognitive complexity levels, ChatGPT demonstrated proficiency in solving C2 MCQs at a rate of 80%, whereas the performance rates for the C1 and C3 categories were 69% and 54%, respectively. In contrast, Google Bard displayed a 33% success rate in solving C2 MCQs while achieving success rates of 95% and 53% in the C1 and C3 categories, respectively.

**Conclusion:** The findings of this study suggest that ChatGPT is a more advantageous tool for students and medical educators than Google Bard. These discerned advantages underscore the potential of ChatGPT to enhance the educational experience within the medical domain.

**Keywords:** Artificial intelligence, multiple choice question, medical education

**IRB:** Approved by the Institutional Review Board of Dar-ul-Sehat Hospital. Ref# DSH/IB/2023/0044. Date: 8<sup>th</sup> December 2023.

**Citation:** Husain S, Shahid S, Ansari Z, Ayoob T, Hussain A, Mujahid R. To Compare the Efficiency of ChatGPT and Bard in Medical Education: An Analysis of MCQ-Based Learning and Assessment [Online]. *Annals ASH & KMDC* 2024;29(1) : 56-63

(ASH & KMDC 2024; 29(1) :56-63)

## Introduction

Artificial Intelligence (AI) represents a pivotal element within the landscape of transformative technologies, significantly impacting the way humans engage with information, make decisions, and navigate the multifaceted world in which we

live. As the relentless progression of technological advancements continues, the incorporation of AI has become all-encompassing, touching nearly every facet of our existence<sup>1</sup>. Among the most recent breakthroughs in A.I. is the development of large language models (LLMs) such as Chat Generative Pre-Trained Transformer (ChatGPT) and Google Bard. These models possess the capacity to generate responses that closely resemble human language and have demonstrated potential in various educational and assessment contexts. The integration of AI into medical education has emerged as a topic of increasing interest and discourse<sup>2</sup>.

ChatGPT is an advanced conversational AI system that incorporates the Generative Pretrained

<sup>1,3,5,6</sup> Department of Internal Medicine; Liaquat College of Medicine and Dentistry & Darul Sehat Hospital,

<sup>2</sup> DHPE (Department of Health Profession and Education), Liaquat College of Medicine and Dentistry; Qamar Dental Hospital

<sup>4</sup> Department of Oral Surgery, Liaquat College of Medicine and Dentistry & Qamar Dental Hospital, Karachi

**Correspondence:** Dr. Sharjeel Husain  
Department of Internal Medicine; Liaquat College of Medicine and Dentistry & Darul Sehat Hospital  
Email: dr\_husain2003@yahoo.com

Date of Submission: 19<sup>th</sup> December 2023

Date of Revision: 26<sup>th</sup> February 2024

Transformer (GPT) series of natural language processing models, which were unveiled by Open AI in November 2022. This model boasts an impressive 175-billion parameter architecture and is capable of generating conversation-style responses to user inputs<sup>3</sup>. It comes in two versions, the free GPT-3.5 and the paid GPT-4, and has already gained significant popularity, with over 1 million users within a week of its release<sup>4</sup>.

Google Bard is an exceptional instance of Artificial Intelligence (AI) that has gained significant attention in the realm of human dialogue generation. This conversational AI was unveiled by Google in March 2023. Unlike ChatGPT, which relies on information updated before September 2021, Bard can gather data from the Internet, thereby showing its potential for application in fields that require more current information. Google Bard, like its counterpart ChatGPT, demonstrates remarkable proficiency in swiftly retrieving, interpreting and delivering information. Its versatility is evident, as it can not only offer information on a specific topic but also generate content tailored to specific requirements<sup>5-6</sup>.

In the field of language translation, article summarization, and draft generation, both Chat GPT and Google Bard are highly advanced and competent tools with promising applications across a range of scholarly pursuits. They are both well-versed in offering recommendations and drafting assignments in a variety of disciplines. The integration of artificial intelligence into medical education has the potential to transform learning and evaluation techniques<sup>7,8</sup>.

In the realm of medical education, multiple-choice questions (MCQs) have long been regarded as an effective means of evaluating students' cognitive abilities, particularly those outlined in Bloom's Taxonomy. The MCQs have been widely adopted as assessment tools that yield valid and reliable scores for undergraduate and postgraduate assessments. They are a useful method for assessing knowledge retention and comprehension as well as the ability to apply empirical knowledge. Compared to open-ended questions, MCQs exhibit a high level

of reliability and are valuable for evaluating a wide range of cognitive skills, including recall, analysis, synthesis, and evaluation<sup>9</sup>.

MCQs are designed to test students' understanding of key concepts, facts, and principles related to a specific topic and their ability to identify and select the correct answer from a range of options. Medical students are frequently required to sit rigorous exams throughout their professional development and must employ effective study strategies to consolidate their core knowledge. Students during their days of course preparation, often encounter a wide range of problems. As a result, they frequently search for multiple ways to reinforce their knowledge and strategies that will aid them in their exam preparation<sup>10</sup>.

Despite their impressive capabilities in content generation and information retrieval, the use of these tools presents unique challenges in the context of research and education. For instance, there is a risk that they may facilitate academic dishonesty during online exams and may be perceived as reducing the emphasis on critical thinking skills. Nevertheless, there are differing perspectives on the use of AI tools in education, research, and healthcare, with varying levels of ambiguity about its acceptability and appropriate applications<sup>11,12</sup>.

To ensure the effective integration of these tools into educational settings, it is essential to carefully evaluate the potential benefits and challenges associated with their use. The dual capacity to enhance productivity and pose challenges to the integrity of assessments demands a thorough comparison of their learning efficiency and assessments<sup>13</sup>.

This study aims to compare ChatGPT (3.5 version) and Google Bard efficiency in the learning and assessment of multiple-choice questions (MCQ) based on varying cognition levels (C1, C2, and C3) in the backdrop of different chapters of internal medicine at the graduate and postgraduate levels.

## **Methodology**

This study was conducted by a diverse research team from June to October 2023 at a private medical and dental college. The research team utilized three MCQ's book as given in table 1. An MCQ bank of 1428 MCQ's was established from three books of MCQ's on subject of Internal Medicine used widely by undergraduate and post graduate medical students. MCQ's were only selected from different subjects of Internal Medicine. As the MCQ's books don't mention the cognitive level of assessment, the senior team members with experience in medical education mark individual MCQ with category of cognition level as C1, C2 and C3 respectively. The research team members carefully reviewed the MCQs' contents and assured that the MCQs were relevant and appropriately challenging. Each question was scenario-based with four sub-stems or had a single correct answer. MCQs were evaluated for quality, and research team concurred with the final answer. Finally, a total of 307 MCQ's were finally selected for the task by the research team.

Each MCQ was scenario-based and had four sub-stems or a single correct answer. The team evaluated the quality of the MCQs and proofread them for errors and inconsistencies. The team excluded any other formats such as pictures, graphs, or flowcharts from the list due to the limitations of the A.I tool utilized in the study. This meticulous process aimed to maintain the integrity and quality of the MCQs used in this study.

The selected MCQs were then evaluated using ChatGPT version 3.5 and Google Bard. The research team manually entered each MCQ into the respective AI tool and started a fresh session for each entry to avoid memory retention bias. The first response obtained from each AI tool was recorded as the final response, and the choice of "regenerate response" was not used. This approach ensured that the responses obtained were not influenced by previous interactions with AI tools.

The study was conducted in compliance with ethical standards and guidelines for research involv

ing human subjects. The research team obtained the necessary approval and permission from the relevant institutional review board. The data collected from the AI tools were securely stored and analyzed to draw meaningful conclusions.

## **Results**

This study involved the administration of 307 multiple-choice questions (MCQs) sourced from textbooks to ChatGPT version 3.5 and Google Bard. The ChatGPT provided 199 correct responses, 98 incorrect responses, and 10 partially correct responses. By contrast, Google Bard offered 181 correct responses, 110 incorrect responses, and 16 partially correct responses (Table 2)

A subdivision was created to further classify the questions based on their level of cognition. The C1 questions represented simple recall, C2 encompassed straightforward question scenarios, and C3 comprised of vignette-style questions with multiple probable answers. Because some C3 cases contained several sub-questions, each sub-question was treated as an individual question, and the AI tool's response to each question was marked separately.

According to the level of cognition, the ChatGPT and Google Bard performance are detailed in Table 3. For C2 level MCQs ChatGPT and Google Bard both achieved 80% accuracy, while the results were low for the C3 category at approximately 54% and 40%, respectively. The results for C1 questions, which were recall-based, had a low percentage of correct answers, at approximately 69.8% for both AI tools.

The replies obtained from Bard were further sequestered as per the subjects in internal medicine, as shown in Table 4. In the Chat GPT, almost all subjects showed healthy responses around the mean, except for endocrinology and hematology, where responses were below 60% and 40%, respectively.

In Google Bard, a low percentage of correct responses was observed for the subjects of GIT (47%), hematology (40%), genetics (20%), and toxicology (38%).

**Tables 1.** MCQ's books utilized to gather MCQ pool by the research team.

Book name	Author	Edition/publisher
1000 Questions and Answers from Kumar & Clark's Clinical Medicine	Parveen Kumar Micheal Clarke	Second edition SaundersElsevier
Rapid review of Clinical Medicine for MRCP Part 2	Sanjay Sharma Rashmi Kaushal	Second Edition Manson publishing
Master the IMM (Medicine)	Usman Muzaffar Ali	4th Edition Nishtar Publications

**Table 2: Marks obtained by Chat GPT & Google Bard after solving selected pool of 307 MCQ's**

	Chat GPT	Google Bard
Correct replies	199 (64.82%)	181 (58.95)
In correct replies	98 (31.92%)	110 (35.83%)
Partially correct	10 (3.25%)	16 (5.21%)

**Table 3:** ChatGPT & Google Bard performance as per cognition level of respective MCQ's

Level of Cognition	Correct(n & %)		In correct(n & %)		Partially correct(n & %)		Total n=307	
	Chat GPT	Google Bard	Chat GPT	Google Bard	Chat GPT	Google Bard	Chat GPT	Google Bard
C1	95 (69.8)	95 (69.8)	40 (29.41)	37 (27.20)	1 (0.73)	4(2.94)	136	136
C2	33 (80.48)	33 (80.48)	8 (19.51)	8 (24.24)	0	0	41	41
C3	71 (54.61)	53 (40.76)	50 (38.46)	65 (50.0)	9 (6.9)	12 (9.23)	130	130

**Table 4 :**Google Bard replies to MCQ's as per chapters of Internal Medicine MCQ Book

Chapters in Internal Medicine	Correct (n)	In Correct(n)	Partially correct (n)	Correct score percentage (%)
CVS	32	12	2	69.56
Rheumatology	11	13	0	45.83
Endocrine	25	15	4	56.81
GIT	14	12	4	46.66
Pulmonology	26	13	1	65.0
Hematology	6	8	1	40.0
Infectious Disease	12	8	2	54.54
Nephrology	25	8	0	75.75
Neurology	23	11	0	67.64
Pharmacology	1	0	0	100
Genetics	1	4	0	20
Toxicology	3	3	2	37.5
A&E	0	1	0	0
Immunology	0	2	0	0
Psychiatry	1	0	0	100
Maternal medicine	1	0	0	100

## Discussion

Assessment of students in medical education is crucial for evaluating their knowledge and skills in various disciplines. The acquisition of knowledge is fundamental in medical education, and the ability to interpret and apply this knowledge effectively in real-life scenarios is essential. Multiple-choice questions (MCQs) are widely used to assess these skills, making it a renowned tool for global medical assessment. MCQs play a significant role in the assessment of medical students, as they are utilized by medical schools and licensing examination bodies during various stages of undergraduate and postgraduate assessments. The use of MCQs in medical education is well-documented and has been a subject of extensive research, focusing on the appropriateness of MCQs in assessing a broad curriculum and the challenges in quality item writing<sup>14</sup>.

The design of MCQs is critical to assessing deep learning effectively, and it is essential to ensure that the questions can assess higher-order thinking. However, creating MCQs that can assess higher-order thinking is challenging and requires adherence to specific guidelines, particularly ensuring that item writers are competent in their fields<sup>15</sup>.

The integration of advanced language models such as ChatGPT and Google Bard has attracted significant attention among various stakeholders, including the public, students, academics, researchers, and the science community. Their ability to swiftly articulate different dimensions of a subject and generate responses with variations to meet the demands of queries has made them a valuable tool in medical education<sup>16</sup>.

ChatGPT is designed to produce human-like responses and engage users in conversational

interactions, thus providing rapid responses in seconds. This technology is guided by a wide range of internet text data, enabling it to understand and produce text in diverse contexts, including answering questions, providing explanations, offering suggestions, creating conversational dialogues, and assisting with multiple tasks<sup>17</sup>.

Similarly, Google Bard, a state-of-the-art natural language processing model, has demonstrated significant potential in understanding the context of words in search queries and providing more relevant search results. Its ability to comprehend the nuances of language and generate accurate responses has made it a valuable tool for information retrieval and knowledge dissemination in medical education. Another aspect of Google Bard is its capability to upgrade continuously with available online data and provides real-time replies to queries making it a continuously updated A.I. tool<sup>18</sup>.

As per the literature search and capability of AI tools, they can help teachers and medical educationists in exam preparation or the formation of assessment tools, such as MCQs or any other format. The comparison of ChatGPT and Google Bard vs. Human teachers is beyond any match as it can generate MCQs ten times faster than that of the entire group of teachers<sup>19</sup>. With such abilities, it was very well expected that when the technology is used for solving the available MCQs, results would be very helpful and encouraging for students to prepare for the assessments minimizing their effort and saving precious time. Moreover, both A.I. tools can provide humanoid responses for each question alleviating the need for textbook referrals regularly.

Bard has a shared conversation function and a double-check function that helps the user fact-check generated results. The researcher team anticipated more efficient responses from it, given its training to utilize online data to generate responses. However, the performance of ChatGPT in answering multiple-choice questions (MCQs) surpassed that of Google Bard. This observation aligns with the potential of ChatGPT demonstrated in previous studies<sup>20</sup>, indicating its proficiency in natural language processing and its ability to provide accurate and coherent responses. The comparison of ChatGPT and Google Bard in the context of assisting students with MCQs during their exam preparation highlights their advanced capabilities in addressing academic queries and supporting students in their learning process.

The findings of this study contribute to the growing body of evidence supporting the

effectiveness of ChatGPT in educational settings, particularly in the domain of medical education, and underscores its potential to enhance students' learning experiences. This research is a continuation of our previous study, in which we established the efficacy of Chat GPT in assisting students in the preparation of MCQs<sup>21</sup>.

The availability of a paid version of ChatGPT, introduced on March 14, 2023, on a subscription basis, has raised concerns about its accessibility, particularly for students in third-world countries. This limitation has prompted a critical examination of its potential usage in medical education, especially when compared to the free version<sup>22</sup>.

This study had several limitations that warrant acknowledgement. Initially, the research predominantly concentrated on Multiple Choice Questions (MCQs) within the domain of Internal Medicine, thereby limiting its scope to other medical specialties and assessment formats. Additionally, the study was conducted within a specific educational context, namely Liaquat College of Medicine and Dentistry, Karachi, Pakistan, which may restrict the applicability of the findings to other medical institutions or regions with distinct curricula and student populations. Finally, the study utilized the free version of ChatGPT, which may have led to the exclusion of its advanced features available in the paid version. Further studies should be conducted to compare the Chat GPT freeware version with the Chat GPT-paid version. The study aimed to evaluate A.I. tools' usage and efficiency in medical education and their utility for students in their pre-examination preparation period. The research team has no conflict of interest with either of the two A.I. tools and their corporations nor received any support of any kind in this regard. All results obtained are per the pre-set guidelines which were established by the research team.

The integration of AI tools, including ChatGPT and Google Bard, into medical education presents a dynamic shift in the learning landscape. The

evolving landscape of AI-driven NLP (natural language processing) tools in medical education presents opportunities and challenges. As these tools continue to advance, it is essential to critically evaluate their performance, limitations, and ethical considerations to ensure responsible integration into educational settings. The findings from the studies referenced in this paper collectively contribute to a nuanced understanding of the implications of integrating AI tools, emphasizing the need for ethical guidelines, academic integrity, and responsible use of AI tools in medical education and knowledge assessment.

### Conclusion

The findings of this study suggest that ChatGPT is a more advantageous tool for students and medical educators than Google Bard. These discerned advantages underscore the potential of ChatGPT to enhance the educational experience within the medical domain.

### References

1. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit Health* 2019;5:1-12. [DOI: 10.1177/2055207619871808]. Available from: <https://journals.sagepub.com/doi/10.1177/2055207619871808>. Accessed on 22nd February 2024.
2. Rahaman MdS, Ahsan MMT, Anjum N, Rahman MdM, Rahman MN. The AI Race is on! Google's Bard and OpenAI's ChatGPT Head to Head: An Opinion Article. *SSRN Electronic Journal* 2023;1-6. Available from: <https://papers.ssrn.com/abstract=4351785>. Accessed on 22nd February 2024.
3. 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):1-14. [DOI: 10.52225/narra.v3i1.103]. Available from: <https://doi.org/10.52225/narra.v3i1.103>. Accessed on 22nd February 2024.
4. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ* 2023. [DOI: 10.1002/ase.2270]. Available from: <https://anatomypubs.onlinelibrary.wiley.com/doi/10.1002/ase.2270>. Accessed on 22nd February 2024.
5. Farhat F, Chaudhry BM, Nadeem M, Sohail SS, Madsen DØ. Evaluating AI Models for the National Pre-Medical Exam in India: A Head-to-Head Analy.

- sis of GPT-3.5, GPT-4, and Bard. *JMIR Med Educ* 2023;6-20. [DOI: 10.2196/preprints.51523]. Available from: [https://www.researchgate.net/publication/372969607\\_Evaluating\\_AI\\_Models\\_for\\_the\\_National\\_Pre-Medical\\_Exam\\_in\\_India\\_A\\_Head-to-Head\\_Analysis\\_of\\_ChatGPT-35\\_GPT-4\\_and\\_Bard\\_Preprint](https://www.researchgate.net/publication/372969607_Evaluating_AI_Models_for_the_National_Pre-Medical_Exam_in_India_A_Head-to-Head_Analysis_of_ChatGPT-35_GPT-4_and_Bard_Preprint). Accessed on 22nd February 2024.
6. Bibault JE, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial. *J Med Internet Res* 2019;21(11):1-7. [DOI: 10.2196/15787]. Available from: <https://pubmed.ncbi.nlm.nih.gov/37428117/>. Accessed on 22nd February 2024
  7. Davis R, Eppler M, Ayo-Ajibola O, Loh-Doyle JC, Nabhani J, Samplaski M, et al. Evaluating the Effectiveness of Artificial Intelligence-powered Large Language Models Application in Disseminating Appropriate and Readable Health Information in Urology. *J Urol* 2023;210(4):688–94. [DOI: 10.1097/JU.0000000000003615]. Available from: <https://pubmed.ncbi.nlm.nih.gov/37428117/>. Accessed on 22nd February 2024.
  8. Iqbal S, Ahmed S, Ali Z, Kashif M, Aslam A. Evaluation of records of oral and maxillofacial surgery cases reported at abbasi shaheed hospital and Karachi medical and dental college, Pakistan. *The International Journal of Frontier Sciences* 2020;4(1):47-51. [DOI: 0.37978/tijfs.v4i1.74]. Available from: <https://publie.frontier-science-associates.com.pk/index.php/tijfs/article/view/74>. Accessed on 22nd February 2024.
  9. Liu Q, Wald N, Daskon C, Harland T. Multiple-choice questions (MCQs) for higher-order cognition: Perspectives of university teachers. *Innovations in Education and Teaching International* 2023; [DOI: 0.1080/14703297.2023.2222715]. Available from: <https://www.tandfonline.com/doi/epdf/10.1080/14703297.2023.2222715?needAccess=true>. Accessed on 22nd February 2024.
  10. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple-choice questions? Research paper. *BMC Med Educ* 2007;7(1):1–7. [DOI: 10.1186/1472-6920-7-49]. Available from: <https://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-7-49>. Accessed on 22nd February 2024.
  11. 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 Med Educ* 2023;9:1-9. [DOI: 10.2196/45312]. Available from: <https://mededu.jmir.org/2023/1/e45312>. Accessed on 22nd February 2024.
  12. Chen L, Chen P, Lin Z. Artificial Intelligence in Education: A Review. *IEEE Access*. 2020;8:75264–78. Available from: <https://ieeexplore.ieee.org/document/9069875>. Accessed on 22nd February 2024.
  13. Chan KS, Zary N. Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. *JMIR Med Educ*. 2019;5(1):1-15. [DOI: 10.2196/13930]. Available from: <https://mededu.jmir.org/2019/1/e13930/>. Accessed on 22nd February 2024.
  14. Douthit NT, Norcini J, Mazuz K, Alkan M, Feuerstein MT, Clarfield AM, et al. Assessment of Global Health Education: The Role of Multiple-Choice Questions. *Front Public Health* 2021;9:1-10. [DOI: 10.3389/fpubh.2021.640204]. Available from: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.640204/full>. Accessed on 22nd February 2024.
  15. DiSantis DJ. A Step-By-Step Approach for Creating Good Multiple-Choice Questions. *Canadian Association of Radiologists Journal* 2020;71(2):131–3. [DOI: 0.1177/084653711988835]. Available from: <https://journals.sagepub.com/doi/10.1177/0846537119888358>. Accessed on 22nd February 2024.
  16. 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):1–8. [DOI: 10.1186/s12909-023-04832-x]. Available from: <https://bmcmeded uc.biomedcentral.com/articles/10.1186/s12909-023-04832-x>. Accessed on 22nd February 2024.
  17. Souza LL de, Fonseca FP, Martins MD, Almeida OP de, Pontes HAR, Coracin FL, et al. ChatGPT and medicine: a potential threat to science or a step towards the future? *J Med Artif Intell* 2023;6:19-19. [DOI: 10.21037/jmai-23-70]. Available from: <https://jmai.amegroups.org/article/view/8197/html>. Accessed on 22nd February 2024.
  18. Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng HT, et al. LaMDA: Language Models for Dialog Applications 2022;1-47. Available from: <https://arxiv.org/abs/2201.08239v3>. Accessed on 22nd February 2024.
  19. Agarwal M, Sharma P, Goswami A, Agarwal M, Sharma P, Goswami A. Analysing the Applicability of ChatGPT, Bard, and Bing to Generate Reasoning-Based Multiple-Choice Questions in Medical Physiology. *Cureus* 2023;15(6):1-23. [DOI: 10.7759/cureus.40977]. Available from: <https://www.cureus.com/articles/165118-analysing-the-applicability-of-chatgpt-bard-and-bing-to-generate-reasoning-based-multiple-choice-questions-in-medical-physiology>. Accessed on 22nd February 2024.

20. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology*. 2023;307(5):e230922. Available from: <https://pubs.rsna.org/doi/epdf/10.1148/radiol.230922>. Accessed on 22nd February 2024.
21. Shirjeel Husain S, Ansari Z, Hussain A, Zubair Abbasi S, Ayoob T, Mujahid R. To Evaluate the Efficiency of ChatGPT in Medical Education: An Analysis of MCQ-Based Learning and Assessment. *Annals ASH & KMDC* 2023;28(4):194–200.
- Available from: <https://www.annals-ashkmdc.org/index.php/ashkmdc/article/view/759>. Accessed on 22nd February 2024
22. Toyama Y, Harigai A, Abe M, Nagano M, Kawabata M, Seki Y, et al. Performance evaluation of ChatGPT, GPT-4, and Bard on the official board examination of the Japan Radiology Society. *Jpn J Radiol* 2024;42:201–7. Available from: <https://link.springer.com/article/10.1007/s11604-023-01491-2>. Accessed on 22nd February 2024



This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 4.0 License (CC BY-NC 4.0). To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>