Indian Journal of Orthopaedics (2025) 59:1413–1419
https://doi.org/10.1007/s43465-025-01430-5

ORIGINAL ARTICLE

How do the Available LLM Platforms Fare as On-the-Go Orthopaedic Referencing Source? A Comparative Analysis

Girinivasan Chellamuthu[1,2] · Sathish Muthu[1,2,3] · Siddeshwar Siddamanickam[1,4] · Naveen Sathiyaseelan[1,5] · Vibhu Krishnan Viswanathan[1,6] · Raviteja Rudraraju[1,7] · Rajkumar Rangabashyam[8]

Received: 1 May 2024 / Accepted: 25 May 2025 / Published online: 11 June 2025
© Indian Orthopaedics Association 2025

Abstract

Introduction On-the-go (OTG) reference is defined as the reference that clinicians make on the go in a narrow time frame during patient care, which tends to play a significant role in decision-making. The common sources of these references over the years, have been changing from hard copies of textbooks and journals to online platforms. With the introduction of artificial intelligence (AI) based large language model (LLM) platforms, they are now being relied upon for these OTG references to provide clinicians with the necessary facts to make good clinical decisions by analyzing the maximum available resources. This study aims to compare the answers given by various LLM platforms with the answers obtained by clinicians using conventional referencing and also grade the relevance of the answers provided by these platforms.

Methods Three commonly used AI-based LLM chat platforms: ChatGPT Version 4(GPT 4), Microsoft Bing chat and Google Bard were selected for the study. 250 OTG clinical queries were collated from orthopaedic practitioners along with their answers and references used. The queries were given to the LLMs and their answers were compared and graded with the human answers for relevance and level of evidence (LOE) of the reference cited to support their answers.

Results We did not find any significant difference between the AI-LLM models tested regarding the relevance of generated answers to the clinical queries raised ($p=0.110$). ChatGPT answers were significantly better to queries that necessitated numerical answers ($p=0.006$) while Bard ($p=0.503$) and Bing ($p=0.545$) did not differ in their performance based on query types. We noted a statistically significant difference concerning the LOE of answers obtained ($p<0.001$). Upon ranking the three LLMs with the LOE of human references, human references ranked the best followed by Bing, ChatGPT, and Bard.

Conclusion Of the three compared tools, Bing Chat used relatively better LOE in answering OTG questions. All three compared AI-LLM tools show promising results concerning OTG referencing. We propose that customization to the medical domain and regulatory policies are needed before their recommended use.

Keywords ChatGPT · Google Bard · Microsoft Bing · LLM · Artificial intelligence · Level of evidence

Introduction

In their day-to-day clinical practice, sometimes the orthopaedicians might not remember numerical or nominal data relevant to a particular clinical scenario. They might search the medical literature to find the same [1, 2]. On-the-go (OTG) reference is defined as the reference that the clinicians make on the go in a narrow time frame during patient care, which tends to play a significant role in decision-making [3]. The common sources of these references over the years, have been changing from hard copies of textbooks and journals to online platforms [2]. This has increased the convenience and shortened the time. However, the accuracy of these online sources is questionable. The quality of the answer depends on the resource availability and awareness of clinicians about trustworthy online platforms [3–5].

With increasing Orthopaedic patient load, finding the right source for a particular OTG question has become cumbersome and quite time-consuming. With the introduction of artificial intelligence (AI) based large language model (LLM) platforms like ChatGPT, Bing and Bard, online

Girinivasan Chellamuthu, Sathish Muthu have contributed equally to the study and are co-first authors.

Research work conducted at Orthopaedic Research Group, Coimbatore.

Extended author information available on the last page of the article

search has been revolutionized [6–8]. We no longer have to analyse and find the best source for our query from thousands of search results that the classical search engines provide. These AI chat boxes are proposed to have the necessary intelligence to analyse various sources and provide us with crisp and clear answers [9].

Can this development also revolutionize OTG reference practises and provide clinicians with the necessary facts to make good clinical decisions by analyzing the maximum available resources? To answer this question, we planned to compare the answers given by various search platforms with the answers obtained by clinicians using conventional means of referencing. We also aimed to grade the relevance of the answers provided by the new-age AI-based LLM platforms to evaluate the possibility of their usage for OTG references.

Methodology

The study methodology was discussed and finalized in the Orthopaedic Research Group (ORG) – Quality Appraisal Committee (QAC) June 2023 convention [10]. Three commonly used AI-based LLM chat platforms—ChatGPT Version 4 (GPT 4), Microsoft Bing chat and Google Bard [6–8] were selected for the study. GPT 4 is the advanced version of GTP available at a monthly subscription price of 20 USD from Open AI [11]. It has access to the data published on the internet till January 2022. Bing Chat from Microsoft also uses GPT 4. It is free and has live access to the internet. Bard is a free AI-based chat box from Google. It is an optimized version of the Language Models for Dialogue Applications (LaMDA), a GPT derivative specialised in dialogues [12].

Five clinicians were enrolled to collect OTG questions for the study. The enrolled clinicians were asked to note down the next consecutive 55 OTG clinical doubts and their answers with the references they used. These questions and answers were collected and collated by two other authors. These questions were then run through the above-mentioned three AI platforms by three authors who were blinded from the answers given by the clinicians on September 12th 2023.

The ORG QAC [10] devised the ROMENS classification to classify the OTG references based on the analysis of questions collected for this study (Supplementary Annexure 1). The ORG-QAC also devised a simple grading system—the ORG Relevance Grading System, to grade the clinical relevance of answers provided by different LLM platforms (Supplementary Annexure 1). The questions were classified according to the ROMENS classification. The Oxford Level of Evidence (LOE) table [13] and ORG Relevance Grading System (ORGS) were used to rank and grade the references and answers derived from different AI platforms, respectively. The answers given by clinicians were graded using the LOE. These answers were assumed to have ORGS A. This grading and classification were done by two authors who were blinded from the rest of the study. Any conflicts during grading were resolved by discussion and the involvement of another author.

The statistical analysis was done by SPSS 25 (IBM, Armonk, NY, USA). The descriptive analytics was made with frequencies and percentages. The chi-square test and Kruskal–Wallis H test were employed to identify the platform that provides superior results for the category of questions put forth to them. A p value less than 0.05 was considered significant.

Results

Among the 275 OTG reference questions collected, duplicate questions and questions from the same orthopaedic topics were excluded. 250 questions were finalized for the analysis. Based on the ROMENS classification, the proportion of the 250 selected questions in each category was R (3.2%), O (4.8%), M (36.8%), E (3.6%), N (39.2%), and S (12.4%). The majority of the OTG questions were made for the numerical data ($n=98$) and management aspects ($n=92$) of orthopaedic ailments.

Relevance Analysis

We graded the relevance of the results given by the LLMs using ORGS. The included LLMs responded to all the questions in less than a minute. The answers given by the ChatGPT were to the point in 66.8% ($n=167$) queries, 17.2% ($n=43$) of them were partially answering the queries, 12.8% ($n=32$) were pertinent to the topic but not answering the queries, and 3.2% ($n=8$) were irrelevant answers. The answers given by the Bard were to the point in 72.8% ($n=182$) queries, 16.8% ($n=42$) of them were partially answering the queries, 6.8% ($n=17$) of them were pertinent to the topic but not answering the queries, and 3.6% ($n=9$) were irrelevant answers. With regards to Bing, the answers were to point in 74.4% ($n=186$) queries, 14.8% ($n=37$) of them were partially answering the queries, 8.4% ($n=21$) of them were pertinent to the topic but not answering the queries raised, and 2.4% ($n=6$) were irrelevant answers as shown in Table 1. We did not find any significant difference between the three AI-LLM models tested regarding the relevance of generated answers to the clinical queries raised ($H=4.410$; $p=0.110$).

On analyzing the LLMs individually based on the type of queries put forth to them, the relevance of the ChatGPT answers was significantly better to queries that necessitated numerical answers ($H=16.450$; $p=0.006$). Bard ($H=4.333$; $p=0.503$) and Bing ($H=4.028$; $p=0.545$) did not differ in their performance based on the type of queries put forth

Table 1 Performance of the large language models based on the relevance of their results to the query given to them

Relevance Grade	ChatGPT *n* (%)	Bard *n* (%)	Bing *n* (%)
A	167 (66.8%)	182 (72.8%)	186 (74.4%)
B	43 (17.2%)	42 (16.8%)	37 (14.8%)
C	32 (12.8%)	17 (6.8%)	21 (8.4%)
D	8 (3.2%)	9 (3.6%)	6 (2.4%)

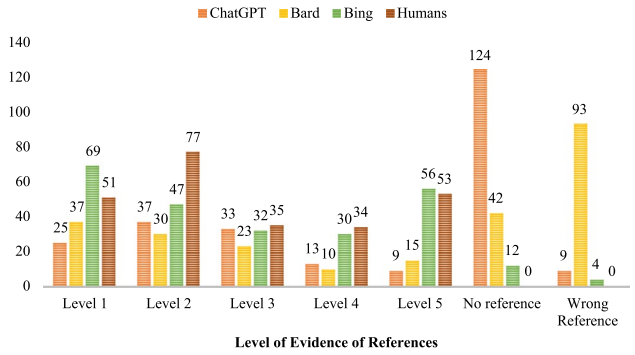


Fig. 1 Level of evidence of the references given by the large language models for the query given to them

to them. However, their best relevance ranks were in queries necessitating numerical answers without statistical significance.

Reference Analysis

On analyzing the references to the answers given by the AI-based LLMs, ChatGPT did not give a specific relevant reference to 49.6% ($n=124$) of the queries raised to it. The common references given were names of orthopaedic textbooks and websites of orthopaedic associations. In 3.6% ($n=9$) queries, the references given were wrong. This was accepted to be a made-up reference by the LLM when asked to verify the authenticity. Among the relevant references obtained for the remaining 46.8% ($n=117$) queries, the LOE distribution was as follows: Level 1 ($n=25$), Level 2 ($n=37$), Level 3 (33), Level 4 (13), and Level 5 ($n=9$) as shown in Fig. 1.

The references given by the Bard were wrong in 37.2% ($n=93$) of the queries while no references were given to 16.8% (n = 42) of the queries. Among the relevant references obtained for the remaining 46% ($n=115$) queries, the LOE distribution was as follows: Level 1 ($n=37$), Level 2 ($n=30$), Level 3 (23), Level 4 (10), and Level 5 ($n=15$) as shown in Fig. 1. On analyzing the references to the answers given by Bing, the queries without any specific relevant reference were 4.8% ($n=12$) and wrong references were provided in 1.6% ($n=4$). Among the relevant references obtained for the remaining 93.6% ($n=234$) queries, the LOE distribution was as follows: Level 1 ($n=69$), Level 2 (47), Level 3 (32), Level 4 (30), and Level 5 ($n=56$) as shown in Fig. 1.

We noted a statistically significant difference among the three LLMs concerning the LOE of the answers obtained ($H=122.399$; $p<0.001$). Upon ranking the three LLMs with the LOE of human references, human references ranked the best followed by Bing, ChatGPT and Bard. ChatGPT ($r=0.407$; $p<0.001$), Bard ($r=0.157$; $p=0.013$) and Bing ($r=0.274$; $p<0.001$) showed significant positive correlations between the LOE of the answers obtained to its ORGS grades. Hence, better relevance scores were obtained with answers with higher LOE references. Upon analyzing the LOE of these LLMs to the nature of the query based on ROMENS grade, we noted a significant difference in the LOE of Bing to its ROMENS type ($H=22.345$; $p<0.001$). We noted significantly higher LOEs to the management-based queries in Bing while the rest of the LLMs and human response did not have their LOEs differ significantly based on the ROMENS type of queries put forth to them as shown in Table 2.

We also compared the efficiency of the LLMs to pick up the highest LOE reference to their answers. Bing significantly outperformed the humans in picking the highest LOE references to its answers ($p<0.001$) while ChatGPT ($p=0.234$) and Bard ($p=0.289$) did not show any statistical significance concerning their reference LOE scores.

Discussion

The LLMs have been evolving rapidly over the years [14] ever since the introduction of Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformers (GPT). They have already been tried in mass medical education, data screening, extraction and summarising. GPT–2 was used to summarize large evidence published during the COVID-19 crisis [15]. Both the BERT and GPT were specifically trained in the medical domain to develop tools like BioBERT, BioGPT, Med-PaLM and PubMedGPT [16–19]. These tools were specialised in answering questions, text mining and extraction. For example, Med-PaLM exceeded expectations in solving the United States Medical Licensing Exam (USMLE) questions[19]. GPT-4, one of the latest versions of GPT exceeded the passing score of USMLE by over 20 points and was superior to the previous LLMs[20] including models tuned to medical domains like Med-PaLM. This finding opens up several possibilities for the usage of GPT-4 in the medical field. Aside from the technical capabilities, the user-friendly interface of the LLMs tested in the study, when compared to the previous

Table 2 Ranking of level of evidence of the large language models and humans based on the type of query given to them

Platform	ROMENS type (*n*)	Mean rank	Kruskal–Wallis *H*	*p*-value
Level of Evidence of ChatGPT	Research-related (8)	153.56	3.437	0.633
	Others (12)	128.67		
	Management-related (92)	131.27		
	Examination-related (9)	126.72		
	Numerical (98)	117.58		
	Surgical technique-related (31)	124.61		
Level of Evidence of Bard	Research-related (8)	186.63	10.461	0.063
	Others (12)	131.42		
	Management-related (92)	132.48		
	Examination-related (9)	140.56		
	Numerical (98)	116.17		
	Surgical technique-related (31)	111.82		
Level of Evidence of Bing	Research-related (8)	102.06	22.345	< 0.001
	Others (12)	79.83		
	Management-related (92)	114.42		
	Examination-related (9)	70.61		
	Numerical (98)	141.29		
	Surgical technique-related (31)	148.15		
Level of Evidence of Human	Research-related (8)	135.75	0.67	0.985
	Others (12)	132.46		
	Management-related (92)	122.92		
	Examination-related (9)	115.17		
	Numerical (98)	126.90		
	Surgical technique-related (31)	126.39		

versions stands out as a distinct improvement allowing non-technical people to explore their possibilities[14]

More than 3000 new articles are published every day in peer-reviewed journals[16]. The evidence on various treatment modalities is proliferating. New insights are being published every day. Against this backdrop, AI seems to offer a possible solution to help clinicians make good clinical decisions, taking into consideration all the recent advancements in a particular field. We aimed to summarize the ability of the commonly used LLM-based chat tools to answer different unique questions posed by clinicians. The important findings of our study were that, out of 250 questions, the LLMs were able to provide clinically relevant answers in 242 questions, and exact answers in 178 questions on average. This is quite significant because these questions were all quite challenging even to trained clinicians and the LLM tools were able to solve them in less than a minute.

While ChatGPT was significantly better in its response to numerical questions, Bing was quite superior to other LLM tools tested with regard to references. It was able to provide a correct reference in 234 questions out of which 69 were level 1 evidence. It was able to provide better references than clinicians and other LLM tools in clinical management-related questions. Overall, its ability to provide correct references was only slightly less than the clinicians. This is again a significant finding considering the less-than-a-minute response generation time by these tools. Another finding was the positive correlation between the level 1 evidence and Grade A ORGS while using the LLM tools. This can be interpreted in many ways. Whenever there is published level 1 evidence on a particular topic, LLM tools were able to access them to provide ORGS grade A answers. Tuning these LLM tools to rely more on level 1 evidence might increase their answer relevance grades.

To summarize, even without customization, current-generation LLM tools were able to adapt to different difficult medical questions and provide relevant answers. Customization like task-specific training, better prompting, and fine-tuning of these LLM tools to a particular domain can improve their relevance and referencing [21, 22]. For example, BERT fine-tuned PubMed abstracts and PubMed Central full-text articles to evolve into BioBERT. BioBERT had superior scores when compared to BERT in functions like named entity recognition, question answering and relation extraction [16].

Gupta et al. [23] explored the possibility of using ChatGPT in research. They concluded that ChatGPT can give novel ideas for systematic reviews. Besides this, they noted that ChatGPT can be used in online consultations, surgical planning, education, and post-surgical care of patients. Lozic

et al. [12] compared the ability of scientific writing by ChatGPT version 4, 3.5, Bard, Bing, Claude 2, and Aria tools. They concluded that these tools need further development to be employed in full-scale scientific writing. Sallam et al. [24] in their systematic review of the usage of ChatGPT in the medical domain noted its ability to cause a paradigm shift in healthcare education, research, and practice. However, they raised ethical concerns about its unregulated use and called for an urgent code of ethics and practice for the use of such LLMs in practice.

The current generation of LLMs, despite being the most advanced, are not without concerns. They are known to "hallucinate" i.e., provide false information. Such misinformation, for example, providing irrelevant answers or wrong references in our study, can have dangerous consequences [9]. The level of conviction that these tools use to convey such false information, worsens this situation. Another potential concern is the feeding of biased input data. This can lead to biased answers. Strict regulation and transparency on input data can solve both of these problems.

Future Directions

New-generation LLMs are here already and it is our responsibility to tame them and use them in cumbersome tasks like OTG referencing. As Clusmann et al. and Mesko et al. [9, 25] point out, National and International agencies such as the Central Drugs Standard Control Organization (CDSCO) of India, The European Medicines Agency and the Food and Drug Administration of the United States of America should step in and form universal regulatory policies and strict framework for the development of medical domain-specific customized LLM tools. The input medical data should be regulated and transparent. The patient-centred tools should be developed independently and released for mass use to prevent misinformation.

Clinicians can effectively utilize large language models (LLMs) to access a broad range of references for clinical queries, provided the authenticity of these references is carefully verified to avoid reliance on inaccurate or fabricated citations. While LLMs, such as Bing, demonstrate proficiency in identifying high-level evidence and can aid in locating literature for diagnoses and treatment protocols, critical appraisal and clinical judgment remain essential for decision-making. The rapid response of LLMs is advantageous, but accuracy and reliability must take precedence. These tools can also enhance team discussions by providing quick supplementary information, though cross-checking within the team is crucial. Ultimately, LLMs should be employed as aids to refine searches and narrow clinical options, complementing evidence-based guidelines and specific patient needs. As the technology evolves rapidly, clinicians are encouraged to stay updated on advancements to maximize benefits and address limitations effectively.

Conclusion

Of the three compared tools, Bing Chat used relatively better LOE in answering OTG questions. All three compared AI-LLM tools show promising results concerning OTG referencing. We propose that customization to the medical domain and regulatory policies are needed before their recommended use.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43465-025-01430-5.

Author Contributions Girinivasan Chellamuthu: conceptualization, data curation, writing—original draft, writing—review and editing. Sathish Muthu: conceptualization, data curation, formal analysis, administration, writing—original draft, writing—review and editing. Siddeshwar Siddamanickam: data curation, writing—review and editing. Naveen Sathiyaseelan: data curation, writing—review and editing. Vibhu Krishnan Viswanathan: data curation, writing—review and editing. Raviteja Rudraraju: data curation, writing—review and editing. Rajkumar Rangabashyam: data curation, writing—review and editing.

Funding There is no funding source.

Availability of Data and Materials The authors will provide the data involved in the study upon reasonable request.

Code availability Not applicable.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Consent for Publication Not applicable.

Informed Consent Not applicable.

Ethical Standard The study does not involve human subjects and is therefore exempt from the necessity of ethical committee approval.

References

1. Bennett, N. L., Casebeer, L. L., Zheng, S., & Kristofco, R. (2006). Information-seeking behaviors and reflective practice. *Journal of Continuing Education in the Health Professions, 26*(2), 120–127. https://doi.org/10.1002/chp.60
2. Covell, D. G., Uman, G. C., & Manning, P. R. (1985). Information needs in office practice: Are they being met? *Annals of Internal Medicine, 103*(4), 596–599. https://doi.org/10.7326/0003-4819-103-4-596
3. Girinivasan, C., & Sathish, M. (2021). Analysis of reference practices among practicing orthopaedicians in India. *Indian Journal of Orthopaedics, 55*(4), 869–878. https://doi.org/10.1007/s43465-021-00350-4

4. Bennett, N. L., Casebeer, L. L., Kristofco, R. E., & Strasser, S. M. (2004). Physicians' Internet information-seeking behaviors. *Journal of Continuing Education in the Health Professions, 24*(1), 31–38. https://doi.org/10.1002/chp.1340240106
5. Bernard, E., Arnould, M., Saint-Lary, O., Duhot, D., & Hebbrecht, G. (2012). Internet use for information seeking in clinical practice: A cross-sectional survey among French general practitioners. *International Journal of Medical Informatics, 81*(7), 493–499. https://doi.org/10.1016/j.ijmedinf.2012.02.001
6. An important next step on our AI journey. (2023). https://blog.google/technology/ai/bard-google-ai-search-updates/. Published February 6, 2023. Accessed 25.12.2023. https://blog.google/technology/ai/bard-google-ai-search-updates/. Accessed 25 Dec 2023
7. Introducing ChatGPT. https://openai.com/blog/chatgpt. Accessed 25.12.2023. https://openai.com/blog/chatgpt. Accessed 25 Dec 2023
8. Bing AI. https://www.bing.com/search?q=Bing+AI&showconv=1. Accessed 25.12.2023. https://www.bing.com/search?q=Bing+AI&showconv=1. Accessed 25 Dec 2023
9. Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine, 6*(1), 1–6. https://doi.org/10.1038/s41746-023-00873-0
10. Orthopaedic Research Group. https://orthopaedicresearchgroup.com/qac.php. Accessed 12.8.2021. https://orthopaedicresearchgroup.com/qac.php. Accessed 12 Aug 2021
11. ChatGPT. https://chat.openai.com. Accessed 26.12.2023. https://chat.openai.com. Accessed 26 Dec 2023
12. Lozić, E., & Štular, B. (2023). Fluent but not factual: A comparative analysis of ChatGPT and other AI Chatbots' proficiency and originality in scientific writing for humanities. *Future Internet, 15*(10), 336. https://doi.org/10.3390/fi15100336
13. OCEBM Levels of Evidence. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. Accessed 25.12.2023. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. Accessed 25 Dec 2023
14. De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., et al. (2023). ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Frontiers in Public Health, 11*, 1166120. https://doi.org/10.3389/fpubh.2023.1166120
15. Batra H., Jain A., Bisht G., Srivastava K., Bharadwaj M., & Bajaj D., et al. (2021). CoVShorts: News Summarization Application Based on Deep NLP Transformers for SARS-CoV-2. In: *2021 9th International Conference on Reliability, Infocom Technologies and Optimization* (Trends and Future Directions) (ICRITO), pp. 1–6. https://doi.org/10.1109/ICRITO51393.2021.9596520
16. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics, 36*(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682
17. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., et al. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics, 23*(6), bbac409. https://doi.org/10.1093/bib/bbac409
18. Stanford CRFM Introduces PubMedGPT 2.7B. (2022). https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b. Published December 15, 2022. Accessed 26.12.2023. https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b. Accessed 26 Dec 2023
19. Singhal K., Azizi S., Tu T., Mahdavi S. S., Wei J., Chung H. W., et al. (2022). Large language models encode clinical knowledge https://doi.org/10.48550/arXiv.2212.13138
20. Nori H., King N., McKinney S. M., Carignan D., & Horvitz E. (2023). Capabilities of GPT-4 on medical challenge problems. https://doi.org/10.48550/arXiv.2303.13375
21. Naveed H., Khan A. U., Qiu S., Saqib M., Anwar S., Usman M., et al. (2023). A comprehensive overview of large language models https://doi.org/10.48550/arXiv.2307.06435
22. Xie, Q., Schenck, E. J., Yang, H. S., Chen, Y., Peng, Y., & Wang, F. (2023). Faithful AI in Medicine: A Systematic Review with Large Language Models and Beyond. *Research Square*. https://doi.org/10.21203/rs.3.rs-3661764/v1
23. Gupta, R., Herzog, I., Weisberger, J., Chao, J., Chaiyasate, K., & Lee, E. S. (2023). Utilization of ChatGPT for plastic surgery research: Friend or foe? *Journal of Plastic, Reconstructive & Aesthetic Surgery, 80*, 145–147. https://doi.org/10.1016/j.bjps.2023.03.004
24. Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6), 887. https://doi.org/10.3390/healthcare11060887
25. The future landscape of large language models in medicine | Communications Medicine. https://www.nature.com/articles/s43856-023-00370-1. Accessed 26.12.2023. https://www.nature.com/articles/s43856-023-00370-1. Accessed 26 Dec 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Girinivasan Chellamuthu[1,2] · Sathish Muthu[1,2,3] · Siddeshwar Siddamanickam[1,4] · Naveen Sathiyaseelan[1,5] · Vibhu Krishnan Viswanathan[1,6] · Raviteja Rudraraju[1,7] · Rajkumar Rangabashyam[8]

✉ Sathish Muthu
drsathishmuthu@gmail.com

Girinivasan Chellamuthu
giri.c.nivasan@gmail.com

Siddeshwar Siddamanickam
drsid08@gmail.com

Naveen Sathiyaseelan
spnaveen17@gmail.com

Vibhu Krishnan Viswanathan
drvibu007@gmail.com

Raviteja Rudraraju
drravitejar@gmail.com

Rajkumar Rangabashyam
rajiam.15@gmail.com

[1] Orthopaedic Research Group, Coimbatore, India

[2] Karpagam Academy of Higher Education, Coimbatore, India

[3] Government Karur Medical College and Hospital, Karur, India

[4] New Cross Hospital, Wolverhampton, UK

[5] Department of Upper Limb Surgery, Ortho One Orthopaedic Speciality Center, Coimbatore, India

[6] Institute of Orthopaedic Research and Accident Surgery, Devadoss Hospital, Madurai, India

[7] Apollo Hospital, Hyderabad, India

[8] Shri Hospitals, Tondiarpet, Chennai, India