An evaluation of orthodontic information quality regarding artificial intelligence (AI) chatbot technologies: a comparison of ChatGPT and google BARD

Can Arslan,[*] Kaan Kahya,[*] Emre Cesur[†] and Derya Germec Cakan[*]
Department of Orthodontics, Faculty of Dentistry, Yeditepe University, Istanbul, Turkey[*]
Private Practice, Ankara, Turkey[†]

Introduction: In recent times, chatbots have played an increasing and noteworthy role in the field of medical practice. The present research was conducted to evaluate the accuracy of the responses provided by ChatGPT and BARD, two of the most utilised chatbot programs, when interrogated regarding orthodontics.
Materials and methods: Twenty-four popular questions about conventional braces, clear aligners, orthognathic surgery, and orthodontic retainers were chosen for the study. When submitted to the ChatGPT and Google BARD platforms, an experienced orthodontist and an orthodontic resident rated the responses to the questions using a five-point Likert scale, with five indicating evidence-based information, four indicating adequate information, three indicating insufficient information, two indicating incorrect information, and one indicating no response. The results were recorded in Microsoft Excel for comparison and analysis.
Results: No correlation was found between the ChatGPT and Google BARD scores and word counts. However, a moderate to significant relationship was observed between the scores and several listed references. No significant association was found between the number of words and references, and a statistically significant difference was observed in both investigators' numerical rating scales using the AI tools ($p = 0.014$ and $p = 0.030$, respectively).
Conclusion: Generally, ChatGPT and BARD provide satisfactory responses to common orthodontic inquiries that patients might ask. ChatGPT's answers marginally surpassed those of Google BARD in quality.
(Aust Orthod J 2024; 40: 149 - 157. DOI: 10.2478/aoj-2024-0012)

Received for publication: January, 2024
accepted: March, 2024.

Can Arslan: can.arslan@yeditepe.edu.tr; Kaan Kahya: kaan.kahya@yeditepe.edu.tr; Emre Cesur: emre-cesur@hotmail.com; Derya Germec Cakan: derya.cakan@yeditepe.edu.tr

Introduction

Artificial Intelligence (AI) may be defined as a sequence of actions to achieve a particular objective.[1] As a sub-field of computer science, it holds the potential to replicate human intelligence to make predictions and complex decisions.[1,2] AI has been increasingly integrated into various modern services, such as content recommendation, facial recognition and smart speakers, and so has created many exciting opportunities.[3]

In the medical and dental fields, substantial progress has been made in utilising Artificial Intelligence, including various technologies related to medical image analysis, data mining and natural language processing. The development of high-performance computing has made it feasible to extract and obtain essential data from large datasets. This data extraction process is commonly called machine learning (ML).[4]

The utilisation of chatbots, such as ChatGPT (Chat Generative Pre-trained Transformer) developed by

Open Access. Published by Sciendo. © 2024 Author(s). This work is licensed under the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/)

OpenAI (OpenAI, San Francisco, CA, USA), can significantly enhance healthcare delivery through simulated natural conversation with human users. ChatGPT, released in November 2022, is currently the most prominent chatbot in the medical field. Additionally, Google (Google, Mountain View, CA, USA) has unveiled a chatbot called BARD, which shares similarities with ChatGPT but boasts real-time internet access and employs the Language Model for Dialogue Applications (LaMDA) as its communication model.[1]

In the context of medical practice, chatbots have a significant influence on the healthcare system.[5] They can be a complementary tool for diagnosing and making decisions in medical fields, including cardiology, radiology, and urology. Sophisticated language models, such as ChatGPT, also possess substantial potential for enhancing clinical applications and research in dentistry. Their prudent use can bring about a transformative change in dental diagnosis and treatment planning. Moreover, integrating diverse medical examination data in investigations can help achieve precision and personalised dentistry and medicine objectives. However, it is crucial to acknowledge the inherent restrictions and potential risks associated with implementing AI technologies. Therefore, the present study aimed to assess the accuracy of the answers provided by ChatGPT and BARD, the most widely used chatbot applications, when questions related to orthodontics were posed. The null hypothesis was that there is no difference in the quality of the answers provided by both tools.

Materials and methods

Based on the study conducted by Perez-Pino et al.,[6] a power analysis[7] determined that a minimum of 20 questions were required for an 80% power test. After clearing all past searches, Google was interrogated to gather the most frequently asked questions about orthodontic treatment, and the questions from the first 50 pages were analysed individually. Of the popular queries about conventional braces, clear aligners, orthognathic surgery, and orthodontic retainers, a total of 24 questions were selected that had been repeated several times (Table I). The answers given by chatGPT and google BARD to an example question are shown in Table II.

An experienced orthodontist (CA, Investigator 1) and an orthodontic resident (KK, Investigator 2) interrogated the two systems and rated their responses using a five-point modified Likert scale: (5) the system responded with evidence-based information; (4) the system responded with adequate information; (3) the system responded but did not provide adequate information; (2) the system responded with the wrong answer; and (1) the system did not know the response to the question. The Likert scale ratings were designed to reflect the utility and efficiency of the systems, with (1) being the worst rating. The results of the queries were recorded on Microsoft Excel (Microsoft Corp., New Mexico, USA) for comparative evaluation. The number of words and listed references in the answers were also recorded (Tables I).

Statistical analysis

Data were analysed using IBM SPSS V23 (IBM, New York, USA). Compliance with normal distribution was analysed using the Shapiro–Wilk test. The independent Samples *t* test was used to compare the parameters conforming to a normal distribution for paired groups, and the Mann–Whitney *U* test was used to compare the parameters not conforming to a normal distribution. Spearman's rho correlation coefficient was applied to examine the relationship between the data that did not fit the normal distribution. The intraclass correlation coefficient (ICC) was used to analyse the intra-observer and inter-observer reliability. The results of the analyses are presented as the mean ± standard deviation and median (minimum-maximum). The significance level was set at $p < 0.05$.

Results

When inter-observer reliability was analysed, a statistically significant and moderate correlation was found between the scores of the two investigators for ChatGPT (ICC = 0.739; $p = 0.733$). Similarly, a statistically significant and solid correlation was found between the BARD scores (ICC =0.784; $p = 0.733$) (Table III).

An examination of the relationships between the word counts, reference counts, and ChatGPT scores revealed no statistically significant correlation between Investigator 1's ChatGPT scores and the ChatGPT word counts ($p = 0.113$). However, a statistically significant and moderate relationship

Table I. Descriptive statistics of the questions asked in the study

	Questions	Investigator 1 (XX) (Mean ± SD)	Investigator 2 (XX) (Mean ± SD)	Number of words (Mean ± SD)	Number of references (Mean ± SD)
1	What iss the best age to visit the orthodontist?	5 ± 0	5 ± 0	240 ± 41.01	2 ± 1.41
2	Will my braces interfere with my sport activities?	4.5 ± 0.71	4.5 ± 0.71	294 ± 48.08	0.5 ± 0.71
3	What can I eat with braces?	5 ± 0	5 ± 0	270 ± 97.58	2 ± 1.41
4	How do I care for my braces?	5.5 ± 0.71	5 ± 0	363.5 ± 94.05	3 ± 2.83
5	Can an adult do orthodontic treatment?	5 ± 1.41	5 ± 1.41	188 ± 89.1	2 ± 2.83
6	Do the braces interfere with playing an instrument?	4.5 ± 0.71	5.5 ± 0.71	381.5 ± 47.38	2.5 ± 0.71
7	How long does orthodontic treatment take?	5 ± 1.41	5 ± 1.41	285 ± 48.08	1.5 ± 2.12
8	Does orthodontic treatment hurt?	4.5 ± 0.71	4.5 ± 0.71	313.5 ± 53.03	1 ± 1.41
9	How to brush your teeth with braces?	5 ± 0	5 ± 0	378.5 ± 78.49	3.5 ± 0.71
10	Do I have to wear orthodontic retainers and how long will I have to wear my retainer?	4 ± 0	3.5 ± 0.71	316.5 ± 111.02	0 ± 0
11	How often will I have to go to the orthodontist?	4 ± 0	4 ± 0	279.5 ± 38.89	0 ± 0
12	How to get orthodontic treatment without metal braces?	5 ± 1.41	4 ± 1.41	387.5 ± 53.03	4.5 ± 2.12
13	What is orthognathic surgery?	5 ± 0	5.5 ± 0.71	381 ± 86.27	3 ± 0
14	How do braces move teeth in orthodontics?	5.5 ± 0.71	5.5 ± 0.71	361.5 ± 2.12	3 ± 0
15	What are the types of braces?	5 ± 1.41	5 ± 1.41	407.5 ± 30.41	15 ± 2.12
16	What is clear aligner treatment?	5.5 ± 0.71	5.5 ± 0.71	436.5 ± 113.84	4 ± 1.41

Table I. Continued

	Questions	Investigator 1 (XX) (Mean ± SD)	Investigator 2 (XX) (Mean ± SD)	Number of words (Mean ± SD)	Number of references (Mean ± SD)
17	What are the risks and side effects of the orthodontic treatment?	4.5 ± 0.71	4 ± 0	507.5 ± 61.52	0 ± 0
18	What should I do if a bracket comes off or an appliance breaks?	4.5 ± 0.71	5 ± 0	373.5 ± 50.2	3 ± 0
19	What should I do if I have pain or other problems with my braces?	4.5 ± 0.71	4.5 ± 0.71	364 ± 1.41	1 ± 1.41
20	Will wisdom teeth cause teeth to shift?	5.5 ± 0.71	6 ± 0	305.5 ± 82.73	2.5 ± 0.71
21	Which treatment is faster ? Clear aligners or braces?	5.5 ± 0.71	5.5 ± 0.71	289.5 ± 142.13	2 ± 0
22	Can I undergo orthognathic surgery with clear aligners?	5 ± 1.41	5 ± 1.41	252.5 ± 26.16	2 ± 2.83
23	Why do some people need extractions before orthodontic treatment?	5.5 ± 0.71	6 ± 0	327.5 ± 12.02	3 ± 0
24	Can I have orthognathic surgery without orthodontic treatment?	4.5 ± 0.71	4.5 ± 0.71	282 ± 38.18	1.5 ± 2.12

was observed between Investigator 1's ChatGPT scores and the number of ChatGPT listed references (r = 0.685, p < 0.001). Additionally, no statistically significant relationship was found between the number of ChatGPT words and the number of references (p = 0.266) (Table IV). Similarly, Investigator 2's ChatGPT scores were not found to have a statistically significant relationship with the number of ChatGPT words (P = 0.536). However, a statistically significant and close correlation was observed between Investigator 2's ChatGPT scores and the number of ChatGPT references (r = 0.708, p < 0.001), and no statistically significant relationship was found between the number of ChatGPT words and the number of references (p = 0.266) (Table V).

In analysing the relationships between word counts, listed reference counts, and Google BARD scores, no significant association was identified between Investigator 1's BARD score and BARD word counts (p = 0.120). However, a moderate and statistically significant relationship was found between Investigator 1's BARD score and the number of references (r = 0.621; p = 0.001). Additionally, no significant association was observed between the number of BARD words and the number of references (p = 0.889) (Table VI). For Investigator 2, no significant relationship was discovered between the BARD score and the number of BARD words (p = 0.382). Instead, a moderate and statistically significant association was found between

Table II. Question and answer format for ChatGPT and BARD

Answers	Question 1. What's the best age to visit the orthodontist?
ChatGPT	The best age to visit the orthodontist can vary depending on individual circumstances, but there are generally two main stages when people seek orthodontic treatment: 1. Early Treatment (Phase I): This usually occurs between the ages of 7 and 10. At this stage, a child's adult teeth are beginning to come in, and an orthodontist can identify potential issues with tooth and jaw development. Early treatment may be recommended for problems like crowding, crossbites, or severe overbites. 2. Adolescent/Teen Years (Phase II): This is the most common time for orthodontic treatment. Typically, it begins between the ages of 11 and 15. By this age, most of the adult teeth have erupted, and any existing dental issues are more easily identified. This is also a period when the jaw is still growing, which makes it an ideal time for more complex treatments. 3. Adult Treatment: Orthodontic treatment can be done at any age. Many adults seek orthodontic care to address issues that were not treated during childhood or to correct new problems that have developed over time. It's important to note that while there are optimal ages for certain types of treatment, there's no strict rule. Each person's case is unique, and the best time to visit the orthodontist ultimately depends on their specific dental situation and the recommendation of their dentist or orthodontist. Regular dental check-ups are essential, and if a child or adult shows signs of orthodontic issues, it's a good idea to consult with an orthodontist as early as possible. They can provide personalized advice on when to start treatment and what type of treatment is needed.
Google Bard	The American Association of Orthodontists (AAO) recommends that all children have their first orthodontic evaluation by the age of 7. This is because most permanent teeth have erupted by this age, and an orthodontist can identify any potential problems with the development of the jaw and teeth. Early intervention can help to prevent more serious problems down the road, and it can also make orthodontic treatment shorter and easier. For example, if a child has a crossbite (where the upper jaw overlaps the lower jaw), early treatment can help to correct the problem before it becomes more severe. Of course, not all children will need orthodontic treatment. But having an early evaluation can give parents and orthodontists the opportunity to identify any potential problems and develop a plan to address them. It is also important to note that it is never too late to see an orthodontist. Adults can also benefit from orthodontic treatment, and there are many different types of braces and other appliances available that can be tailored to meet the needs of individual patients. If you are concerned about the alignment of your child's teeth or jaws, or if you are an adult who is considering orthodontic treatment, schedule a consultation with an orthodontist to discuss your options.

Investigator 2's BARD score and the number of BARD listed references (r = 0.691; p < 0.001). No significant relationship was noted between the number of BARD words and the number of references (p = 0.889) (Table VII).

A statistically significant difference was found in both investigators' numerical rating scales based on the tools used (p = 0.014 and p = 0.030, respectively). The rank mean value was 29.15 in ChatGPT and 19.85 in BARD for Investigator 1. Similarly, the rank

Table III. Inter-observer reliability scores

	ICC (95% CI)	*p*
ChatGPT	0.739 (0.485–0.878)	<0.001*
BARD	0.784 (0.563–0.9)	<0.001*

ICC, Inter-class coefficient.

*$p < 0.001$.

mean value was 28.63 in the ChatGPT and 20.38 in BARD for Investigator 2. However, there was no statistically significant difference in the mean word count ($p = 0.732$) or number of references ($p = 0.796$) between the two groups (Table VIII).

Discussion

AI is a branch of computer science that aims to replicate human intelligence in computer systems. Recent advancements have enabled the automation of tasks once thought to be exclusively performed by humans.[6] Generative AI is one such implementation area, and chatbots are a prominent example. The question of whether computers can think has long been a topic of interest, and chatbots have been at the centre of this debate.[8] Leading companies such as Google, Meta, DeepMind, and OpenAI are actively developing chatbots. The popularity of chatbots has recently increased following OpenAI's decision to make its language models and user interfaces publicly available through ChatGPT.

ChatGPT is a sophisticated natural language processing model that has been pre-trained on a vast dataset to enable the generatation of language that is coherent, contextually appropriate, and often indistinguishable from human-written text. This model utilises a transformer architecture that combines the advantages of generative models with scalability and employs a self-attention mechanism to learn the context of the dialogue and provide more accurate replies.[9] BARD AI is a cloud-based conversational AI platform powered by LaMDA which is a collection of conversational neural language models developed by Google. This platform

Table IV. An analysis of the relationship between the variables (word and reference counts) with Investigator 1's rating scale for ChatGPT

		Investigator 1's ChatGPT scores (Mean)	Number of words (ChatGPT) (Mean)
Number of words (ChatGPT)	*r*	0.332	
	p	0.113	
Number of references (ChatGPT)	*r*	0.685	0.237
	p	<0.001*	0.266

r: Spearman's rho Correlation Coefficient.

*$p<0.001$.

Table V. An analysis of the relationship between the variables (word and reference counts) with Investigator 2's rating scale for ChatGPT

		Investigator 2's ChatGPT scores (Mean)	Number of words (ChatGPT) (Mean)
Number of words (ChatGPT)	*r*	0.133	
	p	0.536	
Number of references (ChatGPT)	*r*	0.708	0.237
	p	<0.001*	0.266

r: Spearman's rho Correlation Coefficient.

*$p<0.001$.

Table VI. An analysis of the relationship between the variables (word and reference counts) with Investigator 1's rating scale for BARD

		Investigator 1's BARD scores (Mean)	Number of words (BARD) (Mean)
Number of words (BARD)	*r*	-0.326	
	p	0.120	
Number of References (BARD)	*r*	0.621	-0.030
	P	0.001*	0.889

r: Spearman's rho Correlation Coefficient.

*$p = 0.001$.

Table VII. An analysis of the relationship between the variables (word and listed reference counts) with Investigator 2's rating scale for BARD

		Investigator 2's BARD Scores (Mean)	Number of Words (BARD) (Mean)
Number of words (BARD)	*r*	-0.187	
	p	0.382	
Number of references (BARD)	*r*	0.691	-0.030
	P	<0.001*	0.889

r: Spearman's rho Correlation Coefficient.

*$p < 0.001$.

allows organisations to build and deploy chatbots that communicate with consumers through various channels, such as websites, messaging applications, and voice assistants. BARD AI has been used and will continue to be used in a number of fields, including academic writing, healthcare, education, and law, among others.[10] Although the reliability of these environments is widely debated, to date, there are no studies identified that evaluate the reliability of BARD and ChatGPT in the field of orthodontics.

Upon evaluation of the results of the present study, it is apparent that ChatGPT outperforms BARD in providing higher-quality answers to popular orthodontic questions posed by patients. This finding aligns with the those of Rahsepar et al.[10] who assessed the reliability of ChatGPT and BARD in answering common lung cancer questions. Additionally, recent studies suggest that ChatGPT surpasses BARD in answering cancer-related questions and neurosurgery oral board preparation questions, a conclusion that is consistent with the present analysis.[11] The superior performance of ChatGPT can be attributed to the variation in their respective training data and the strength of their models in addressing medical enquiries.[10,11] Previous studies have reported that Google BARD provides answers with more words, whereas ChatGPT offers better quality responses.[10–12] However, in the present study, it was noted that both applications offered answers containing a similar number of words. The higher reliability of ChatGPT was attributed to the provision of answers with more references, as evidence-based information is considered more reliable and acceptable by scientists.

In March 2023, OpenAI introduced the latest version of ChatGPT, GPT 4. Research has demonstrated that GPT 4 outperforms GPT 3.5 in addressing medical queries.[5,13] This superiority may be attributed to GPT 4's increased access to current resources. However, in the present study, GPT 3.5 was chosen due to the belief that free chatbot applications, such as GPT 3.5 and Google BARD, will be more appealing to patients seeking online information about orthodontics compared to paid applications.

Despite the recent and significant advancements in chatbot applications such as ChatGPT and

Table VIII. Comparison of the numerical rating scale, word and reference count values according to the tool used

	Tool					
	ChatGPT		BARD			
	Mean ± SD	Median (Min–Max)	Mean ± SD	Median (Min–Max)	Mann Whitney *U* test	*p*
Investigator 1	5.17 ± 0.76	5 (4–6)	4.63 ± 0.65	5 (4–6)	*U* = 176.500	0.014*
Investigator 2	5.17 ± 0.87	5 (3 - 6)	4.67 ± 0.76	5 (3 - 6)	U = 189.000	0.030*
Number of Words	337.04 ± 82.81	349.5 (125 - 464)	328.46 ± 89.89	321.5 (189 - 551)	t = 0.344	0.732
Number of References	2.13 ± 1.51	2.5 (0 - 6)	1.96 ± 1.76	2.5 (0 - 5)	U = 276.000	0.796

U, Mann–Whitney *U* test; *t*, Independent sample *t* test.

*$p < 0.05$.

BARD, their use in the health field remains limited without professional support. A recent study aimed to compare the performance of orthopaedic residents, ChatGPT 3.5 and GPT 4, on orthopaedic assessment examinations.[14] It was reported that the residents outperformed ChatGPT 3.5 and GPT 4 on the assessment examinations, and it was therefore concluded that it is unlikely that GPT 4 or ChatGPT 3.5 would pass the American Board of Orthopaedic Surgery written examination. While the answers provided by ChatGPT and BARD applications in the present study were generally of good quality, it should be noted that patients posed these questions. Therefore, it is not surprising that more superficial answers were acceptable in the study. The quality of answers provided by chatbot applications, when asked to answer orthodontic questions at the graduate level, remains an open topic for discussion.

Conclusions

- Generally, ChatGPT and BARD provided satisfactory responses to the common orthodontic inquiries posed by patients.
- ChatGPT's answers surpassed those of BARD in quality.

More research is needed to comprehensively assess the capabilities and constraints of these models and determine the most effective methods for their incorporation into clinical practice. While AI language models may become more prevalent, healthcare professionals remain responsible for ultimately making decisions regarding diagnoses, investigations, and treatments.

Limitations

This study endeavours to provide a comprehensive evaluation of the quality of orthodontic information dispensed by Artificial Intelligence (AI) chatbot technologies, specifically focusing on ChatGPT and Google BARD. While the findings offer significant insights into the capabilities of these technologies, the study is subject to limitations that underscore the importance of future research in this area.

The evaluation of AI chatbot responses was conducted with input from two orthodontic professionals. Future studies should aim to include a more diverse group of professionals to enrich the assessment with a broader range of clinical perspectives and expertise. Such an expanded evaluation is crucial for understanding the applicability and accuracy of AI-generated information in orthodontics across different clinical scenarios and patient needs.

Additionally, the present study's focus on ChatGPT and Google BARD, while instrumental, represents only a subset of the AI technologies available. Incorporating a wider array of AI chatbot applications in future research will provide a more holistic view of the landscape of AI in orthodontics. Moreover, the rapid evolution of AI technology and the diverse global contexts in which orthodontic care is provided highlight the need for ongoing research that keeps pace with technological advancements and reflects the global diversity of orthodontic practice.

In summary, despite its contributions, the present study's limitations emphasise the need for continued and expanded research involving more orthodontic professionals and a wider range of AI technologies. Such efforts are essential to fully leverage AI in improving orthodontic care and patient outcomes globally.

Conflict of interest

The authors declare that there is no conflict of interest.

Corresponding author

Kaan Kahya, DDS, Research Assistant, Department of Orthodontics, Faculty of Dentistry, Yeditepe University, Bagdat cad. No:238 Kadikoy/Istanbul 34728 Turkey, Phone: +905399360576 Email: kaan.kahya@yeditepe.edu.tr

References

1. Chen YW, Stanley K, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020;51(3):248–57.
2. Aminoshariae A, Kulild J, Nagendrababu V. Artificial intelligence in endodontics: current applications and future directions. J Endod. 2021;47(9):1352–7.
3. Heo MS, Kim JE, Hwang JJ, Han SS, Kim JS, Yi WJ, et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible? Dentomaxillofacial Radiol. 2020;50(3):20200375–20200375.
4. Carrillo FJG, Carrillo-Perez F, Pecho OE, Morales JC, Paravina RD, Della Bona A, et al. Applications of artificial intelligence in dentistry: a comprehensive review. J Esthet Restor Dent. 2021;33:7–18.

5. Schwendicke F, Samek W, Samek W, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769–74.
6. Perez-Pino A, Yadav S, Upadhyay M, Cardarelli L, Tadinada A. The accuracy of artificial intelligence-based virtual assistants in responding to routinely asked questions about orthodontics. Angle Orthod. 2023;93(4):427–32.
7. Faul F, Erdfelder E, Lang AG, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–91.
8. Rapaport WJ. Book Review: The Turing Test: Verbal Behavior as the Hallmark of Intelligence, edited by Stuart Shieber. 2005
9. Wu C, Yin S, Qi W, Wang X, Tang Z, Duan N. Visual ChatGPT: talking, drawing and editing with visual foundation models. 2023.
10. Rahsepar A, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT versus Google Bard. Radiology. 2023;307:5.
11. Tang OY, Zadnik Sullivan PL, Telfeian AE. Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank. medRxiv. 2023;93(5):1090–8.
12. Patil NS, Huang RS, Van Der Pol CB, Larocque N. Comparative performance of ChatGPT and Bard in a text-based radiology knowledge assessment. Can Assoc Radiol J J Assoc Can Radiol. 2023; 8465371231193716.
13. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. Diagn Interv Imaging. 2023;104(6):269–74.
14. Massey PS, Montgomery C, Zhang AS. Comparison of ChatGPT-3.5, ChatGPT-4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. J Am Acad Orthop Surg. 2023;31:1173–9.