RESEARCH   **Open Access**

Assessment of various artificial intelligence applications in responding to technical questions in endodontic surgery

Sevda Durust Baris[1*] and Kubilay Baris[1]

Abstract

Background The objective of this study was to evaluate the performance of ScholarGPT, ChatGPT-4o and Google Gemini in responding to queries pertaining to endodontic apical surgery, a subject that demands advanced specialist knowledge in endodontics.

Methods A total of 30 questions, including 12 binary and 18 open-ended queries, were formulated based on information on endodontic apical surgery taken from a well-known endodontic book called Cohen's pathways of the pulp (12th edition). The questions were posed by two different researchers using different accounts on the ScholarGPT, ChatGPT-4o and Gemini platforms. The responses were then coded by the researchers and categorised as 'correct', 'incorrect', or 'insufficient'. The Pearson chi-square test was used to assess the relationships between the platforms.

Results A total of 5,400 responses were evaluated. Chi-square analysis revealed statistically significant differences between the accuracy of the responses provided applications ($x^2 = 22.61$; $p < 0.05$). ScholarGPT demonstrated the highest rate of correct responses (97.7%), followed by ChatGPT-4o with 90.1%. Conversely, Gemini exhibited the lowest correct response rate (59.5%) among the applications examined.

Conclusions ScholarGPT performed better overall on questions about endodontic apical surgery than ChatGPT-4o and Gemini. GPT models based on academic databases, such as ScholarGPT, may provide more accurate information about dentistry. However, additional research should be conducted to develop a GPT model that is specifically tailored to the field of endodontics.

Keywords Artificial intelligence, ChatGPT, Endodontic apical surgery, Gemini, ScholarGPT

*Correspondence:
Sevda Durust Baris
svdedrst@hotmail.com
[1]Kırıkkale University, Kırıkkale, Turkey

Introduction

Artificial intelligence (AI), which is at the forefront of modern technological progress, is having a significant impact on our lives today, and its areas of application are expanding. Natural language processing (NLP) is a branch of AI that allows computers to understand and interpret human language [1]. NLP has been instantiated in a special type of AI known as the large language model (LLM) since November 2022. This has revolutionised the way information is searched for and retrieved, as it is capable of creating, translating and summarising human-like text [1, 2]. ChatGPT, developed by OpenAI, is the most well-known and popular LLM today [3, 4].

Gemini, launched by Google DeepMind in 2023, is another LLM [5]. It carries out tasks such as text generation, translation and summarisation as well as creative content production [5, 6]. In May 2024, OpenAI unveiled ChatGPT-4o, the new update to its intelligent chatbot [7]. GPT-4o is a multimodal LLM that combines several different models that understand audio, video and text into a single model [7, 8]. OpenAI's ScholarGPT, which is the product of the development of GPT-4, is a language model that has been trained on academic texts and customised for academic use [9]. ScholarGPT has important advantages, such as the ability to quickly find sources for academic research and to summarise and analyse articles.

Today, the use of AI in healthcare, as in many other fields, is becoming increasingly important. Although scientific publications on AI in medicine have increased significantly in recent years, the literature on AI in dentistry is still limited [3, 4, 10, 11]. In endodontics, AI is used for various purposes, such as detecting the presence of periapical lesions, calculating working lengths, detecting differences in root and canal morphology, detecting root fractures, assessing pain after endodontic treatment and predicting treatment outcomes [11, 12]. Despite all these developments, dentists and researchers should be aware that AI may have ethical limitations such as the potential for discrimination and bias, data privacy, and security issues, as well as technical limitations such as providing incorrect, insufficient or obsolete information [13]. The reliability of AI in providing scientific information on various topics requires further evaluation.

The aim of root canal treatment is to clean the infected pulp tissue in the root canal system, shape and hermetically fill the root canal with a biocompatible material and attempt to prevent reinfection [14]. Retreatment is the most commonly recommended solution to unsuccessful root canal treatment. However, in cases in which retreatment is also unsuccessful or not possible, treatments such as endodontic apical surgery or intentional replantation are recommended [15]. Endodontic apical surgery, also known as apicectomy or apical resection, is a surgical procedure performed to preserve a tooth that has not healed following conventional root canal treatment or retreatment. Its primary goal is to prevent tooth loss by resolving persistent periapical pathology. The procedure involves making an incision, elevating a flap, and creating an osteotomy to access and curette the periapical lesion. This is followed by root-end resection, preparation of a retrograde cavity, and sealing of the cavity with a biocompatible filling material [16].

As AI is being used more and more in medicine and dentistry, it is crucial to evaluate its accuracy and reliability. An examination of the extant literature reveals that few studies have sought to determine the capacity of AI chatbots to furnish precise responses to queries in the domain of endodontics [11, 17–23]. In addition, Balel [24]. , has previously evaluated the performance of ScholarGPT in answering technical questions in the field of oral and maxillofacial surgery by comparing it with ChatGPT. The present study is the first to compare the performance of ScholarGPT with that of different AI robots in providing endodontic information. The aim was to evaluate and compare the accuracy of the answers given by ScholarGPT, Gemini and ChatGPT-4o AI chatbots to questions that dentists may ask about endodontic apical surgery. The null hypothesis of this study was that there would be no significant difference in the accuracy and completeness of information related to endodontic apical surgery provided by ScholarGPT, ChatGPT-4o, and Gemini.

Methods

This study was conducted in accordance with the Declaration of Helsinki. Since the study exclusively evaluated publicly available AI-generated data and did not involve human participants, it was deemed exempt from ethical approval. A total of 30 questions about endodontic apical surgery (Table 1), comprising 12 dichotomous and 18 open-ended queries, were developed. The list of questions is also provided in the supplementary file 1. The questions were designed to guide professionals in endodontic apical surgery. We used a systematic approach to develop the questions to avoid the complexity of answers in artificial intelligence applications. All questions were developed with scientific accuracy and clinical relevance, based on Cohen's Pathways of the Pulp (12th edition) in Chap. 11: Periradicular Surgery [25]. The questions focused on practical aspects of endodontic apical surgery, such as applicability, healing stages, complications, materials used, application technique and patient selection. An endodontist and a periodontist evaluated and contributed to the applicability and scientific nature of the questions.

The questions were posed by two different researchers using different accounts on the ScholarGPT, ChatGPT-4o and Gemini platforms between 25th of Nov 2024 and

Table 1 Questions

1. What are the indications for endodontic apical surgery?
2. What are the contraindications for endodontic apical surgery?
3. It is not necessary to request a cone-beam computerized tomography (CBCT) from the patient prior to endodontic apical surgery. Yes or no?
4. Local anaesthesia provides a haemostatic effect in endodontic apical surgery. Yes or no?
5. Performing root canal resurfacing treatment before endodontic apical surgery makes surgery more successful. Yes or no?
6. Giving patients non-steroidal anti-inflammatory drugs (NSAIDs) before endodontic apical surgery helps control their pain during the procedure. Yes or no?
7. How much of the root tip needs to be cut during endodontic apical surgery?
8. In endodontic apical surgery, root-end resection should be performed parallel to the long axis of the root. Yes or no?
9. What are the benefits of performing a right-angle root resection in endodontic apical surgery?
10. What is the ideal root-end cavity depth in endodontic apical surgery?
11. What are the benefits of using ultrasonic tips to remove the root tip during endodontic apical surgery?
12. What are the disadvantages of using ultrasonic tips to remove the root tip during endodontic apical surgery?
13. Why is it important to remove 3 mm of the root tip during endodontic apical surgery?
14. Why is it important to create a retrograde root-end cavity in endodontic apical surgery?
15. Which retrograde filling materials can be used for endodontic apical surgery?
16. Mineral trioxide aggregate (MTA) is an ideal retrograde filling material for use in endodontic apical surgery. Yes or no?
17. Using MTA and Biodentin as retrograde filling materials in endodontic apical surgery is better than using amalgam and glass ionomer cement. Yes or no?
18. What local haemostatic agents can be used in endodontic apical surgery?
19. What are the ideal suture materials for endodontic apical surgery?
20. What are the different types of flaps used in endodontic apical surgery?
21. What phases does soft tissue healing after endodontic apical surgery involve?
22. Granulation tissue forms in the proliferative phase after endodontic apical surgery. Yes or no?
23. After endodontic apical surgery, fibrous tissues are formed in the maturation phase. Yes or no?
24. After endodontic apical surgery, the first bone formation can be seen after six days. Yes or no?
25. When does the periodontal ligament form after endodontic apical surgery?
26. What are the postoperative recommendations for the patient after endodontic apical surgery?
27. What complications can occur after endodontic apical surgery?
28. Every patient should be prescribed antibiotics after apical endodontic treatment. Yes or no?
29. Asking patients to use chlorhexidine mouthwash after endodontic apical surgery reduces the risk of postoperative infection. Yes or no?
30. What painkillers can be prescribed after endodontic apical surgery?

Table 2 Distribution and comparison of accuracy of responses of AI apps

| AI App	Results			χ^2	p
	Incorrect %	Referral %	Correct %		
Scholar GPT	1.2	1.1	97.7	22.612	
Gemini	19.4	21.1	59.5		**0.000**
Chat GPT 4.0	2.9	7.1	90.1		

4th of Dec 2024. The questions were asked 3 times a day – in the morning, afternoon and evening – for 10 days. A new conversation option was chosen each time to minimise the influence of previous answers. Thus, a total of 60 answers were obtained for each question. The responses were then coded by the researchers, who categorised them as 'correct', 'incorrect' or 'insufficient'. Each answer was then compared with the correct answers as provided in the reference book, Cohen's Pathways of the Pulp in Chap. 11: Periradicular Surgery. The responses were carefully documented in an Excel spreadsheet (Microsoft, Redmond, WA). The distribution of the responses was analysed. Since the chi-square test is a statistical method used to evaluate the significance of the difference between categorical data, the Pearson chi-square test was used to evaluate the relationships between platforms. Inter-rater agreement was assessed through Cohen's Kappa Test.

Results

A total of 5,400 responses were evaluated – 1,800 from each of the AI applications. The chi-square analysis revealed statistically significant differences between the accuracy of the responses provided by applications ($\chi^2 = 22.61$; $p < 0.05$) (Table 2). The analysis demonstrated that 90.1% of ChatGPT-4o's responses were accurate, 2.9% were erroneous and 7.1% were incomplete. On the other hand, Gemini's responses were 59.5% accurate, 19.4% erroneous and 21.1% incomplete. In contrast, ScholarGPT demonstrated superior performance compared to the other two applications, as its responses were 97.7% correct, 1.2% incorrect and 1.1% insufficient (Fig. 1). The weighted kappa value for inter-rater agreement was 0.85.

Discussion

Substantial evidence suggests AI has undergone rapid development in recent years and will become a widely used tool in modern dentistry in the near future [19, 26, 27]. It is imperative to acknowledge that the use of AI in dentistry is still in its developmental stage, and the benefits it offers vary according to the particular use case and application. The use of non-expert educational data, the potential risks associated with the use of outdated information, and the ethical and legal concerns surrounding patient confidentiality require careful consideration [28].
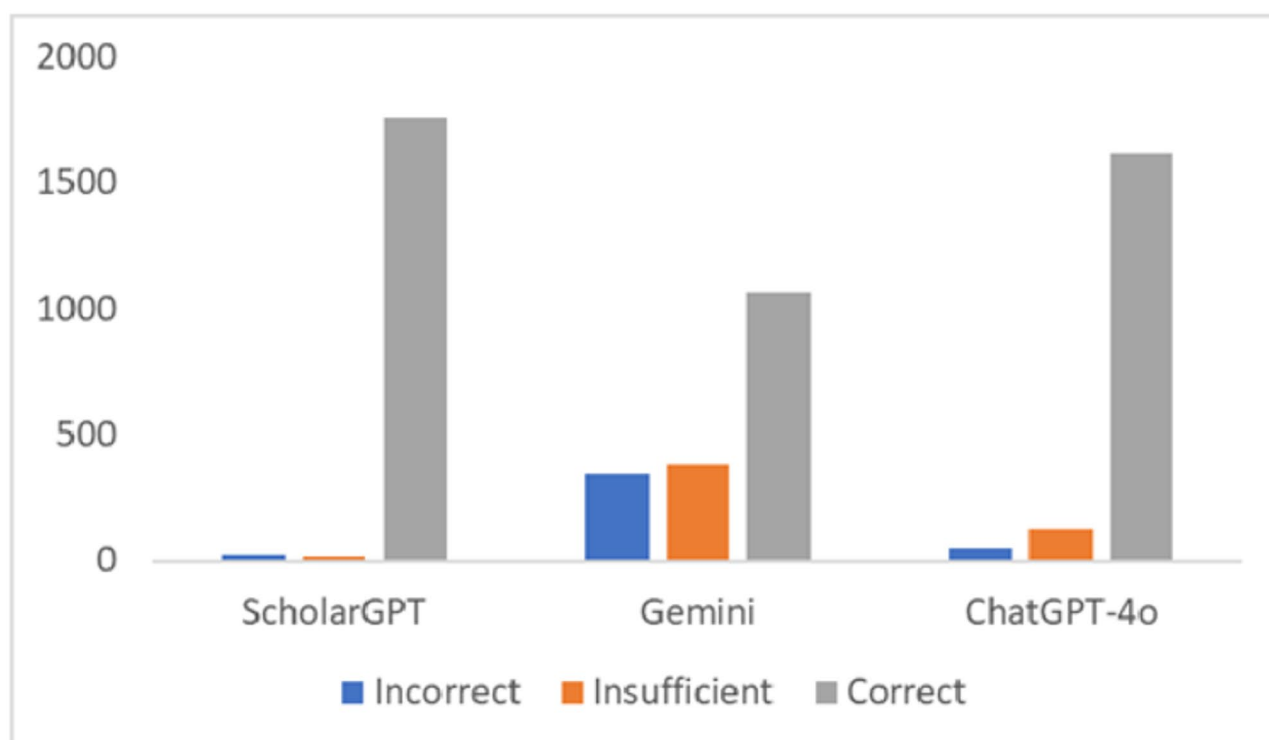
Fig. 1 Distribution of answers produced by the AI applications

The most prevalent and well-regarded AI products in contemporary use are those that fall under the category of language models, which use NLP algorithms [4]. Language models, including prominent examples such as ChatGPT, Gemini and Meta LLaMA, provide users with the benefit of AI access without the need for advanced technological expertise [29]. Nevertheless, while multimodal LLMs have made considerable progress in various domains, further research is necessary due to their current limitations, particularly in medical and dental research [30]. Therefore, the present study evaluated the capacity of the ChatGPT-4o, Gemini and ScholarGPT platforms to address enquiries concerning endodontic apical surgery, which requires advanced specialist knowledge in endodontics and represents a challenging subject to manage clinically and theoretically.

ChatGPT-4o and Gemini were chosen for this study primarily because their multimodal structures facilitate the management of health problems and because they are the most widely used and easily accessible AI chatbots today. ScholarGPT is an AI model developed for academic and scientific use [9]. It performs functions such as analysing and summarising articles, producing texts that comply with conventions of academic language and providing field-specific academic information. In this study, we evaluated the accuracy rate of ScholarGPT's responses by comparing them with the AI applications Gemini and ChatGPT-4o. To our knowledge, this is the first academic

study to evaluate the performance of ScholarGPT in the field of endodontics.

It has been argued that the acceptable limit of accuracy for artificial intelligence applications should be above 90%, in order to ensure safety and efficacy [31, 32]. In this study, the answers given by the AI robots were coded as 'insufficient' if they were not completely correct or not completely incorrect according to Cohen's pathways of the pulp, one of the most important resources in the field of endodontics. According to our findings, ScholarGPT exceeded the accuracy threshold and achieved the highest correct response rate, with 97.7% correct responses, while ChatGPT-4o followed with a 90.1% correct response rate. Gemini exhibited the lowest correct response rate by far (59.5%) among the chatbots examined. Therefore, the null hypothesis was rejected.

In the present study, ChatGPT-4o demonstrated superior accuracy and performance in comparison to Gemini. This finding is consistent with a study by Doshi et al. [33], who compared Gemini and ChatGPT in the domain of radiology. Quah et al. [34] evaluated the accuracy of the GPT-4, GPT-3.5, Llama 2, Gemini and Copilot chatbots in answering multiple-choice questions in the field of oral and maxillofacial surgery. Their study reported that GPT-4 demonstrated the highest performance with 76.8%, followed by Copilot with 72.6%, GPT-3.5 with 62.2%, Gemini with 58.7% and Llama 2 with 42.5%. In addition, Ekmekci and Durmazpinar [19] evaluated the

accuracy of responses by the Gemini, ChatGPT-4o and ChatGPT-4 chatbots with PDF plugins to questions posed by dentists about regenerative endodontic treatment. The findings revealed that ChatGPT-4 with a PDF plugin exhibited the highest accuracy, with a correct response rate of 98.1%, while ChatGPT-4o demonstrated an accuracy rate of 86.2%, and Gemini exhibited the lowest accuracy at 48%. These findings support the results of our study.

In their study investigating the accuracy and consistency of ChatGPT's responses to three different levels of dichotomous (yes/no) endodontic questions posed to ChatGPT and human experts, Suarez et al. [18] found that ChatGPT's responses exhibited an accuracy rate of 57.33%, and that the accuracy varied significantly with the question difficulty. Ozden et al. [21] asked ChatGPT and Gemini dichotomous (yes/no) questions about dental trauma over 10 days. They found that both applications gave the right answer to 57.5% of the questions. We believe that the higher accuracy rate of ChatGPT in our study compared to these two studies is due to the fact that the version used in our study (ChatGPT-4o) has a more advanced database.

In a study [25] evaluating Gemini's performance in answering questions about diagnosing and treating dental problems in endodontics, Gemini's responses were reported to be accurate 37.11% of the time. This level of accuracy is evidently inadequate. Furthermore, it is imperative to acknowledge that medical information provided by ChatGPT and Gemini does not constitute academic knowledge. The information provided should not be regarded as a substitute for medical advice.

In the inaugural study [24] conducted to appraise the performance of ScholarGPT in the domain of healthcare, Balel used a modified Global Quality Scale to evaluate the responses of ScholarGPT. The study reported that ScholarGPT exhibited strong performance compared to ChatGPT in addressing technical inquiries related to oral and maxillofacial surgery. These findings are consistent with the results of the present study. Based on these results, we can say that GPT models developed on the basis of academic databases can provide more accurate and reliable information. As no other studies have evaluated the performance of ScholarGPT in medicine and dentistry, a comparison of the present findings with the results of previous studies is not possible.

ScholarGPT is capable of retrieving information from a variety of databases, including Google Scholar, PubMed and Arxiv [9]. The insufficient or incorrect results in the responses of ScholarGPT may be due to the fact that only the abstracts, and not the full texts, of some articles in these databases are used [24]. Another reason for this situation may be that ScholarGPT does not have full access to the most important and popular databases in medical science, such as Elsevier, Wiley, Nature, Oxford Academic, Scopus & Web of Science and Cambridge University Press which are reliable sources of evidence-based information. The development of an academic GPT model tailored to the field of endodontics would be a significant advancement. However, it is imperative to consider the potential legal and ethical implications of such a model [23, 27].

In the present study, the accuracy of AI applications was evaluated through the use of both open-ended and dichotomous (yes/no) questions. The decision to not employ only a 'yes/no' format for the questions was motivated by the need to mirror the multidimensional nature of clinical practice [35]. This is a significant strength of our study. Although temporal variation was overcome by having two different researchers simultaneously ask the same questions 3 times a day for 10 days, this study does have some limitations. First, the study appraised the performance of three different AI applications in the context of endodontic apical surgery. A plethora of endodontic topics was not evaluated. Subsequent studies encompassing a more extensive range of subjects and incorporating a greater number of questions are needed to evaluate the performance of AI applications in the domain of endodontics. A further limitation of the current study is that the responses provided by the AI applications to the questions were not compared with the responses of general dentists and endodontists. A comparison of answers provided by AI applications with those offered by general dentists and endodontists would provide valuable information about the performance of AI applications. Further research in this area is recommended. The AI applications whose accuracy performance was evaluated in the current study are chatbots for the general audience, not specifically trained in the field of endodontics. Therefore, there may be certain biases in the responses. This is another limitation of our study.

In conclusion, ScholarGPT demonstrated strong performance in the domain of endodontic apical surgery, exhibiting a higher accuracy rate than ChatGPT-4o and Gemini. However, none of the three applications are entirely secure, necessitating caution during their use. These applications should be regarded as an adjunct to clinical knowledge and experience. GPT models based on academic databases have the potential to provide more accurate and reliable information in the medical field. In addition, the development of a dedicated GPT model for the field of endodontics with full access to the most important and popular databases in medical science could provide higher quality and more accurate informations.

Supplementary Information

Supplementary Material 1

Author contributions
Design of the work; S.D.B., K.B., acquisition; S.D.B., K.B., interpretation of data; S.D.B., drafted the work; S.D.B., revised; K.B. All the authors have read and approved the final version of the manuscript.

Data availability
Availability of data and materialThe datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Abbreviations
AI Artificial intelligence.
CBCT Cone-beam computerized tomography.
ChatGPT Chat Generative Pre-Trained Transformer.
GPT Generative Pre-trained Transformer.
LLMs Large Language Models.
LLM Large Language Model.
MTA Mineral trioxide aggregate.
NSAIDs Non-steroidal anti-inflammatory drugs.
NLP Natural language processing.

Ethics approval
and consent to partcipate.
Not applicable.

References
1. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, Scales N, Tanwani A, Cole-Lewis H, Pfohl S, et al. Large Language models encode clinical knowledge. Nature. 2023;620(7972):172–80. https://doi.org/10.1038/s41586-023-06291-2.
2. Park YJ, Pillai A, Deng J, Guo E, Gupta M, Paget M, Naugler C. Assessing the research landscape and clinical utility of large Language models: a scoping review. BMC Med Inf Decis Mak. 2024;24(1):72. https://doi.org/10.1186/s12911-024-02459-6.
3. Silva TP, Andrade-Bortoletto MFS, Ocampo TSC, Alencar-Palha C, Bornstein MM, Oliveira-Santos C, Oliveira ML. Performance of a commercially available generative pre-trained transformer (GPT) in describing radiolucent lesions in panoramic radiographs and Establishing differential diagnoses. Clin Oral Investig. 2024;28(3):204. https://doi.org/10.1007/s00784-024-05587-5.
4. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? J Stomatol Oral Maxillofac Surg. 2023;124:101471. https://doi.org/10.1016/j.jormas.2023.101471.
5. Team G, Anil R, Borgeaud S, Wu Y, Alayrac JB, Yu J, et al. Gemini: a family of highly capable multimodal models. ArXiv Preprint. 2023. https://doi.org/10.48550/arXiv.2312.11805. arXiv:2312.11805.
6. Saab K, Tu T, Weng WH, Tanno R, Stutz D, Wulczyn E, et al. Capabilities of gemini models in medicine. ArXiv Preprint. 2024. https://doi.org/10.48550/arXiv.2404.18416. arXiv:2404.18416.
7. Kerner SM. GPT-4o explained: everything you need to know. 2024. https://www.techtarget.com/whatis/feature/GPT-4o-explained-Everything-you-need-to-know. Accessed 26 May 2024.
8. Doyle K. 'The o is for omni' and other things you should know about GPT-4o. https://www.jasper.ai/blog/what-is-gpt-4o. Accessed 26 May 2024.
9. OpenAI. ScholarGPT. 2024.
10. Khanagar SB, Al-Ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al. Developments, application, and performance of artificial intelligence in dentistry–a systematic review. J Dent Sci. 2021;16(1):508–22. https://doi.org/10.1016/j.jds.2020.06.019.
11. Mohammad-Rahimi H, Dianat O, Abbasi R, et al. Artificial intelligence for detection of external cervical resorption using Label-Efficient Self-Supervised learning method. J Endod. 2024;50(2):144–e1532. https://doi.org/10.1016/j.joen.2023.11.004.
12. Aminoshariae A, Kulild J, Nagendrababu V. Artificial intelligence in endodontics: current applications and future directions. J Endod. 2021;47(9):1352–7. https://doi.org/10.1016/j.joen.2021.06.003.
13. Umer F, Khan M. A call to action: concerns related to artifcial intelligence. Oral Surg Oral Med Oral Pathol Oral Radiol. 2021;132(2):255. https://doi.org/10.1016/j.oooo.2021.04.056.
14. Siqueira Junior JF, Rôças IDN, Marceliano-Alves MF, Pérez AR, Ricucci D. Unprepared root Canal surface areas: causes, clinical implications, and therapeutic strategies. Braz Oral Res. 2018;32:65. https://doi.org/10.1590/1807-3107bor-2018.vol32.0065.
15. Ricucci D, Siqueira JF Jr. Biofilms and apical periodontitis: study of prevalence and association with clinical and histopathologic findings. J Endod. 2010;36(8):1277–88. https://doi.org/10.1016/j.joen.2010.04.007.
16. Lieblich SE. Current concepts of periapical surgery: 2020 update. Oral Maxillofac Surg Clin North Am. 2020;32(4):571–. https://doi.org/10.1016/j.coms.2020.07.0071. 82.
17. Mohammad-Rahimi H, Ourang SA, Pourhoseingholi MA, Dianat O, Dummer PMH, Nosrat A. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. Int Endod J. 2024;57(3):305–14. https://doi.org/10.1111/iej.14014.
18. Suárez A, Díaz-Flores García V, Algar J, Gómez Sánchez M, Llorente de Pedro M, Freire Y. Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers. Int Endod J. 2024;57(1):108–13. https://doi.org/10.1111/iej.13985.
19. Ekmekci E, Durmazpinar PM. Evaluation of different artificial intelligence applications in responding to regenerative endodontic procedures. BMC Oral Health. 2025;25(1):53. https://doi.org/10.1186/s12903-025-05424-5.
20. Portilla ND, Garcia-Font M, Nagendrababu V, Abbott PV, Sanchez JAG, Abella F. Accuracy and consistency of gemini responses regarding the management of traumatized permanent teeth. Dent Traumatol. 2024. https://doi.org/10.1111/edt.13004. Advance online publication.
21. Ozden I, Gokyar M, Ozden ME, Sazak Ovecoglu H. Assessment of artificial intelligence applications in responding to dental trauma. Dent Traumatology: Official Publication Int Association Dent Traumatol. 2024;40(6):722–9. https://doi.org/10.1111/edt.12965.
22. Künzle P, Paris S. Performance of large Language artificial intelligence models on solving restorative dentistry and endodontics student assessments. Clin Oral Investig. 2024;28(11):575. https://doi.org/10.1007/s00784-024-05968-w.
23. Maltarollo TFH, Strazzi-Sahyon HB, Amaral RR, Sivieri-Araújo G. Is the field of endodontics prepared to utilise ChatGPT? Aust Endod J. 2024;50(1):176–7. https://doi.org/10.1111/aej.12821.
24. Balel Y. ScholarGPT's performance in oral and maxillofacial surgery. J Stomatol Oral Maxillofac Surg. 2024;126(4):102114. https://doi.org/10.1016/j.jormas.2024.102114. Advance online publication.
25. Hargreaves KM, Cohen S. Cohen's pathways of the pulp. Elsevier; 2021.
26. Díaz-Flores García V, Freire Y, Tortosa M, Tejedor B, Estevez R, Suárez A. Google Gemini's performance in endodontics: a study on answer precision and reliability. Appl Sci. 2024;14(15):6390. https://doi.org/10.3390/app14156390.
27. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci. 2023;3(4):100324. https://doi.org/10.1016/j.xops.2023.100324.

28. Gromova EA, Ferreira DB, Begishev IR. ChatGPT and other intelligent Chatbots: legthical and dispute resolution concerns. Revista Brasileira de Alternative Dispute Resolution-Brazilian Journal of Alternative Dispute Resolution-RBADR. 2023;5(10):153–75. https://doi.org/10.52028/rbadr.v5i10.ART07.RU

29. Shervin M, Mikolov T, Nikzad N, Chenaghlu MA, Socher R, Amatriain X, Gao. J. Large Language models: a survey. ArXiv Preprint arXiv:2402. 2024;06196:1–43. https://doi.org/10.48550/arXiv.2402.06196.

30. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large Language model. Int J Oral Sci. 2023;15(1):29. https://doi.org/10.1038/s41368-023-00239-y.

31. Labkoff S, Oladimeji B, Kannry J, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inf Assoc. 2024;31(11):2730–39. https://doi.org/10.1093/jamia/ocae209.

32. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368. https://doi.org/10.1136/bmj.l6927.

33. Doshi R, Amin K, Khosla P, Bajaj S, Chheang S, Forman HP. Utilizing large Language models to simplify radiology reports: a comparative analysis of ChatGPT3. 5, ChatGPT4. 0, Google bard, and Microsoft Bing. medRxiv. 2023;2023-06. https://doi.org/10.1101/2023.06.04.23290786

34. Quah B, Yong CW, Lai CWM, Islam I. Performance of large Language models in oral and maxillofacial surgery examinations. Int J Oral Maxillofac Surg. 2024;53(10):881–6. https://doi.org/10.1016/j.ijom.2024.06.003.

35. Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. Front Digit Health. 2021;3:645232. https://doi.org/10.3389/fdgth.2021.645232.

Publisher's note

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