

RESEARCH

Open Access



# Evaluation of the performance of large language models in clinical decision-making in endodontics

Yağız Özbay<sup>1\*</sup>, Deniz Erdoğan<sup>2</sup> and Gözde Akbal Dinçer<sup>3</sup>

## Abstract

**Background** Artificial intelligence (AI) chatbots are excellent at generating language. The growing use of generative AI large language models (LLMs) in healthcare and dentistry, including endodontics, raises questions about their accuracy. The potential of LLMs to assist clinicians' decision-making processes in endodontics is worth evaluating. This study aims to comparatively evaluate the answers provided by Google Bard, ChatGPT-3.5, and ChatGPT-4 to clinically relevant questions from the field of Endodontics.

**Methods** 40 open-ended questions covering different areas of endodontics were prepared and were introduced to Google Bard, ChatGPT-3.5, and ChatGPT-4. Validity of the questions was evaluated using the Lawshe Content Validity Index. Two experienced endodontists, blinded to the chatbots, evaluated the answers using a 3-point Likert scale. All responses deemed to contain factually wrong information were noted and a misinformation rate for each LLM was calculated (number of answers containing wrong information/total number of questions). The One-way analysis of variance and Post Hoc Tukey test were used to analyze the data and significance was considered to be  $p < 0.05$ .

**Results** ChatGPT-4 demonstrated the highest score and the lowest misinformation rate ( $P = 0.008$ ) followed by ChatGPT-3.5 and Google Bard respectively. The difference between ChatGPT-4 and Google Bard was statistically significant ( $P = 0.004$ ).

**Conclusion** ChatGPT-4 provided more accurate and informative information in endodontics. However, all LLMs produced varying levels of incomplete or incorrect answers.

**Keywords** Chat GPT, Chatbot, Large Language model, Endodontics, Endodontology

\*Correspondence:

Yağız Özbay

yagiz\_ozbay@hotmail.com

<sup>1</sup>Department of Endodontics, Faculty of Dentistry, Karabük University, Karabük, Türkiye

<sup>2</sup>Private Dentist, Ankara, Türkiye

<sup>3</sup>Department of Endodontics, Faculty of Dentistry, Okan University, İstanbul, Türkiye



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

The term “artificial intelligence” (AI) was introduced in the 1950s and refers to the concept of creating machines that can perform tasks typically carried out by humans [1]. AI employs algorithms to simulate intelligent actions with little human input. Predominantly reliant on machine learning, AI's scope includes data retrieval and analysis of datasets and images [2, 3]. AI has been growing rapidly in dental practices over the past few years, enabling computers to handle tasks previously requiring human involvement and enabling healthcare professionals to provide better oral health care.

Large language models (LLMs) are AI-based software that simulates human language processing abilities such as understanding the meaning of a phrase or responding and creating new content after being trained with massive datasets. Thus, an LLM can generate an article on any subject, answer a question, or translate a text after being accordingly instructed [4]. In dentistry, LLMs can contribute to research, clinical decision-making, and personalized patient care, enhancing efficiency and reducing errors [1]. They can assist dentists in diagnosing oral diseases by analyzing patient records, medical histories, and imaging reports. They also support treatment planning by synthesizing vast amounts of scientific literature and providing evidence-based recommendations [5]. As these models continue to evolve, they hold significant potential to transform healthcare delivery and improve patient outcomes.

Released in November 2022, Chatbot Generative Pre-trained Transformer (ChatGPT) (OpenAI, San Francisco, CA, USA), is an LLM that can answer questions quickly and fluently and interact with the user in a way that mimics human communication [6]. ChatGPT can write code or articles, summarise text, remember the previous user's input and response in the thread, and elaborate its answers with further queries [7, 8]. ChatGPT is shown to have a basic understanding of oral potentially malignant disorders with potential limitations such as inaccurate content and references [9]. Currently, there is an older version, ChatGPT-3.5 available for free and the latest version ChatGPT-4 which is subject to a paid subscription. A recent study by Suárez et al. showed that ChatGPT-3.5 can consistently answer dichotomous questions, but is not able to give correct answers to all questions [10]. Unlike its previous model, ChatGPT-4 can also retrieve information from the internet. On the other hand, The Google Bard chatbot (Alphabet, Mountain View, California), was launched in March 2023 as a potential competitor, can also complete language-related tasks and answer questions with detailed information [11]. Another study demonstrated that Google Bard holds the potential to aid evidence-based dentistry; nevertheless, it gives

occasional wrong answers or no answers despite its live access to the internet [12].

There are studies examining the ability of LLMs to answer multiple choice questions and the ability to answer open-ended questions in various clinical scenarios [10, 13–16]. It has been observed that success varies depending on the LLM used, field and type of question. Unlike previous ones, the present study focuses on the ability of LLMs to provide scientific knowledge in endodontics.

Endodontics is a branch of dentistry that requires clinicians to have up-to-date knowledge in many areas such as pathologies, stages of root canal treatment, vital pulp treatments, dental trauma, etc. The application of LLMs in endodontics has the potential to enhance diagnostic accuracy, treatment planning, and patient management. These AI-driven models can analyze vast amounts of endodontic literature, clinical guidelines, and patient records to assist clinicians in identifying complex root canal pathologies and suggesting evidence-based treatment approaches [17]. Moreover, LLMs can aid in the interpretation of radiographic images by integrating natural language processing (NLP) with deep learning techniques, improving the detection of periapical lesions and root fractures [18].

In one study, periapical X-rays were analyzed by ChatGPT and its performance was evaluated based on its ability to accurately identify a variety of dental conditions, including tooth decay, endodontic treatments, dental restorations, and other oral health problems. Endo-oral lesions were missed in 56% of cases. The AI's overall correct interpretation rate was 11%, indicating limited clinical utility in its current form. Misinterpretations included incorrect tooth identification and failure to recognize certain dental lesions. Although ChatGPT is promising in endodontics, it has been reported that it has significant gaps in its diagnostic accuracy [19].

The potential of LLMs to assist clinicians in decision-making in the endodontic clinic is worth evaluating so that, if adequate, they can be utilized and, if not, so that clinicians are aware of the limitations that exist and can contribute to improving the deficiencies of these language models. To the best of the authors' knowledge, there are no prior studies that compare LLMs such as ChatGPT and Google Bard as potential sources of information in Endodontics. Therefore, this study aimed to compare the accuracy and completeness of the answers generated by ChatGPT-3.5, ChatGPT-4, and Google Bard to open-ended questions related to Endodontics.

## Materials and methods

The study was conducted under the Declaration of Helsinki. Since our study had no human subjects involved, ethical approval was not required. 57 open-ended

questions covering different areas of endodontics within the content of the Guidelines and Position Statements of the European Society of Endodontology, and the American Association of Endodontists were prepared (Table 1) [20–24]. These articles were chosen by two endodontists (Y.Ö. and G.A.D.) since articles represent a consensus achieved by expert committees in the field. The validity of the questions was evaluated by 8 volunteer endodontists using the Lawshe's Content Validity Index, a widely accepted method [25].

The experts rated each question as “essential,” “useful but not essential,” or “not essential.” The critical content

validity ratio value obtained according to Wilson et al., is 0.69 for 8 volunteers at a significance level of 0.05 [25]. Hence, questions with a content validity ratio value of  $\geq 0.69$  were selected for inclusion. As a result, 40 open-ended, clinically relevant questions that require text-based responses were included. The sample size was calculated at the significance level of 0.05, effect size of 0.3 and power of 0.85 using G\* Power v3.1 (Heinrich Heine, Universität Düsseldorf, Germany). Questions aimed to measure broad endodontic knowledge, such as endodontic indications, antibiotic use, root canal treatment procedures, symptoms of acute apical abscesses,

**Table 1** Questions directed to the LLMs

Questions	
1.	What are the indications for root canal treatment?
2.	What are the contra-indications for root canal treatment?
3.	What are the indications for nonsurgical root canal retreatment?
4.	What are the considerations in the preparation of the access cavity?
5.	Which pulp tests are applied to assess the health of dental pulp?
6.	Why is rubber dam used in Endodontics?
7.	How is working length determined during root canal treatment?
8.	What are the considerations in the obturation of root canals?
9.	What are the considerations in the apical preparation of root canals?
10.	What are the considerations in the delivery of irrigants in root canals?
11.	How is an acute apical abscess with systemic involvement endodontically managed?
12.	How is the glide path in the root canal created?
13.	What are the preoperative and intraoperative considerations of intentional replantation of the tooth?
14.	What are the clinical symptoms of reversible pulpitis?
15.	What are the clinical symptoms of irreversible pulpitis?
16.	When and how is direct pulp capping performed?
17.	When and how is full pulpotomy performed?
18.	Which materials are used in vital pulp therapies?
19.	How is lateral luxation of permanent tooth diagnosed and treated?
20.	How is avulsion of permanent tooth diagnosed and treated?
21.	How is a complicated crown fracture of permanent teeth diagnosed and treated?
22.	How is a horizontal root fracture of a permanent tooth diagnosed and treated?
23.	What are the clinical features of external cervical resorption?
24.	Which intracanal medicaments are currently used in Endodontics?
25.	What are the applications of calcium hydroxide in Endodontics?
26.	How is internal resorption of the tooth treated?
27.	What is a cracked tooth?
28.	What are the indications for periradicular surgery?
29.	What are the criteria for the use of CBCT in Endodontics?
30.	What are the indications for systemic antibiotics in Endodontics?
31.	What are the contra-indications for systemic antibiotics in Endodontics?
32.	What are the clinical symptoms of chronic apical abscesses?
33.	What are the clinical steps of the revitalization of teeth?
34.	What are the clinical symptoms of acute apical abscesses?
35.	What are the radiographic features of condensing osteitis?
36.	When using engine-driven NiTi files, which principles should be followed?
37.	How is inadvertent extrusion of sodium hypochlorite beyond the end of the root into the periradicular tissues managed?
38.	Which clinical and radiological findings indicate a favorable outcome after root canal treatment?
39.	Which local anesthetics are commonly used for pain control before root canal treatment?
40.	Which irrigants are currently used in Endodontics?

and common conditions that may cause concern during endodontic treatment, were presented. All questions and responses were in English.

The questions were introduced into Google Bard, ChatGPT-3.5, and ChatGPT-4 using the “new chat” option on 20th December 2023. Each question was presented only once to each LLM, and there was no follow-up, rephrasing, or annotation reflecting real-world situations for general dentists. Examples of answers to the questions are presented in Figs. 1, 2 and 3. All answers were recorded for subsequent analysis. Two endodontists, (Y.Ö. and D.E.) with ten years of experience [26], blindly evaluated the answers using a 3-point (incorrect, partially correct/incorrect or correct). The Likert scale, which is very popular in social research and has a simple measurement process, was used [27]. Likert scale adapted from Suárez et al. (Table 2) [14]. When disagreements arose in some evaluations, these were resolved by the intervention of a third researcher. Additionally, all responses deemed to contain factually wrong information were noted and a misinformation rate for each LLM was calculated (number of answers containing wrong information/total number of questions). Scores were saved in an Excel® spreadsheet (Microsoft, Redmond, Washington, USA).

### Statistical analysis

The statistical analysis was performed via MiniTab 17 (Minitab Inc., USA). The normality of the data was assessed with Ryan-Joiner test and the normal distribution of the data was confirmed. One-way analysis of variance (ANOVA) and Post Hoc Tukey test were performed. A significance level of  $P < 0.05$  was used to determine statistical significance.

### Results

The mean values and standard deviations of each group are presented in Table 3; Fig. 4. ChatGPT-4 demonstrated the highest score and the lowest misinformation rate followed by ChatGPT-3.5 and Google Bard respectively. The difference between ChatGPT-4 and Google Bard was statistically significant ( $P = 0.004$ ). However, there was no statistically significant difference between ChatGPT-3.5 and ChatGPT-4, and ChatGPT-3.5 and Google Bard.

### Discussion

Interest in LLMs has increased in studies in the field of dentistry as in other fields. Some of these studies have examined the ability of LLMs to answer multiple-choice questions, and some studies have examined the ability to answer open-ended questions in various clinical scenarios [12, 14–16, 28]. It has been observed that success varies according to the LLM used, the field and the type

of question. This study rather focuses on ability of LLMs to provide scientific knowledge in endodontics.

One of the important issues to consider when using chatbots as a source of medical information is the formulation of questions, which significantly influences chatbot responses [29]. Open-ended questions better capture the nuances of the medical decision-making process [30]. The limitation of this study is that the questions are narrow in scope and only related to the field of endodontics. Conversational LLMs perform well in summarizing health-related texts, answering general questions in health care, and collecting information from patients [31].

It was reported that ChatGPT-3 was generally effective and could be used as an aid when an oral radiologist needed additional information about pathologies. However, it was also emphasized that it could not be used as a main reference source [32]. In a diagnostic accuracy study, ChatGPT had the most difficulty recognizing dental lesions. In some cases, images incorrectly identified by ChatGPT as tooth decay were reported to be actually caused by overlapping tooth ridges, and were found to be inadequate for detecting bone loss associated with periodontal disease. It was also shown that when interpreting some radiographs, it did not properly recognise ceramic crowns placed on teeth that were slightly radiopaque [19]. On the other hand, in a study comparing the answers given by AI and humans in the exam, ChatGPT-3 and ChatGPT 4 were found to outperform humans, and were reported that the newer version performed better [33]. AI has the potential to provide significant practicality in clinical applications such as treatment recommendations in the field of endodontics, and more scientific studies are needed to evaluate its possible benefits and risks in a balanced manner.

Similar to a study evaluating the ability of ChatGPT to be a potential source of information for patients' questions about periodontal diseases [34], it was also observed that all LLMs suggested to seek professional help from a dentist or medical professional especially when it comes to questions related to diagnosis, diseases, and treatment indications. A comprehensive review reported that conversational LLMs showed promising results in summarizing and providing general medical information to patients with relatively high accuracy. However, it was noted that conversational LLMs such as ChatGPT cannot always provide reliable answers to complex health-related tasks that require expertise (e.g., diagnosis). Although bias or confidentiality issues are often cited as concerns, no research has yet been found in articles that carefully examine how conversational LLMs lead to these problems in health services research. Furthermore, various ethical concerns have been reported regarding the application of LLMs in human health care, including

### Why is rubber dam used in Endodontics?

The rubber dam serves several critical functions in endodontics, making it a valuable tool for both dentists and patients. Here are its main benefits:

- 1. Sterile environment:** Endodontic procedures involve cleaning and disinfecting the inside of the tooth, where bacteria can easily infiltrate and compromise the success of the treatment. The rubber dam acts as a barrier, isolating the tooth from saliva and oral bacteria, thereby creating a sterile environment essential for successful root canal treatment.
- 2. Improved visibility:** Saliva, blood, and tissue debris can obscure the dentist's view of the tooth during treatment. The rubber dam keeps the operating field clear and dry, allowing the dentist to see the tooth structure and canals more precisely, leading to more accurate and efficient treatment.
- 3. Enhanced control and access:** The rubber dam holds the tongue and cheeks away from the treatment area, providing the dentist with better control over instruments and improved access to the tooth. This translates to smoother, more efficient procedures and minimizes the risk of accidental instrument contact with surrounding tissues.
- 4. Patient safety:** During endodontic procedures, small instruments and debris can potentially be swallowed or inhaled. The rubber dam acts as a safety net, catching any dislodged materials and preventing accidental ingestion or aspiration, enhancing patient safety and comfort.
- 5. Improved success rate:** Studies have shown that using a rubber dam during endodontics significantly improves the success rate of root canal treatments. This is primarily due to the improved visibility, control, and sterile environment it provides.
- 6. Reduced risk of contamination:** Endodontic procedures involve using potentially irritating or harmful chemicals like disinfectants. The rubber dam protects the surrounding tissues and the patient's airway from exposure to these materials, enhancing patient comfort and safety.

While some dentists may choose alternative methods for isolation in certain situations, the rubber dam remains the gold standard in endodontics due to its comprehensive benefits for both patient and practitioner.

Additionally:

- Using a rubber dam can also make the procedure more comfortable for the patient by reducing gagging and tongue fatigue.
- It can help conserve water and other materials used during the procedure.

Overall, the rubber dam is a crucial tool in endodontics, contributing to successful treatment outcomes, improved patient safety, and enhanced comfort.



**Fig. 1** Sample answer from Google Bard



ChatGPT

Paylaş

YA

How is inadvertent extrusion of sodium hypochlorite beyond the end of the root into the periradicular tissues managed?

Inadvertent extrusion of sodium hypochlorite beyond the root apex during root canal treatment can lead to potential complications and tissue damage. It's crucial to manage this situation promptly. Here are some general guidelines, but it's important to consult with a dental professional for specific advice based on the individual case:

1. Immediate Action:

- **Stop Irrigation:** Cease the irrigation immediately to prevent further extrusion.
- **Aspiration:** Use a high-volume suction to remove excess solution from the oral cavity.

2. Patient Management:

- **Patient Comfort:** Ensure the patient is comfortable and informed about the situation.
- **Monitoring:** Monitor the patient for any signs of adverse reactions or complications.

3. Evaluation:

- **Radiographic Evaluation:** Take radiographs to assess the extent of extrusion and any potential damage.
- **Clinical Examination:** Evaluate the patient clinically for signs of swelling, pain, or other complications.

4. Management of Periradicular Tissues:

- **Topical Anesthetics:** Apply topical anesthetics to alleviate any discomfort.
- **Cold Compress:** Apply a cold compress to control swelling and provide pain relief.

5. Pharmacological Management:

- **Analgesics:** Prescribe appropriate analgesics to manage pain.
- **Anti-inflammatory Medications:** If necessary, prescribe anti-inflammatory medications to reduce swelling.

6. Referral:

- **Specialist Consultation:** Consider referring the patient to an endodontic specialist or oral and maxillofacial surgeon for further evaluation and management.

7. Documentation:

- **Record Keeping:** Document the incident in the patient's chart, including details of the extrusion, actions taken, and any symptoms reported by the patient.

8. Prevention:

- **Review Protocols:** Evaluate and revise root canal treatment protocols to prevent similar incidents in the future.
- **Training:** Ensure that dental professionals are adequately trained in the proper use of irrigation solutions and techniques.

It's important to note that the management of sodium hypochlorite extrusion is highly case-specific, and the above guidelines are general recommendations. Prompt and appropriate action is crucial to minimize potential complications and ensure the best possible outcome for the patient. Dental professionals should stay informed about the latest guidelines and best practices in endodontics.

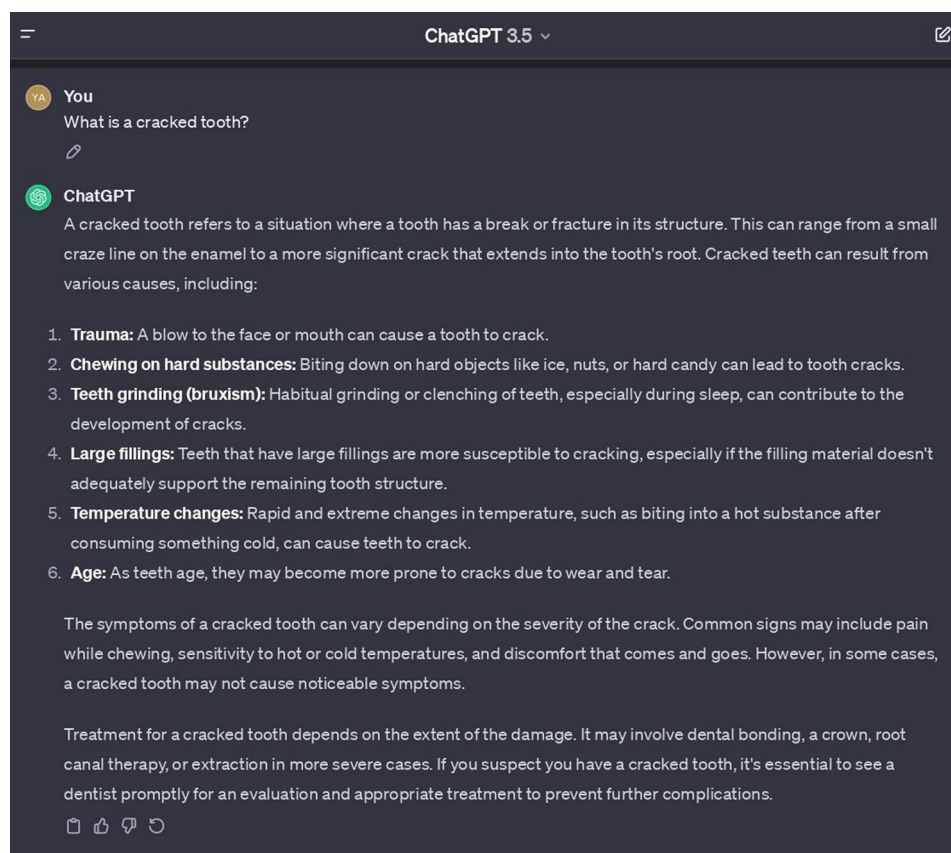
**Fig. 2** Sample answer from ChatGPT-4

reliability, bias, confidentiality and public acceptance [31]. Clear policies should be established to manage the use of sensitive data in the use of AI, ensure transparency in decision-making, and protect against biased results, and to mitigate security risks that may arise from the use of LLMs [35]. It is also critical to develop guidelines on ethical issues that balance the need to protect individual privacy and autonomy, especially in cases where erroneous or harmful advice is provided.

To simplify the evaluation process, the answers given by the language models were scored only in terms of accuracy and completeness and no other evaluation criteria were used. The predominance of ChatGPT-4 over other LLMs in providing accurate information has also

been shown in studies using general clinical dentistry and Japanese National Dentist Examination questions in agreement with our study [12, 16]. However, a drawback of ChatGPT-4 is not being free of charge and the pricing is the same for all countries, which lessens the accessibility to low-income countries. Suárez et al. reported that the correct answer rate was 57.33% when they directed dichotomous questions in endodontics to ChatGPT-3.5 software [10]. The relatively low correct answer rate may be because this language model does not have internet access and the trained data set does not include scientific publications in the field of endodontics.

The questions were asked only once without rephrasing to simulate a real clinical consultation. An important



**Fig. 3** Sample answer from ChatGPT-3.5

**Table 2** Likert scale description

Experts' grading	Description
Incorrect (0)	The answer provided is completely incorrect or unrelated to the question. It does not show an adequate understanding or knowledge of the topic.
Partially correct or Incomplete (1)	The answer shows some understanding or knowledge of the topic, but there are significant errors or missing elements. Although not entirely incorrect, the answer is not sufficiently accurate or complete to be considered confident or appropriate.
Correct (2)	The answer is completely correct and shows a sound and accurate understanding of the topic. All key elements are addressed accurately and comprehensively.

**Table 3** Mean, standard deviation of the likert scales cores

Language Model	N	Mean	StDev	Misinformation Rate
Google Bard <sup>A</sup>	40	1,3000	0,5164	25%
ChatGPT-3.5 <sup>A, B</sup>	40	1,4500	0,5524	15%
ChatGPT-4 <sup>B</sup>	40	1,7000	0,5164	10%

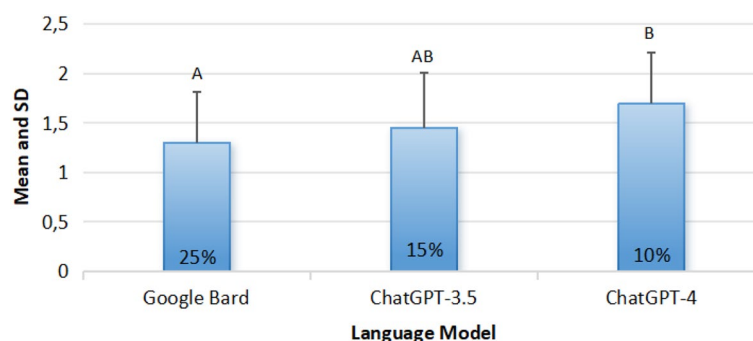
Different letters indicate statistically significant difference ( $P < .05$ )

finding of this is, the capability of all LLMs to perfectly understand the questions and generating an answer in the context. The answers of LLMs were found to be well-structured and written convincingly while containing

misleading or wrong information. This finding, a phenomenon so-called "hallucination", has been also reported in previous studies [12, 36]. For instance, in a study on head and neck surgery, it was observed that 46.4% of the references provided by ChatGPT4 did not actually exist [37]. Such made-up or misleading information could mislead dentists and potentially harm patients.

A limitation of the study may be that the prepared questions may not cover all clinical scenarios in endodontics. Since the questions and answers were limited to guidelines and position statements issued by scientific societies, points not covered in these documents were not included in the questions. In the study, questions were asked with specific prompts as much as possible to obtain the response and to evaluate the usefulness of the information obtained in clinical decision-making. Further studies that will evaluate LLMs as a source of information for patients may need to use different question patterns using less scientific terms.

It can be presumed that LLMs will continue to improve in terms of usability, precision, and ability to provide reliable information. Therefore, evaluation of LLMs as potential sources of information in the future might be valuable. Alternatively, and as suggested earlier, LLMs can be integrated into databases such as Web of Science



**Fig. 4** Mean, standard deviation and misinformation rates of the Likert Scales cores. Different letters indicate statistically significant difference ( $P < 0.05$ ) (Mean  $\pm$  SD,  $n = 40$ ). Misinformation rates of each language model were given inside columns with % symbol

and PubMed and made available for academic use. A chatbot that can provide information with real scientific references can thus provide convenience and save time for clinicians and academics [15]. In addition to the existing LLMs, software to be developed with the input of scientific publications for use in the field of dentistry can help clinicians and researchers. With proper training and development of LLMs, both diagnostic accuracy can be increased, and treatment planning can be facilitated. For example, AI can help quickly gain insight into adverse events encountered during endodontic procedures or decide on appropriate root canal filling material instead of detecting endodontic lesions on radiographs. Scientific guidance from dental associations will undoubtedly play a major role in the development of such software. Once validated, such software can serve both as a chairside assistant to dentists and be used in the training of dental students.

## Conclusion

This study represents the evaluation of the accuracy and completeness of different LLMs as potential sources in endodontics. Compared to other LLMs, ChatGPT-4 seems to be ahead in providing accurate and informative information in the field of endodontics. While this study, in agreement with the literature, proves that LLMs have utilizable features in dentistry, they are not free of flaws. Although LLMs are not designed to impart knowledge in dentistry and do not claim to do so, with further improvements in place, LLMs are promising tools for assisting dentists in endodontics.

## Abbreviations

AI	Artificial intelligence
LLMs	Large language models
USMLE	United States Medical Licensing Exam
ANOVA	One-way analysis of variance

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12903-025-06050-x>.

Supplementary Material 1

## Acknowledgements

Not applicable.

## Author contributions

Y.Ö., designed the study. D.E. and G.A.D. acquired data. Y.Ö., D.E. and G.A.D. completed the analysis and writing of the manuscript. All authors reviewed the manuscript and approved the final submission.

## Funding

Not applicable.

## Data availability

Data is provided within the manuscript or supplementary information files.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 20 January 2025 / Accepted: 23 April 2025

Published online: 28 April 2025

## References

- Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020;99(7):769–74.
- Howard J. Artificial intelligence: implications for the future of work. *Am J Ind Med*. 2019;62(11):917–26.
- Deng L. Artificial intelligence in the rising wave of deep learning: the historical path and future outlook [perspectives]. *IEEE Signal Process Mag*. 2018;35(1):180–177.
- Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, Damseh R, Alabed Alrazak S, Sheikh J. Large Language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ*. 2023;9:e48291.
- Revilla-Leon M, Gomez-Polo M, Vyas S, Barmak BA, Galluci GO, Att W, Krishnamurthy VR. Artificial intelligence applications in implant dentistry: A systematic review. *J Prosthet Dent*. 2023;129(2):293–300.
- Plebani M. ChatGPT: Angel or Demond? Critical thinking is still needed. In., vol. 61: De Gruyter; 2023: 1131–1132.
- Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, Ozdemir H, Tolios A, Carobene A, Padoan A. Potentials and pitfalls of ChatGPT and



- natural-language artificial intelligence models for the Understanding of laboratory medicine test results. An assessment by the European federation of clinical chemistry and laboratory medicine (EFLM) working group on artificial intelligence (WG-AI). *Clin Chem Lab Med*. 2023;61(7):1158–66.
8. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large Language models in medicine and medical research. *Lancet Digit Health*. 2023;5(6):e333–5.
  9. Diniz-Freitas M, Rivas-Mundina B, Garcia-Iglesias JR, Garcia-Mato E, Diz-Dios P. How ChatGPT performs in oral medicine: the case of oral potentially malignant disorders. *Oral Dis* 2023.
  10. Suarez A, Diaz-Flores Garcia V, Algar J, Gomez Sanchez 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.
  11. Cheong RCT, Unadkat S, McNeillis V, Williamson A, Joseph J, Randhawa P, Andrews P, Paleri V. Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google bard. *Eur Arch Otorhinolaryngol* 2023.
  12. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. Evaluation of generative artificial intelligence large Language models chatGPT, Google bard, and Microsoft Bing chat in supporting Evidence-based dentistry: A comparative Mixed-Methods study. *J Med Internet Res* 2023.
  13. Giannakopoulos K, Kavadella A, Stamatopoulos V, Kaklamanos E. Evaluation of generative artificial intelligence large Language models chatGPT, Google bard, and Microsoft Bing chat in supporting Evidence-based dentistry: A comparative Mixed-Methods study. *J Med Internet Res*. 2023;25(1):e51580.
  14. Suárez A, Jiménez J, de Pedro ML, Andreu-Vázquez C, García VD-F, Sánchez MG, Freire Y. Beyond the scalpel: assessing ChatGPT's potential as an auxiliary intelligent virtual assistant in oral surgery. *Comput Struct Biotechnol J*. 2024;24:46–52.
  15. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? *J Stomatology Oral Maxillofacial Surg*. 2023;124(5):101471.
  16. Ohta K, Ohta S. The performance of GPT-3.5, GPT-4, and bard on the Japanese National dentist examination: A comparison study. *Cureus*. 2023; 15(12).
  17. Ourang SA, Sohrabniya F, Mohammad-Rahimi H, Dianat O, Aminoshariae A, Nagendrababu V, Dummer PMH, Duncan HF, Nosrat A. Artificial intelligence in endodontics: fundamental principles, workflow, and tasks. *Int Endod J*. 2024;57(11):1546–65.
  18. Setzer FC, Li J, Khan AA. The use of artificial intelligence in endodontics. *J Dent Res*. 2024;103(9):853–62.
  19. Bragazzi NL, Szarpak Ł, Piccotti F. Assessing ChatGPT's Potential in Endodontics: Preliminary Findings from A Diagnostic Accuracy Study. Available SSRN 4631017 2023.
  20. Segura-Egea JJ, Gould K, Sen BH, Jonasson P, Cotti E, Mazzoni A, Sunay H, Tjaderhane L, Dummer PMH. European society of endodontology position statement: the use of antibiotics in endodontics. *Int Endod J*. 2018;51(1):20–5.
  21. Patel S, Brown J, Semper M, Abella F, Mannocci F. European society of endodontology position statement: use of cone beam computed tomography in endodontics: European society of endodontology (ESE) developed by. *Int Endod J*. 2019;52(12):1675–8.
  22. Galler K, Krastl G, Simon S, Van Gorp G, Meschi N, Vahedi B, Lambrechts P. European society of endodontology position statement: revitalization procedures. *Int Endod J*. 2016;49(8):717–23.
  23. by, Duncan ESE, Galler H, Tomson K, Simon P, El-Karim S, Kundzina I, Krastl R, Dammaschke G, Fransson T. European society of endodontology position statement: management of deep caries and the exposed pulp. *Int Endod J*. 2019;52(7):923–34.
  24. by, Krastl ESE, Weiger G, Filippi R, Van Waes A, Ebeleseder H, Ree K, Connert M, Widbiller T, Tjaderhane M. European society of endodontology position statement: endodontic management of traumatized permanent teeth. *Int Endod J*. 2021;54(9):1473–81.
  25. Wilson FR, Pan W, Schumsky DA. Recalculation of the critical values for Lawshe's content validity ratio. *Meas Evaluation Couns Dev*. 2012;45(3):197–210.
  26. Snigdha NT, Batul R, Karobari MI, Adil AH, Dawasaz AA, Hameed MS, Mehta V, Noorani TY. Assessing the performance of ChatGPT 3.5 and ChatGPT 4 in operative dentistry and endodontics: an exploratory study. *Hum Behav Emerg Technol*. 2024;24(1):1119816.
  27. Tanujaya B, Prahmana RCI, Mumu J. Likert scale in social sciences research: problems and difficulties. *FWU J Social Sci*. 2022;16(4):89–101.
  28. Suarez A, Diaz-Flores Garcia V, Algar J, Sanchez MG, de Pedro ML, Freire Y. Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers. *Int Endod J*. 2023.
  29. Alhaidry HM, Fatani B, Alrayes JO, Almana AM, Alhaed NK. ChatGPT in dentistry: A comprehensive review. *Cureus*. 2023;15(4):e38317.
  30. Goodman RS, Patrinely JR, Stone CA Jr., Zimmerman E, Donald RR, Chang SS, Berkowitz ST, Finn AP, Jahangir E, Scoville EA, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open*. 2023;6(10):e2336483.
  31. Wang L, Wan Z, Ni C, Song Q, Li Y, Clayton E, Malin B, Yin Z. Applications and concerns of ChatGPT and other conversational large Language models in health care: systematic review. *J Med Internet Res*. 2024;26:e22769.
  32. Mago J, Sharma M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. *Cureus*. 2023;15(7):e42133.
  33. Newton P, Xiromeriti M. ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assess Evaluation High Educ*. 2024;49(6):781–98.
  34. Alan R, Alan BM. Utilizing ChatGPT-4 for providing information on periodontal disease to patients: A DISCERN quality analysis. *Cureus*. 2023;15(9):e46213.
  35. Villena F, Véliz C, García-Huidobro R, Aguayo S. Generative artificial intelligence in dentistry: current approaches and future challenges. *ArXiv Preprint arXiv:240717532* 2024.
  36. Acar AH. Can natural Language processing serve as a consultant in oral surgery? natural Language processing in oral surgery?. *J Stomatology Oral Maxillofacial Surg* 2023;101724.
  37. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltrami GA, Bergonzani M, Bolzoni A, Committeri U, Crimi S et al. Accuracy of ChatGPT-Generated information on head and neck and oromaxillofacial surgery: A multicenter collaborative analysis. *Otolaryngol Head Neck Surg* 2023.

## Publisher's note

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