Original Article

Potential Use of Artificial Intelligence (Chatbots) for Assessing Patient Satisfaction in Endodontics- Cross-sectional Comparative Study

Sitara Subhas Naduvinamani[1], Raghavendra Ainapur[2], Balaram Naik[3], Prashant Moogi[4], Mahantesh Yeli[5], Sharmila Tapashetti[6]

From, [1]Post Graduate, [2]Addl. Professor, [3]Principal, [4]Professor & HOD, [5]Professor, [6]Addl. Professor, Department of Conservative Dentistry and Endodontics, SDM Dental College and Hospital, Sattur , Dharwad, Karnataka, India.

ABSTRACT

Aim: To evaluate and compare patient satisfaction with responses generated by Generative Pre-trained Transformer 4 (ChatGPT-4) and Generative Multimodal Intelligence 2 (GEMINI-2) to frequently asked questions in the Outpatient Department of Conservative Dentistry and Endodontics**. Methodology:** Seventy literate patients were enrolled over seven consecutive days. Each patient posed one question about their symptoms or treatment plan to both AI models using their personal device. After receiving AI responses, the treating clinician provided the correct explanation. Patients scored their satisfaction with each AI response using a five-point Likert scale (1 = Not at all satisfied; 5 = Very much satisfied). The Kolmogorov–Smirnov test confirmed non-normal data distribution, and the Mann–Whitney U test was used for group comparison. **Results:** The mean satisfaction score for ChatGPT-4 was 3.91 and for GEMINI-2 was 4.30. The Mann–Whitney U test revealed a statistically significant difference (U = 1829.500, P = .005), with GEMINI-2 receiving significantly higher patient satisfaction scores. **Conclusion:** GEMINI-2 demonstrated better patient satisfaction than ChatGPT-4 in an endodontic outpatient setting. AI chatbots show promise in supporting patient communication but cannot replace clinical expertise.

Key words: *Artificial Intelligence, Chatbots, Patient Satisfaction, Endodontics, Dental Informatics*

Artificial Intelligence (AI) refers to the capacity of machines to perform tasks that typically require human cognitive functions, including reasoning, learning, and decision-making. Among the most significant recent advances in AI are large language models (LLMs) and deep learning algorithms trained on vast text corpora that can comprehend, generate, and contextualize natural language with remarkable fluency. Prominent examples include Generative Pre-trained Transformer (ChatGPT; OpenAI, San Francisco, USA) and Generative Multimodal Intelligence (GEMINI; Google Ireland Limited, Dublin, Ireland), which have rapidly become widely accessible tools with significant implications across healthcare and dentistry [1,2].

ChatGPT was first released publicly in November 2022 and reached over 100 million users within two months, making it the most rapidly adopted consumer software application in history [2]. GEMINI-2 represents the latest evolution of Google's LLM family, incorporating multimodal content generation, advanced contextual reasoning, and enhanced conversational capabilities [1]. In dentistry and endodontics, AI tools, including ChatGPT and GEMINI, have been studied for applications spanning differential diagnosis, clinical decision-making, image analysis, disease prevention, dissemination of treatment information, and research support [7,8]. AI-powered systems have further demonstrated utility in detecting periapical pathologies, determining working lengths, analyzing root canal morphology, and predicting treatment prognosis [18].

As patients increasingly seek to be active participants in their own care, AI chatbots have emerged as a primary point of contact for health information, offering round-the-clock access to detailed, non-judgmental responses [5,6]. Despite their promise, AI systems can generate inaccurate or misleading information with apparent authority, a concern of particular significance in health-sensitive domains where reliability directly influences patient compliance and the doctor–patient relationship [9,10]. While prior studies have largely evaluated AI chatbot responses through expert-assessed accuracy, completeness, and readability [1,11,12], there is a critical gap in the literature regarding patient-perceived satisfaction with AI-generated clinical responses. Patient satisfaction encompasses perceived relevance, clarity, empathy, and reassurance, dimensions that may differ substantially between patient and expert evaluators. The

Access this article online

Received – 08th May 2026 Initial Review – 15th May 2026
Accepted – 19th May 2026

Quick response code

Correspondence to: Dr. Sitara Subhas Naduvinamani, Department of Conservative Dentistry and Endodontics, SDM Dental College and Hospital, Sattur, Dharwad-580009. Karnataka, India.

Email: sitaranaduvinamani1@gmail.com

present study, therefore, aims to directly measure patient satisfaction with ChatGPT-4 and GEMINI-2 responses in a live endodontic outpatient setting, where patients contextualize AI answers against their treatment clinician's explanation, providing a novel, patient-centered contribution to this growing field.

MATERIALS AND METHODS

2.1 Study Design and Setting

This was a cross-sectional comparative observational study conducted over seven consecutive days in the Outpatient Department of Conservative Dentistry and Endodontics at SDM Dental College and Hospital, Dharwad, Karnataka, India. Ethical approval was obtained from the Institutional Ethics Committee, and the study was conducted in accordance with the Declaration of Helsinki (2013 revision). Written informed consent was obtained from all participants.

2.2 Participants

Seventy literate patients (aged ≥18 years) who attended the Outpatient Department during the study period were enrolled by purposive sampling. Inclusion required literacy in English or the local language, possession of a personal internet-connected smartphone or tablet, and comfort with independent AI chatbot use. Patients were excluded if they were illiterate, lacked a personal device, had prior formal training in dentistry or healthcare, were below 18 years of age, or declined written informed consent. The sample size was determined based on feasibility within the study period and consistency with published literature in this domain [1,11,12].

2.3 Procedure

The study was conducted in four structured steps, as summarized in Table 1. Each enrolled patient independently formulated one question about their symptoms, diagnosis, or treatment plan and posed it to both ChatGPT-4 and GEMINI-2 on their personal device. Both AI models generated responses via their publicly available interfaces without modification. The treatment clinician then provided the patient with a correct, evidence-based explanation of the same question, serving as the reference standard. Patients subsequently scored their satisfaction with each AI response on a five-point Likert scale by comparing it with the clinician's explanation. Each patient provided two scores, one for ChatGPT-4 and one for GEMINI-2, yielding 140 total satisfaction ratings.

Table 1: Study Procedure and Satisfaction Scoring

Step	Action	Detail
1	Patient enrolment	Eligible patients identified; written informed consent obtained
2	Query generation	Patient independently poses one question to ChatGPT-4 and GEMINI-2 on a personal device
3	AI response	Both models generate unmodified responses via public interfaces
4	Clinician explanation	The treatment clinician provides an evidence-based explanation as a reference standard
5	Satisfaction scoring	Patient scores each AI response against clinician's explanation using a five-point Likert scale: 1 = Not at all satisfied; 2 = Not really satisfied; 3 = Undecided; 4 = Somewhat satisfied; 5 = Very much satisfied

2.4 Statistical Analysis

Data were entered in Microsoft Excel and analyzed using SPSS version 20 (IBM Corp., Armonk, NY, USA). Descriptive statistics (frequency, percentage, mean, median, interquartile range [IQR]) were calculated for each group. The Kolmogorov–Smirnov test confirmed non-normal distribution ($P < .05$), and the Mann–Whitney U test was used to compare satisfaction scores between groups. A P value < .05 was considered statistically significant.

RESULTS

All 70 patients completed evaluations for both AI models (100% response rate). The Kolmogorov–Smirnov test confirmed non-normal data distribution, warranting non-parametric analysis. The frequency distribution of satisfaction scores and comparative statistics is presented in Tables 2 and 3, respectively. GEMINI-2 attracted a higher proportion of top scores, with 45.7% of patients rating it 'Very much satisfied' compared to 25.7% for ChatGPT-4. Conversely, combined satisfaction (scores 4–5) was 87.1% for GEMINI-2 versus 72.8% for ChatGPT-4, and dissatisfaction (scores 1–2) was lower for GEMINI-2 (2.9%) than ChatGPT-4 (7.1%). The Mann–Whitney U test confirmed this difference to be statistically significant ($U = 1829.500$, $P = .005$), with GEMINI-2 demonstrating significantly higher patient satisfaction scores.

Table 2: Frequency Distribution of Patient Satisfaction Scores

Likert Score	Description	ChatGPT-4 (n)	ChatGPT-4 (%)	GEMINI-2 (n)	GEMINI-2 (%)
1	Not at all satisfied	0	0.0	0	0.0
2	Not really satisfied	5	7.1	2	2.9
3	Undecided	14	20.0	7	10.0
4	Somewhat satisfied	33	47.1	29	41.4
5	Very much satisfied	18	25.7	32	45.7
Total		70	100%	70	100%

Table 3: Comparison of Patient Satisfaction Scores — Mann–Whitney U Test

Group	N	Mean Score	Median (IQR: Q1, Q3)	Mann-Whitney U	P Value	Significance
ChatGPT-4	70	3.91	4 (3, 5)	1829.500	.005	Significant
GEMINI-2	70	4.30	4 (4, 5)			

DISCUSSION

The present study evaluated patient satisfaction with ChatGPT-4 and GEMINI-2 responses in a live endodontic outpatient setting, one of the few studies in the endodontic literature to adopt this patient-centered approach rather than expert-based accuracy assessment. The higher satisfaction with GEMINI-2 may reflect its enhanced contextual reasoning, conversational tone, and multimodal generation capabilities, qualities that render responses more accessible and reassuring to lay patients, rather than superior factual accuracy per se.

These findings stand in instructive contrast to several expert-assessed studies. Zengin et al. found no significant difference between ChatGPT-4 and GEMINI in accuracy or readability for endodontic questions but noted higher completeness scores for ChatGPT-4 by one expert evaluator [1]. Mohammad-Rahimi et al. found that GPT-3.5 provided significantly higher validity than Google Bard and Bing for endodontic questions [19]. Künzle and Paris reported ChatGPT-4.0 to demonstrate the highest accuracy on a restorative and endodontic question pool [20], and Dursun and Bilici Geçer found ChatGPT-4 to have superior accuracy scores while Gemini showed higher readability in an orthodontic context [12].

The consistent observation across these studies, that ChatGPT-4 tends to outperform on expert-assessed accuracy while GEMINI demonstrates advantages in readability, aligns with and may explain the present study's patient satisfaction findings, as lay patients may weigh clarity and comprehensibility more heavily than technical completeness. Supporting this, Lv et al. found lay users' AI response ratings to differ substantially from those of expert dentists [27], and Fattah et al. noted Gemini's superiority in certain clinical domains despite ChatGPT's generally greater accuracy [28]. The present findings therefore underscore that patient-reported outcomes and expert-assessed accuracy are complementary but non-interchangeable evaluation frameworks, both of which should be incorporated into assessments of clinical AI tools.

This study has several limitations. Only literate, digitally enabled patients were included, introducing selection bias that excludes older, less educated, or lower-socioeconomic-status patients who may form a substantial proportion of endodontic outpatients in India. Each patient asked a single, self-formulated question, introducing heterogeneity not controllable by a standardized question bank. AI model responses were not independently verified for clinical accuracy, preventing the determination of whether GEMINI-2's higher satisfaction reflected better accuracy or merely response style.

Additionally, AI models continuously update their algorithms, limiting the reproducibility of findings beyond the study period, and patient satisfaction may have been influenced by uncontrolled confounders such as digital confidence, prior expectations, or the manner of the clinician's explanation. Future studies should include patients with limited digital literacy through mediated AI interaction, assessing both accuracy and satisfaction within a single design, conducting evaluations in regional languages, and longitudinally tracking satisfaction as model versions evolve.

CONCLUSION

Within the limitations of this study, GEMINI-2 demonstrated significantly higher patient satisfaction than ChatGPT-4 when patients compared AI-generated responses to their treating clinician's explanation in an endodontic outpatient setting. AI chatbots show considerable promise in supporting patient communication; however, they cannot replace the clinical expertise and personalized care of a qualified dental professional.

REFERENCE

1. Zengin B, Güzelce Sultanoğlu E, Sultanoğlu E. Potential use of ChatGPT and GEMINI for patient information in endodontics. J Oral Med Dent Res. 2025;6(1):1-9.
2. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. Adv Neural Inf Process Syst. 2020; 33:1877-1901.
3. Brickley MR, Shepherd JP, Armstrong RA. Neural networks: a new technique for development of decision support systems in dentistry. J Dent. 1998;26(4):305-309.
4. Thurzo A, Urbanova W, Novak B, Czako L, Siebert T, Stano P, et al. Where is the artificial intelligence applied in dentistry? Systematic review and literature analysis. Healthcare (Basel). 2022;10(7):1269.
5. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology. 2023;307(2):e230163.
6. Mokkink DH, Dijkstra R, Verbakel D. Surfende patiënten. Huisarts Wet. 2008;51(3):138-41.
7. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. J Esthet Restor Dent. 2023;35(7):1098-1102.
8. Strunga M, Urban R, Surovková J, Thurzo A. Artificial intelligence systems assisting in the assessment of the course and retention of orthodontic treatment. Healthcare (Basel). 2023;11(5):683.
9. Shi W, Zhuang Y, Zhu Y, Iwinski H, Wattenbarger M, Wang MD. Retrieval-augmented large language models for adolescent idiopathic scoliosis patients in shared decision-making. In: Proceedings of the 14th ACM International Conference on

Bioinformatics, Computational Biology, and Health Informatics. New York: ACM; 2023.
10. Boreak N. Effectiveness of artificial intelligence applications designed for endodontic diagnosis, decision-making, and prediction of prognosis: a systematic review. J Contemp Dent Pract. 2020;21(7):796-804.
11. Güzelce Sultanoğlu E. Can natural language processing (NLP) provide consultancy to patients about edentulism teeth treatment? Cureus. 2024;16(10):e70945.
12. Dursun D, Bilici Geçer R. Can artificial intelligence models serve as patient information consultants in orthodontics? BMC Med Inform Decis Mak. 2024;24(1):211.
13. Meyer MKR, Kandathil CK, Davis SJ, Durairaj KK, Patel PN, Pepper JP, et al. Evaluation of rhinoplasty information from ChatGPT, Gemini, and Claude for readability and accuracy. Aesthet Plast Surg. 2024. Epub ahead of print.
14. Snigdha NT, Batul R, Karobari MI, Adil AH, Dawasaz AA, Hameed MS, et al. Assessing the performance of ChatGPT 3.5 and ChatGPT 4 in operative dentistry and endodontics: an exploratory study. Hum Behav Emerg Technol. 2024; 2024:8.
15. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? JMIR Med Educ. 2023;9(1):e45312.
16. Alser M, Waisberg E. Concerns with the usage of ChatGPT in academia and medicine: a viewpoint. Am J Med Open. 2023; 9:100036.
17. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology. Ophthalmol Sci. 2023;3(3):100324.
18. Karobari MI, Adil AH, Basheer SN, Abduljabbar T, Marya A, Noorani TY. Evaluation of the diagnostic and prognostic accuracy of artificial intelligence in endodontic dentistry: a comprehensive review. Comput Math Methods Med. 2023; 2023:7049360.
19. 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):305-314.
20. 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.
21. Acar AH. Can natural language processing serve as a consultant in oral surgery? J Stomatol Oral Maxillofac Surg. 2024;125(3):101724.
22. Gill GS, Tsai J, Moxam J, Li J, Qiu H, Dhaliwal G. Comparison of Gemini Advanced and ChatGPT 4.0's performances on the Ophthalmology Resident Ophthalmic Knowledge Assessment Program (OKAP) Examination Review Question Banks. Cureus. 2024;16(9):e69612.
23. 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. 2024;75(2):344-350.
24. Barker FG, Rutka JT. Editorial: Generative artificial intelligence, chatbots, and the Journal of Neurosurgery Publishing Group. J Neurosurg. 2023;139(3):901-903.
25. Alhaidry HM, Fatani B, Alrayes JO, Almana AM, Alfhaed NK. ChatGPT in dentistry: a comprehensive review. Cureus. 2023;15(4):e38317.
26. 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):108-113.
27. Lv X, Zhang X, Li Y, Ding X, Lai H, Shi J. Leveraging large language models for improved patient access and self-management: assessor-blinded comparison between expert- and AI-generated content. J Med Internet Res. 2024;26:e55847.
28. Fattah IM, Salih AM, Salih RQ, Asaad F, Ghafour D, Bapir B, et al. Comparative analysis of ChatGPT and Gemini (Bard) in medical inquiry: a scoping review. Front Artif Intell. 2025; 8:1450843.
29. Sivaramakrishnan G, Almuqahwi M, Ansari S, Lubbad M, Alagamawy E, Sridharan K. Assessing the power of AI: a comparative evaluation of large language models in generating patient education materials in dentistry. npj Dent Med. 2025;2(1):42.
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.
31. Farhadi Nia M, Ahmadi M, Irankhah E. Transforming dental diagnostics with artificial intelligence: advanced integration of ChatGPT and large language models for patient care. Front Dent Med. 2025; 5:1456208.
32. Dufey-Portilla N, Billik Frisman A, Gallardo Robles M, Peña-Bengoa F, Cabrera Ávila C, Nagendrababu V, et al. Assessing the validity of ChatGPT-4o and Google Gemini Advanced when responding to frequently asked questions in endodontics. J Appl Oral Sci. 2025;33:e20250321.
33. Francis L, Deepthi VS, Viswambharan P, Nair VV. Comparative evaluation of accuracy, completeness and readability of common patient queries related to prosthodontic treatment by two artificial intelligence models (ChatGPT-4o and Gemini). Cureus. 2025;17(12):e98458.
34. Taşyürek M, Adıgüzel Ö, Ortaç H. Comparative evaluation of responses from ChatGPT-5, Gemini 2.5 Flash, Grok 4, and Claude Sonnet-4 chatbots to questions about endodontic iatrogenic events. Healthcare (Basel). 2025;13(20):2615.
35. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023;388(13):1233-1239.
36. Jabbar MA, Iqbal H, Chawla U. Patient satisfaction: the role of artificial intelligence in healthcare. J Health Manag. 2024. doi:10.1177/09720634241246331.
37. Ekmekçi O, Kılınç MS, Özyurt M. Evaluation of responses given by artificial intelligence applications to questions about regenerative endodontic procedures. J Endod. 2024. Epub ahead of print.
38. Vassis S, Powell H, Petersen E, Barkmann A, Noeldeke B, Kristensen KD, et al. Large-language models in orthodontics: assessing reliability and validity of ChatGPT in pretreatment patient education. Cureus. 2024;16(8):e67423.
39. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769-774.
40. Dave M, Patel N. Artificial intelligence in healthcare and education. Br Dent J. 2023;234(10):761-764.

How to cite this article: Naduvinamani SS, Ainapur R, Naik B, Moogi P, Yeli M, Tapashetti S. Potential Use of Artificial Intelligence (Chatbots) for Assessing Patient Satisfaction in Endodontics- Cross-sectional Comparative Study. Indian J Integr Med. 2026; 6(2):83-86.

Fundina: None; *Conflicts of Interest: None Stated*