RESEARCH PAPER

THE ACCURACY OF AI-BASED CHATBOTS IN RESPONDING TO ROUTINELY ASKED QUESTIONS ABOUT ORTHODONTICS: A COMPARATIVE STUDY

Dr. Aditi Singh[1], Dr. Divya Shetty[2*], Dr. Prakash VS[3], Dr. Payal Sharma[4], Dr. Piush Kumar[5]

[1]PG Student, Department of Orthodontics and dentofacial Orthopaedics, ITS-CDSR, Muradnagar, Ghaziabad, UP-201206, India

[2]Professor, Department of Orthodontics and dentofacial, Orthopaedics ITS-CDSR, Muradnagar, Ghaziabad, UP-201206, India

[3]Assistant Professor, Department of Orthodontics and dentofacial, Orthopaedics ITS-CDSR, Muradnagar, Ghaziabad, UP-201206, India

[4]Professor and head, Department of Orthodontics and dentofacial Orthopaedics, ITS-CDSR, Muradnagar, Ghaziabad, UP-201206, India

[5]Professor, Department of Orthodontics and dentofacial Orthopaedics, ITS-CDSR, Muradnagar, Ghaziabad, UP-201206, India

Corresponding Author

Dr. Aditi Singh

ABSTRACT

Objective: This study aimed to evaluate and compare the accuracy and effectiveness of ChatGPT and Gemini in responding to frequently asked orthodontic questions.

Methods: Ten experienced orthodontists assessed the responses of ChatGPT and Gemini to a standardized set of patient questions using a five-point modified Likert scale. Data were statistically analyzed using non-parametric tests due to non-normal distribution.

Results: Both chatbots generally provided satisfactory responses, with ChatGPT achieving a slightly higher average accuracy score (1.53 ± 0.47) compared to Gemini (1.59 ± 0.52), though the difference was not statistically significant. ChatGPT's most accurate answer was to the question about the advantages of braces vs. Invisalign, while Gemini performed best on dietary recommendations for patients with braces.

Conclusion: ChatGPT and Gemini demonstrated similar performance in addressing patient queries in orthodontics, suggesting that both can serve as supplementary virtual assistants in clinical settings. While their responses are mostly accurate, clinicians should remain aware of the limitations and potential for misinformation. Continued refinement and integration of feedback mechanisms are essential to enhance their utility in orthodontic care.

Keywords: Artificial Intelligence, Chatbots, ChatGPT, Gemini, Orthodontics, Patient Communication, Healthcare Technology

How to cite this article: Singh A, Shetty D, Prakash VS, Sharma P, Kumar P. The Accuracy of AI-Based Chatbots in Responding to Routinely Asked Questions About Orthodontics: A Comparative Study. Int J Drug Deliv Technol. 2026;16(59s): 43-52. DOI: 10.25258/ijddt.16.59s.6

Source of support: Nil.

Conflict of interest: Nil.

INTRODUCTION

Artificial intelligence (AI) has rapidly expanded its role and importance across various industries, including dentistry. Its recent integration into dental practice has led to noteworthy progress. AI is proving to be a powerful tool for interpreting clinical dental data. Innovations in machine learning and deep learning have revolutionized how information is retrieved and communicated, particularly in healthcare.[1]

These technological advancements have led to the creation of AI-driven chatbots specifically designed for medical use. In contrast to traditional search engines that provide broad information, AI-powered chatbots offer more user-friendly interactions by presenting details in a conversational manner. This approach helps patients and their families better grasp complex healthcare concepts. Additionally, these models not only convey known information but also generate insightful questions and offer reliable references.[2]

On November 30, 2022, OpenAI introduced ChatGPT to the public via a free online platform accessible to anyone with a registered account. ChatGPT is a large language model powered by artificial intelligence, trained on an extensive dataset that spans a wide array of subjects, including medical literature. When users ask questions, it generates clear, conversational, and seemingly well-informed responses that are easy to comprehend.[3]

Formerly known as Google Bard, Gemini is an artificial intelligence chatbot developed by Google in 2021, ongoing research and development hold significant promise for responsible and ethical applications of these technologies to improve healthcare delivery and patient outcomes.

Although a study evaluating ChatGPT's reliability in delivering medical information recognized the potential

*Author for Correspondence:- **Dr. Aditi Singh**

of such chatbots to benefit both patients and healthcare professionals, it also pointed out the limitations that need to be addressed before widespread adoption. The main concern is the generation of inaccurate information ("hallucinations") which can be dangerous. The solution involves international standards, fact-checking procedures, and user education about these risks.[3]

Given the broad range of treatment options in orthodontics, it is essential to evaluate whether AI-based chatbots like ChatGPT and Gemini can provide accurate and reliable answers to domain-specific questions. If research demonstrates that these generative AI tools consistently offer trustworthy and timely responses to frequently asked questions, they could become valuable virtual assistants for general dentists, supporting clinical decision-making and increasing efficiency in orthodontic care.

Previous studies were done wherein the responses generated by the chatbots regarding endodontics treatment and orthognathic surgeries were done, but none of the studies were done regarding orthodontics until this study was started. Hence in this study, we will assess and compare the responses generated by the two AI-based chatbots i.e. ChatGPT and Gemini, to some commonly asked questions regarding orthodontic treatment.

MATERIALS AND METHODS

Ten orthodontists used a standardized set of questions (Table 1) to evaluate the performance of the two most widely used AI chatbots: ChatGPT and Gemini. Each orthodontist had a minimum 2years of work experience. The orthodontists were instructed to open the AI chatbots i.e., ChatGPT and Gemini on their devices. Then to type and ask the questions mentioned one at a time, each response generated by the chatbot was rated as per the Fivepoint modified Likert scale provided.

Each participant was given the worksheet, providing the instructions scoring scale and the questions to be searched and scored. Five-point modified Likert scale: 1 responded with adequate information; 2 responded but did not provide adequate information; 3 did not directly answer the question, but provided a list of accurate websites that address the question; 4 did not directly answer the question, but provided a list of inaccurate websites that address the question; 5 did not know the response to the question.

STATISTICAL ANALYSIS

Data was tabulated in MS Excel software and analyzed using SPSS v23 software. The validity of the questionnaire was tested by Pearson's correlation and reliability by Cronbach's Alpha. The level of significance was kept at 5%. Data was subjected to normality testing using the Shapiro-Wilk test. Results showed that data did not follow the normal distribution. Therefore, non-parametric tests were applied. The distribution of the responses in the ChatGPT group in the Gemini group was presented using frequency and percentage. Intergroup comparison of the mean scores per question among two groups was done using Mann-Whitney test.

RESULTS

The aim of the study was to assess the usefulness and effectiveness of AI-based chatbots, specifically ChatGPT and Gemini, in responding to commonly asked patient questions in orthodontic practices.

Figure1 presents the distribution of responses to the individual questions in the ChatGPT group. The best response generated by ChatGPT was for the question, "What is the advantage of braces vs. Invisalign?" as it got score of 1 by all the 10 participants.

Figure 2 displays the distribution of responses to the individual questions in the Gemini group. "What should I eat with braces?" got the best scores, i.e. 1 by 9 investigators. Meanwhile "Do braces hurt?", got the worst score of 5 by 5 analyst.

Figure 3 presents the descriptive statistical comparison of mean score of the responses per questions among two groups. For each question assessed, there was a non-significant difference in the mean scores of the two groups (ChatGPT & Gemini).

Table 2 presents the descriptive statistical comparison of combined mean score of the responses among two groups. The combined mean for ChatGPT was 1.53 ± 0.47 and for Gemini, it was 1.59 ± 0.52. There was a non-significant difference in the combined mean score of the two groups (ChatGPT & Gemini).

The accuracy of the generated answers was assessed in the table 3. The accuracy of the responses for each method was determined by summing all the correct responses for each question and then obtaining the average percentage of correct responses. Although the mean percentage accuracy was greater for ChatGPT, there was a non-significant difference in the accuracy of the two groups.

DISCUSSION

Artificial Intelligence, a field within computer science, focuses on creating systems that can mimic human intelligence. In recent years, it has made significant progress by automating tasks that were once considered beyond the capabilities of machines. An important feature highlighted in the development of chatbots is their capacity to understand context and maintain meaningful, consistent dialogue.

Our goal was to assess whether these generative AI platforms can provide reliable and prompt answers to common questions, potentially acting as helpful virtual assistants for general dentists and enhancing decision-making efficiency in orthodontic practice. Hence, we conducted an assessment comparing the responses generated by two AI-based chatbots, namely ChatGPT and Gemini, to common inquiries concerning orthodontic treatment.

In this study, the majority of responses were deemed highly commendable by the ten evaluators. Particularly, the inquiry "Do braces hurt?" by Gemini received the highest scores, indicating dissatisfaction among evaluators with the responses provided for this question (with a mean score of 3.2). Conversely, for ChatGPT, the question "Does insurance cover braces?" received the highest mean score of 2.30. This indicates that although chatbots are capable of identifying patterns and structuring data, they struggle to fully understand the deeper meaning and context behind the information. Notably, for the question "What is the advantage of braces vs. Invisalign?" directed to ChatGPT, the AI chatbot received the lowest mean score of 1, with unanimous agreement among all evaluators. This unanimity suggests that the AI chatbot provided adequate information according to all assessments. Gemini received the lowest score for the question "What can I eat with braces?" with a mean of 1.20. Table 4 shows that the difference in the combined mean scores between the two groups (ChatGPT and Gemini) was not statistically significant. Both ChatGPT (1.53) and Gemini (1.59) provided satisfactory responses on average for the predetermined list of questions. Therefore, they may be deemed reliable for addressing patients' frequently asked questions in orthodontic offices.

The results aligned with earlier research conducted by Perez-Pino et al.[17], who assessed the effectiveness and efficacy of four voice-activated, AI-powered virtual assistants (Alexa, Google Assistant, Siri, and Cortana) in handling frequently asked patient queries in orthodontic settings. The study concluded that a smart virtual assistant capable of delivering evidence-based responses specific to orthodontics could effectively address this issue. All participating orthodontists unanimously agreed that this type of technology would be beneficial to orthodontic practice.

As per Tanaka et al.[19] ChatGPT has demonstrated effectiveness in furnishing high-quality responses concerning clear aligners, temporary anchorage devices, and digital imaging within the orthodontic domain. However, in a separate investigation by Arqub et al.[28], it was concluded that the software's capacity to deliver current and precise information regarding clear aligner therapy was limited. Hence, clinicians and patients should exercise caution regarding potentially inaccurate assertions and important details omitted in the responses provided by ChatGPT.

Until the completion of this study, only a few investigations had explored the intersection of orthodontics and chatbots. Nonetheless, comparable research had been conducted in other branches of dentistry.

Yurdakurban et al.[2] suggested that AI-powered chatbots, equipped with diverse functionalities, consistently deliver answers of high quality and reliability, often presenting complex medical information on orthognathic surgery. Despite the quality of information provided by these chatbots, they advised individuals to seek guidance from healthcare professionals for comprehensive consultation on such matters.

Likewise, Suárez et al.[12] agreed that ChatGPT attained an average accuracy of 57.33% regarding answer precision. Nonetheless, notable variations in accuracy were noted depending on the complexity of questions, with lower accuracy recorded for simpler inquiries. Their findings led them to conclude that ChatGPT cannot currently supplant dentists in clinical decision-making. However, they anticipated that as ChatGPT advances through deep learning, it will likely become increasingly valuable and proficient in realm of endodontics.

According to Freire et al.[(29)], current evidence suggests that ChatGPT has a limited capacity to provide accurate information regarding removable dental prostheses (RDPs) and toothsupported fixed dental prostheses (FDPs). Consequently, they emphasized that ChatGPT cannot serve as a substitute for a dentist, and if utilized by professionals, they must be cognizant of its constraints.

The need of reviewing, a special modality of treatment has become one of keystone features. On the basis of which, patient gets motivated for getting himself treated by that modality.

With the coming of A.I, it becomes more convenient and faster for the patient to get in touch with the chatbots on internet, to get an answer to their queries with detailed explanations. This study not only provides us with the answer to question that whether chatbots are reliable source of information, but also opens multiple windows regarding ways how new chatbots can be improved and utilized in the field of orthodontics.

Including the option to rate the answer and submit feedback is a common feature in many chatbots used for research purposes. User ratings provide valuable insights into the helpfulness and accuracy of the chatbot's responses. This data can be used to refine the chatbot's training and improve future responses. Feedback allows researchers to understand specific areas where the chatbot might be lacking, such as missing information, unclear explanations, or irrelevant responses. This feedback helps tailor the chatbot's capabilities better. Offering the ability to rate and provide feedback increases user engagement and makes them feel like their input matters. This can lead to more accurate data

collection and a more positive user experience. Overall, incorporating answer rating and feedback mechanisms is crucial for ensuring the chatbot is providing valuable and accurate information and continuously improving the chatbot's performance and capabilities.

CONCLUSION

The results of this study suggest that orthodontic practices can confidently adopt AI-powered chatbots to enhance patient communication. By improving the precision and clarity of chatbot responses, practices can boost patient satisfaction, streamline operations, and elevate the overall quality of care.

As AI continues to evolve, future research should aim to fine-tune chatbot algorithms and broaden their functionalities. Integrating features like voice recognition and machine learning can further improve the chatbot's ability to handle orthodontic-related questions. With advancements in deep learning, chatbot performance is expected to grow more effective and beneficial in orthodontics. Continuous research is necessary to evaluate the AI's progress and ensure it continues to learn and improve over time. Consistent monitoring is essential to maintain accuracy, reliability, and safety in clinical applications.

In a rapidly changing world, AI offers convenience and reduces stress; however, it's important to remember that the "A" in AI stands for "artificial." Given the mixed legacy of man-made technologies, complete reliance on AI should be approached thoughtfully, with an awareness of its limitations

REFERENCES

1. Agrawal P, Nikhade P. Artificial intelligence in dentistry: Past, present, and future. Cureus 2022

2. Yurdakurban E, Topsakal KG, Duran GS. A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient information. J Stomatol Oral Maxillofac Surg 2024;125(5):101757

3. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, Srinivasan N, Park J, Burch M, Watson R, Liran O, Samakar K. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obes Surg 2023;33(6):1790–
6. Available from: http://dx.doi.org/10.1007/s11695-023-06603-5.

4. Bharti U, Bajaj D, Batra H, Lalit S, Lalit S, Gangwani A. Medbot: Conversational artificial intelligence powered chatbot for delivering Tele-health after COVID-19. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE; 2020.

5. Wang D, Fang H. Length adaptive regularization for retrieval-based chatbot models. In: Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval. New York, NY, USA: ACM; 2020.

6. Oh YJ, Zhang J, Fang M-L, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021;18(1).

7. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information (Basel) 2022;13(1):41.

8. Suárez A, Adanero A, Díaz-Flores García V, Freire Y, Algar J. Using a virtual patient via an artificial intelligence chatbot to develop dental students' diagnostic skills. Int J Environ Res Public Health 2022;19(14):8735.

9. Pandey S, Sharma S, Wazir S. Mental healthcare chatbot based on natural language processing and deep learning approaches: Ted the therapist. Int J Inf Technol 2022;14(7):3757–66.

10. Xue J, Zhang B, Zhao Y, Zhang Q, Zheng C, Jiang J, Li H, Liu N, Li Z, Fu W, Peng Y, Logan J, Zhang J, Xiang X. Evaluation of the current state of chatbots for digital health:
Scoping review. J Med Internet Res 2023;25: e47217. Available from: http://dx.doi.org/10.2196/47217.

11. Hossein Mohammad-Rahimi, Seyed Amir Hossein Ourang, Mohamad Amin Pourhoseingholi, Omid Dianat, Michael P, Nosrat A. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. . Int Endod J. 2023. 2023 Dec 20;57(3):305–14.

12. Suárez A, Víctor Díaz-Flores García, Algar J, Margarita Gómez Sánchez, Pedro, Freire Y. Unveiling the ChatGPT phenomenon: Evaluating the consistency and accuracy of endodontic question answers. Int Endod J. 2023 Oct 9;57(1):108–13.

13. Dashti M, Londono J, Ghasemi S, Moghaddasi N. How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? J Prosthet Dent 2023

14. Iqbal J, Cortés Jaimes DC, Makineni P, Subramani S, Hemaida S, Thugu TR, Butt AN, Sikto JT, Kaur P, Lak MA, Augustine M, Shahzad R, Arain M. Reimagining healthcare: Unleashing the power of artificial intelligence in medicine. Cureus 2023;15(9):e44658. Available from: http://dx.doi.org/10.7759/cureus.44658

15. 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):1098–102.

16. Gilson A, Safranek CW, Huang T, Socrates V, Taylor CL. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9.

17. 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.

18. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, Chang S, Berkowitz S, Finn A, Jahangir E, Scoville E, Reese T, Friedman D, Bastarache J, van der Heijden Y, Wright J, Carter N, Alexander M, Choe J, Chastain C, Zic J, Horst S, Turker I, Agarwal R, Osmundson E, Idrees K, Kieman C, Padmanabhan C, Bailey C, Schlegel C, Chambless L, Gibson M, Osterman T, Wheless L. Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the chat-GPT model. Res Sq. 2023; Available from: http://dx.doi.org/10.21203/rs.3.rs-2566942/v1;

19. Tanaka OM, Gasparello GG, Hartmann GC, Casagrande FA, Pithon MM. Assessing the reliability of ChatGPT: a content analysis of self-generated and self-answered questions on clear aligners, TADs and digital imaging. Dental Press J Orthod 2023;28(5): e2323183.

20. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589–96. Available from:http://dx.doi.org/10.1001/jamainternmed.2023.1838

21. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. J Med Internet Res 2023;25: e40789

22. 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. Available from: http://dx.doi.org/10.2196/48291

23. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(6):887

24. Bhardwaz S, Kumar J. An extensive comparative analysis of chatbot technologies - ChatGPT, Google BARD and Microsoft Bing. In: 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE; 2023.

25. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. J Med Syst 2023;47(1):33.

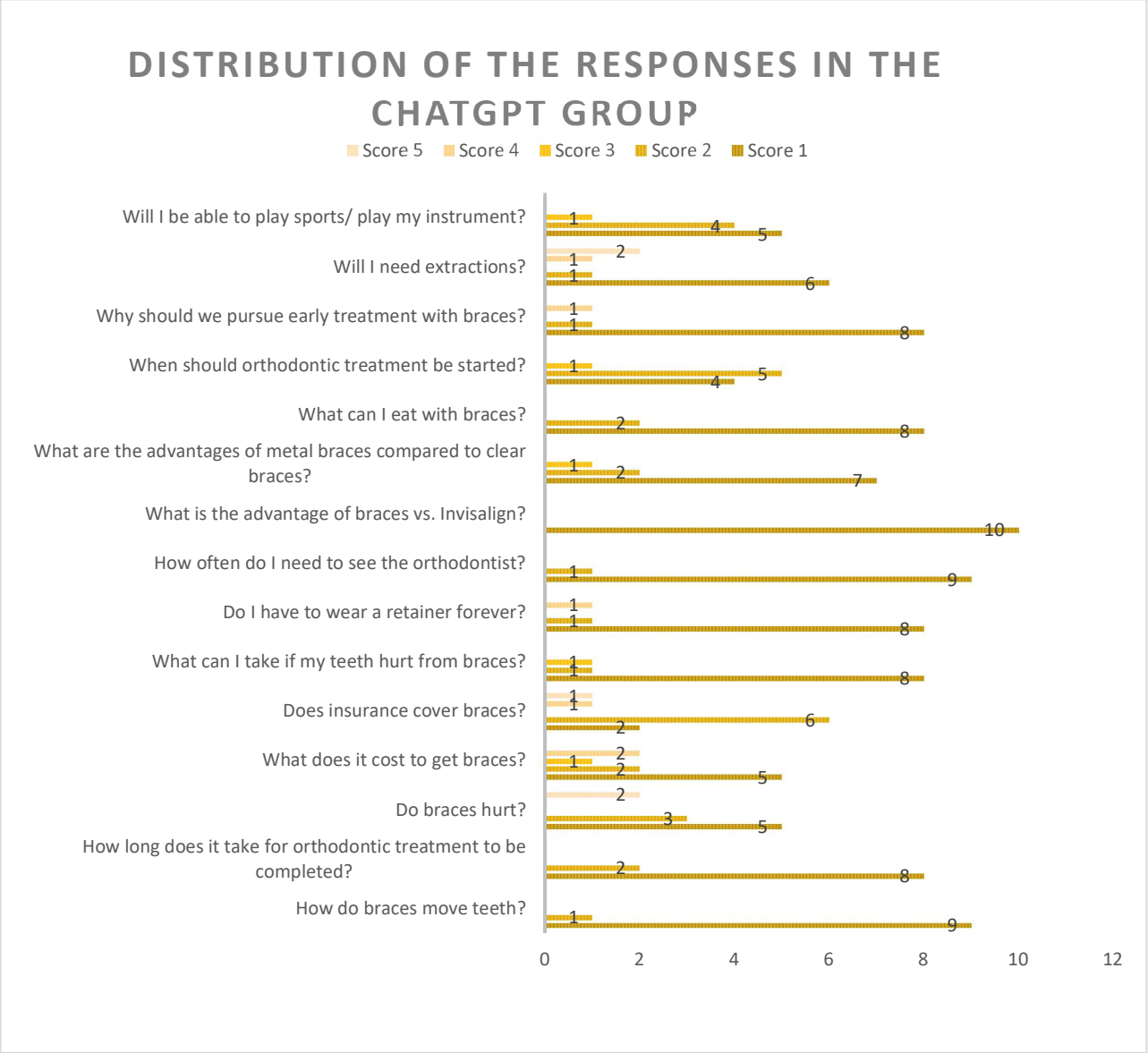
26. Doshi R, Amin K, Khosla P, Bajaj S, Chheang S, Forman HP. Utilizing Large Language Models to simplify radiology reports: A comparative analysis of ChatGPT3.5, ChatGPT4.0, Google Bard, and Microsoft Bing bioRxiv. 2023.

27. Madadi Y, Delsoz M, Khouri AS, Boland M, Grzybowski A, Yousefi S. Applications of artificial intelligence-enabled robots and chatbots in ophthalmology: recent advances and future trends. Curr Opin Ophthalmol 2024;35(3):238–43

28. Abu Arqub S, Al-Moghrabi D, Allareddy V, Upadhyay M, Vaid N, Yadav S. Content analysis of AI-generated (ChatGPT) responses concerning orthodontic clear aligners. Angle Orthod 2024;94(3):263–72

29. Freire Y, Santamaría Laorden A, Orejas Pérez J, Gómez Sánchez M, Díaz-Flores García V, Suárez A. ChatGPT performance in prosthodontics: Assessment of accuracy and repeatability in answer generation. J Prosthet Dent 2024;131(4): 659.e1-659.e6

Figure 1: Distribution of the responses in the ChatGPT group



*Author for Correspondence:- **Dr. Aditi Singh**

Figure 2: Distribution of the responses in the Gemini group

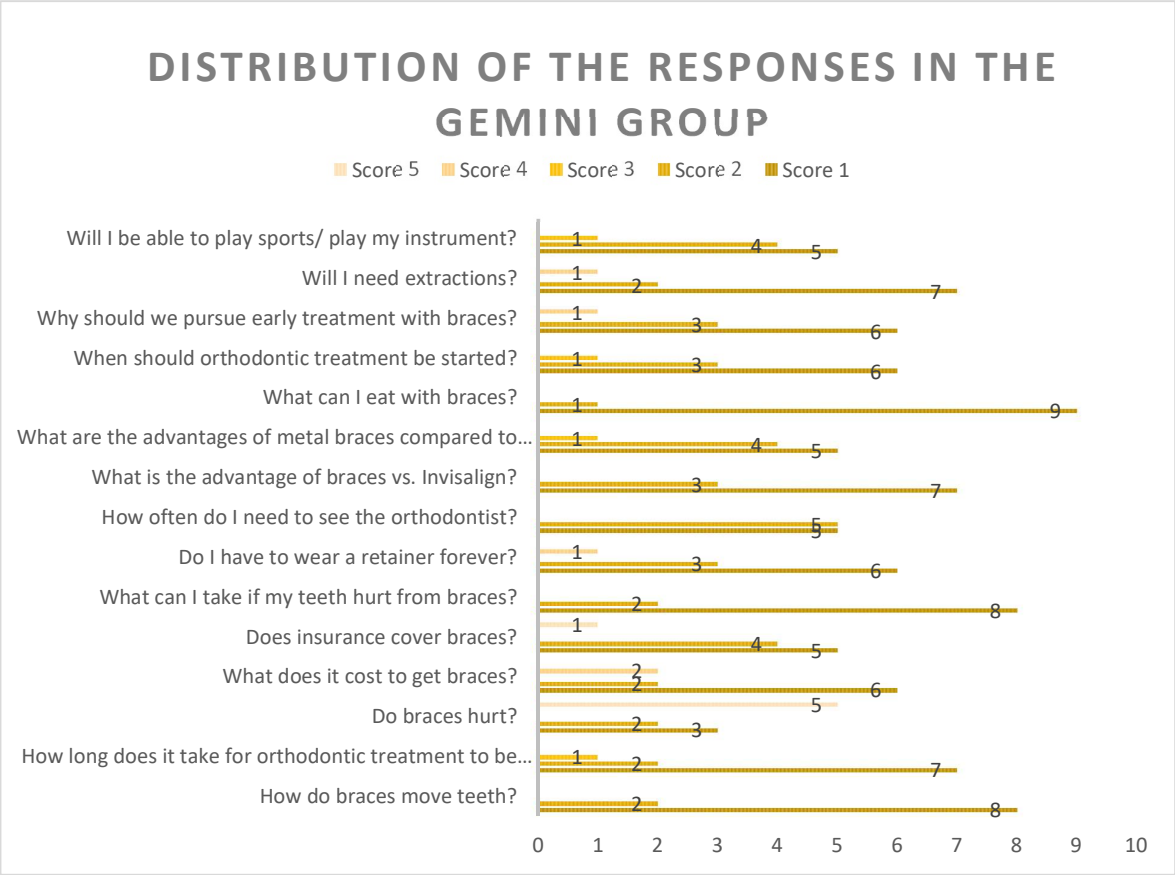


Figure 3: Comparison of mean scores per question among two groups

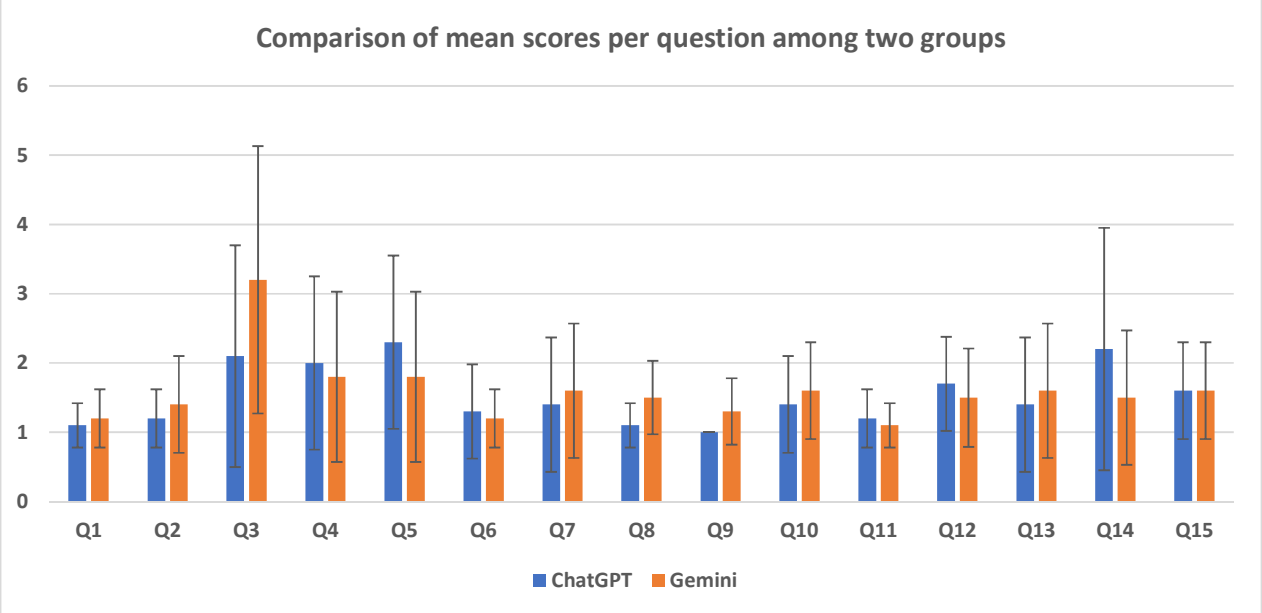


Table 1 . Most frequently asked questions about orthodontic treatment

Sno.	Questions	ChatGPT	GEMINI
1.	How do braces move teeth?		
2.	How long does it take for orthodontic treatment to be completed?		
3.	Do braces hurt?		
4.	What does it cost to get braces?		
5.	Does insurance cover braces?		
6.	What can I take if my teeth hurt from braces?		
7.	Do I have to wear a retainer forever?		
8.	How often do I need to see the orthodontist?		
9.	What is the advantage of braces vs. Invisalign?		
10.	What are the advantages of metal braces compared to clear braces?		
11.	What can I eat with braces?		
12.	When should orthodontic treatment be started?		
13.	Why should we pursue early treatment with braces?		
14.	Will I need extractions?		
15.	Will I be able to play sports/ play my instrument?		

TABLE 2. Comparison of mean scores per question among two groups

Question	ChatGPT		Gemini		p-value
	Mean	SD	Mean	SD	
How do braces move teeth?	1.10	0.32	1.20	0.42	0.542
How long does it take for orthodontic treatment to be completed?	1.20	0.42	1.40	0.70	0.549
Do braces hurt?	2.10	1.60	3.20	1.93	0.212
What does it cost to get braces?	2.00	1.25	1.80	1.23	0.676
Does insurance cover braces?	2.30	1.25	1.80	1.23	0.199
What can I take if my teeth hurt from braces?	1.30	0.68	1.20	0.42	0.914
Do I have to wear a retainer forever?	1.40	0.97	1.60	0.97	0.399
How often do I need to see the orthodontist?	1.10	0.32	1.50	0.53	0.057
What is the advantage of braces vs. Invisalign?	1.00	0.00	1.30	0.48	0.067
What are the advantages of metal braces compared to clear braces?	1.40	0.70	1.60	0.70	0.435
What can I eat with braces?	1.20	0.42	1.10	0.32	0.542
When should orthodontic treatment be started?	1.70	0.68	1.50	0.71	0.450
Why should we pursue early treatment with braces?	1.40	0.97	1.60	0.97	0.399
Will I need extractions?	2.20	1.75	1.50	0.97	0.449
Will I be able to play sports/ play my instrument?	1.60	0.70	1.60	0.70	1.000

TABLE 3. Comparison of combined mean score of the responses among two groups

Group	Mean	SD	p-value
ChatGPT	1.53	0.47	0.542
Gemini	1.59	0.52	

Mann Whitney test