BMC Oral Health

RESEARCH

Open Access

Artificial intelligence in pediatric dental trauma: do artificial intelligence chatbots address parental concerns effectively?

Mihriban Gökcek Taraç[1*] and Tuğba Nale[2]

Abstract

Background This study focused on two Artificial Intelligence chatbots, ChatGPT 3.5 and Google Gemini, as the primary tools for answering questions related to traumatic dental injuries. The aim of this study to evaluate the reliability, understandability, and applicability of the responses provided by these chatbots to commonly asked questions from parents of children with dental trauma.

Methods The case scenarios were developed based on frequently asked questions that parents commonly ask their dentists or Artificial Intelligence chatbots regarding dental trauma in children. The quality and accuracy of the information obtained from the chatbots were assessed using the DISCERN Instrument. The understandability and actionability of the responses obtained from the Artificial Intelligence chatbots were assessed using the Patient Education Materials Assessment Tool for Printed Materials. In statistical analysis; categorical variables were analyzed in terms of frequency and percentage. For numerical variables, skewness and kurtosis values were calculated to assess normal distribution.

Results Both Artificial Intelligence chatbots performed similarly, although Google Gemini provided higher quality and more reliable responses. Based on the mean scores, ChatGPT 3.5 had a higher understandability. Both chatbots demonstrated similar levels of performance in terms of actionability.

Conclusion Artificial Intelligence applications can serve as a helpful starting point for parents seeking information and reassurance after dental trauma. However, they should not replace professional dental consultations, as their reliability is not absolute. Parents should use Artificial Intelligence applications as complementary resources and seek timely professional advice for accurate diagnosis and treatment.

Keywords Artificial Intelligence, Chat GPT, Discern Instrument, Patient Education Materials Assessment Tool for Printed Materials

Introduction

Traumatic dental injuries (TDIs) represent a significant proportion of all head and neck injuries, particularly among children and young adults [1]. If not properly treated, dental trauma can result in both functional and aesthetic complications. Given that most negative outcomes are caused by inadequate or inappropriate emergency interventions, the timing of treatment is as crucial as the treatment itself [2, 3].

*Correspondence:
Mihriban Gökcek Taraç
gokcekmihriban@karabuk.edu.tr
[1] Department of Pediatric Dentistry, Karabuk University School of Dentistry, Karabük, Turkey
[2] Antalya Oral and Dental Health Hospital, Antalya, Turkey

Studies indicate that TDIs most commonly occur at home in settings where children are with their families [4]. However, public awareness regarding the emergency management of TDIs remains generally insufficient [5, 6]. Moreover, many individuals face challenges in securing timely appointments and accessing specialist consultations [7]. Consequently, the trend of seeking online advice after dental trauma has notably increased in recent years [8].

In today's technological era, artificial intelligence (AI) chatbots are increasingly being used to provide patients with access to vital information [9]. AI is defined as the ability of a system to replicate human-like intelligence [10]. AI chatbots simulate human cognitive processes using advanced algorithms and extensive datasets to generate instant, human-like responses. This capability enables AI chatbots to engage effectively with users using natural language, thus allowing for quick and accurate answers to the user's queries [11, 12].

The integration of AI into dentistry marks a considerable advancement and are a valuable tool for predicting dental diseases, diagnosing conditions, and developing treatment plans [13]. AI chatbots can save time in emergencies by offering instant, comprehensive, and personalized responses [14]. Unlike search engines that present general information from various sources, AI-based chatbots deliver information in a conversational format and simplify complex topics for the user [15, 16]. This approach is particularly beneficial in cases of TDIs as it eliminates the need to navigate multiple web pages [14].

Chat Generative Pre-Trained Transformer (ChatGPT) is a large language model that leverages both AI and machine learning to engage in conversations with users. ChatGPT uses deep learning techniques to generate human-like responses to natural language inputs and provides meaningful and contextually relevant answers based on an extensive knowledge base [17].

Another AI application introduced as an alternative to ChatGPT is Gemini, which is an AI-powered chatbot that answers user questions, engages in conversations, and generates creative content [18]. These advanced AI applications that are free to access provide enhanced capabilities in text-based interactions, delivering quick and effective solutions to users' requests [19, 20].

In health care, patients who cannot communicate directly with their doctor are often driven to seek information from various online sources [21, 22]. While all chatbots aim to provide accurate feedback and address concerns via user interaction, there are ongoing debates, particularly in health care, about the validity and reliability of AI chatbot responses [8]. A review of the literature identifies only one study that evaluates the performance of AI chatbots in answering hypothetical, case scenario-based questions related to dental trauma [14]. As such, there is a need for additional research that simulates real patient inquiries. This study aimed to assess and compare the accuracy and applicability of responses from two different AI chatbots, based on case scenarios derived from common questions that parents of children with dental trauma typically ask their doctors or AI chatbots.

The null hypothesis of the study is that the reliability, understandability, and actionability of the responses provided by the ChatGPT 3.5 and Gemini chatbots to questions related to TDIs are not significantly different.

Material and methods

This study focused on two AI chatbots, ChatGPT 3.5 and Gemini, as the primary tools for answering questions related to TDIs. The aim was to evaluate the reliability, understandability, and applicability of the responses provided by these chatbots to commonly asked questions from parents of children with TDIs.

Question design

The case scenarios used as queries for the chatbots were developed based on frequently asked questions that parents commonly ask their doctors or AI chatbots regarding dental trauma in children. These questions were developed based on inquiries frequently directed to pediatric dentists during clinical examinations regarding dental trauma. They were formulated through literature reviews and the insights of the pediatric dentists who authored the study.To assess the understandability of the case scenarios, five pediatric dentists who were selected independently of the study reviewed them. These pediatric dentists were selected based on having at least five years of experience and holding educational roles in university hospitals. 17 questions were sent electronically via Google Forms, and pediatric dentists were asked to provide feedback on each question as either "Clear" or "Needs Revision." For the questions marked as "Needs Revision," participants were asked to specify the necessary modifications in the comment section. This iterative process continued—refining and resubmitting the questions based on the participants' feedback—until all questions were clearly understood.

Chatbot processing

The prepared questions were proposed to the AI chatbots by a single individual (one of the authors). To avoid any influence from previous searches, new accounts were created on each platform. After inputting the case scenarios into the chatbots, the questions were asked using the phrase: "What should I do in such a situation?" The "new conversation" option was selected each time,

ensuring that the questions were asked independently and to prevent follow-up questions. To maintain consistency, all questions were inputted to both chatbots by the same user on the same day.

Quality assessment

The quality and accuracy of the information obtained from the AI chatbots was independently assessed by two authors using the Discern instrument (DI) and the mean scores were calculated. Before the evaluation, the authors gained proficiency in the assessment process by reviewing the literature on DI. Inter-rater and intra-rater reliability were analyzed using Cohen's kappa coefficient. The kappa coefficient ranged from 0.72 to 0.78, indicating a substantial level of agreement. During this evaluation, the accuracy and quality of the responses obtained from the chatbots were determined based on the International Association of Dental Traumatology (IADT) guidelines, and the DI scoring was conducted accordingly [23, 24]. Developed in 1999, the DISCERN tool is a 16-item measurement designed to evaluate the quality and reliability of health-care websites [25]. The DI is divided into three sections: Sect." Introduction"with eight questions that assess the reliability of the information; Sect." Material and Methods"focuses on treatment information with seven questions; and Sect." Results"includes one question that evaluates the overall quality of the information. Each question is scored from 1–5 with the total score being the sum of all 16 questions. Based on this scoring system, the total average score is used to categorize the quality of the information: a score between 16–26 is considered very poor, 27–38 is weak, 39–50 is average, 51–62 is good, and a score higher than 63 is considered excellent [26].

Evaluation of the understandability and actionability of the responses

The understandability and actionability of the responses obtained from the AI chatbots were independently assessed by two authors using the Patient Education Assessment Tool for Printed Materials (PEMAT-P) and the mean scores were calculated. Before the evaluation, the authors gained proficiency in the assessment process by reviewing the literature on PEMAT-P. PEMAT-P consists of 24 questions, divided into two sections: one that evaluates understandability (questions 1–17) and the other that evaluates actionability (questions 18–24). During the evaluation, responses are scored as 0 for "Disagree", 1 for "Agree", and NA for "Not Applicable". As no visual content was included in the questions posed to the AI chatbots, the visual questions in this analysis were marked as NA. Following the PEMAT-P guidelines, the scores for understandability and actionability were calculated separately [27]. Additionally, the understandability and actionability scores were evaluated separately for primary and permanent teeth, and a comparison was made based on the type of tooth.

Ethical committee approval

Since this study did not involve humans or animals, ethical approval was not required. The data were collected in full compliance with the terms of service of the relevant AI platforms.

Statistical analysis

Data analysis was performed with IBM SPSS Statistics version 26. Categorical variables were analyzed in terms of frequency and percentage. For numerical variables, skewness and kurtosis values were calculated to assess normal distribution, as shown in Table 1. According to the rules for normal distribution, skewness values should fall between ± 1.5, and kurtosis values (calculated by dividing the statistical value by the standard error) should be between ± 7 [28]. Based on these criteria, all values presented in Table 1exhibit normal distribution. Consequently, parametric tests (independent sample T test, one-way ANOVA, and Pearson's correlation analysis) were applied. The significance levels for the study were set at 0.05 and 0.01 [29].

Table 1 Skewness and kurtosis values of the data

Skewness			Kurtosis	
	Statistics	Std. Error	Statistics	Std. Error
Discern I Chat GPT	−0.774	0.550	0.484	1.063
Discern I GG	−1.405	0.550	3.720	1.063
PEMAT-P GG understandability	−0.593	0.550	−0.706	1.063
PEMAT-P GG actionability	−0.656	0.550	−1.419	1.063
PEMAT-P Chat GPT 3.5 understandability	−1.594	0.550	0.803	1.063
PEMAT-P Chat GPT 3.5 actionability	0.000	0.550	−2.200	1.063

Table 2 Comparison of DISCERN Scores by AI Chatbots

Value	AI Chatbots		t	*p*
	Chat GPT 3.5	Google Gemini		
Mean ± s.d	48.6 ± 12.4	51.6 ± 10.0	−0.265	0.793
Median (min–max)	51.5 (19.5–65)	51.5 (21.5–65.5)		

s.d Standard deviation, *min* Minimum, *max* Maximum, *t* Independent sample t-test

Results

A total of 17 case scenario questions containing prepared inquiries were proposed to both chatbots, resulting in a total of 34 detailed responses. The complete list of responses is provided in Appendix 1.

Accuracy and quality of the responses

The quality and accuracy of the information obtained from the AI chatbots were assessed using the DI. The mean DI scores for each chatbot are presented in Table 2. The analysis revealed no statistically significant difference between the two groups (t = −0.265, $p = 0.793$). These results suggest that both AI chatbots performed similarly, although Gemini provided more reliable and higher quality responses.

The comparison of DI scores of AI chatbots for different tooth types is presented in Table 3. No statistically significant difference was seen between the scores for permanent teeth and primary teeth for either ChatGPT 3.5 or Gemini ($p > 0.05$). These results suggest that both AI chatbots offer similar levels of quality and accuracy across different tooth types. However, the mean scores for Gemini were higher than ChatGPT 3.5 for both types, indicating that ChatGPT 3.5 might provide better overall quality and accuracy.

Understandability and actionability of the responses

The comparison of PEMAT-P scores between the AI chatbots is presented in Table 4. The results indicate that there was no statistically significant difference between ChatGPT 3.5 and Gemini in terms of understandability or actionability ($p > 0.05$). However, based on the mean scores, ChatGPT 3.5 had a higher understandability score. Both chatbots showed similar levels of performance in terms of actionability.

The comparison of PEMAT-P understandability scores of AI chatbots for different tooth types is presented in Table 5. No statistically significant difference was seen between the scores for permanent teeth and primary teeth for either ChatGPT 3.5 or Gemini ($p > 0.05$). These results suggest that both AI chatbots offer similar levels

Table 3 Comparison of DI Scores by Tooth Types and AI Chatbots

Variable	Value	AI Chatbots		t	p
		Chat GPT 3.5	Google Gemini		
Permanent teeth	Mean ± s.d.	49.3±7.2	52.8±6.3	-0.362	0.673
	Median (min-max)	51.5 (19.5-65)	51.5 (21.5-65.5)		
Primary teeth	Mean ± s.d.	47.8±5.4	50.4±6.2	0.283	0.314
	Median (min-max)	51.5 (19.5-65)	51.5 (21.5-65.5)		
Statistics[a]		t=0.413	t=0.216		
		p=0.654	p=0.451		

s.d. Standard deviation, *min* Minimum, *max* Maximum, *t* Independent sample t-test

[a] Dependent sample t-test

Table 4 Comparison of PEMAT-P Scores by AI Chatbots

Variable	Value	AI Chatbots		t	p
		Chat GPT 3.5	Google Gemini		
PEMAT understandability	Mean ± s.d.	66.0±6.1	62.0±7.4	1.744	0.091
	Median (min-max)	69.2 (53.8-69.2)	61.5 (46.2-69.2)		
PEMAT actionability	Mean ± s.d.	40±20	39±26	0.148	0.883
	Median (min-max)	40 (20-60)	60 (0-60)		

s.d. Standard deviation, *min* Minimum, *max* Maximum, *t* Independent sample t-test

Table 5 Comparison of PEMAT-Understandability Scores by Tooth Types and AI Chatbots

Variable	Value	AI Chatbots		t	p
		Chat GPT 3.5	Google Gemini		
Permanent teeth	Mean ± s.d.	66.3±5.7	62.5±8.6	1.047	0.313
	Median (min-max)	69.2 (53.8-69.2)	61.5 (46.2-69.2)		
Primary teeth	Mean ± s.d.	65.8±6.8	61.5±6.7	1.348	0.540
	Median (min-max)	69.2 (53.8-69.2)	61.5 (53.8-69.2)		
Statistics[a]		t=0.174	t=0.262		
		p=0.864	p=0.797		

s.d. Standard deviation, *min* Minimum, *max* Maximum, *t* Independent sample t-test

[a] Dependent sample t-test

of understandability across different tooth types. However, the mean scores for ChatGPT 3.5 were higher than Gemini for both types, indicating that ChatGPT 3.5 might provide better overall understandability.

The comparison of PEMAT-P actionability scores by AI chatbots for different tooth types is Table 6. Overall, there was no statistically significant difference between ChatGPT 3.5 and Gemini in terms of actionability scores for both permanent and primary teeth ($p > 0.05$). However, it is worth noting that ChatGPT 3.5 had a higher mean score for primary teeth, while Gemini achieved higher mean scores for permanent teeth. ChatGPT 3.5 demonstrated a more consistent performance, whereas Gemini showed greater variability in its results. These findings suggest that both AI chatbots exhibit similar levels of actionability, but their performance can differ depending on the tooth type included in the query.

Discussion

In health care, the inability of patients to communicate with their doctors at all times or learn from the experiences of individuals with similar medical histories often leads patients to seek information from various online sources [21, 22]. Information from the internet is valuable due to its constant accessibility and the variety of perspectives it provides [30]. Prominent examples of AI applications include large language models, such as ChatGPT 3.5 and Gemini. The accuracy and performance of these models in clinical settings are crucial for both patients and clinicians [31]. Consequently, this study evaluated the responses of AI chatbots to frequently asked questions from families regarding simulated dental trauma case scenarios.

The reliability of AI chatbots largely depends on their ability to provide accurate, clear, and contextually appropriate information. Research evaluating the responses of AI chatbots in health care scenarios, including dental trauma, has highlighted both the strengths and limitations of these technologies. While most studies assessing the validity and reliability of AI chatbots as information sources in dentistry have focused on endodontics [32, 33], research on dental traumatology remains limited. Given that dental trauma is one of the most common dental issues encountered in daily life and is frequently searched online, there is a clear need for further research in this area [34].

When evaluating a chatbot, the language used and the types of questions addressed are important; however, frequently asked questions during clinical visits also play a crucial role. These questions, which cover the types of traumas and the psychological and physiological effects of dental trauma on patients, help assess how effectively the chatbot responds to specific situations [8]. In this study, similar to the work by Güven et al., (2024) [14] simulated case scenarios were created and the responses from AI chatbots were evaluated. This approach led to more personalized answers regarding dental trauma. Furthermore, when parents express their concerns in greater detail, the quality and reliability of the responses from AI chatbots improve.

Çalışmamızda cevapların doğruluğu ve kalitesi DI ile değerlendirilmiştir. DI, başlangıçta yazılı sağlık bilgilerinin kalitesini değerlendirmek amacıyla geliştirilmiş olsa da, web siteleri, YouTube videoları ve son zamanlarda yapay zekâ sohbet botları gibi çevrim içi hizmetlerin değerlendirilmesi için de uyarlanmıştır [14]. İnternetteki tıbbi bilgilerin güvenilirliğini, makullüğünü ve kullanılabilirliğini belirlemek amacıyla HONcode ve JAMA kriterleri ile Ensuring Quality Information for Patient (EQIP) aracı da kullanılmaktadır. McCool ve arkadaşları tarafından yapılan çalışmalarda DI, Ensuring Quality Information for Patients (EQIP) aracına kıyasla daha yüksek iç tutarlılık ve gözlemciler arası güvenilirlik, daha iyi uyum ve daha hassas değerlendirme yeteneği sergilemiştir [35]. Doubleday ve ark. HONcode'un web tabanlı bilgilerin kalitesini değerlendirmek için yeterli bir araç olmadığını çalışmalarında belirtmiştir. Ayrıca HONcode ise 15 Aralık 2022 tarihinden itibaren kalıcı olarak

Table 6 Comparison of PEMAT-Actionability Scores by AI Chatbots

Variable	Value	AI Chatbots		t	p
		Chat GPT 3.5	Google Gemini		
Permanent teeth	Mean ± s.d.	35±20.7	40±28.0	-0.403	0.316
	Median (min-max)	20 (20-60)	60 (0-60)		
Primary teeth	Mean ± s.d.	44.44±19.4	37.8±25.4	0.626	0.540
	Median (min-max)	60 (20-60)	40 (0-60)		
Statistics[a]		t=0.970	t=0.171		
		p=0.347	p=0.867		

s.d. Standard deviation, *min* Minimum, *max* Maximum, *t* Independent sample t-test

[a] Dependent sample t-test

sonra erdirilmiş ve artık güncellenmemektedir [36]. JAMA kriterlerinin de dereceli bir ölçek yerine evet/hayır şeklinde ikili sorular kullanmasından dolayı, bazı web siteleri için ne derece geçerli olduğu konusunda da soru işaretleri mevcuttur [37].

When assessing the quality of responses according to the DI categorization, responses from ChatGPT 3.5 were classified as "moderate", while Gemini responses were "good". These results suggest higher quality and more reliable responses from Gemini compared with ChatGPT 3.5, although the data remains somewhat limited. A similar study by Güven et al., (2024) found that Gemini had higher DI scores when compared with three other AI chatbots [14]. In contrast, a study by Behers et al., (2024) evaluating the quality of patient education materials on cardiac catheterization from AI chatbots found that ChatGPT had higher DI scores [38]. AI chatbots effectively engage with users by processing text and previous dialogue data, adopting a dialogue-based approach, and learning from user interaction. However, continuous learning and training on the latest information are essential for improving the quality of chatbot responses [39, 40]. As AI technology evolves, the continuous development of response generation capabilities might explain the variations in response quality and reliability observed in studies conducted at different times [41].

Tariq et al., (2024) emphasized the need for a standardized evaluation framework to assess AI models in health care. This framework should not only measure the accuracy and scope of the information but also ensure that responses address patients' diverse needs and backgrounds, and that they are both understandable and actionable [42]. In this study, PEMAT-P was used to evaluate the understandability and actionability of responses. No statistically significant differences were observed in terms of understandability, although ChatGPT 3.5 provided better understandability. Regarding actionability, both AI applications showed similar levels of competence. In contrast to our findings, Güven et al., (2024) [14] reported that the understandability of ChatGPT 3.5 was lower compared with other AI chatbots, while actionability results were consistent with those observed in this study.

Suárez et al., (2024), in their assessment of ChatGPT's accuracy and consistency, concluded that while AI applications show promise, they are not yet sufficient for clinical decision making [33]. In this study, a detailed evaluation of the responses from ChatGPT 3.5 and Gemini revealed several shortcomings. For instance, neither chatbot addressed the role of an open or closed apex in root canal treatment for dental trauma. Additionally, in question 16, while GG suggests that the hardness felt in the lip may result from a broken tooth fragment embedded in the lip, ChatGPT does not mention this possibility. Furthermore, in question 4, ChatGPT 3.5 explained that: "For a 4-year-old child, since the anterior teeth are most likely primary teeth, replantation is generally not recommended" while Gemini did not mention about the replantation procedure of primary teeth. However, upon comparative evaluation, GG's responses were found to be more comprehensive and valid than those of ChatGPT 3.5. Güven et al., (2024) also noted that ChatGPT mistakenly recommended the replantation of primary teeth in cases of avulsion injuries [14].

These differences are believed to arise from the AI chatbots interpreting open-ended clinical questions in different ways. Additionally, responses to the same questions might vary when asked at different times. Özden et al., (2024) also reported that Gemini did not achieve a sufficient level of consistency and provided different responses at various times [43]. Furthermore, large language models are not specifically trained in specialized fields, such as endodontics and dental traumatology, which can affect the accuracy and applicability of their responses in clinical settings. These models are typically trained with general language and knowledge datasets,

limiting their ability to provide detailed information on specific dental subspecialties.

Although AI chatbots generally achieve highly valid scores, critical errors in some responses and the potential to mislead the public on certain topics can lead to considerable issues, especially in health care contexts. ChatGPT 3.5 is more popular among patients than other AI platforms because of its free access, lack of registration requirements, and its longer presence in the market compared with other AI platforms [44]. However, since this model was trained on a broad range of online sources, including books, scientific articles, and websites up until 2021, there is a risk of it providing outdated or limited medical advice [17]. In contrast, while Gemini is free, it requires logging in with a Google account [44]. Med-Gemini, a health care-specialized model developed by Google, holds greater potential for the future. Equipped with the ability to seamlessly integrate web searches and an evolving database, Med-Gemini is seen as a promising tool for accurate medical dialogues, medical research, and health education [45]. These advancements could considerably improve the validity and reliability of Gemini in providing medical advice, positioning it as a more effective AI platform than ChatGPT 3.5.

Güven et al., (2024) noted that ChatGPT 3.5 struggled to accurately assess the relationship between a pediatric patient's age and their dental development stage based on information provided by parents, resulting in more superficial responses [14]. In this study, the questions were specifically tailored to the patients' ages, and the AI chatbots were expected to accurately identify dental types based on age and provide appropriate suggestions. In this context, the understandability and actionability of responses related to primary and permanent teeth were evaluated separately. Although no statistically significant difference was found, ChatGPT 3.5 showed more consistent performance across both primary and permanent teeth, whereas Gemini showed more variability in its responses. This finding suggests that both AI chatbots have similar capabilities in terms of actionability, but their performance might differ depending on the type of teeth that is queried.

This study has several limitations. Although the case scenarios were developed based on real expressions from patients' relatives to accurately reflect the patient's condition, AI chatbots might not capture all the nuances of the situation, which typically requires a clinical assessment by a dentist. Furthermore, the AI chatbots used in this study were not specifically trained in dental or dental trauma topics, meaning that their responses might lack the depth and accuracy needed for clinical contexts. Finally, while the authors followed specific guidelines when evaluating the responses, subjective judgments could have influenced the scoring.

Dental trauma, especially in children, can be a major source of concern for parents. However, consulting a specialist is not always possible. In such cases, parents might turn to AI-powered chatbots to alleviate their concerns and obtain information. While AI chatbots have the potential to guide parents regarding dental trauma and possible complications; however, certain shortcomings have been observed in the responses provided.

This comparison offers valuable insights into the effectiveness of AI chatbots and contributes to a deeper understanding of their applications in dentistry. Specifically, the data highlights the potential advantages and limitations of AI in dental traumatology and general dental practices. These findings underscore the importance of further research to improve and refine AI technologies in the health-care sector, ensuring that their integration leads to more reliable, accurate, and actionable information for both practitioners and patients.

Conclusion

AI applications can serve as a helpful starting point for parents seeking information and reassurance after dental trauma. However, they should not replace professional dental consultations as their reliability is not absolute. Parents should use AI applications as a complementary resource and seek timely professional advice for accurate diagnosis and treatment. Additionally, parents should be trained by pediatric dentists on dental trauma.

Abbreviations

AI	Artificial Intelligence
TDI	Traumatic dental injuries
ChatGPT	Chat Generative Pre-Trained Transformer
DI	Discern Instrument
PEMAT-P	Patient Education Assessment Tool for Printed Materials

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12903-025-06105-z.

Additional file 1.

Acknowledgements

I would like to express my sincere gratitude to the 5 pediatric dentists who participated in the pilot study for their valuable insights, which contributed to the clarity of the questions.

Human ethics and consent to participate declarations

Not applicable.

Clinical trial number

Not applicable.

Authors' contributions
Study concepts: MGT and TN Study design: MGT and TN Data acquisition: MGT and TN Quality control of data and algorithms: MGT Data analysis and interpretation: MGT and TN Statistical analysis: MGT Manuscript preparation: MGT, TN Manuscript editing: MGT and TN All authors reviewed the manuscript.

Funding
No financial support was received from any institution or person.

Data availability
The data generated in this study are available upon request from the authors.

Declarations

Ethics approval and consent to participate
Since this study did not involve human or animal subjects, ethical approval was not required.

Consent for publication
Not applicable

Competing interests
The authors declare no competing interests.

Received: 9 February 2025 Accepted: 5 May 2025
Published online: 17 May 2025

References

1. Andersson L, Petti S, Day P, Kenny K, Glendor U, Andreasen JO. Classification, Epidemiology and Etiology. In: Andreasen JO, Andreasen FM, Andersson L, editors. Textbook and Color Atlas of Traumatic Injuries to the Teeth. Copenhagen: Wiley Blackwell; 2019. p. 252–94.
2. Al-Jundi SH. Dental emergencies presenting to a dental teaching hospital due to complications from traumatic dental injuries. Dent Traumatol. 2002;18(4):181–5. https://doi.org/10.1034/j.1600-9657.2002.02081.x.
3. Zuhal K, Semra OE, Hüseyin K. Traumatic injuries of the permanent incisors in children in southern Turkey: a retrospective study. Dent Traumatol. 2005;21(1):20–5. https://doi.org/10.1111/j.1600-9657.2004.00265.x.
4. Azami-Aghdash S, Ebadifard Azar F, Pournaghi Azar F, Rezapour A, Moradi-Joo M, Moosavi A, Ghertasi OS. Prevalence, etiology, and types of dental trauma in children and adolescents: systematic review and meta-analysis. Med J Islam Repub Iran. 2015;29(4):234.
5. Bamashmous N, Cunningham SJ, Parekh S. Information seeking behaviour of dental trauma patients and their parents. Dent Traumatol. 2020;36(6):590–7. https://doi.org/10.1111/edt.12581. (Epub 2020 Jul 15).
6. Tewari N, Goel S, Srivastav S, Mathur VP, Rahul M, Haldar P, et al. Global Status of Knowledge of Parents for Emergency Management of Traumatic Dental Injuries: A Systematic Review and Meta-Analysis. Evid Based Dent. 2023;24(2):91. https://doi.org/10.1038/s41432-023-00883-7.
7. Erwin J, Horrell J, Wheat H, Axford N, Burns L, Booth J, Witton R, Shawe J, Doughty J, Kaddour S, Boswell S, Devalia U, Nelder A, Paisi M. Access to dental care for children and young people in care and care leavers: a global scoping review. Dent J (Basel). 2024;12(2):37. https://doi.org/10.3390/dj12020037.
8. Johnson AJ, Singh TK, Gupta A, Sankar H, Gill I, Shalini M, et al. Evaluation of validity and reliability of AI Chatbots as public sources of information on dental trauma. Dent Traumatol 2024 Oct 17. Epub ahead of print. https://doi.org/10.1111/edt.13000.
9. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol. 2023;9(10):1437–40. https://doi.org/10.1001/jamaoncol.2023.2947.
10 Katne T, Kanaparthi A, Gotoor S, Muppirala S, Devaraju R, Gantala R. Artificial intelligence: demystifying dentistry – The future and beyond. Int J Contemp Med Surg Radiol. 2019;4:D6-9. https://doi.org/10.21276/ijcmsr.2019.4.4.2.
11. Rodrigues JA, Krois J, Schwendicke F. Demystifying artificial intelligence and deep learning in dentistry. Braz Oral Res. 2021;13(35):e094. https://doi.org/10.1590/1807-3107bor-2021.vol35.0094.
12. Khanagar SB, Alfadley A, Alfouzan K, Awawdeh M, Alaqla A, Jamleh A. Developments and performance of artificial intelligence models designed for application in endodontics: a systematic review. Diagnostics (Basel). 2023;13(3):414. https://doi.org/10.3390/diagnostics13030414.
13. Agrawal P, Nikhade P. Artificial intelligence in dentistry: past, present, and future. Cureus. 2022;14(7):e27405. https://doi.org/10.7759/cureus.27405.
14. Guven Y, Ozdemir OT, Kavan MY. Performance of Artificial Intelligence Chatbots in Responding to Patient Queries Related to Traumatic Dental Injuries: A Comparative Study. Dent Traumatol 2024 Nov 22. Epub ahead of print. https://doi.org/10.1111/edt.13020.
15. Pandey S, Sharma S. A comparative study of retrieval-based and generative-based chatbots using deep learning and machine learning. Healthc Anal. 2023;3:100198. https://doi.org/10.1016/j.health.2023.100198.
16. Chiesa-Estomba CM, Lechien JR, Vaira LA, Brunet A, Cammaroto G, Mayo-Yanez M, Sanchez-Barrueco A, Saga-Gutierrez C. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. Eur Arch Otorhinolaryngol. 2024;281(4):2081–6. https://doi.org/10.1007/s00405-023-08267-4.
17. Kim HW, Shin DH, Kim J, Lee GH, Cho JW. Assessing the performance of ChatGPT's responses to questions related to epilepsy: A cross-sectional study on natural language processing and medical information retrieval. Seizure. 2024;114:1–8. https://doi.org/10.1016/j.seizure.2023.11.013.
18 Koçyiğit A, Darı AB. ChatGPT in artificial intelligence communication: the future of humanized digitalization. J Strateg Soc Res. 2023;7(2):427–38. https://doi.org/10.30692/sisad.1311336.
19. Vaswani A, Shazeer N, Parmar N, et al., "Attention Is all You Need," *arXiv* 7 (2023): 1–15.
20. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT vs Google bard. Radiology. 2023;307(5):e230922. https://doi.org/10.1148/radiol.230922.
21. Cobb RJ, Scotton WJ. The reliability of online information regarding orthognathicsurgery available to the public. Oral Surg. 2013;6(2):56–60. https://doi.org/10.1111/ors.12020.
22. Engelmann J, Fischer C, Nkenke E. Quality assessment of patient information onorthognathic surgery on the internet. J Craniomaxillofac Surg. 2020;48(7):661–5. https://doi.org/10.1016/j.jcms.2020.05.004.
23. Bourguignon C, Cohenca N, Lauridsen E, Flores MT, O'Connell AC, Day PF, Tsilingaridis G, Abbott PV, Fouad AF, Hicks L, Andreasen JO, Cehreli ZC, Harlamb S, Kahler B, Oginni A, Semper M, Levin L. International Association of Dental Traumatology guidelines for the management of traumatic dental injuries: 1. Fractures and luxations Dent Traumatol. 2020;36(4):314–30. https://doi.org/10.1111/edt.12578.
24. Day PF, Flores MT, O'Connell AC, Abbott PV, Tsilingaridis G, Fouad AF, Cohenca N, Lauridsen E, Bourguignon C, Hicks L, Andreasen JO, Cehreli ZC, Harlamb S, Kahler B, Oginni A, Semper M, Levin L. International Association of Dental Traumatology guidelines for the management of traumatic dental injuries: 3. Injuries in the primary dentition. Dent Traumatol. 2020;36(4):343–59. https://doi.org/10.1111/edt.12576.
25. Tan DJY, Ko TK, Fan KS. The Readability and Quality of Web-Based Patient Information on Nasopharyngeal Carcinoma: Quantitative Content Analysis. JMIR Form Res. 2023;27(7):e47762. https://doi.org/10.2196/47762.
26. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health. 1999;53(2):105–11. https://doi.org/10.1136/jech.53.2.105.
27. Shoemaker SJ, Wolf MS, Brach C. The Patient Education Materials Assessment Tool (PEMAT) and User's Guide. An Instrument To Assess the Understandability and Actionability of Print and Audiovisual Education Materials (Version 1.0) Agency for Healthcare Research and Quality. Publication No.14–0002-EF November 2013 Updated August 2014. (Rockville, MD:, 2013). Available from: https://www.ahrq.gov/sites/default/files/publications/files/pemat_guide.pdf
28. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. Boston: Pearson; 2013.
29. Büyüköztürk Ş. Sosyal Bilimler İçin Veri Analiz El Kitabı. 27th Ed. Ankara: PEGEM Yayınları, 2020
30. Ghanem YK, Rouhi AD, Al-Houssan A, Saleh Z, Moccia MC, Joshi H, Dumon KR, Hong Y, Spitz F, Joshi AR, Kwiatt M. Dr. Google to Dr. ChatGPT:

assessing the content and quality of artificial intelligence-generated medical information on appendicitis. Surg Endosc. 2024;38(5):2887-2893. https://doi.org/10.1007/s00464-024-10739-5.
31. 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.
32. 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.
33. 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. https://doi.org/10.1111/iej.13985.
34. Petti S, Glendor U, Andersson L. World traumatic dental injury prevalence and incidence, a meta-analysis-One billion living people have had traumatic dental injuries. Dent Traumatol. 2018;34(2):71–86. https://doi.org/10.1111/edt.12389.
35. McCool ME, Wahl J, Schlecht I, Apfelbacher C. Evaluating written patient information for eczema in German: comparing the reliability of two instruments, DISCERN and EQIP. PLoS ONE. 2015;10(10):e0139895. https://doi.org/10.1371/journal.pone.0139895.
36. Doubleday AR, Novin S, Long KL, Schneider DF, Sippel RS, Pitt SC. Online information for treatment for low-risk thyroid cancer: assessment of timeliness, content, quality, and readability. J Cancer Educ. 2021;36(4):850–7. https://doi.org/10.1007/s13187-020-01713-5.
37. Bharmal RV, Johal A. Development and assessment of reliable patient-based hypodontia website. Patient Prefer Adherence. 2018;19(12):1065–70. https://doi.org/10.2147/PPA.S153438.
38. Behers BJ, Stephenson-Moe CA, Gibons RM, Vargas IA, Wojtas CN, Rosario MA, et al. Assessing the quality of patient education materials on cardiac catheterization from artificial intelligence chatbots: an observational cross-sectional study. Cureus. 2024;16(9):e69996. https://doi.org/10.7759/cureus.69996.
39. Ouyang L, Wu J, Jiang X, et al. Training Language Models to Follow Instructions With Human Feedback. *arXiv* 1 (2022):1–68.
40. Thoppilan R, Freitas DD, Hall J, et al. LaMDA: Language Models for Dialog Applications. *arXiv* 3 (2022):1–47.
41. Sharma D, Vidhate DA, Osei-Asiamah J, Kumari M, Mahajan V, Rajagopal K. Exploring the Evolution of Chatgpt: From Origin to Revolutionary Influence. Educ Adm Theory Pract. 2024;30(5):2685–92. https://doi.org/10.53555/kuey.v30i5.3329.
42. Tariq R, Khanna S. Artificial Intelligence-Powered Patient Education for Comprehensive and Individualized Understanding for Patients. Clin Gastroenterol Hepatol. 2024;22(7):1550–1. https://doi.org/10.1016/j.cgh.2023.10.027.
43. Ozden I, Gokyar M, Ozden ME, Sazak OH. Assessment of artificial intelligence applications in responding to dental trauma. Dent Traumatol. 2024;40(6):722–9. https://doi.org/10.1111/edt.12965.
44. 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. https://doi.org/10.1111/jerd.13046.
45. Saab K, Tu T, Weng WH, Tanno R, Stutz D, Wulczyn E, et al. Capabilities of Gemini Models in Medicine. *arXiv* 2024;2:1–58. https://arxiv.org/abs/2404.18416.

Publisher's Note