RESEARCH   **Open Access**

Can ChatGPT be guide in pediatric dentistry?

Canan Bayraktar Nahir[1]*

Abstract

Background The use of ChatGPT in the field of health has recently gained popularity. In the field of dentistry, ChatGPT can provide services in areas such as, dental education and patient education. The aim of this study was to evaluate the quality, readability and originality of pediatric patient/parent information and academic content produced by ChatGPT in the field of pediatric dentistry.

Methods A total of 60 questions were asked to ChatGPT for each topic (dental trauma, fluoride, and tooth eruption/oral health) consisting of pediatric patient/parent questions and academic questions. The modified Global Quality Scale (the scoring ranges from 1: poor quality to 5: excellent quality) was used to evaluate the quality of the answers and Flesch Reading Ease and Flesch-Kincaid Grade Level were used to evaluate the readability. A similarity index was used to compare the quantitative similarity of the answers given by the software with the guidelines and academic references in different databases.

Results The evaluation of answers quality revealed an average score of 4.3 ± 0.7 for pediatric patient/parent questions and 3.7 ± 0.8 for academic questions, indicating a statistically significant difference ($p < 0.05$). Academic questions regarding dental trauma received the lowest scores ($p < 0.05$). However, no significant differences were observed in readability and similarity between ChatGPT answers for different question groups and topics ($p > 0.05$).

Conclusions In pediatric dentistry, ChatGPT provides quality information to patients/parents. ChatGPT, which is difficult to readability for patients/parents and offers an acceptable similarity rate, needs to be improved in order to interact with people more efficiently and fluently.

Keywords Artificial intelligence, Fluorides, Tooth injuries, Tooth eruption, Oral health, Public health informatics, Pediatric dentistry

Background

The use of Artificial Intelligence (AI)-supported tools in the delivery of healthcare services has gained popularity recently. Among these tools, Chat Generative Pre-trained Transformer (ChatGPT), released in November 2022, is one of the most advanced Natural Language Processing (NLP) models developed by OpenAI (OpenAI, L. L. C.,

San Francisco, CA, USA) [1]. ChatGPT mimics human language processing abilities using deep learning and neural networks to generate human-like [2].

In the field of dentistry, ChatGPT can provide services in areas such as digital data recording, image analysis, diagnosis and treatment planning, dental telemedicine, dental education and patient education [3–5]. Technological advances have led both healthcare professionals and patients to increasingly turn to ChatGPT as a convenient source for medical and dental information [6]. ChatGPT is reported to be beneficial for patients as it provides instant feedback and provides information about their health status [7]. Balel emphasised in his

*Correspondence:
Canan Bayraktar Nahir
canannbayraktarr@gmail.com
[1]Present address: Department of Pediatric Dentistry, Faculty of Dentistry, Tokat Gaziosmanpaşa University, Tokat, Türkiye

study that ChatGPT has significant potential as a patient information tool in oral and maxillofacial surgery, but its use in education may not be completely safe at this time [8]. Berkant et al., when comparing the official answers to the questions people frequently ask about fluoride with ChatGPT's answers, reported that ChatGPT answers were sufficient and comprehensive and that patients/parents could access accurate information this way [9].

According to the American Academy of Pediatric Dentistry (AAPD), it is advocated that children's first dentist visit should be within the first year of life [10]. Thus, together with parental counseling, evaluation of symptoms related to tooth eruption, oral health risk assessment and oral hygiene training of children can be provided [11]. During this period, early childhood caries (ECC) is the most common dental problem encountered in children. During children's dental visits, carious lesions can be identified and parents can be educated about preventive measures. Fluoride, one of these preventive measures and a key element of successful caries prevention, has been used for many years to prevent dental caries [12]. Despite its benefits, overexposure to fluoride can lead to dental fluorosis, skeletal fluorosis, and other health issues, such as endocrine disruption and neurotoxicity [13]. Additionally, misinformation about fluoride can contribute to public fear and decreased use of fluoride, potentially leading to higher rates of dental caries [14]. Therefore, patients are confused about the amount of fluoride considered safe to use. In particular, anti-fluoride information is widely shared on social media, misleading people about the benefits of fluoride [15]. It is of great importance that both parents and clinicians form their perspectives with accurate information. In addition, it has been reported that traumatic dental injuries constitute 5–17% of all injuries and are the 5th most common disease in the World [16]. In these injuries where emergency intervention is important, the knowledge of the parents/carers who will provide first aid at the time of trauma and the ability of the clinician to control the emergency situation are of great importance. Parents may encounter such problems related to oral and dental health from the moment their children's first teeth begin to erupt. Parents who cannot be in constant contact with doctors about their children's dental health or simply out of curiosity may use internet-based information. The use of ChatGPT is already a reality. In a rapidly evolving AI, the quality of prompts provided to AI tools such as ChatGPT is crucial in ensuring relevant, accurate and detailed responses. With ChatGPT it is common for logical errors or hallucinations - the creation of information that has no factual basis [17]. Creating a general structure for the model to incorporate into its response through prompt engineering or providing cues such as keywords requires the model to explain its responses [18]. However, there is still uncertainty about accessing high-quality and reliable information through ChatGPT without the use of prompt engineering [19]. Currently, there is no study evaluating the effectiveness of ChatGPT in providing patient information and academic information in the field of pediatric dentistry. In this respect, this research can provide important data on the reliability and effectiveness of AI-supported information assurance systems and provide insight into future applications by filling the knowledge gaps in this field.

The aim of this study is to evaluate the quality, readability, and originality of pediatric patient/parent information and academic content generated by ChatGPT in the field of pediatric dentistry. The null hypothesis of this study is that there is no significant difference between the quality, readability and originality of pediatric patient/parent information and academic content produced by ChatGPT on children's oral and dental health and existing standard information sources.

Methods
Study design and ethics
This cross-sectional study was conducted in accordance with the Declaration of Helsinki. Ethical approval was not required because publicly available data was assessed and did not include any biological materials obtained from humans or animals.

Study protocols
In the study, ChatGPT-3.5 (OpenAI, 2022) was utilized to acquire information provided by an artificial intelligence model. ChatGPT-3.5 was chosen because this AI model is free of charge and thus accessible to everyone. The questions that pediatric patients/parents ask to artificial intelligence about the field of pediatric dentistry and academic questions that can be used for dental educational purposes have been determined. A total of 60 questions were asked to ChatGPT for each topic (dental trauma, fluoride, and tooth eruption/oral health) consisting of pediatric patient/parent questions and academic questions (Supplementary material 1).

In determining the patient questions, potential questions that paediatric patients/parents could ask the software were investigated. For this purpose, hospital and clinic websites, guides and articles for pediatric patients and parents, where information about pedodontics procedures are provided in a non-terminological language, were examined. Clinical textbooks and literature were used to determine academic questions. The questions were randomly selected according to clinical experience, considering the questions frequently asked by parents in the clinic and the practices frequently performed in the clinic. Questions were not systematically categorised according to varying levels of difficulty.

The ChatGPT-3.5 implementation was accessed through Google's (Google Inc., California, United States) search engine query entry. In order to avoid the influence of search algorithms, access (https://chat.openai.com) was provided on October 29, 2023, with a new account created for the purposes of this study. All search history and cookies on the computer were cleared before the question-answer session. In order to minimize the impact of previous answers, a new chat window has been opened for each category asked. According to the given command, ChatGPT's first answer was recorded and no answer was requested again with extra commands. The questions consist of pediatric patient/parent questions and academic questions. A typical user interaction without the use of prompt engineering was evaluated, given that people directly communicate their problem to chatbots in emergencies. The answers were recorded in a word document for later evaluation (Supplementary material 2).

Experienced pedodontists were contacted via publicly accessible e-mail links to evaluate the quality of the answers given by ChatGPT. The 12 pedodontists who agreed to the evaluation were trained online to use the modified Global Quality Scale (GQS).

Evaluation criteria

The quality of the answers provided by ChatGPT was evaluated using the modified GQS. According to this scale, the pedodontists used a score between 1 and 5 points to assess the quality of the answers. Answers of poor quality, with poor information flow and most of the information missing were given a score of 1, while answers of excellent quality and information flow were given a score of 5. In the ranging of this scoring, especially the lack of information in the content was taken into consideration. The scores with the highest frequency for each question were considered as the score for that question (Fig. 1) [8].

Flesch Reading Ease and Flesch-Kincaid Grade Level tests were used to evaluate the readability level of the answers given by the software. Essentially, the formulas for both tests are based on calculating the average number of words per sentence and the average number of syllables per word by assigning different weights to each of these measures. The calculations are as follows [20, 21]:

Flesch Reading Ease Formula=206.835−1.015×(total words/total sentences) − 84.6×(total syllables/total words).

Flesch-Kincaid Grade Level Formula=0.39×(total words/total sentences)+11.8×(total syllables/total words) − 15.59.

The Flesch Reading Ease score rates text on a 100-point scale. Values of 90–100 indicate that it is easily understandable by 5th grade students; values of 80–90 indicate 6th grade students; values of 70–80 indicate 7th grade students; values of 60–69 indicate 8th or 9th grade students; values of 50–60 indicate 10th or 12th grade students; values of 30–50 indicate college students; values of 30−10 indicate college graduates; values of 10−0 indicate professional [20, 22].

The Flesch-Kincaid Grade Level score rates the text at the U.S. elementary school level. The values vary from 0 to 18, where 18 represents the most difficult text. Values between 0 and 6 indicate basic reading level, values between 6 and 12 indicate average reading level and values between 12 and 18 indicate advanced reading level. Values of 0–3 indicate that it is easily understandable by kindergartes or elementary school students; values of 3–6 indicate elementary students; values of 6–9 indicate middle school students; values of 9–12 indicate high school students; values of 12–15 indicate college students; values of 15–18 indicate college graduates [21, 23].

Original Global Quality Scale
Score 1: Poor quality, poor flow of the video, most information missing, not at all useful for patients
Score 2: Generally poor quality and flow, some information listed but many important topics missing, of very limited use to patients
Score 3: Moderate quality, suboptimal flow, some important information adequately discussed but others poorly discussed, somewhat useful for patients
Score 4: Good quality and generally good flow. Most of the relevant information is listed but some topics are not listed. useful for patients
Score 5: Excellent quality and flow, very useful for patients

Modified Global Quality Scale
Score 1: Poor quality, poor flow of the information, most information missing, not at all useful for patients or education
Score 2: Generally poor quality and flow, some information listed but many important topics missing, of very limited use to patients or education
Score 3: Moderate quality, suboptimal flow, some important information adequately discussed but others poorly discussed, somewhat useful for patients or education
Score 4: Good quality and generally good flow. Most of the relevant information is listed but some topics are not listed. useful for patients or education
Score 5: Excellent quality and flow, very useful for patients or education

Fig. 1 Global quality scale and modified global quality scale

Table 1 Demographics characteristics of the pediatric dentists

	Gender		Test Statistics	p
	Female	**Male**		
Age			0.646	0.536*
N (%)	7 (70)	3 (30)		
Mean ± sd	35.9 ± 3.7	34.0 ± 5.3		
Professional Experience Years			1.076	0.313*
N (%)	7 (70)	3 (30)		
Mean ± sd	10.0 ± 2.3	8.0 ± 3.6		
Artificial Intelligence Experience			-	0.333**
Yes (%)	5 (71.4)	1 (33.3)		
No (%)	2 (28.6)	2 (66.7)		

*Independent Samples Test

**Fisher's Exact Test

N: Number of the pediatric dentists sd: Standard deviation

Table 2 Evaluation of the quality, readability and similarity of answers to pediatric patient/parent questions and academic questions

	Answer to pediatric patient/parent questions		Answer to academic questions		*p*
	Mean ± sd	**Median (Min-Max)**	**Mean ± sd**	**Median (Min-Max)**	
modified GQS	4.3 ± 0.7	4.0 (2–5)	3.7 ± 0.8	4.0 (2–5)	**0.005***
Flesch Reading Ease	41.5 ± 12.5	40.4 (14.6–67.8)	34.0 ± 11.3	32.8 (9.5–57.7)	**0.017****
Flesch-Kincaid Grade Level	12.7 ± 2.4	12.8 (8.4–17.9)	13.7 ± 1.9	13.7 (10.2–17.7)	0.091**
Similarity index	8.4% ± 8.4	4.0% (0–25)	5.7% ± 5.1	4.5% (0–21)	0.383*

*Mann-Whitney U Test

**Independent Sample T

GQS: Global Quality Scale, N: Number of the answer sd: Standard deviation, Min: Minimum, Max: Maximum

N=30 answers for each category Answer to Pediatric Patient/Parent and Academic Questions

For most standard documents, Microsoft recommends aiming for a Flesch Reading Ease score of approximately 60 to 70 and a Flesch-Kincaid Grade Level score of approximately 7.0 to 8.0 [24, 25].

Similarity Index was used to determine the quantitative similarity between the answers given by the software and the written texts in different databases. In order to determine the possible plagiarism rate and the originality level of the answers, all answers given by the artificial intelligence model were transferred to the plagiarism detection program (Turnitin, http://www.turnitin.com). Similarity rates were calculated as % and divided into four categories (0%, 1–24%, 25–49%, 50–74% and 75–100%).

Statistical analysis

All statistical calculations were performed using IBM SPSS Statistics version 26.0. Descriptive statistics including mean, median, standard deviation, maximum and minimum values were calculated for the data obtained from the evaluation criteria. The Kolmogorov-Smirnov test was used to evaluate the normality of the data. In the comparison of pediatric patient/parent and academic questions (2 groups), Independent T-test was employed when the data exhibited a normal distribution, whereas the Mann-Whitney U test was used when the data did not follow a normal distribution. For the comparison of subgroups of questions (3 groups), One-way ANOVA (post hoc: Tukey test) was conducted when the data displayed a normal distribution, and the Kruskal-Wallis H test was utilized when the data did not conform to a normal distribution. The relationship between two categorical variables was examined using Fisher's Exact Test. A p-value of <0.05 was considered statistically significant in all tests.

Results

A total of 60 questions were asked to ChatGPT for each topic (dental trauma, fluoride, and tooth eruption/oral health in children), consisting of pediatric patient/parent questions and academic questions. The questions asked to ChatGPT and their answers are provided in Supplementary Material 1.

Twelve pediatric dentists agreed to participate in the study. However, 2 pediatric dentists were excluded from the study due to missing data. All demographic data of the participants are presenred in Table 1. Of the 10 pediatric dentists included in the study, 7 (70%) were female (mean age: 35.9±3.7) and 3 (30%) were male (mean age: 34±5.3). While 6 (60%) of the participants had experience in artificial intelligence, 4 (40%) had no experience.

Information about the quality (modified GQS), readability (Flesch Reading Ease, Flesch-Kincaid Grade Level) and similarity of the answers given to pediatric patient/parent questions and academic questions are presented in Table 2. While the average modified GQS score of the answers to the pediatric patient/parent questions is 4.3±0.7, the average score of the answers to the academic questions is 3.7±0.8. The difference between the two groups was found to be statistically significant (*p*<0.05). To evaluate the readability of the answers given, according to the Flesch Reading Ease score, the average of the answers to the pediatric patient/parent questions is 41.5±12.5, while the average of the answers to the academic questions is 34±11.3. The difference between the two groups was found to be statistically significant (*p*<0.05). The averages being between 30 and 50 indicate

Table 3 Evaluation of the quality of answers to pediatric patient/parent questions and answers to academic questions by topic

Topic	Answer to pediatric patient/parent questions			Answer to academic questions		
	Mean ± sd	Median (Min-Max)	p	Mean ± sd	Median (Min-Max)	p
Dental Trauma	4.6 ± 0.5	5.0 (4–5)	0.120	3.0 ± 0.8	3.0 (2–4)[a]	**0.003**
Fluoride	3.9 ± 0.9	4.0 (2–5)		4.1 ± 0.6	4.0 (3–5)[b]	
Tooth eruption/oral health	4.4 ± 0.7	4.5 (3–5)		4.1 ± 0.6	4.0 (3–5)[b]	

Kruskal Wallis-H Test

sd: Standard deviation, Min: Minimum, Max: Maximum

Table 4 Evaluation of the readability of answers to pediatric patient/parent questions and answers to academic questions by topic

	Topic	Answer to pediatric patient/parent questions			Answer to academic questions		
		Mean ± sd	Median (Min-Max)	p	Mean ± sd	Median (Min-Max)	p
Flesch Reading Ease	Dental Trauma	43.3 ± 13.6[ab]	44.5 (25.3–61.0)	**0.040**	33.4 ± 9.8	32.0 (18.2–51.5)	0.956
	Fluoride	33.9 ± 9.4[a]	32.0 (14.6–47.2)		33.7 ± 12.6	32.8 (10.0–57.7)	
	Tooth eruption/oral health	47.4 ± 11.1[b]	45.3 (34.5–67.8)		34.9 ± 12.4	35.9 (9.5–52.6)	
Flesch-Kincaid Grade Level	Dental Trauma	12.1 ± 2.2	11.7 (8.6–15.0)	0.127	14.0 ± 2.4	14.4 (10.2–16.7)	0.734
	Fluoride	14.0 ± 1.8	13.8 (11.4–17.9)		13.8 ± 1.6	13.8 (11.8–16.5)	
	Tooth eruption/oral health	12.1 ± 2.8	11.8 (8.4–16.8)		13.3 ± 1.9	12.8 (11.2–17.7)	

One-way ANOVA, post hoc: Tukey test

sd: Standard deviation, Min: Minimum, Max: Maximum

that it is understandable at the college students level. According to the Flesch-Kincaid Grade Level score, the average score of the answers to the pediatric patient/parent questions was 12.7 ± 2.4, while the average score of the answers to the academic questions was 13.7 ± 1.9 and the difference was found to be statistically insignificant ($p > 0.05$). The averages being between 12 and 15 indicate that it is understandable at the college students level. While the average similarity rate of the answers to the pediatric patient/parent questions was 8.4% ± 8.4, the average similarity rate of the answers to the academic questions was 5.7% ± 5.1 and the difference was found to be statistically insignificant ($p > 0.05$).

The distribution of the quality of the answers given to pediatric patient/parent questions and academic questions by subject is given in Table 3. There was no statistically significant difference between the average quality scores of the answers given to the pediatric patient/parent questions regarding dental trauma (4.6 ± 0.5), fluoride (3.9 ± 0.9) and tooth eruption/oral health (4.4 ± 0.7) ($p > 0.05$). When the average quality scores of the answers to academic questions were evaluated, it was found that the highest average scores were in the subjects fluoride (4.1 ± 0.6) and tooth eruption/oral health (4.1 ± 0.6) and were statistically significantly higher than dental trauma (3 ± 0.8) ($p < 0.05$).

Distributions of Flesch Reading Ease and Flesch-Kincaid Grade Level scores, which are used to evaluate the readability of answers to pediatric patient/parent questions and academic questions, by subject are shown in Table 4. When the readability of the answers to the pediatric patient/parent questions is evaluated according to the Flesch Reading Ease scoring, the highest score is on

tooth eruption/oral health (47.4 ± 11.1) and the lowest score is on fluoride (33.9 ± 9.4). The difference between the two subjects was found to be statistically significant ($p < 0.05$). All scores between 30 and 50 (dental trauma: 43.3; fluoride:33.9; tooth eruption/oral health: 47.4) are understandable at college student level. When the readability of the answers to academic questions was evaluated, no statistically significant difference was observed in the subjects of dental trauma (33.4 ± 9.8), fluoride (33.7 ± 12.6) and tooth eruption/oral health (34.9 ± 12.4). ($p > 0.05$). All scores between 30 and 50 (dental trauma: 33.4; fluoride: 33.7; tooth eruption/oral health: 34.9) are understandable at college student level. According to Flesch-Kincaid Grade Level scoring, there was no statistically significant difference between the subjects in the readability of the answers to both the pediatric patient/parent questions and the academic questions ($p > 0.05$). The averages of the answers given to both academic and pediatric patient/parent questions were between 12 and 15 and were comprehensible at the level of college students.

When the similarity rate of the answers given to the questions with the literature is evaluated, the highest similarity rate among the answers given to the pediatric patient/parent questions is observed in the field of fluoride (12% ± 9.4), and in the academic questions, the rate of similarity is observed in the field of tooth eruption/oral health (7.4% ± 6). However, no statistically significant difference was observed between the subjects in the answers to both the pediatric patient/parent questions and the answers to the academic questions ($p > 0.05$) (Table 5).

Table 5 Evaluation of the similarity of answers to pediatric patient/parent questions and answers to academic questions by topic

Topic	Answer to pediatric patient/parent questions			Answer to academic questions		
	Mean % ± sd	Median % (Min-Max)	p	Mean % ± sd	Median % (Min-Max)	p
Dental Trauma	4.0% ± 3.4	4.0% (0–12)	0.175*	3.2% ± 2.9	3.0% (0–9)	0.151**
Fluoride	12.0%± 9.4	11.5% (0–25)		6.5% ± 5.3	5.05 (0–16)	
Tooth eruption/oral health	9.2% ± 9.5	6.5% (0–24)		7.4% ± 6.0	7.0% (0–21)	

* Kruskal Wallis-H Test

**One-way ANOVA

sd: Standard deviation, Min: Minimum, Max: Maximum

Discussion

ChatGPT is an application released towards the end of 2022 and is becoming popular in presenting health information. Especially when it comes to the health of patients, the accuracy of the information provided is undoubtedly very important [26]. In a study by Maltarollo et al., it was suggested that clinicians in the field of paediatric dentistry should be aware of ChatGPT in supporting the patient's health and make the best use of this tool [27]. Therefore, the accuracy and quality of ChatGPT's answers to questions about dentistry should be examined through comprehensive research. However, as in many fields, a comprehensive evaluation of the answers provided by ChatGPT has not yet been made in the field of pediatric dentistry. This study is the first in the literature to evaluate the quality, readability and originality of ChatGPT's responses at both patient and academic levels in the field of pediatric dentistry.

Considering the limited number of studies in the literature where the accuracy and quality of the answers given by ChatGPT in the field of dentistry are evaluated, in a study conducted in the field of orthodontics, the general accuracy level of the answer given by ChatGPT to 111 questions about clear aligner treatment was found to be insufficient [19]. In another study comparing official answers to questions people frequently ask about fluoride with ChatGPT answers, it was stated that ChatGPT answers were sufficient and comprehensive [9]. In a study conducted by another researcher in the field of oral and maxillofacial surgery, ChatGPT was found to be sufficient in providing patient information, but was not found sufficient in answering advanced technical questions. It is emphasized that the current version of ChatGPT should not be relied upon for academic use as the answers are not based on scientific references [8]. According to the results of this study, ChatGPT provided reasonably accurate and useful answers to pediatric patient/parent-centered questions, but did not perform as well as pediatric patient/parent-centered questions in answering advanced academic questions. No statistically significant difference was observed in pediatric patient/parent questions regarding tooth eruption, fluoride and dental trauma. However, the differences in the answers to academic questions were statistically significant and

the lowest score was observed to be on dental trauma. Dental trauma is an area where technical knowledge for pediatric dentists is intense and case-specific information is complex. For this reason, it has been observed that ChatGPT is inadequate in answering technical and case questions, especially such as "decoronation". Additionally, there was a limitation in providing up-to-date and clear information. This suggests that ChatGPT need to be trained to improve their ability to answer questions from an evidence-based perspective. In addition, Balel supports the view that a new version of ChatGPT called ChatGPT-Academic can be offered by working integrated with scientific databases [8].

Readability is generally defined as the ease of understanding or comprehension due to the writing style [28]. Flesch Reading Ease score and Flesch-Kincaid Grade Level score, which are recommended methods to measure readability, measure readability by evaluating word and sentence lengths with certain formulas [29]. In a study evaluating the readability of websites and brochures in the field of patient information in the field of endodontics, the average Flesch Reading Ease score was found to be 60 and the Flesch-Kincaid Grade Level score was 8.1–9.3. However, despite this, it was reported that some educational materials analysed in the field of endodontics were difficult to be understood by patients [24]. Hendrickson et al., evaluated brochures providing information about preventive oral health in children in terms of readability. According to the results of this study, the readability of the evaluated content varies from 2nd grade to 9th grade [29]. It is observed that there are many educational materials in the field of dentistry and these materials vary in terms of content and readability [24, 29, 30]. However, there is no study yet that evaluates the readability of ChatGPT's answers. In this study, the readability level of ChatGPT's answers was significantly more understandable in its answers to pediatric patient/parent questions than in its answers to academic questions, according to the Fresch Reading Ease score. Among the topics, it was determined that the fluoride topic was more complex for pediatric patients/parents, and the most readable topic was tooth eruption/oral health. According to the Flesch-Kinciad Grade Level, the readability of the answers to both pediatric patient/parent and academic

questions was similar, and no significant difference was observed in terms of readability between the subjects. In this study, it was determined that the readability level of the answers was at college level. This shows that while the answers may be understandable to dental students/clinicians, they may be difficult to appeal to parents with different educational levels.

According to the results of the plagiarism checking software, the similarity index of the questions asked to ChatGPT varies between 0% and 25%, and there is no significant difference between the answers to pediatric patient/parent questions and academic questions and between the subjects. Based on these findings, ChatGPT is believed to have an acceptable similarity rate. In a study conducted by Khalil and Er, the originality of 50 articles on the ChatGPT program was evaluated using plagiarizm checking software and the program was reported to have an acceptable similarity rate, which is similar to the results of this study [31].

A strength of this study is the inclusion of 60 questions developed based on outcomes for clinicians and pediatric patients/parents on topics involving dental trauma, fluoride, and tooth eruption/oral health in pediatric dentistry. In addition, this study, in which the quality, readability and originality of the answers given by ChatGPT to these questions are evaluated, will shed light on many future studies.

The limitations of the study include the fact that the selection of questions created for artificial intelligence to answer is based on the clinical experience of the author and the lack of standardization. Secondly, since ChatGPT has limited knowledge of scientific developments that occurred after September 2021, it may provide incomplete and misleading information in response to current questions. Third, the quality and relevance of responses may have been affected by the lack of prompt engineering in posing questions to ChatGPT. Fourth, the use of only ChatGPT-3.5 in the study may not reflect the current state of the literature, which could be supported by various advanced AI language models. Finally, the language used for questions and answers is English, which limits the accuracy of the data obtained to English speakers only.

In today's rapidly developing AI language models, a new one is added every day. For this reason, there is a need for future studies in which the scientific knowledge levels of these new AI language models are evaluated comparatively.

Conclusions
ChatGPT has significant potential as a patient information tool in pediatric dentistry. However, pediatric dentists should be careful when using ChatGPT and consider this in addition to their own clinical knowledge and experience. Attempts should be made to ensure the accuracy of this artificial intelligence model's responses before it is integrated into pediatric dentistry at the academic level. In addition, according to this study, ChatGPT, which is of high quality in the field of pediatric dentistry but is difficult to read and offers an acceptable similarity rate, should be developed to interact with people more efficiently and fluently.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12903-024-05393-1.

Supplementary Material 1

Supplementary Material 2

References
1. OpenAI ChatGPT: optimizing language models for dialogue. 2022. At: https://openai.com/blog/chatgpt/. Accessed: December 5, 2023.
2. Scott K. Microsoft teams up with OpenAI to exclusively license GPT-3 language model. 2020. At: https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/. Accessed: December 5, 2023.
3. Eggmann F, Weiger R, Zitzmann NU, Markus BB. Implications of large language models such as ChatGPT for dental medicine. J Esthet Restor Dent. 2023;35:1098–102.
4. Huang H, Zheng O, Wang D et al. ChatGPT for Shaping the Future of Dentistry: The Potential of Multi-Modal Large Language Model. At: https://doi.org/10.48550/arXiv.2304.03086. Accessed: December 5, 2023.

5. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769–74.
6. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. RSNA. 2023;307(2):e230163.
7. Alhaidry HM, Fatani B, Alrayes JO, Almana AM, Alfhaed NK. ChatGPT in dentistry: a comprehensive review. Cureus. 2023;15(4):e38317.
8. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? J Stomatol Oral Maxillofac Surg. 2023;124(5):101471.
9. Buldur M, Sezer B. Can Artificial Intelligence effectively respond to frequently asked questions about fluoride usage and effects? A qualitative study on ChatGPT. Fluoride. 2023;56(3):201–16.
10. American Academy of Pediatric Dentistry (AAPD). Guideline on Infant Oral Health Care 2004. At: https://www.aapd.org/assets/1/7/g_infantoralhealthcare.pdf. Accessed: December 5, 2023.
11. Meera R, Muthu MS, Phanibabu M, Rathnaprabhu V. First dental visit of a child. J Indian Soc Pedod Prev Dent. 2008;26(6):68–71.
12. Toumba KJ, Twetman S, Splieth C, Parnell C, Van Loveren C, Lygidakis NA. Guidelines on the use of fluoride for caries prevention in children: an updated EAPD policy document. Eur Arch Paediatr Dent. 2019;20:507–16.
13. National Research Council. Fluoride in drinking water: a scientific review of EPA's standards. Washington DC: National Academies; 2006. pp. 1–530.
14. Buzalaf MAR, Levy SM. Fluoride intake of children: considerations for dental caries and dental fluorosis. Monogr Oral Sci. 2011;22:1–19.
15. Mertz A, Allukian M. Community water fluoridation on the internet and social media. J Mass Dent Soc. 2014;63(2):32–6.
16. Petti S, Andreasen JO, Glendor U, Andersson L. The fifth most prevalent disease is being neglected by public health organisations. Lancet Glob Health. 2018;6(10):e1070–1.
17. Teel A, Wang T, Lund B. ChatGPT conundrums: probing plagiarism and parroting problems in higher education practices. Coll Res Libr News. 2023;84(6):205.
18. Jung J, Qin L, Welleck S, Brahman F, Bhagavatula C, Bras RL, Choi Y. Maieutic prompting: logically consistent reasoning with recursive explanations. arXiv preprint arXiv:2205.11822. 2022.
19. 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.
20. Flesch R. A new readability yardstick. J Appl Psychol. 1948;32(3):221.
21. Kincaid JP, Fishburne RP Jr, Rogers RL. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading ease Formula) for Navy enlisted Personnel. Naval Technical Training Command Millington TN Research Branch; 1975.
22. Flesch R, How to Write Plain English. 2016. At: https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml. Accessed: November 20, 2024.
23. Laura Kelly. What is the Flesch-Kincaid Grade Level? 2017. At: https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/. Accessed: November 20, 2024.
24. Woodmansey K. Readability of educational materials for endodontic patients. J Endod. 2010;36(10):1703–6.
25. Using Microsoft Word's Readability Program. At: https://www.michbar.org/file/barjournal/article/documents/pdf4article1467.pdf. Accessed: December 18, 2023.
26. Kessels RPC. Patients' memory for medical information. J R Soc Med. 2003;96(5):219–22.
27. Maltarollo TFH, Shitsuka C, Sivieri-Araújo G. ChatGPT and its use in paediatric dentistry. Eur Arch Paediatr Dent. 2023;24(5):675–6.
28. Alexander RE. Readability of published dental educational materials. J Am Dent Assoc. 2000;131(7):937–42.
29. Hendrickson RL, Huebner CE, Riedy CA. Readability of pediatric health materials for preventive dental care. BMC Oral Health. 2006;6(14):1–9.
30. Boles CD, Liu Y, November-Rider D. Readability levels of dental patient education brochures. Am Dent Hygienists' Association. 2016;90(1):28–34.
31. Khalil M, Er E. Will ChatGPT get you caught? Rethinking of plagiarism detection. At: https://arxiv.org/abs/2302.04335. Accessed: December 18, 2023.