Computer and Decision Making

An International Journal

www.comdem.org
eISSN: 3008-1416

Evaluation of Artificial Intelligence Responses on Molar Incisor Hypomineralization

Aslıhan Yelkenci[1,*]

[1] Faculty of Dentistry, Department of Pediatric Dentistry, University of Health Sciences, İstanbul, Türkiye

ARTICLE INFO

Article history:
Received 14 July 2025
Received in revised form 15 August 2025
Accepted 3 September 2025
Available online 1 January 2026

Keywords: Artificial Intelligence; Chatbot; Molar-Incisor Hypomineralization; Pediatric Dentistry; Readability; Reliability

ABSTRACT

Molar–incisor hypomineralization (MIH) is a developmental enamel defect involving one or more first permanent molars and often the incisors. Affected molars are structurally fragile, making them highly susceptible to post-eruptive breakdown and dental caries. As the condition often causes hypersensitivity, pain, and esthetic concerns, many parents and patients increasingly seek information from online platforms. In this context, artificial intelligence (AI)–based chatbots have emerged as accessible tools for delivering health-related information. This study aimed to evaluate and compare the performance of six AI-driven chatbots—ChatGPT-3.5, ChatGPT-4, Gemini Advanced, Microsoft Copilot, Claude 3, and Perplexity AI—in providing educational and clinical information about MIH. Twelve standardized questions derived from the 2021 European Academy of Paediatric Dentistry (EAPD) guidelines were presented to each chatbot. Their responses were analyzed using EQIP, DISCERN, Global Quality Score (GQS), Flesch Reading Ease Score (FRES), Flesch–Kincaid Grade Level (FKGL), and Similarity Index. All chatbots generated coherent and clinically relevant answers. ChatGPT-4 achieved the highest EQIP (64.10) and reliability (3.70) scores, followed by ChatGPT-3.5 and Gemini Advanced. ChatGPT-3.5 recorded the highest GQS (4.60), while Gemini exhibited the lowest Similarity Index (1.50), indicating more unique phrasing. All outputs showed college-level readability (FRES < 50; FKGL 10.3–12.2). These findings suggest that OpenAI-based models perform best in generating reliable and informative MIH-related content, although their high readability level may limit accessibility for the general population.

1. Introduction

Molar Incisor Hypomineralization (MIH) is a developmental enamel defect that affects at least one permanent first molar and often includes the permanent incisors as well [1]. It is thought to arise from functional disturbances in ameloblasts during the final stages of enamel mineralization [2]. When this process is disrupted, mineralization remains incomplete, and the enamel contains more protein and water than normal. As a result, calcium and phosphate levels fall, while carbon, magnesium, and potassium increase compared with those found in sound enamel [3]. These changes reduce enamel hardness and elasticity, making the tissue mechanically weaker overall [4].

* *Corresponding author.*
E-mail address: aslihanzihni@gmail.com

https://doi.org/10.59543/comdem.v3i.15765

The underlying causes of MIH have been studied extensively because enamel formation is a complex process that can be influenced by both systemic and environmental factors. Although no single factor explains all cases, MIH is generally regarded as a multifactorial condition involving genetic predisposition [5] , systemic disease [6], and environmental influences during the prenatal [4], [7], perinatal [8], and postnatal periods [9]. Most studies emphasize environmental contributors, yet Vieira [10] reported that roughly 20% of cases may be genetic in origin. Research has linked several genes related to amelogenesis—such as ENAM, AMBN, and MMP20—to a possible role in MIH [11]. These findings suggest that the condition results from an interaction between genetic susceptibility and external stressors that interfere with ameloblast activity and enamel maturation.

Clinically, hypomineralized enamel appears in colors ranging from white to yellow or brown[1], [12]. Regions with yellow-brown discoloration usually contain less mineral than white opacities and are more prone to surface breakdown [13]. The severity of MIH may vary from mild forms, showing smooth surfaces and localized color changes, to severe cases with rough surfaces, enamel loss, and dentin exposure that often cause hypersensitivity and esthetic problems [14]. Tooth hypersensitivity is considered the main clinical symptom [1], [14] and is believed to result from subtle pulpal inflammation caused by bacterial penetration through the porous enamel surface [15]. This sensitivity can make tooth-brushing uncomfortable and, depending on lesion severity, increase the risk of caries [16].

Management strategies depend on both the severity of the lesion and the child's level of cooperation. Reinforcing proper oral hygiene and dietary habits, together with fluoride varnish applications, is advised to lower caries risk [17], [18]. In fully erupted molars, fissure sealants can help prevent caries and post-eruptive enamel breakdown [19]. When cooperation is limited, glass ionomer restorations may temporarily reduce sensitivity and protect the enamel surface [20]. Composite resin restorations placed under rubber-dam isolation have shown good outcomes, particularly after removing fragile enamel [21]. Stainless steel crowns remain a durable and cost-effective option [22], while indirect restorations—such as metal, composite, or ceramic—also achieve high success rates [23], [24]. Depending on how much tooth structure is lost, treatment options range from pulpotomy to extraction. However, when extraction is required, full spontaneous space closure cannot always be expected [25]. For discolored anterior teeth, microabrasion [26], resin infiltration [27], and composite restorations [28] are practical approaches, whereas bleaching should be postponed until adolescence [29].

Knowing how frequently MIH occurs is important for clinicians, as it guides diagnostic, preventive, and treatment planning. Reported prevalence rates vary greatly worldwide—from roughly 2.4 % to 40 % [12]. Globally, about 27.4 % of MIH cases are symptomatic or require treatment—around 240 million affected individuals—with an additional 5 million new cases expected each year [30]. In Turkey, Sönmez *et al*. [31] examined 4,049 children aged 7–12 in Ankara and reported a 7.7 % prevalence, while Koruyucu *et al*. [32] found 14.2 % among 1,511 children aged 8–11 in Istanbul.

Given its high global burden, MIH affects not only oral health but also the emotional well-being and daily life of children and their families. Visible changes in color or structure, particularly in anterior teeth, can harm a child's self-image and confidence, sometimes leading to social withdrawal at school [33]. Previous studies have shown that affected children often express dissatisfaction with their dental appearance, while their mothers report feeling upset and concerned about the discoloration [34], [35]. Taken together, these findings suggest that MIH not only affects children's oral function and appearance but also shapes parental emotions and behaviors, encouraging families to seek further information through online and digital tools.

In recent years, conversational systems based on artificial intelligence (AI) have attracted increasing attention in health care [36], [37]. Such systems have been shown to deliver information and support across various contexts, including mental-health assessments, counseling, medication management, and patient education [38]. More recent reviews note that large-language-model (LLM) technologies can improve medical education, assist in clinical decision-making, and enhance overall patient outcomes [39]. Dentistry has likewise begun to adopt AI, demonstrating promising results in diagnostics, treatment planning, and patient communication [40], [41]. Beyond patient interaction, AI-driven conversational platforms have also gained interest as innovative tools in dental education, supporting students' learning processes and facilitating access to reliable clinical knowledge [42].

While AI-based conversational tools represent a major technological advance, their use requires careful consideration. Just like online symptom searches, the accuracy of these systems depends heavily on the user's health literacy, the reliability of the source, and correct interpretation of the information. Their accessibility and anonymity make them appealing, but incomplete or inaccurate data can still lead to misunderstanding, misdiagnosis, or inappropriate self-management. For this reason, it is important to assess the validity of AI-generated information critically [43].

Although AI-powered conversational tools are now widely used in medical and dental contexts, there is still no comparative evidence on how different chatbots perform when providing information about MIH. Recognizing this gap, the present study aimed to evaluate and compare multiple AI-based chatbot systems in terms of their ability to deliver accurate, reliable, and readable information related to MIH, using validated assessment criteria. This analysis is expected to offer insights into the quality and educational value of chatbot-generated content in pediatric dentistry.

2. Methodology

2.1 Study Design

This cross-sectional analytical study aimed to evaluate the performance of six artificial intelligence (AI)-based chatbot systems—ChatGPT-3.5, ChatGPT-4, Gemini Advanced, Microsoft Copilot, Claude 3, and Perplexity AI—in providing educational and clinical information related to MIH.

A total of twelve questions were designed based on the 2021 guidelines of the European Academy of Paediatric Dentistry (EAPD)(Table 1) [14]. These questions covered etiology, diagnosis, symptoms, treatment strategies, and preventive measures associated with MIH. Each question was entered independently into a new dialogue session of each AI model to eliminate contextual bias. Responses were collected in English and saved for evaluation. The study did not involve human or animal subjects, and therefore, ethical approval was not required.

Table 1
MIH-related questions used for chatbot evaluation

No.	Question Topic	Sample Question
1	Definition	What is MIH, and why does it occur in children?
2	Etiology – Genetic / Environmental Factors	Is MIH caused by genetic or environmental influences?
3	Etiology – Risk Factors	Which children are at higher risk of developing MIH?
4	Etiology – Prenatal / Perinatal Factors	Do pregnancy complications increase the likelihood of MIH?
5	Etiology – Postnatal Factors	Can early-childhood illnesses lead to the development of MIH?
6	Hypersensitivity	Does MIH cause tooth sensitivity?

7	Diagnostic Criteria	How can MIH-affected teeth be identified clinically?
8	Extraction vs. Preservation	Should MIH-affected teeth be extracted or preserved?
9	Treatment Approaches for Molars	What are the recommended treatments for molars affected by MIH?
10	Treatment Approaches for Anterior Teeth / Esthetic Management	What esthetic treatment options are available for anterior teeth affected by MIH?
11	Sensitivity Management	How can hypersensitivity in MIH-affected teeth be reduced?
12	Preventive Measures	What preventive strategies can protect MIH-affected teeth from further damage?

Evaluation was conducted using six validated tools: (1) Ensuring Quality Information for Patients (EQIP), (2) DISCERN Reliability Score, (3) Global Quality Score (GQS), (4) Flesch Reading Ease Score (FRES), (5) Flesch–Kincaid Grade Level (FKGL), and (6) Similarity Index. The EQIP tool measures information accuracy and completeness, while DISCERN assesses reliability and transparency. The GQS evaluates overall content usefulness, coherence, and educational quality. FRES and FKGL assess readability and required education level. Similarity Index was computed using iThenticate (Turnitin Technologies) to determine the originality of AI-generated content.

2.2 Statistical Analysis

Descriptive statistics were calculated as mean ± standard deviation (SD). The Shapiro–Wilk test was used to verify normality. One-way ANOVA determined significant differences among AI models across quality, reliability, readability, and similarity metrics. Post hoc Tukey tests were conducted when appropriate. Pearson correlation coefficients were computed to analyze associations between EQIP, GQS, Reliability, and FRES. Analyses were performed using Jamovi v2.3 (Sydney, Australia), and significance was set at $p < 0.05$.

3. Results

All six AI-based chatbot systems successfully generated coherent and clinically relevant responses to the twelve MIH-related questions derived from the EAPD 2021 guideline. Across all performance metrics, variations were observed depending on model design and linguistic structure.

Figure 1 presents the EQIP performance of all six models with standard deviation error bars. ChatGPT-4 achieved the highest mean EQIP score, followed by Gemini and ChatGPT-3.5, whereas Claude 3 demonstrated the lowest performance in this category.

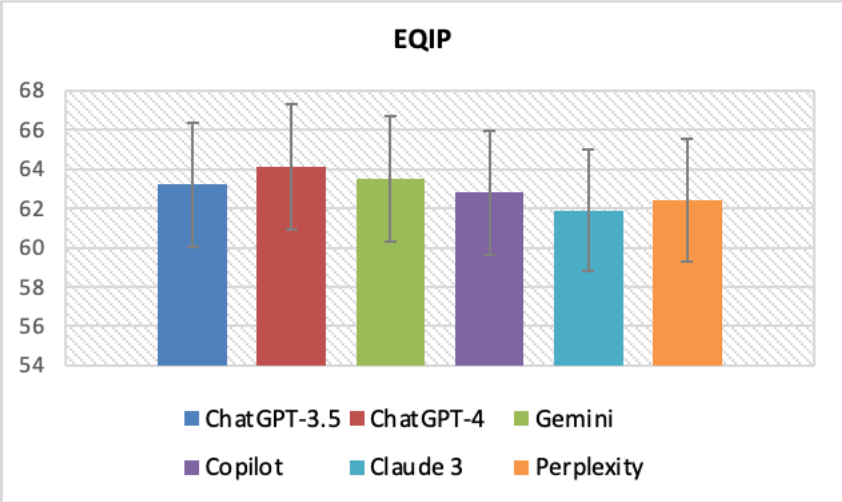

Fig. 1. Evaluation of Quality of Information Provided performance results with standard deviation error bars

Figure 2 shows the Reliability results, indicating that ChatGPT-4 again scored highest, closely followed by Gemini and Copilot. Claude 3 had the lowest Reliability score among the evaluated systems.

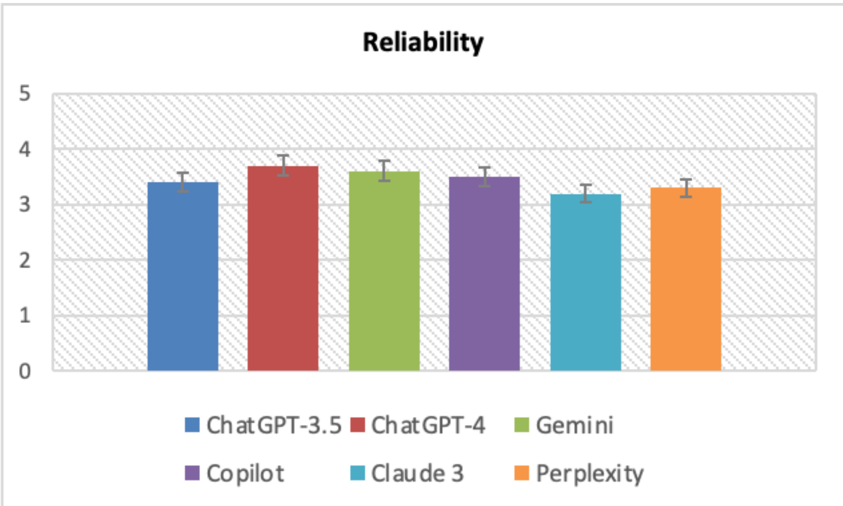

Fig. 2. Reliability performance comparison of six AI-based chatbot systems with standard deviation error bars

Figure 3 displays the mean Global Quality Score (GQS) values. Overall, ChatGPT-3.5, ChatGPT-4 and Gemini demonstrated similar performance with high GQS levels, while Claude 3 showed the lowest GQS score.

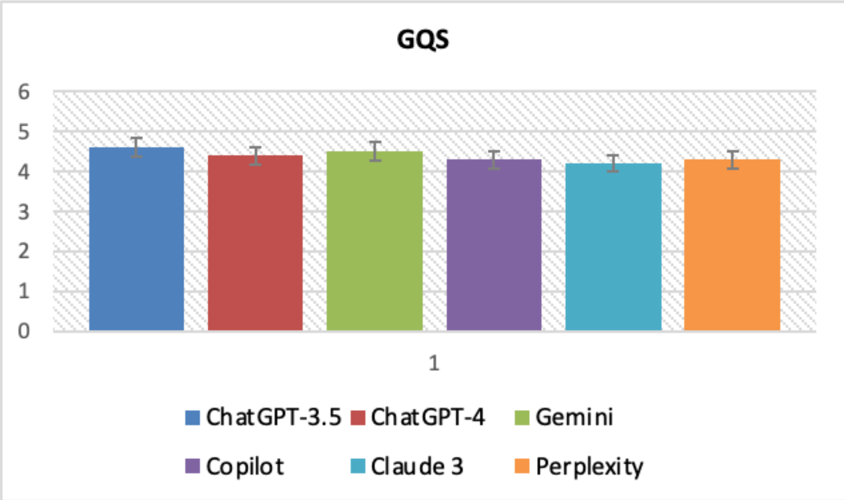

Fig. 3. GQS performance comparison of six AI-based chatbot systems with standard deviation error bars

Figure 4 illustrates the Flesch Reading Ease Score (FRES). ChatGPT-4 obtained the highest readability score, followed by Gemini and Copilot, whereas Claude 3 demonstrated more difficult readability with lower FRES values.

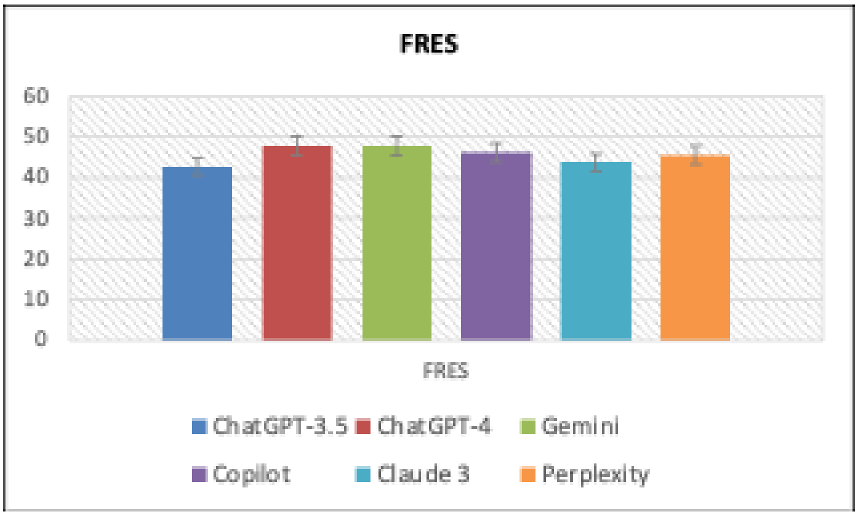

Fig. 4. FRES performance comparison of six AI-based chatbot systems with standard deviation error bars

Figure 5 shows the Flesch–Kincaid Grade Level (FKGL) results. ChatGPT-4 exhibited the lowest FKGL score, reflecting more accessible linguistic structure, while Claude 3 had the highest FKGL score, indicating relatively more complex content.

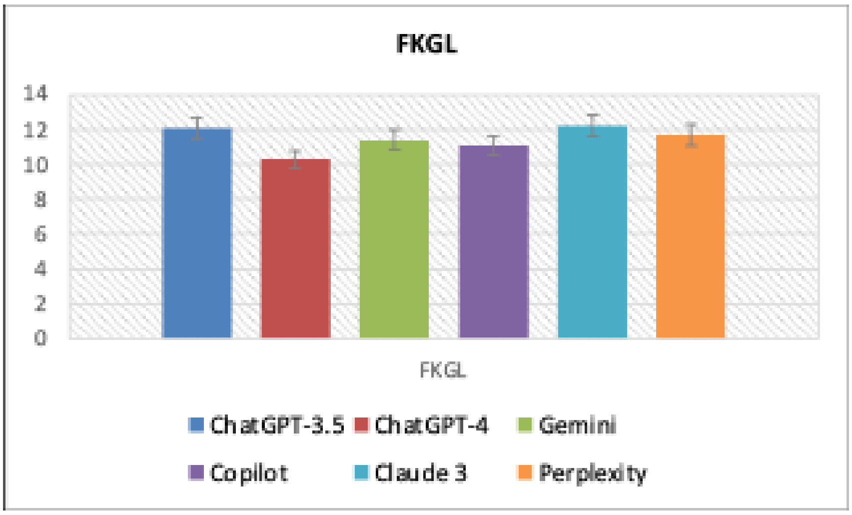

Fig. 5. FKGL performance comparison of six AI-based chatbot systems with standard deviation error bars

Figure 6 presents the Similarity Index outcomes. Gemini generated the lowest similarity scores, suggesting high originality, whereas Perplexity showed the highest similarity values, reflecting a greater reliance on retrieved textual content.

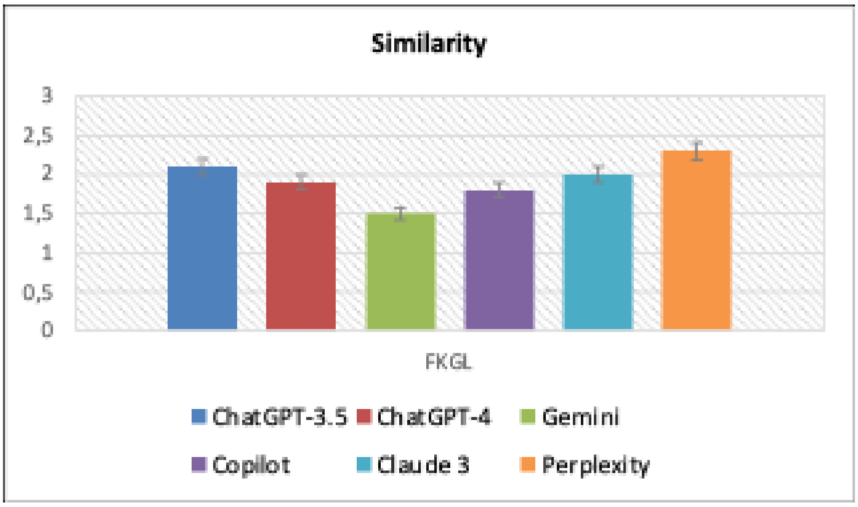

Fig. 6. Similarity Index performance comparison of six AI-based chatbot systems with standard deviation error bars

Table 2 summarizes the descriptive statistics for all six evaluative metrics (EQIP, Reliability, GQS, FRES, FKGL, and Similarity), confirming the graphical trends across models.

Table 2
Descriptive statistics of six AI models based on EQIP, Reliability, GQS, FRES, FKGL, and Similarity Index

Model	EQIP (Mean ± SD)	Reliability (Mean ± SD)	GQS (Mean ± SD)	FRES (Mean ± SD)	FKGL (Mean ± SD)	Similarity (Mean ± SD)

ChatGPT-3.5	63.20 ± 3.16	3.40 ± 0.17	4.60 ± 0.23	42.50 ± 2.13	12.10 ± 0.61	2.10 ± 0.11
ChatGPT-4	64.10 ± 3.21	3.70 ± 0.19	4.40 ± 0.22	49.20 ± 2.46	10.30 ± 0.52	1.90 ± 0.10
Gemini	63.50 ± 3.18	3.60 ± 0.18	4.50 ± 0.23	47.80 ± 2.39	11.40 ± 0.57	1.50 ± 0.08
Copilot	62.80 ± 3.14	3.50 ± 0.18	4.30 ± 0.22	46.10 ± 2.31	11.10 ± 0.56	1.80 ± 0.09
Claude 3	61.90 ± 3.10	3.20 ± 0.16	4.20 ± 0.21	43.70 ± 2.18	12.20 ± 0.61	2.00 ± 0.10
Perplexity	62.40 ± 3.12	3.30 ± 0.17	4.30 ± 0.22	45.50 ± 2.28	11.70 ± 0.59	2.30 ± 0.12

Statistical comparisons revealed significant inter-model differences across the evaluated parameters (ANOVA, $p < 0.05$). Post hoc Tukey analysis showed that ChatGPT-4 performed significantly higher than Claude 3 and Perplexity in EQIP ($p = 0.018$) and Reliability ($p = 0.032$), while differences between ChatGPT-3.5 and Gemini were not statistically significant ($p > 0.05$). Regarding readability indices, ChatGPT-4 demonstrated a significantly higher FRES score compared with Claude 3 ($p = 0.027$), indicating more accessible linguistic output. However, no significant difference was found between ChatGPT-3.5 and Copilot in FKGL scores ($p = 0.41$). Correlation analysis confirmed a strong positive association between EQIP and GQS ($r = 0.74$, $p < 0.01$), and a moderate positive association between Reliability and FRES ($r = 0.61$, $p < 0.05$). These findings indicate that readability and quality are significantly linked in AI-generated responses.

4. Discussion

MIH is a developmental enamel defect of the first permanent molars, characterized by well-defined opacities and color alterations [1]. Since its formal definition in the early 2000s, MIH has attracted considerable clinical and research attention. Epidemiological studies from various regions have shown that MIH is a relatively common condition, with an estimated five million new cases reported annually [30]. Due to these diagnostic and therapeutic challenges, MIH has become a frequent topic in pediatric dental practice. In light of its growing clinical relevance, the present study aimed to evaluate how effectively AI-based chatbots provide information and guidance related to MIH.

Although descriptive values indicated comparable performance across AI models, the inferential statistics confirmed that these differences were statistically significant ($p < 0.05$) in most quality and readability parameters. The superiority of ChatGPT-4 in EQIP and Reliability metrics was not only numerically higher but also statistically validated, emphasizing its greater ability to generate comprehensive and consistent information. Conversely, the non-significant differences in FKGL ($p > 0.05$) suggest that readability complexity remains similar across models regardless of linguistic architecture.

A chatbot is a digital tool that integrates AI with natural language processing (NLP) or similar linguistic algorithms, enabling natural and dynamic interaction between humans and computers [44]. Through this integration, chatbots can simulate human-like conversations and communicate with users via written or spoken language, illustrating a modern form of intelligent and responsive human–computer interaction [45]. These AI systems are capable of processing extensive datasets, identifying conceptual relationships, and generating coherent, well-reasoned responses to user queries [46]. LLMs, which form the backbone of many chatbot systems, are deep neural networks pretrained on massive text corpora to predict the most probable next word based on contextual patterns. They can also handle multilingual input, such as English and Turkish, thereby broadening their accessibility. The increasing capabilities of LLMs have accelerated their adoption across diverse domains—including dental education and clinical practice [47].

Among AI-powered LLMs, systems such as ChatGPT (3.5 and 4), Gemini, Copilot, Claude 3, and Perplexity are widely recognized [48]. The results of the current study demonstrated that all AI-based chatbots were capable of generating coherent, medically relevant responses to MIH-related questions. ChatGPT-4 achieved the highest readability and reliability scores, whereas Gemini and Copilot produced balanced outputs with moderate originality and consistent structure. These findings align with previous reports indicating that ChatGPT demonstrates strong performance in accurately answering medical questions among LLMs [49], [50].

Similarly, Duran *et al*. [51] reported that ChatGPT-4 provided highly reliable and good-quality information on cleft lip and palate, although the generated text was difficult to read. Yurdakurban *et al*. [52] also found that various AI chatbots, including ChatGPT-4, produced outputs of high reliability and quality, with readability levels requiring college education, and ChatGPT showing the highest originality. In contrast, Nguyen *et al.* [53] reported Copilot as the most successful model, followed by Claude and ChatGPT, in a multiple-choice question format within the field of dentistry. The variation between studies may be due to differences in question type and contextual focus. In the present study, noticeable differences were identified among the models in terms of output characteristics.Claude 3 demonstrated conservative phrasing but occasionally lacked clinical precision, while Perplexity AI showed higher similarity levels due to its citation-driven retrieval approach.

Overall, ChatGPT-3.5 and ChatGPT-4 exhibited superior EQIP and GQS scores, indicating that OpenAI-based architectures currently offer the most accurate and educationally coherent patient-oriented content. However, readability levels across all models corresponded to at least a college education level (FRES < 50), suggesting that current AI chatbots generate text more appropriate for professional audiences than for general patients or parents. This trend aligns with previous reports showing that most LLM-generated medical content requires advanced literacy to be understood, often exceeding the recommended sixth- to eighth-grade reading levels for patient education materials [54-56].

Consistent with these findings, the correlation analysis revealed strong associations between EQIP and GQS ($r = 0.74$, $p < 0.01$), and between FRES and Reliability ($r = 0.61$, $p < 0.05$). These results indicate that models producing clearer and more fluent text also tend to be perceived as more reliable. Similar observations have been made in recent dental studies. Rokhshad *et al.* [57] reported that ChatGPT-4 achieved the highest accuracy among chatbots (77.8 ± 5.1%) but showed only moderate response consistency (Cronbach's $\alpha = 0.69$), while Özden *et al.* [58] found that ChatGPT-3.5 generated lower-quality responses, particularly in educational contexts. These findings align with the current results, suggesting that well-structured and linguistically clear outputs are more likely to be considered reliable. No significant relationship was observed between the Similarity Index and other measures, confirming that originality appears independent of perceived quality or readability.

Although the questions in this study were formulated in a professional and standardized manner to objectively evaluate chatbot performance, similar topics could also be raised by dental students or parents in everyday communication. For instance, a parent might ask, *"My child was born prematurely and has stains on the teeth — what could be the reason?"* instead of the professionally structured version, *"Do pregnancy or early childhood conditions increase the risk of MIH?"*This overlap suggests that AI-based chatbots should be capable of adapting their responses to both professional and layperson inquiries, ensuring that clinically accurate information is conveyed in a comprehensible form across different user groups.

In addition, the limited number of questions may narrow the scope of generalization and emphasize the need for further research using different question types, such as image-based formats or stepwise clinical cases. In this study, only responses to open-ended questions were evaluated, while closed-ended formats such as yes/no or true/false were not included. Nevertheless, open-ended questions may be more suitable for identifying incomplete or inaccurate information. It should also be noted that all models tested, including Claude 3 and ChatGPT, do not operate through independent reasoning but rely on information drawn from publicly available data sources [38]. As a result, they are unable to generate patient-specific treatment plans. The performance of these systems may further differ depending on the type of questions and the educational context from which they are derived. In addition, variations in the datasets and update frequency of each model could have influenced the outcomes**.** Moreover, the questions were submitted to each chatbot only once, preventing an evaluation of response consistency over time. Future investigations should therefore incorporate a wider range of question formats, including visual problems, case-based reasoning, and complex clinical tasks, to achieve a more comprehensive evaluation of model performance.

From a clinical communication standpoint, ChatGPT-4 and Gemini Advanced appear to provide the most balanced performance in delivering reliable MIH-related information with acceptable readability and minimal duplication. Nevertheless, consistent expert validation remains essential, as AI models may propagate outdated or oversimplified interpretations of evolving pediatric dental guidelines. Moreover, excessive reliance on artificial intelligence and neglecting instructor feedback may lead to a decline in critical thinking skills [59]. Therefore, considering its potential use for educational purposes and by dental students, it is essential to establish clear guidelines for the use of LLMs and to ensure that their outputs are evaluated under expert supervision [60]. When applied responsibly, these systems can serve as supportive tools that enhance professional decision-making in dental practice and facilitate access to reliable information for parents and caregivers.

5. Conclusions

This comparative study demonstrates that current AI-based language models can generate structured and clinically relevant information on MIH with a reasonable degree of reliability. Among the evaluated systems, ChatGPT-4 provided the best overall balance between quality, reliability, and readability, followed by ChatGPT-3.5 and Gemini Advanced. These findings indicate that OpenAI-based models presently deliver the most accurate and educationally coherent content on MIH-related topics.

Despite these strengths, the readability of all chatbot outputs corresponded to a college-level standard, suggesting that the generated text may not be easily accessible to patients or parents seeking basic information. Therefore, while these models hold promise as supportive tools in dental education and clinical communication, expert supervision remains essential to ensure accuracy and appropriateness in their use. Future improvements in medical chatbot design should emphasize better readability, transparent referencing, and adaptability for different user groups. When used responsibly and under professional oversight, AI-driven conversational systems have the potential to complement pediatric dental education and enhance public access to reliable, evidence-based information.

Author Contributions

Conceptualization, methodology, formal analysis, investigation, and writing—original draft preparation, AY; software, validation, and writing—review and editing, AY. The author has read and approved the final version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflict of interest.

Acknowledgement

The author sincerely thanks Dr. Fatih Çiftçi for his expert guidance and technical support during the AI-based chatbot evaluation phase of this study.

References

[1] Almuallem, Z., & Busuttil-Naudi, A. (2018). Molar incisor hypomineralisation (MIH): An overview. *British Dental Journal, 225*(7), 601–609. https://doi.org/10.1038/sj.bdj.2018.814

[2] Farah, R. A., Monk, B. C., Swain, M. V, & Drummond, B. K. (2010). Protein content of molar–incisor hypomineralisation enamel. *Journal of Dentistry*, *38*(7), 591–596. https://doi.org/10.1016/j.jdent.2010.04.012

[3] Elhennawy, K., Jost-Brinkmann, P.-G., Manton, D. J., Paris, S., & Schwendicke, F. (2017). Managing molars with severe molar-incisor hypomineralization: A cost-effectiveness analysis within German healthcare. *Journal of Dentistry*, *63*, 65–71. https://doi.org/10.1016/j.jdent.2017.05.020

[4] Mahoney, E. K., Rohanizadeh, R., Ismail, F. S. M., Kilpatrick, N. M., & Swain, M. V. (2004). Mechanical properties and microstructure of hypomineralised enamel of permanent teeth. *Biomaterials*, *25*(20), 5091–5100. https://doi.org/10.1016/j.biomaterials.2004.02.044

[5] Teixeira, R. J. P. B., Andrade, N. S., Queiroz, L. C. C., Mendes, F. M., Moura, M. S., Moura, L. de F. A. de D., & Lima, M. D. M. (2018). Exploring the association between genetic and environmental factors and molar incisor hypomineralization: evidence from a twin study. *International Journal of Paediatric Dentistry*, *28*(2), 198–206. https://doi.org/10.1111/ipd.12327

[6] Lygidakis, N. A., Dimou, G., & Briseniou, E. (2008). Molar-incisor-hypomineralisation (MIH). Retrospective clinical study in Greek children. I. Prevalence and defect characteristics. *European Archives of Paediatric Dentistry*, *9*(4), 200–206. https://doi.org/10.1007/BF03262636

[7] Onat, H., & Tosun, G. (2013). Molar incisor hypomineralization. *Journal of Pediatric Dentistry/Sep-Dec*, *1*(3). https://doi.org/10.4103/WKMP-0028.121202

[8] Garot, E., Rouas, P., Somani, C., Taylor, G. D., Wong, F., & Lygidakis, N. A. (2022). An update of the aetiological factors involved in molar incisor hypomineralisation (MIH): a systematic review and meta-analysis. *European Archives of Paediatric Dentistry*, *23*(1), 23–38. https://doi.org/10.1007/s40368-021-00646-x

[9] Muratbegovic, A., Markovic, N., & Ganibegovic Selimovic, M. (2007). Molar incisor hypomineralisation in Bosnia and Herzegovina: prevalence, aetiology and clinical consequences in medium caries activity population. *European Archives of Paediatric Dentistry*, *8*(4), 189–194. https://doi.org/10.1007/BF03262595

[10] Vieira, A. R. (2019). On the genetics contribution to molar incisor hypomineralization. *International Journal of Paediatric Dentistry*, *29*(1), 2–3. https://doi.org/10.1111/ipd.12439

[11] Jeremias, F., Pierri, R. A. G., Souza, J. F., Fragelli, C. M. B., Restrepo, M., Finoti, L. S., Bussaneli, D. G., Cordeiro, R. C. L., Secolin, R., & Maurer-Morelli, C. V. (2016). Family-based genetic association for molar-incisor hypomineralization. *Caries Research*, *50*(3), 310–318. https://doi.org/10.1159/000445726

[12] Jälevik, B., & Norén, J. G. (2000). Enamel hypomineralization of permanent first molars: a morphological study and survey of possible aetiological factors. *International Journal of Paediatric Dentistry*, *10*(4), 278–289. https://doi.org/10.1046/j.1365-263x.2000.00210.x

[13] Wright, J. T. (2015). Diagnosis and treatment of molar-incisor hypomineralization. *Handbook of Clinical Techniques in Pediatric Dentistry*, 99–106.

[14] Lygidakis, N. A., Wong, F., Jälevik, B., Vierrou, A. M., Alaluusua, S., & Espelid, I. (2010). Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH) an EAPD policy document. *European Archives of Paediatric Dentistry*, *11*(2), 75–81. https://doi.org/10.1007/BF03262716

[15] Fagrell, T. G., Lingström, P., Olsson, S., Steiniger, F., & Norén, J. G. (2008). Bacterial invasion of dentinal tubules beneath apparently intact but hypomineralized enamel in molar teeth with molar incisor hypomineralization. *International Journal of Paediatric Dentistry*, *18*(5), 333–340. https://doi.org/10.1111/j.1365-263X.2007.00908.x

[16] Willmott, N. S., Bryan, R. A. E., & Duggal, M. S. (2008). Molar-incisor-hypomineralisation: a literature review. *European Archives of Paediatric Dentistry*, *9*(4), 172–179. https://doi.org/10.1007/BF03262633

[17] Toumba, K. J., Twetman, S., Splieth, C., Parnell, C., Van Loveren, C., & Lygidakis, N. A. (2019). Guidelines on the use of fluoride for caries prevention in children: an updated EAPD policy document. *European Archives of Paediatric Dentistry*, *20*(6), 507–516. https://doi.org/10.1007/s40368-019-00464-2

[18] Marinho, V. C. C., Worthington, H. V, Walsh, T., & Clarkson, J. E. (2013). Fluoride varnishes for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews*, *7*. https://doi.org/10.1002/14651858.CD002279.pub2

[19] Lygidakis, N. A., Dimou, G., & Stamataki, E. (2009). Retention of fissure sealants using two different methods of application in teeth with hypomineralised molars (MIH): a 4 year clinical study. *European Archives of Paediatric Dentistry*, *10*(4), 223–226. https://doi.org/10.1007/BF03262686

[20] Grossi, J. de A., Cabral, R. N., Ribeiro, A. P. D., & Leal, S. C. (2018). Glass hybrid restorations as an alternative for restoring hypomineralized molars in the ART model. *BMC Oral Health*, *18*(1), 65. https://doi.org/10.1186/s12903-018-0528-0

[21] Lygidakis, N. A., Chaliasou, A., & Siounas, G. (2003). Evaluation of composite restorations in hypomineralized permanent molars: a four year clinical study. *European Journal of Paediatric Dentistry*, *4*, 143–148.

[22] Koleventi, A., Sakellari, D., Arapostathis, K. N., & Kotsanos, N. (2018). Periodontal impact of preformed metal crowns on permanent molars of children and adolescents: a pilot study. *Pediatric Dentistry*, *40*(2), 117–121.

[23] Zagdwon, A. M., Fayle, S. A., & Pollard, M. A. (2003). A prospective clinical trial comparing preformed metal crowns and cast restorations for defective first permanent molars. *European Journal of Paediatric Dentistry*, *4*, 138–142.

[24] Dhareula, A., Goyal, A., Gauba, K., Bhatia, S. K., Kapur, A., & Bhandari, S. (2019). A clinical and radiographic investigation comparing the efficacy of cast metal and indirect resin onlays in rehabilitation of permanent first molars affected with severe molar incisor hypomineralisation (MIH): a 36-month randomised controlled clinical trial. *European Archives of Paediatric Dentistry*, *20*(5), 489–500. https://doi.org/10.1007/s40368-019-00430-y

[25] Ashley, P., & Noar, J. (2019). Interceptive extractions for first permanent molars: a clinical protocol. *British Dental Journal*, *227*(3), 192–195. https://doi.org/10.1038/s41415-019-0561-7

[26] Bhandari, R., Thakur, S., Singhal, P., Chauhan, D., Jayam, C., & Jain, T. (2019). In vivo comparative evaluation of esthetics after microabrasion and microabrasion followed by casein phosphopeptide–amorphous calcium fluoride phosphate on molar incisor hypomineralization-affected incisors. *Contemporary Clinical Dentistry*, *10*(1), 9–15. https://doi.org/10.4103/ccd.ccd_852_17

[27] Kim, S., KIM, E., JEONG, T., & KIM, J. (2011). The evaluation of resin infiltration for masking labial enamel white spot lesions. *International Journal of Paediatric Dentistry*, *21*(4), 241–248. https://doi.org/10.1111/j.1365-263X.2011.01126.x

[28] Welbury, R. R. (1991). A clinical study of a microfilled composite resin for labial veneers. *International Journal of Paediatric Dentistry*, *1*(1), 9–15. https://doi.org/10.1111/j.1365-263X.1991.tb00315.x

[29] Monteiro, J., Ashley, P. F., & Parekh, S. (2020). Vital bleaching for children with dental anomalies: EAPD members' survey. *European Archives of Paediatric Dentistry*, *21*(5), 565–571. https://doi.org/10.1007/s40368-019-00494-w

[30] Schwendicke, F., Elhennawy, K., & Krois, J. (2020). Prevalence, incidence, and burden of molar incisor hypomineralization. In *Molar Incisor Hypomineralization: A Clinical Guide to Diagnosis and Treatment* (pp. 21–31). Springer. https://doi.org/10.1007/978-3-030-31601-3_3

[31] Sönmez, H., Yıldırım, G., & Bezgin, T. (2013). Putative factors associated with molar incisor hypomineralisation: an epidemiological study. *European Archives of Paediatric Dentistry*, *14*(6), 375–380. https://doi.org/10.1007/s40368-013-0012-0

[32] Koruyucu, M., Özel, S., & Tuna, E. B. (2018). Prevalence and etiology of molar-incisor hypomineralization (MIH) in the city of Istanbul. *Journal of Dental Sciences*, *13*(4), 318–328. https://doi.org/10.1016/j.jds.2018.05.002

[33] Scheffel, D. L. S., Jeremias, F., Fragelli, C. M. B., dos Santos-Pinto, L. A. M., Hebling, J., & de Oliveira Jr, O. B. (2014). Esthetic dental anomalies as motive for bullying in schoolchildren. *European Journal of Dentistry*, *8*(01), 124–128. https://doi.org/10.4103/1305-7456.126266

[34] Leal, S. C., Oliveira, T. R. M., & Ribeiro, A. P. D. (2017). Do parents and children perceive molar–incisor hypomineralization as an oral health problem? *International Journal of Paediatric Dentistry*, *27*(5), 372–379. https://doi.org/10.1111/ipd.12271

[35] Dias, F., Gradella, C. M. F., Ferreira, M. C., & Oliveira, L. B. (2021). Molar–incisor hypomineralization: parent's and children's impact perceptions on the oral health-related quality of life. *European Archives of Paediatric Dentistry*, *22*(2), 273–282. https://doi.org/10.1007/s40368-020-00556-4

[36] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*(4). https://doi.org/10.1136/svn-2017-000101

[37] Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., & Lau, A. Y. S. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072

[38] Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of Medical Systems*, *47*(1), 33. https://doi.org/10.1007/s10916-023-01925-4

[39] Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887. https://doi.org/10.3390/healthcare11060887

[40] Panahi, O., & Zeinaldin, M. (2024). Digital Dentistry: Revolutionizing Dental Care. *J Dent App*, *10*(1), 1121.

[41] Panahi, O., & Safaralizadeh, R. (2024). How Artificial Intelligence and Biotechnology are Transforming Dentistry. *Adv Biotech & Micro*, *18*, 555981. https://doi.org/10.19080/AIBM.2024.18.555981

[42] Kaur, A., Singh, S., Chandan, J. S., Robbins, T., & Patel, V. (2021). Qualitative exploration of digital chatbot use in medical education: A pilot study. *Digital Health*, *7*, 20552076211038150. https://doi.org/10.1177/20552076211038151

[43] Shamszare, H., & Choudhury, A. (2023). The impact of performance expectancy, workload, risk, and satisfaction on trust in ChatGPT: Cross-sectional survey analysis. *ArXiv Preprint ArXiv:2311.05632*. https://doi.org/10.2196/55399

[44] Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549–1565. https://doi.org/10.1002/cae.22326

[45] Bansal, H., & Khan, R. (2018). A review paper on human computer interaction. *International Journal of Advanced Research in Computer Science and Software Engineering*, *8*(4), 53. https://doi.org/10.23956/ijarcsse.v8i4.630

[46] Huynh, L. M., Bonebrake, B. T., Schultis, K., Quach, A., & Deibert, C. M. (2023). RETRACTED: New Artificial Intelligence ChatGPT Performs Poorly on the 2022 Self-assessment Study Program for Urology. *Urology Practice*, *10*(4), 409–415. https://doi.org/10.1097/UPJ.0000000000000406

[47] Noda, R., Izaki, Y., Kitano, F., Komatsu, J., Ichikawa, D., & Shibagaki, Y. (2024). Performance of ChatGPT and Bard in self-assessment questions for nephrology board renewal. *Clinical and Experimental Nephrology*, *28*(5), 465–469. https://doi.org/10.1007/s10157-023-02451-w

[48] Kerimbayev, N., Menlibay, Z., Garvanova, M., Djaparova, S., & Jotsov, V. (2024). A comparative analysis of generative AI models for improving learning process in higher education. *Nternational Conference Automatics and Informatics*, 271–276. https://doi.org/10.1109/ICAI63388.2024.10851491

[49] Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). 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 Medical Education*, *9*(1), e45312. https://doi.org/10.2196/45312

[50] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., & Maningo, J. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health*, *2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

[51] Duran, G. S., Yurdakurban, E., & Topsakal, K. G. (2025). The quality of CLP-related information for patients provided by ChatGPT. *The Cleft Palate Craniofacial Journal*, *62*(4), 588–595. https://doi.org/10.1177/10556656231222387

[52] Yurdakurban, E., Topsakal, K. G., & Duran, G. S. (2024). A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient information. *Journal of Stomatology, Oral and Maxillofacial Surgery*, *125*(5), 101757. https://doi.org/10.1016/j.jormas.2023.101757

[53] Nguyen, H. C., Dang, H. P., Nguyen, T. L., Hoang, V., & Nguyen, V. A. (2025). Accuracy of latest large Language models in answering multiple choice questions in dentistry: A comparative study. *PLoS One*, *20*(1), e0317423. https://doi.org/10.1371/journal.pone.0317423

[54] Kayabaşı, M., Köksaldı, S., & Durmaz Engin, C. (2025). Evaluating the reliability of the responses of large language models to keratoconus-related questions. *Clinical and Experimental Optometry*, *108*(7), 784–791. https://doi.org/10.1080/08164622.2024.2419524

[55] Patel, A. V, Panakam, A., Amin, K., Doshi, R. H., Patil, A., & Sheth, S. S. (2024). Comparative Readability Assessment of Four Large Language Models in Answers to Common Contraception Questions [Id 2683638]. *Obstetrics & Gynecology*, *143*(5S), 12S. https://doi.org/10.1097/01.AOG.0001013004.01563.47

[56] Strzalkowski, P., Strzalkowska, A., Chhablani, J., Pfau, K., Errera, M.-H., Roth, M., Schaub, F., Bechrakis, N. E., Hoerauf, H., & Reiter, C. (2024). Evaluation of the accuracy and readability of ChatGPT-4 and Google Gemini in providing information on retinal detachment: a multicenter expert comparative study. *International Journal of Retina and Vitreous*, *10*(1), 61. https://doi.org/10.1186/s40942-024-00579-9

[57] Rokhshad, R., Zhang, P., Mohammad-Rahimi, H., Pitchika, V., Entezari, N., & Schwendicke, F. (2024). Accuracy and consistency of chatbots versus clinicians for answering pediatric dentistry questions: A pilot study. *Journal of Dentistry*, *144*, 104938. https://doi.org/10.1016/j.jdent.2024.104938

[58] Ozden, I., Gokyar, M., Ozden, M. E., & Sazak Ovecoglu, H. (2024). Assessment of artificial intelligence applications in responding to dental trauma. *Dental Traumatology*, *40*(6), 722–729. https://doi.org/10.1111/edt.12965

[59] Uribe, S. E., Maldupa, I., Kavadella, A., El Tantawi, M., Chaurasia, A., Fontana, M., Marino, R., Innes, N., & Schwendicke, F. (2024). Artificial intelligence chatbots and large language models in dental education: Worldwide survey of educators. *European Journal of Dental Education*, *28*(4), 865–876. https://doi.org/10.1111/eje.13009

[60] Büttner, M., Leser, U., Schneider, L., & Schwendicke, F. (2024). Natural language processing: chances and challenges in dentistry. *Journal of Dentistry*, *141*, 104796. https://doi.org/10.1016/j.jdent.2023.104796