



Assessing the Accuracy and Reliability of Artificial Intelligence in First Permanent Molar Extraction Decisions

Aslıhan Yelkenci^{1,*}

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

ARTICLE INFO

Article history:

Received 01 December 2025

Received in revised form 5 February 2026

Accepted 15 February 2026

Available online 28 February 2026

Keywords:

Artificial intelligence; Pediatric dentistry;
First permanent molar; Extraction
decisions; Decision support

ABSTRACT

The first permanent molar (M1) plays a critical role in maintaining functional occlusion during mixed dentition. When this tooth is severely compromised by deep caries, structural weakness, or developmental enamel defects, extraction may be required. However, the decision to extract M1 is influenced by multiple prognostic factors, including the child's age, dental development stage, second molar (M2) angulation, and the presence of third molars. In this context, artificial intelligence (AI)-based conversational systems can support clinical judgment and patient education by offering quicker access to relevant information. This descriptive-analytical study evaluated the performance of two AI-powered language models (ChatGPT-3.5 and ChatGPT-4) in providing clinical guidance on M1 extraction. Ten standardized questions were formulated based on established prognostic parameters associated with spontaneous space closure. Both AI systems produced coherent and clinically relevant content. However, ChatGPT-4 tended to produce longer responses with greater sentence complexity, resulting in lower readability and higher academic demand. By comparison, ChatGPT-3.5 produced shorter, more straightforward responses that may be more appropriate for communication with parents or patients. Future research incorporating varied question structures, such as case-based scenarios, radiographic-based questions, and real-user feedback, may help define how conversational AI can be safely and effectively integrated into pediatric dental workflows. Until then, AI-based tools should be used as supportive resources rather than as independent clinical decision-makers, and their outputs should be reviewed by qualified professionals to ensure patient safety.

1. Introduction

The first permanent molar (M1), often referred to as the six-year molar, plays an important role in the development of a stable occlusal relationship [1]. It is also considered one of the permanent teeth most susceptible to caries [2]. Because it erupts early, the tooth is exposed to the oral environment for a longer period. This prolonged exposure increases the risk of demineralization, particularly in the presence of complex occlusal anatomy [3]. This risk may be further increased in teeth affected by developmental disturbances, such as molar-incisor hypomineralization (MIH). In

* Corresponding author.

E-mail address: aslihanzihnil@gmail.com

<https://doi.org/10.59543/bcx2xp86>

these cases, enamel weakness can lead to more rapid structural breakdown [4]. In severely compromised M1s, extraction may be considered when repeated restorative attempts fail or endodontic treatment is not successful [5]. Unplanned loss of the M1 during the mixed dentition period has been associated with several adverse clinical outcomes. These include displacement or overeruption of adjacent teeth, loss of arch integrity, midline deviation, temporomandibular joint problems, and altered craniofacial growth [6].

Because the M1 plays an important role in arch integrity and functional occlusal stability, the timing of its extraction warrants careful clinical consideration. Spontaneous space closure is more commonly observed when M1 extraction is carried out between approximately 8 and 11.5 years of age and occurs prior to the eruption of the second permanent molar (M2) [7]. Higher success rates of spontaneous mesial drift are reported when the M2 is at Demirjian stages E and F [4], [5], [6]. Additionally, extraction performed between 8 and 9 years of age facilitates contact between the erupting M2 and the second premolar. In patients predisposed to crowding, this timing may reduce the risk of anterior crowding and help maintain arch form [8].

Beyond age and root development, several other factors influence the success of space closure. Mesial or upright angulation of the M2 is associated with higher rates of space closure [9], [10]. The presence of third molar (M3) may also contribute to more favorable closure outcomes [11]. Tooth movement tends to occur more readily in the maxillary arch than in the mandible, which may be related to differences in bone density [12], [13]. Although M3 can be helpful during treatment planning, their radiographic identification at younger ages is often limited. In the mandible, calcification typically begins between 7 and 10 years of age and commonly remains insufficient for reliable radiographic detection until adolescence [14]. When M3s are not yet radiographically visible, extraction may be deferred after evaluation of other clinical factors [15]. At the same time, early loss of the M1 has been reported to be associated with accelerated M3 development on the extraction side compared with the contralateral side [16]. Space closure generally occurs through mesial movement of the M2, while distal movement of the second premolar may also contribute to this process. [17]. Considering these factors, dental development and occlusal relationships should be evaluated in a careful and comprehensive manner before planning M1 extraction.

Compensating extraction of the opposing maxillary molar may be considered after mandibular M1 extraction to support inter-arch alignment. However, extraction of a sound mandibular M1 solely because the maxillary M1 has been extracted is usually avoided. Balancing extraction of the contralateral M1 in the same arch has been considered for arch symmetry, but it is rarely supported when midline preservation is the only objective. In patients with Class III malocclusion, where treatment planning is more complex, both balancing and compensating extractions are usually avoided [15].

As extraction-related decisions can be complex, Artificial intelligence (AI) based chatbots may support clinical decision-making and patient education. AI refers to a field of computer science focused on enabling machines to carry out tasks that typically require human cognitive abilities, including learning, reasoning, and language processing [18]. Within this domain, chatbots are AI-based tools used to process natural language and generate context-appropriate responses [19].

Early chatbot development began with ELIZA, a rule-based program created by Joseph Weizenbaum in 1966 to model basic conversational interaction [20]. Over the years, companies such as Apple, Google, IBM, and Microsoft have contributed significantly to the development of this concept through ongoing technological work. Today, systems such as ChatGPT (OpenAI) [21] use advanced natural language processing (NLP) methods to interpret user input and generate relevant responses, with growing applications in education and healthcare. Within educational settings,

chatbots contribute to interactive and personalized learning experiences by responding to questions, supporting assignment evaluation, and tracking student progress[21]. In healthcare settings, chatbots support patients by providing tailored health information, offering preliminary guidance on possible diagnoses or treatment options, and directing them to appropriate services. More broadly, disciplines including cardiology, radiology, and urology have begun to consider chatbots as complementary tools for diagnostic reasoning and clinical decision-making. Nevertheless, limitations related to accuracy, possible algorithmic bias, and the need for human oversight remain important barriers to their use without clinical supervision [22].

Dentistry has similarly undergone a rapid phase of digital transformation, shaped by advances in robotics, machine learning, and AI technologies. In pediatric dentistry, these tools have been applied across a range of tasks, including the detection of dental plaque and caries, the identification of fissure sealants, mesiodens, and supernumerary teeth, as well as the classification of primary and young permanent dentition, chronological age estimation, and the recognition of anatomical landmarks [23]. Building on these developments, structured consultation systems, including chatbots, are now being explored within clinical dentistry. In complex clinical situations like M1 extraction, these systems can assist decision-making by offering quick access to relevant, guideline-based information. For clinicians, students, and parents, chatbots offer a practical way to obtain essential information aligned with current evidence and clinical protocols. Therefore, this study examined responses generated by ChatGPT-3.5 and ChatGPT-4 to ten standardized questions related to first permanent molar (M1) extraction, in order to explore potential differences between the two model versions with respect to information quality, reliability, readability, and linguistic features.

2. Methodology

2.1 Study Design

This cross-sectional analytical study examined how two AI-based language models, ChatGPT-3.5 and ChatGPT-4, respond to educational and clinical questions related to first permanent molar (M1) extraction. Ten standardized questions were developed using two previously published systematic reviews [24], [25] that identified the main prognostic factors influencing spontaneous space closure after M1 extraction (Table 1). The questions focused on major clinical considerations such as extraction indications, potential complications, optimal timing, arch-related differences, M2 angulation, M3 presence, and the need for space maintenance or orthodontic treatment. Each question was entered independently in a separate session for both models to minimize cross-context influence. All responses were collected in English for analysis. Readability indices (FRES, FKGL) and linguistic metrics such as word count, average sentence length, lexical diversity, and medical terminology density were calculated programmatically. The outcomes were further compared using publication-quality plots generated at 300 dpi with serif typography for enhanced visibility. As no human or animal participants were involved, ethical approval was not required.

Table 1

Standardized clinical questions and associated prognostic factors used for chatbot evaluation

No.	Original Question	Associated Clinical / Prognostic Factor	Purpose of Evaluation
1	What are the most common reasons for M1 extraction in children?	Extraction indication	To assess whether the chatbot identifies main etiological causes such as dental caries and molar–incisor hypomineralization (MIH).
2	What possible complications may occur after M1 extraction?	Post-extraction consequences	To evaluate whether the chatbot mentions potential problems such as drifting or overeruption of adjacent teeth, midline shift, or TMJ dysfunction.
3	At which Demirjian developmental stage does M1 extraction most favorably increase the rate of spontaneous space closure?	M2 developmental stage / Extraction timing	To assess whether the chatbot identifies Demirjian stages E and F as optimal for extraction, associated with the highest spontaneous closure rates, and recognizes that stage H is linked to poor outcomes.
4	Does the child’s chronological age affect the outcome?	Chronological age	To determine if the chatbot indicates 8–10 years as the ideal age range for successful spontaneous closure.
5	Is there a difference between the maxilla and mandible in spontaneous space closure?	Arch-related differences	To check whether the chatbot distinguishes between arches, noting higher success in the maxilla.
6	How does the angulation of the second molar (M2) influence the result?	M2 angulation	To assess if the chatbot explains that mesial inclination of M2 enhances closure rates.
7	Does the presence of the third molar (M3) influence the extraction decision or closure outcome?	M3 presence	To evaluate whether the chatbot recognizes that M3 presence may improve closure, though evidence is inconsistent.
8	What are the reported spontaneous space-closure rates after M1 extraction?	Clinical outcome rates	To test if the chatbot provides realistic closure percentages (~ 52–94 % maxilla; 39–89 % mandible).
9	Should a space maintainer be used, or can spontaneous closure be expected?	Space-maintenance requirement	To assess whether the chatbot integrates developmental factors when recommending use of a space maintainer.
10	Is orthodontic treatment always required after extraction?	Orthodontic need	To assess whether the chatbot relates ideal clinical conditions, including Demirjian stage, age, and arch characteristics, to a reduced need for orthodontic intervention.

2.2 Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (Version 28.0; IBM Corp., Armonk, NY, USA) and GraphPad Prism software (Version 9.0). Quantitative variables, including word count, sentence length, lexical diversity, and medical terminology density, were summarized with descriptive statistics and expressed as mean \pm standard deviation (SD). Before comparing the ChatGPT-3.5 and ChatGPT-4 models, the normality of each linguistic parameter was tested with the Shapiro–Wilk test. Parameters that followed a normal distribution were analyzed with the paired-sample t-test, while those that did not meet normality assumptions were tested with the Wilcoxon signed-rank test. Differences in readability indices (FRES and FKGL) between models were examined using one-way repeated-measures analysis of variance (ANOVA). Post-hoc pairwise comparisons were adjusted with the Bonferroni correction to control for type I error. Correlations between word count, sentence length, and FKGL scores were analyzed using Pearson’s correlation coefficient (r) for parametric data and Spearman’s rho (ρ) for non-parametric data. The strength of these correlations was interpreted based on Cohen’s criteria (0.10–0.29 weak, 0.30–0.49 moderate, ≥ 0.50 strong). Effect sizes were calculated as Cohen’s d for t-tests and eta squared (η^2) for ANOVA models to estimate practical significance. For all analyses, statistical significance was defined as $p < 0.05$.

3. Results

Across all ten questions, ChatGPT-4 showed consistently lower FRES scores than ChatGPT-3.5, indicating lower readability. This pattern is associated with a higher clause density and more frequent use of technical vocabulary in ChatGPT-4 responses. For several items, negative FRES values indicated a highly academic style that would require simplification for parent-facing informational materials. These differences are shown in Figure 1. Differences in FRES scores between ChatGPT-3.5 and ChatGPT-4 were statistically significant (ANOVA, $p < 0.05$).

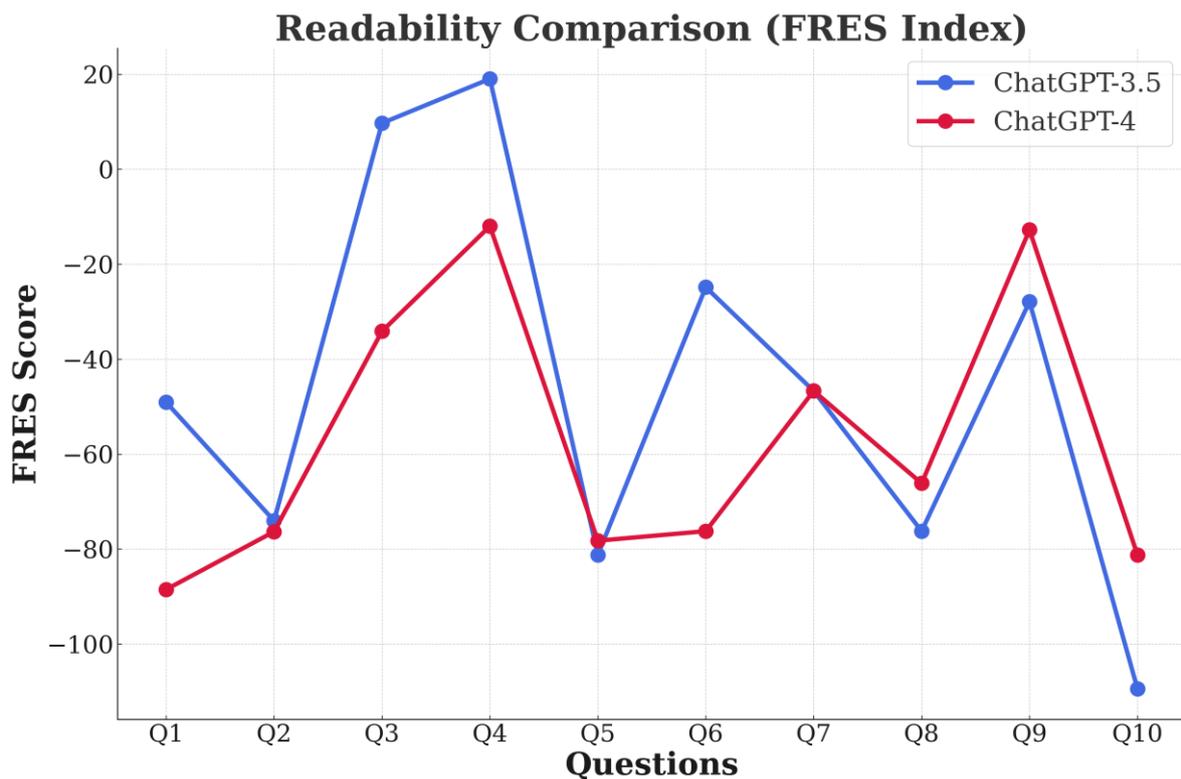


Fig. 1. Readability Comparison (FRES Index)

Higher FKGL scores were observed for ChatGPT-4 compared with ChatGPT-3.5, indicating a higher required reading level. The largest differences between models were observed for items Q2 and Q10, in which ChatGPT-4 responses exhibited greater contextual scope and multi-clause constructions, corresponding to higher grade-level estimates. This supports the view that ChatGPT-4 is better suited for clinician-to-clinician communication while ChatGPT-3.5 is more accessible for lay audiences. These differences are shown in Fig. 2. The difference in FKGL scores between the two models was statistically significant (ANOVA, $p < 0.05$).

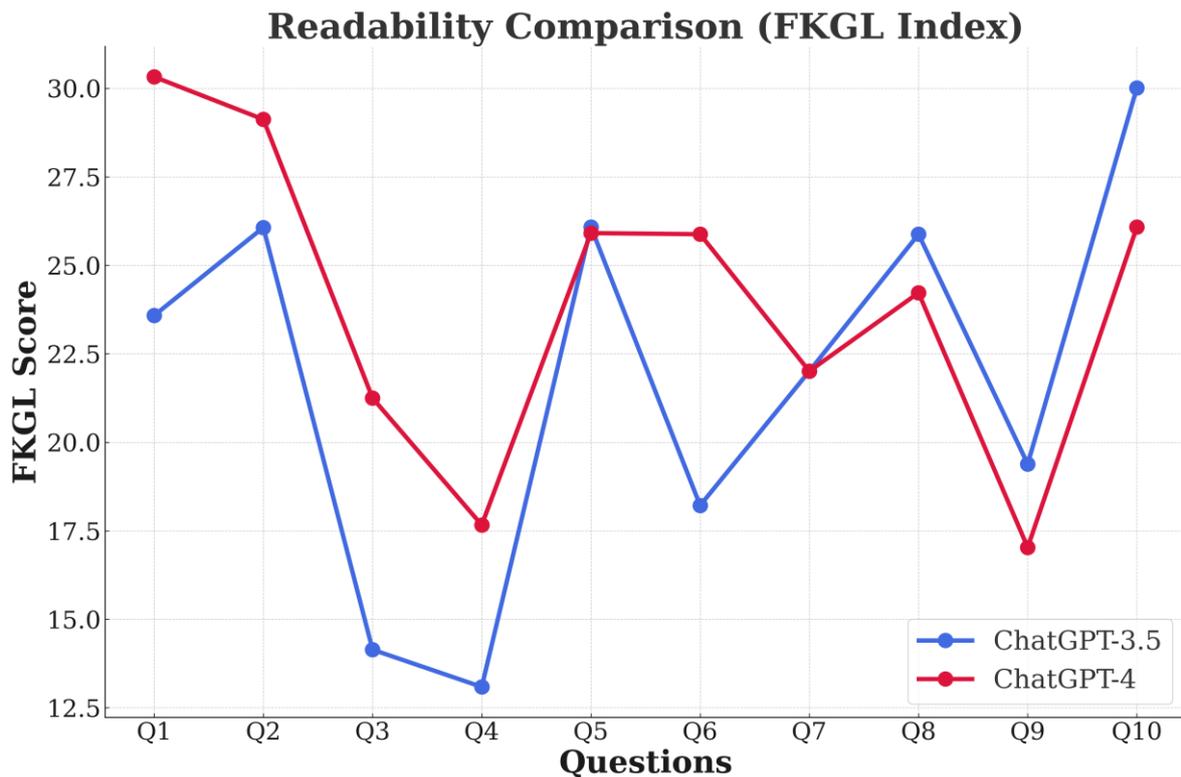


Fig. 2. Readability Comparison (FKGL Index)

On average, ChatGPT-4 produced longer responses (Word Count: 14.30 ± 5.78) with longer sentences (14.60 ± 5.49) than ChatGPT-3.5 (8.70 ± 2.66 ; 8.90 ± 2.41 , respectively). Lexical diversity values were near ceiling for both models, as expected given the short length of the responses; however, term density was comparable between models (ChatGPT-3.5: 0.22 ± 0.14 vs ChatGPT-4: 0.21 ± 0.09). The higher difficulty observed for ChatGPT-4 was therefore associated with greater syntactic expansion rather than increased jargon frequency. These results are summarized in Table 2. Paired comparisons showed statistically significant differences in word count and sentence length between the two models ($p < 0.05$).

Table 2

Summary Statistics (Mean \pm SD) Across AI Models

Metric	ChatGPT-3.5	ChatGPT-4
Word Count	8.70 ± 2.66	14.30 ± 5.78
Sentence Length	8.90 ± 2.41	14.60 ± 5.49
Lexical Diversity	1.00 ± 0.00	0.98 ± 0.04
Term Density	0.22 ± 0.14	0.21 ± 0.09

The largest inter-model divergence in output length occurs at Q2 (WC: 10 vs 26) and Q10 (6 vs 14), where ChatGPT-4 expands on functional and multidisciplinary considerations (e.g., occlusal balance, orthodontic referral). Term density was highest in questions focusing on anatomical or developmental criteria (Q5–Q6, Q10), reflecting the greater informational demands of decision-oriented content. Identical word counts observed for Q8 suggest the presence of a well-established and narrowly defined fact set (reported spontaneous closure rates), resulting in similar response lengths across models. These item-specific differences are presented in Table 3.

Table 3

Per-Question Linguistic Metrics

Question	WC_3.5	WC_4	SentLen_3.5	SentLen_4	LexDiv_3.5	LexDiv_4	TermDen_3.5	TermDen_4
Q1	14	19	14.0	19.0	1.0	1.0	0.14	0.0
Q2	10	26	10.0	26.0	1.0	0.92	0.0	0.12
Q3	9	13	9.0	13.0	1.0	1.0	0.0	0.08
Q4	10	14	10.0	14.0	1.0	0.86	0.2	0.21
Q5	6	7	6.0	7.0	1.0	1.0	0.33	0.29
Q6	6	11	6.0	11.0	1.0	1.0	0.33	0.27
Q7	9	11	9.0	11.0	1.0	1.0	0.11	0.18
Q8	8	8	8.0	8.0	1.0	1.0	0.25	0.25
Q9	9	10	9.0	10.0	1.0	1.0	0.11	0.1
Q10	6	14	6.0	14.0	1.0	0.93	0.5	0.29

ChatGPT-4 consistently produced longer responses than ChatGPT-3.5 across most items, with the largest increases observed for Q2 and Q10. This corresponds to the higher FKGL scores and reflects more extensive elaboration of risk–benefit context in clinically nuanced prompts. This pattern is shown in Fig. 3.

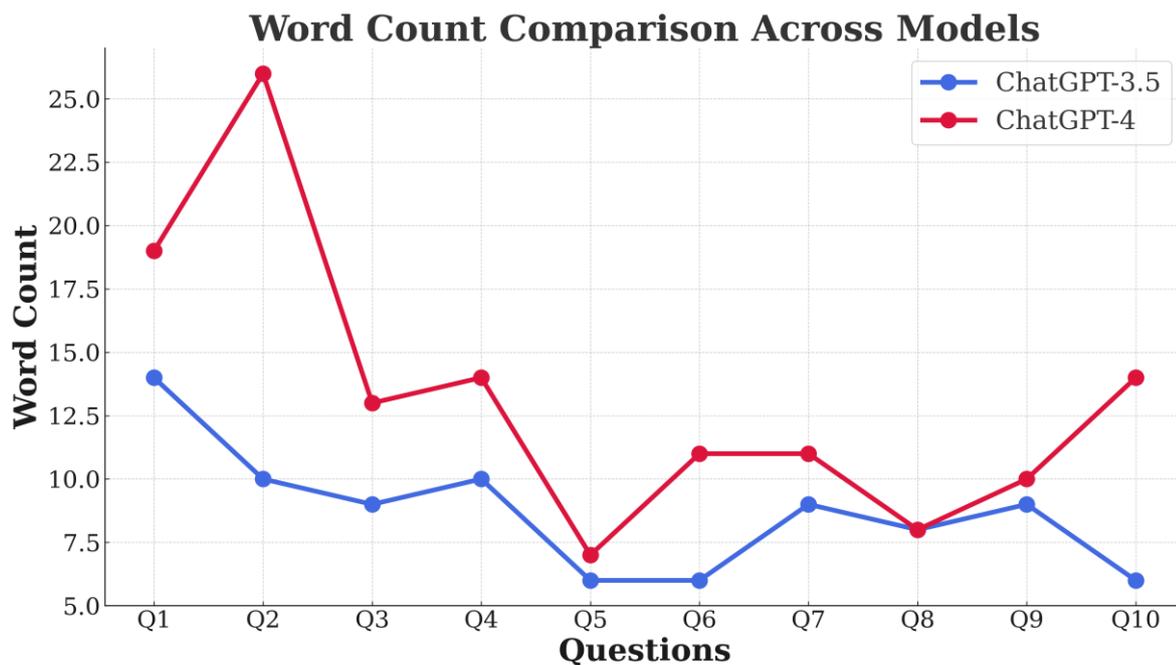


Fig.3. Word Count Comparison Across Models

Medical terminology density was similar across models, with peaks in items related to anatomical planes and developmental stages, such as mesial inclination and Demirjian staging. Accordingly, the observed readability differences were associated more with clause stacking, subordinate structures, and connective framing in ChatGPT-4 outputs than with differences in jargon frequency. These observations are shown in the heatmap provided in Fig. 4.

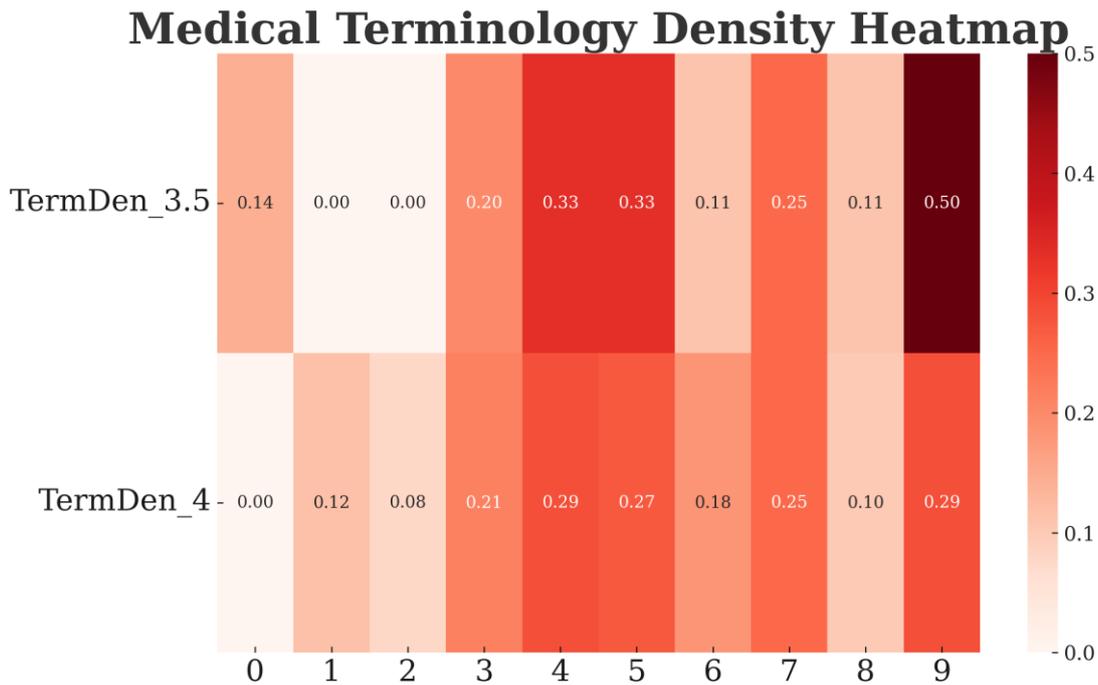


Fig. 4. Medical Terminology Density (Heatmap)

4. Discussion

Extraction of the M1 is often unavoidable when the tooth is severely compromised due to deep caries, extensive restorations, endodontic or periodontal problems, apical pathology, or MIH [24]. However, the timing of extraction is a decisive factor influencing the likelihood of spontaneous space closure. Previous studies have shown that extractions performed between approximately 8 and 11 years of age provide the most favorable outcomes, particularly when conducted before the eruption of the M2 [11], [26]. At these ages, M2 is usually in Demirjian stages E or F, which are associated with increased mesial drift and higher rates of spontaneous closure [27], [28]. In addition to chronological age, factors such as dental developmental stage, tooth angulation, and the presence of the M2 and M3 have been identified as important prognostic determinants of successful space closure. Studies indicate that mesial angulation of M2 and the presence of M3 are linked to higher space closure rates [11], [29]. These findings underscore that decisions regarding M1 extraction timing are multifactorial and should be based on a comprehensive clinical and radiographic assessment. In the present study, ten questions were developed to address key prognostic factors, including extraction indications, timing, arch-related differences, M2 angulation, and M3 presence, to assess how AI-driven chatbots respond to evidence-based clinical situations.

This study examined differences between ChatGPT-3.5 and ChatGPT-4 in delivering clinically relevant information related to M1 extraction in pediatric dentistry, focusing on readability, linguistic features, and medical terminology use. While both models produced largely accurate and coherent answers, they differed clearly in their style of presentation. ChatGPT-4 used longer and syntactically

more complex sentences, which resulted in lower FRES and higher FKGL scores. ChatGPT-4 communicates in a style that is more appropriate for dental professionals, whereas ChatGPT-3.5 tends to present information in a form that is easier for parents or students to understand.

The present results are consistent with previous studies that have examined the performance of AI-driven conversational agents in medical settings. Studies have shown that ChatGPT-4 can achieve diagnostic reasoning performance comparable to experts or exceed average scores in medical examinations [30], [31]. On the other hand, reported variability in answer consistency and notable error rates in case-based decision-making tasks [32] indicate that these systems are not yet reliable enough to independently guide clinical care. The potential for misinformation has been emphasized particularly in situations where chatbots are used directly by patients without professional supervision [33].

Similar findings have also been reported in the field of dentistry. Although high levels of accuracy have been reported in periodontology and oral and maxillofacial applications, performance decreases substantially in multiple-choice formats and more complex clinical scenarios [34-36]. In pediatric dentistry, chatbots have been shown to perform less effectively than expert clinicians, despite offering an acceptable baseline of information [37]. Consistent with these findings, the present study indicates that conversational AI tools can support basic education and preliminary counseling. ChatGPT-4's detailed and professional style may help clinicians quickly revisit core principles of M1 extraction. However, its higher level of linguistic complexity may make it less suitable for caregivers who require simplified explanations. Therefore, selecting a model that is appropriate for the target audience may improve communication effectiveness. Nevertheless, clinical accuracy should always be confirmed through expert review.

While these findings provide meaningful insight into the current use of conversational AI in pediatric dentistry, several limitations should be noted. Only open-ended questions were included; yes/no formats, multiple-choice items, and case-based scenarios were not used. As a result, the findings may not be broadly generalizable and may not reflect the full diversity of decision-making scenarios in pediatric dentistry. Future research may explore more diverse question formats, such as X-ray-based questions or stepwise clinical cases, to better assess the performance of conversational AI in complex clinical situations. Taken together, the results indicate that conversational AI systems can support patient communication and provide faster access to information in pediatric dental practice. At the same time, limitations related to accuracy consistency, source transparency, and ethical responsibility remain important considerations. Controlled use supported by professional oversight will be necessary to ensure the safe and responsible integration of these technologies into clinical workflows.

5. Conclusion

Both ChatGPT-3.5 and ChatGPT-4 generated clinically coherent content in the context of M1 extraction. ChatGPT-4 tended to provide more detailed and professionally oriented responses suitable for expert communication and documentation, while ChatGPT-3.5 offered simpler explanations that may be more useful for parent-focused education.

However, the greater linguistic complexity observed in ChatGPT-4 indicates that readability measures, such as limiting sentence length or providing brief glossaries, may be necessary when adapting its output for patient-facing use. Therefore, AI-generated information should be considered as a supportive tool rather than a standalone source of clinical decision-making, with content reviewed and refined under professional oversight.

Future studies including varied question structures, such as case-based scenarios or decision-focused prompts, along with direct user assessments, will help better determine how conversational AI systems can be safely and effectively incorporated into pediatric dental decision-making and communication.

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] Taylor, G. D., & Bulmer, V. (2025). Advances in knowledge and practice benefiting the health and management of first permanent molars in children. *British Dental Journal*, 238(2), 92-98. <https://doi.org/10.1038/s41415-024-8012-5>
- [2] Broadbent, J. M., Foster-Page, L. A., Thomson, W. M., & Poulton, R. (2013). Permanent dentition caries through the first half of life. *British Dental Journal*, 215(7), E12. <https://doi.org/10.1038/sj.bdj.2013.991>
- [3] 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>
- [4] Steffen, R., Krämer, N., & Bekes, K. (2017). The Würzburg MIH concept: The MIH treatment need index (MIH TNI): A new index to assess and plan treatment in patients with molar-incisor hypomineralisation (MIH). *European Archives of Paediatric Dentistry*, 18(5), 355-361. <https://doi.org/10.1007/s40368-017-0301-0>
- [5] Normando, A. D. C., Maia, F. A., da Silva Ursi, W. J., & Simone, J. L. (2010). Dentoalveolar changes after unilateral extractions of mandibular first molars and their influence on third molar development and position. *World Journal of Orthodontics*, 11(1), 55-60. <https://doi.org/10.1007/s40510-010-0004-6>
- [6] Saber, A. M., Altoukhi, D. H., Horaib, M. F., El-Housseiny, A. A., Alamoudi, N. M., & Sabbagh, H. J. (2018). Consequences of early extraction of compromised first permanent molar: A systematic review. *BMC Oral Health*, 18(1), 59. <https://doi.org/10.1186/s12903-018-0528-2>
- [7] Rathee, M., & Sapra, A. (2024). *Dental Caries [Updated 2023 Jun 21]*. In StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK551699/>
- [8] Richardson, A. (1979). Spontaneous changes in the incisor relationship following extraction of lower first permanent molars. *British Journal of Orthodontics*, 6(2), 85-90.
- [9] Schwendicke, F., Elhennawy, K., & Krois, J. (2020). Prevalence, incidence, and burden of molar-incisor hypomineralization. In K. Bekes (Ed.), *Molar Incisor Hypomineralization: A Clinical Guide to Diagnosis and Treatment*(pp. 21-31). Springer. https://doi.org/10.1007/978-3-030-31601-3_4
- [10] D. Ong and J. E. Bleakley, "Compromised first permanent molars: an orthodontic perspective," *Aust Dent J*, vol. 55, no. 1, pp. 2-14, 2010.
- [11] Teo, T. K.-Y., Ashley, P. F., & Derrick, D. (2016). Lower first permanent molars: Developing better predictors of spontaneous space closure. *European Journal of Orthodontics*, 38(1), 90-95. <https://doi.org/10.1093/ejo/cjv082>

- [12] Deguchi, T., Takano-Yamamoto, T., Yabuuchi, T., Ando, R., Roberts, W. E., & Garetto, L. P. (2008). Histomorphometric evaluation of alveolar bone turnover between the maxilla and the mandible during experimental tooth movement in dogs. *American Journal of Orthodontics and Dentofacial Orthopedics*, 133(6), 889-897.
- [13] Chugh, T., Jain, A. K., Jaiswal, R. K., Mehrotra, P., & Mehrotra, R. (2013). Bone density and its importance in orthodontics. *Journal of Oral Biology and Craniofacial Research*, 3(2), 92-97. <https://doi.org/10.1016/j.jobcr.2013.01.001>
- [14] Massler, M., & Schour, I. (1946). Growth of the child and the calcification pattern of the teeth. *American Journal of Orthodontics and Oral Surgery*, 32(9), 495-517.
- [15] Cobourne, M. T., Williams, A., & Harrison, M. (2014). National clinical guidelines for the extraction of first permanent molars in children. *British Dental Journal*, 217(11), 643-648. <https://doi.org/10.1038/sj.bdj.2014.11>?
- [16] Yavuz, I., Baydaş, B., İkbāl, A., Dağsuyu, İ. M., & Ceylan, I. (2006). Effects of early loss of permanent first molars on the development of third molars. *American Journal of Orthodontics and Dentofacial Orthopedics*, 130(5), 634-638.
- [17] Çağlaroğlu, M., Kılıç, N., & Erdem, A. (2008). Effects of early unilateral first molar extraction on skeletal asymmetry. *American Journal of Orthodontics and Dentofacial Orthopedics*, 134(2), 270-275. <https://doi.org/10.1016/j.ajodo.2006.07.036>
- [18] Krishnan, C., Gupta, A., Gupta, A., & Singh, G. (2022). Impact of artificial intelligence-based chatbots on customer engagement and business growth. In T. P. Hong, L. Serrano-Estrada, A. Saxena, & A. Biswas (Eds.), *Deep Learning for Social Media Data Analytics* (pp. 195-210). Springer. https://doi.org/10.1007/978-3-031-10869-3_11
- [19] Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, 7(2), 139-172. <https://doi.org/10.1080/23270012.2020.1756939>
- [20] Wang, H., Zhang, Q., Ip, M., & Lau, J. T. F. (2018). Social media-based conversational agents for health management and interventions. *Computer (Long Beach Calif)*, 51(8), 26-33.
- [21] Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, 6(7). <https://doi.org/10.14569/IJACSA.2015.060712>
- [22] Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006.
- [23] Tuzoff, D. V., et al. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofacial Radiology*, 48(4), 20180051. <https://doi.org/10.1259/dmfr.20180051>
- [24] Alqanas, S., Alsaheim, J., Aljami, A., Alsudairi, N., Ahmad, S., Sharma, S., Rajinder, S., Alamri, A., Alhazmi, H., & Hegazi, F. (2025). Factors related to spontaneous space closure following early first permanent molar extraction: A systematic review. *International Journal of Paediatric Dentistry*, 35(3), 625-638. <https://doi.org/10.1111/ipd.13273>
- [25] Hamza, B., Papageorgiou, S. N., Patcas, R., & Schätzle, M. (2024). Spontaneous space closure after extraction of permanent first molars in children and adolescents: A systematic review and meta-analysis. *European Journal of Orthodontics*, 46(6), cjae054. <https://doi.org/10.1093/ejo/cjae054>
- [26] Brusevold, I. J., Kleivene, K., Grimsøen, B., & Skaare, A. B. (2022). Extraction of first permanent molars severely affected by molar-incisor hypomineralisation: A retrospective audit. *European Archives of Paediatric Dentistry*, 23(1), 89-95.
- [27] Aldahool, Y., Sonesson, M., & Dimberg, L. (2024). Spontaneous space closure in patients treated with early extraction of the first permanent molar: A retrospective cohort study using radiographs. *The Angle Orthodontist*, 94(2), 180-186. <https://doi.org/10.2319/12345.2023.68>
- [28] Serindere, G., Bolgul, B., Parlar, T., & Cosgun, A. (2019). Effects of first permanent molar extraction on space changes observed in the dental arch using data-mining method. *Nigerian Journal of Clinical Practice*, 22(7), 936-942.
- [29] Nordeen, K. A., Kharouf, J. G., Mabry, T. R., Dahlke, W. O., Beiraghi, S., & Tasca, A. W. (2022). Radiographic evaluation of permanent second molar substitution after extraction of permanent first molar: Identifying predictors for spontaneous space closure. *Pediatric Dentistry*, 44(2), 123-130.
- [30] Huang, L., Hu, J., Cai, Q., Ye, A., Chen, Y., Xiao-Zhi, Z. Y., Liu, Y., Zheng, J., & Meng, Z. (2024). Preliminary discrimination and evaluation of clinical application value of ChatGPT4o in bone tumors. *Journal of Bone Oncology*, 48, 100632. <https://doi.org/10.1016/j.jbo.2024.100632>
- [31] Nicikowski, J., Szczepański, M., Miedziaszczyk, M., & Kudliński, B. (2024). The potential of ChatGPT in medicine: An example analysis of nephrology specialty exams in Poland. *Clinical Kidney Journal*, 17(8), sfae193. <https://doi.org/10.1093/ckj/sfae193>

- [32] G. ÇELİK, T. KARAÇOCUK, F. ÇİFTÇİ, and F. YÜLEK, "Nöro-Oftalmolojide Yapay Zeka Programları Bilgi Düzeylerinin Karşılaştırmalı Olarak Değerlendirilmesi.," *MN Ophthalmology/MN Oftalmoloji*, vol. 31, no. 4, 2024.
- [33] Davis, R. J., Ayo-Ajibola, O., Lin, M. E., Swanson, M. S., Chambers, T. N., Kwon, D. I., & Kokot, N. C. (2024). Evaluation of oropharyngeal cancer information from revolutionary artificial intelligence chatbot. *Laryngoscope*, 134(5), 2252-2257. <https://doi.org/10.1002/lary.31191>
- [34] Azadi, A., Gorjinejad, F., Mohammad-Rahimi, H., Tabrizi, R., Alam, M., & Golkar, M. (2024). Evaluation of AI-generated responses by different artificial intelligence chatbots to the clinical decision-making case-based questions in oral and maxillofacial surgery. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 137(6), 587-593. <https://doi.org/10.1016/j.oooo.2024.02.018>
- [35] Tomo, S., Lechien, J. R., Bueno, H. S., Cantieri-Debortoli, D. F., & Simonato, L. E. (2024). Accuracy and consistency of ChatGPT-3.5 and -4 in providing differential diagnoses in oral and maxillofacial diseases: A comparative diagnostic performance analysis. *Clinical Oral Investigations*, 28(10), 544.
- [36] Chatzopoulos, G. S., Koidou, V. P., Tsalikis, L., & Kaklamanos, E. G. (2024). Large language models in periodontology: Assessing their performance in clinically relevant questions. *Journal of Prosthetic Dentistry*. Advance online publication. <https://doi.org/10.1016/j.prosdent.2024.10.020>
- [37] Jung, Y. S., Chae, Y. K., Kim, M. S., Lee, H. S., Choi, S. C., & Nam, O. H. (2024). Evaluating the accuracy of artificial intelligence-based chatbots on pediatric dentistry questions in the Korean national dental board exam. *Journal of the Korean Academy of Pediatric Dentistry*, 51(3), 299–309.