Cumhuriyet Dental Journal, 28(2): 237-245, 2025
DOI: https://doi.org/10.7126/cumudj.1634513

Cumhuriyet Dental Journal

| cdj.cumhuriyet.edu.tr | Founded: 1998 Available online, ISSN: 1302-5805 Publisher: Sivas Cumhuriyet Üniversitesi

Evaluation of the Performance of Different Chatbots' Responses to Restorative Dentistry-Related Questions

Cansu Yıkıcı Çöl[1,a], Merve Nezir[2,b,*], Suat Özcan[2,c]

[1]*Panorama Ankara Oral and Dental Health Clinic, Çankaya, Ankara, Turkiye.*
[2]*Department of Restorative Dentistry, Faculty of Dentistry, Gazi University, Çankaya, Ankara, Turkiye.*
**Corresponding author*

Research Article

History

Received: 06/02/2025
Accepted: 06/03/2025

ABSTRACT

Objectives: To evaluate and compare the performance of the responses given by three different chatbots to questions related to the field of restorative dentistry for undergraduate education, specialist education, and which are considered controversial currently.

Materials and Methods: Thirty-five questions total was created by two dentists. In these questions, many different topics such as terminology, treatment procedures, technical details, material and application procedure, post-procedure care, indications, contraindications, approach in the presence of medical problems are touched upon. Three different chatbots (Copilot, Gemini and ChatGPT) were used in the study. Evaluation was done using a 5-point Likert Scale. The statistical significance level was determined as 0.05.

Results: When the correlation of the sub-dimensions among themselves is evaluated, there is a very strong positive statistically significant correlation between the questions about undergraduate education and the questions about specialist education ($p < 0.001$). Out of a total of 105 responses, Copilot produced 48 valid responses and 57 invalid responses. While Gemini produced 54 valid and 51 invalid responses, ChatGPT produced 58 valid and 47 invalid responses.

Conclusions: We think that this study may provide ideas for further studies in terms of evaluating the responses to various questions, including especially controversial questions regarding different areas of dentistry.

Keywords: *Artificial intelligence, ChatGPT, Copilot, Gemini, Restorative Dentistry.*

Restoratif Diş Hekimliği ile İlgili Sorulara Verilen Farklı Sohbet Botu Yanıtlarının Performansının Değerlendirilmesi

Araştırma Makalesi

Süreç

Geliş: 06/02/2025
Kabul: 06/03/2025

ÖZET

Amaç: Lisans eğitimi, uzmanlık eğitimi ve günümüzde tartışmalı kabul edilen restoratif diş hekimliği alanıyla ilgili sorulara üç farklı chatbot tarafından verilen yanıtların performansını değerlendirmek ve karşılaştırmak.

Gereç ve Yöntemler: Toplam 35 soru iki diş hekimi tarafından oluşturuldu. Bu sorularda terminoloji, tedavi prosedürleri, teknik detaylar, materyal ve uygulama prosedürü, işlem sonrası bakım, endikasyonlar, kontrendikasyonlar, tıbbi sorunlar varlığında yaklaşım gibi birçok farklı konuya değinilmektedir. Çalışmada üç farklı chatbot (Copilot, Gemini ve ChatGPT) kullanıldı. Değerlendirme 5 puanlık Likert Ölçeği kullanılarak yapıldı. İstatistiksel anlamlılık düzeyi 0,05 olarak belirlendi.

Bulgular: Alt boyutların kendi aralarındaki korelasyonu değerlendirildiğinde lisans eğitimiyle ilgili sorularla uzmanlık eğitimiyle ilgili sorular arasında çok güçlü pozitif istatistiksel olarak anlamlı bir korelasyon vardır ($p < 0,001$). Toplam 105 yanıttan Copilot 48 geçerli yanıt ve 57 geçersiz yanıt üretti. Gemini 54 geçerli ve 51 geçersiz yanıt üretirken, ChatGPT 58 geçerli ve 47 geçersiz yanıt üretti.

Sonuçlar: Bu çalışmanın, özellikle diş hekimliğinin farklı alanlarıyla ilgili tartışmalı sorular da dahil olmak üzere çeşitli sorulara verilen yanıtları değerlendirmek açısından daha ileri çalışmalar için fikir sağlayabileceğini düşünüyoruz.

Anahtar Kelimeler: *Yapay zeka, ChatGPT, Copilot, Gemini, Restoratif Diş Hekimliği.*

[a] *cansu.yikici@gmail.com* *0000-0001-8855-7417*
[b] *mervenezir@gazi.edu.tr* *0000-0001-8902-5471*
[c] *suatozcan@gazi.edu.tr* *0000-0001-8782-2899*

How to Cite: Yıkıcı Çöl C, Nezir M, Özcan S. (2025) Evaluation of the Performance of Different Chatbots' Responses to Restorative Dentistry-Related Questions, Cumhuriyet Dental Journal, 28(2): 237-245.

Introduction

These days, artificial intelligence (AI) chatbots have completely changed digital communication by sharing knowledge and giving people the chance to ask tailored questions and get answers that are both original and targeted. Chatbots have the potential to enhance their answers over time, owing to their ability to extract information from large amounts of data thanks to deep learning algorithms.[1] In late 2022, there was a groundbreaking advancement in artificial intelligence technology: OpenAI Inc. ChatGPT program offered by (San Francisco, CA, USA) has been introduced. This chatbot reached a staggering 100 million new users in the first 3 months. Over the past few years, the field of dentistry has also experienced remarkable growth in artificial intelligence applications and tools.[2] With the use of computational linguistics and vast amounts of data, ChatGPT is an artificial intelligence computer software that can mimic human speech in response to user requests, hence enhancing responsiveness and communication skills. Through text-based interfaces, methods like machine learning and deep learning are applied. ChatGPT is a huge language model that generates human-like replies to natural language questions using deep learning AI techniques. For patients, healthcare professionals, and educators alike, ChatGPT offers a wide range of services. For instance, it may help students with their homework and tutoring by providing explanations and answers to help them understand complicated concepts. Additionally, by serving as a teaching tool, it has the power to fundamentally alter how students study biomedical science. Healthcare professionals working in dentistry and other medical fields can benefit from ChatGPT's many services, which include improved decision-making support, digital data recording, diagnosis, disease prevention, image analysis, treatment, fewer treatment interruptions, and research and discovery opportunities.[3]

Apart from ChatGPT, artificial intelligence robots also include applications such as Google Gemini and Microsoft Copilot. The healthcare sector is paying close attention to AI models, and initial research has shown positive results. AI programs excel at emulating human speech patterns and generating answers to spoken text that seem natural. The application of ChatGPT and related models in healthcare services has piqued the interest of many individuals. AIs will probably be utilized, in particular, to assist physicians and patients in understanding how to interpret laboratory results and use clinical laboratory services.[4] However, the dangers of misuse, including the spread of misinformation, ethical concerns, and factual errors, should all be considered in writing for academia and healthcare practices. Human intelligence (HI) is more adaptable than artificial intelligence due to its natural evolutionary history, versatility, creativity, psychological intelligence, and capacity to understand complex abstract ideas.[5] In the light of all this information, this study aimed to evaluate and compare the performance of the responses given by three different chatbots (Microsoft Copilot, Google Gemini and ChatGPT 3.5) to questions that are considered controversial today for undergraduate students and specialists in the field of restorative dentistry. The first null hypothesis of this study was that the responses given by different chatbots to the questions are consistent with each other. The second null hypothesis is that each of the three responses to the same questions is consistent with each other.

Materials and Methods

No ethics committee approval is required for this study.

There was not an appeal for ethical approval because the research was done using data that was freely accessible online. Two practitioners in restorative dentistry produced a comprehensive set of questions about education for the field. Thirty questions total was created by two dentists (C.Y.C. and M.N.), who are full-time specialists in restorative dentistry working by both public and private institutions. Twenty-five questions were designed undergraduate education, five questions for specialist education, and five questions associated with controversial issues (Figure 1). Table 1 displays the list of questions that the restorative dentistry specialist chosen. In these questions, many different topics such as terminology, treatment procedures, technical details, material and application procedure, post-procedure care, indications, contraindications, approach in the presence of medical problems are touched upon. In this study, the responses given to questions by three different chatbots were evaluated. The chatbots evaluated are summarized in Table 2. A new e-mail account was created to avoid being affected by past activity. All search history and cookies on the computer have been cleared. Three different chatbots were used in the study. These are summarized in Table 2. Accounts were opened with new e-mail addresses in all three applications. A total of 35 questions were asked to each chatbot on the same day (22.03.2024). In order to evaluate its reliability, each question was repeated 3 times and a total of 270 responses were obtained. Before asking questions for each answer, a 'new topic' or 'new chat' was opened. All responses received were recorded in a separate file. Evaluation was done using a 5-point Likert Scale Scores[6,7] from 1 to 5 are given on the scale (Table 3).

The responses received were evaluated independently by two restorative dentistry specialists. Two raters then reviewed the responses. Differences were discussed based on evidence, and disagreements were resolved by consulting a third evaluator (S.Ö.). All responses were noted in a separate excel file. Descriptive statistics were used to describe continuous variables (mean, standard deviation, minimum, median, and maximum). Frequency and percentage data were acquired for descriptive statistics pertaining to categorical variables. Fleiss Kappa agreement overall test was used to compare the evaluations of chatbots on a categorical feature on a question basis and in subcategories with randomness and to measure the fit that is above or below randomness according to the category. To investigate the link between quantitative data that differ from the normal distribution, the Spearman's rho correlation test was used. When evaluating the validity of the responses, the threshold value was accepted as 4. Questions that received 4 or more points on a question basis and in subcategories were considered valid.[1] The reliability and internal consistency of each Chatbot were evaluated using Spearman-Brown Coefficient, Cronbach Alpha, and Guttman Split Half Coefficient coefficients. The statistical significance level was determined as 0.05.

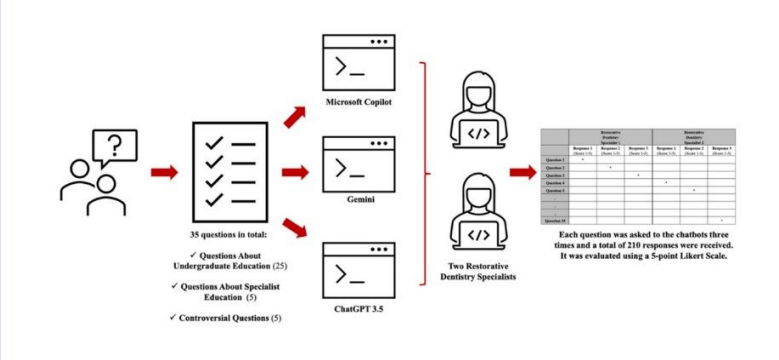

Figure 1. Study design

Table 1. List of questions asked to the chatbots

Category of Questions	Questions Asked to Chatbots
Questions About Undergraduate Education	U1. What type of composite resin material should be used in restorations of front and back teeth? U2. What should be taken into consideration when applying dental bonding agent? U3. Why is matrix band applied to the tooth when filling? U4. What are the types of dental matrix band? U5. Why is wedge application used when filling teeth? U6. What are the application stages of amalgam restoration? U7. What are the application stages of composite resin restoration? U8. What should be taken into consideration in composite resin polymerization? U9. How should the finishing and polishing of composite restorations be done? U10. What advice should be given to patients after amalgam restoration application? U11. What advice should be given to patients after composite resin restoration application? U12. What should be taken into consideration during dental treatment in a patient with diabetes? U13. What should be taken into consideration during dental treatment in a patient with heart problems? U14. What should be taken into consideration during dental treatment in a patient with rheumatic disease? U15. What should be taken into consideration during dental treatment in pregnant women? U16. In which patients should antibiotic prophylaxis be applied? U17. How is antibiotic prophylaxis performed? U18. What is vital bleaching and how is it applied? U19. What is devital bleaching and how is it applied? U20.What should be taken into consideration in patients who will undergo tooth whitening treatment? U21. What should be recommended to patients after teeth whitening treatment? U22. When should filling be applied after teeth whitening treatment? U23. What is indirect pulp capping treatment and how is it applied? U24. What is direct pulp capping treatment and how is it applied? U25. What is atraumatic restorative treatment and why is it used?
Questions About Specialist Education	S1. What is the gold standard dental bonding agent? S2. Which inorganic filler type shows the lowest wear on the tooth structure in dental composites? S3. Why and how to apply chlorhexidine after dental cavity preparation? S4. How to do deep margin elevation? S5. What are the advantages of the ethanol based etch amd rinse system?
Controversial Questions	C1. Can teeth whitening be applied to pregnant women? C2. Is whitening treatment harmful to health? C3. Should fluoride applications be recommended to patients? C4. Can we recommend fluoride toothpaste to pregnant patients? C5. Which patients should not have amalgam fillings?

Table 2. The attributes of the chatbots used in the study

Chatbots	Developer	Language model	Versions	Free/paid versions	Version used in the study
Microsoft Copilot	Microsoft	GPT-4 and DALLE	Copilot Pro, Copilot for Microsoft 365, Free Mocrosoft Copilot	Copilot, Copilot Pro	Copilot
Gemini	Google AI	LaMDA	Nano, Pro and Ultra	Gemini, Gemini Advanced	Gemini
ChatGPT	OpenAI	GPT-4, GPT-3 and DALLE	GPT-1 GPT-2 GPT-3 GPT-4	ChatGPT, ChatGPT plus	ChatGPT 3.5

Table 3. 5-point Likert Scale

Score	Meaning
Score 5 (Strongly Agree)	The answer is correct, and the content is comprehensive.
Score 4 (Agree)	The answer is correct and most of the content is correct, but it lacks information, or contains incorrect information.
Score 3 (Neutral)	The answer is somewhat correct, but details are primarily incorrect, missing, or irrelevant.
Score 2 (Disagree)	The answer is incorrect, but the content includes some correct elements.
Score 1 (Strongly Disagree)	The answer and the entire content are incorrect or irrelevant.

The Fleiss Kappa overall agreement test was used to compare the evaluations of more than two observers on a categorical feature at the same time with randomness and to measure the agreement that was above or below randomness according to the category. Spearman's rho correlation test was used to examine the relationship between quantitative variables that do not comply with normal distribution. The statistical significance level was determined as 0.05.

Results

When the consistency between 3 different ChatBots is evaluated on a question basis, there is a strong statistically significant consistency in 13 out of 35 questions (37%) in total $p < 0.001$. These questions are distributed as 8 out of 25 (32%) in questions about undergraduate education, 1 out of 5 (20%) in questions related to specialist Education, and 4 out of 5 (80%) in controversial questions. For the question 'How to do deep margin elevation (S4)', there is a moderate negative statistically significant agreement between the AI tools ($p < 0.05$). When compared on the basis of sub-dimensions, no statistically significant consistency was found between the chatbots. ($p > 0.05$). Table 4 summarizes the compatibility between ChatBots according to questions and sub-dimensions.

When the correlation of the sub-dimensions among themselves is evaluated, there is a very strong positive statistically significant correlation between the questions about undergraduate education and the questions about specialist education ($p < 0.001$).

When evaluating the validity of the responses, the threshold value was accepted as 4. Questions that received 4 or more points on a question basis and in subcategories were considered valid.[1]

When the validity of the responses was evaluated, a total of 35 questions were repeated 3 times, so 105 responses were received from each chatbot. Out of a total of 105 responses, Copilot produced 48 valid responses (45%) and 57 invalid responses (54%). While Gemini produced 54 valid (51%) and 51 invalid (48%) responses, ChatGPT produced 58 valid (55%) and 47 invalid (44%) responses.

When the valid/invalid responses of the sub-dimensions were evaluated in general, all three chatbots produced wrong responses in questions about undergraduate education and questions about specialist education. On the contrary, it was determined that all three chatbots produced correct responses on controversial issues.

When the internal consistency and reliability of the chatbots were analyzed, Copilot and Gemini showed good reliability and good internal consistency. ChatGPT demonstrated good internal consistency and good-moderate reliability. Cronbach's Alpha, Spearman-Brown and Guttman Split-Half Coefficients of chatbots are shown in Figure 2.

Table 4. Compatibility between chatbots by question and for sub-dimensions

	Question	Copilot_mean	Gemini_mean	ChatGPT_mean	Kappa κp
Questions About Undergraduate Education	What type of composite resin material should be used in restorations of front and back teeth?	2.33±0.58 2- (2-3)	2±0 2- (2-2)	3.33±1.15 4- (2-4)	-0.350 0.173
	What should be taken into consideration when applying dental bonding agent?	3±0 3- (3-3)	3±0 3- (3-3)	3.67±0.58 4- (3-4)	-0.286 0.391
	Why is matrix band applied to the tooth when filling?	5±0 5- (5-5)	5±0 5- (5-5)	5±0 5- (5-5)	1.000 **<0.001**
	What are the types of dental matrix band?	3±0 3- (3-3)	4±0 4- (4-4)	3±0 3- (3-3)	-0.500 0.134
	Why is wedge application used when filling teeth?	4±0 4- (4-4)	3±0 3- (3-3)	4.33±0.58 4- (4-5)	-0.370 0.158
	What are the application stages of amalgam restoration?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What are the application stages of composite resin restoration?	3±0 3- (3-3)	3±0 3- (3-3)	3.33±0.58 3- (3-4)	-0.125 0.708
	What should be taken into consideration in composite resin polymerization?	3±0 3- (3-3)	3±0 3- (3-3)	4±0 4- (4-4)	-0.500 0.134
	How should the finishing and polishing of composite restorations be done?	2±0 2- (2-2)	4±0 4- (4-4)	4±0 4- (4-4)	-0.500 0.134
	What advice should be given to patients after amalgam restoration application?	3±0 3- (3-3)	3±0 3- (3-3)	3.33±0.58 3- (3-4)	-0.125 0.708
	What advice should be given to patients after composite resin restoration application?	4±0 4- (4-4)	3±0 3- (3-3)	4±0 4- (4-4)	-0.500 0.134
	What should be taken into consideration during dental treatment in a patient with diabetes?	4±0 4- (4-4)	4±0 4- (4-4)	5±0 5- (5-5)	-0.500 0.134
	What should be taken into consideration during dental treatment in a patient with heart problems?	4±0 4- (4-4)	4±0 4- (4-4)	4±0 4- (4-4)	1.000 **<0.001**
	What should be taken into consideration during dental treatment in a patient with rheumatic disease?	3±0 3- (3-3)	3±0 3- (3-3)	4±0 4- (4-4)	-0.500 0.134
	What should be taken into consideration during dental treatment in pregnant women?	3.67±0.58 4- (3-4)	3.67±0.58 4- (3-4)	4±0 4- (4-4)	0.357 0.284
	In which patients should antibiotic prophylaxis be applied?	4±0 4- (4-4)	3±0 3- (3-3)	3±0 3- (3-3)	-0.500 0.134
	How is antibiotic prophylaxis performed?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What is vital bleaching and how is it applied?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What is devital bleaching and how is it applied?	3.33±0.58 3- (3-4)	3±0 3- (3-3)	3±0 3- (3-3)	-0.125 0.708
	What should be taken into consideration in patients who will undergo tooth whitening treatment?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What should be recommended to patients	4±0 4- (4-4)	4.33±0.58 4- (4-5)	4.33±0.58 4- (4-5)	-0.286 0.391

	after teeth whitening treatment?				
	When should filling be applied after teeth whitening treatment?	4±0 4- (4-4)	4±0 4- (4-4)	3.33±0.58 3- (3-4)	-0.286 0.391
	What is indirect pulp capping treatment and how is it applied?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What is direct pulp capping treatment and how is it applied?	3±0 3- (3-3)	3±0 3- (3-3)	3±0 3- (3-3)	1.000 **<0.001**
	What is atraumatic restorative treatment and why is it used?	3±0 3- (3-3)	3±0 3- (3-3)	3.67±0.58 4- (3-4)	-0.286 0.391
Questions About Specialist Education	What is the gold standard dental bonding agent?	1±0 1- (1-1)	1±0 1- (1-1)	1±0 1- (1-1)	1.000 **<0.001**
	Which inorganic filler type shows the lowest wear on the tooth structure in dental composites?	4±0 4- (4-4)	3.33±0.58 3- (3-4)	4±0 4- (4-4)	-0.286 0.391
	Why and how to apply chlorhexidine after dental cavity preparation?	3±0 3- (3-3)	4±0 4- (4-4)	4±1 4- (3-5)	-0.312 0.236
	How to do deep margin elevation?	3±0 3- (3-3)	4±0 4- (4-4)	1±0 1- (1-1)	-0.500 **0.034**
	What are the advantages of the ethanol based etch amd rinse system?	4±0 4- (4-4)	3±0 3- (3-3)	3.67±2.31 5- (1-5)	-0.397 0.052
Controversial Questions	Can teeth whitening be applied to pregnant women?	4±0 4- (4-4)	4.33±0.58 4- (4-5)	4.67±0.58 5- (4-5)	-0.500 0.134
	Is whitening treatment harmful to health?	4±0 4- (4-4)	4±0 4- (4-4)	4±0 4- (4-4)	1.000 **<0.001**
	Should fluoride applications be recommended to patients?	4±0 4- (4-4)	4±0 4- (4-4)	4±0 4- (4-4)	1.000 **<0.001**
	Can we recommend fluoride toothpaste to pregnant patients?	4±0 4- (4-4)	4±0 4- (4-4)	4±0 4- (4-4)	1.000 **<0.001**
	Which patients should not have amalgam fillings?	4±0 4- (4-4)	4±0 4- (4-4)	4±0 4- (4-4)	1.000 **<0.001**

Fleiss Kappa Agreement

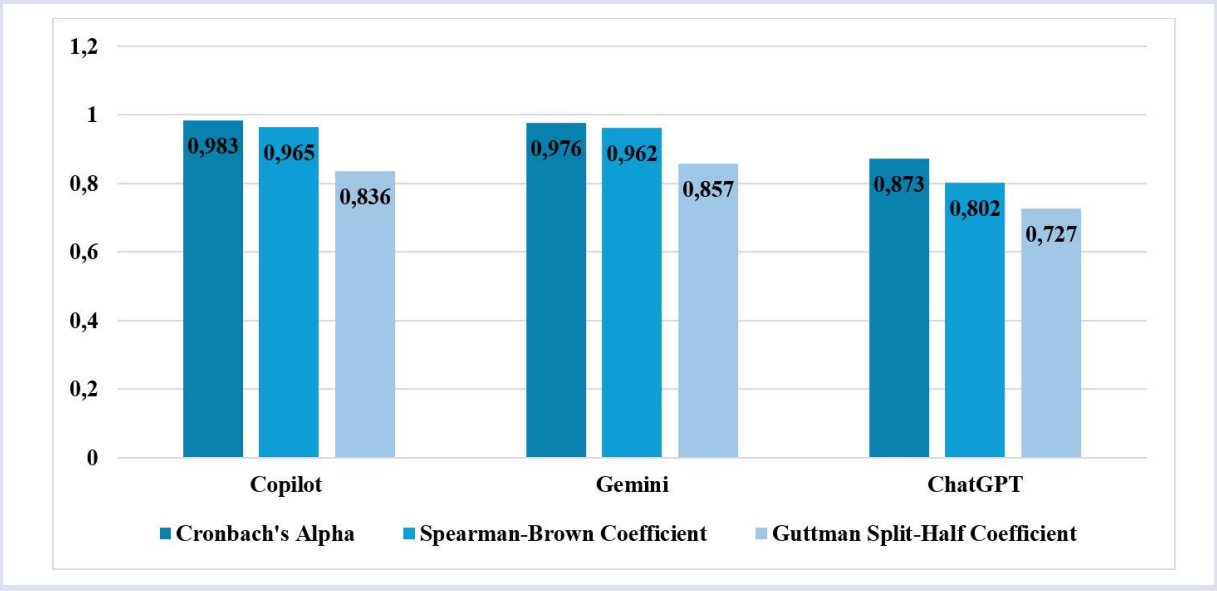

Figure 2. Cronbach's Alpha, Spearman-Brown and Guttman Split-Half Coefficients of chatbots

Discussion

Artificial intelligence is an interdisciplinary field that draws on logic, cybernetics, statistics, decision theory, cognitive psychology, linguistics, neuroscience, and computer engineering to create computer algorithms that simulate intelligent behavior with little human intervention. Information retrieval, speech and image recognition, robotics, sensor technologies, and cognitive decision support systems are just a few of the applications of artificial intelligence that employ machine learning as their primary technology. Artificial intelligence is already having an influence on the whole world and is quickly changing every aspect of contemporary life, including business, social media, healthcare, space technology, and many government responsibilities.[8] Additionally, artificial intelligence chatbots, which have recently started to attract a lot of attention, can also play an active role in ensuring that people obtain the detailed information they want to learn in various fields. AI systems are capable of perceiving and reacting to objects, events, and phenomena similarly to how humans do. technology is becoming more and more crucial to comprehend how AI is seen and used as technology continues to enter every part of daily life. AI is being used extensively in dentistry, especially in fields like clinical prediction, practice management, implantology, caries detection, teledentistry, and x-ray and diagnostics.[9]

AI-based chatbots have a great deal of promise to advance dental clinical practice. Through thorough analysis, their right usage may facilitate the formulation of precision and customized medical goals, as well as help with diagnostic and treatment planning.[7] This study evaluated and compared the performance of the responses given by three different chatbots to questions related to the field of restorative dentistry for undergraduate students and specialist students and which are considered controversial currently.

In this study, when the consistency between three different chat robots was evaluated on a question-by-question basis, it was observed that there was consistency in 13 out of 35 questions. Consequently, the first null hypothesis of the study was rejected. According to researchs, a number of factors, such as the underlying technology, the particular algorithms employed, the datasets used for training, and the objectives the chatbot is intended to accomplish, might be responsible for variations in the replies provided by chatbots. The accuracy, fluency, and responsiveness of the chatbot can be impacted by significant variations in the quantity, quality, and diversity of training data. The variety of chatbot responses might be caused by these variables.[7]

Chatbot technology can create a more personalized and interesting learning environment for students. The chatbot facilitates real-time interaction by comprehending and identifying user questions, constructing contextually appropriate sentences, and responding instantly.[10] Studies have focused on chatbot technology, especially its application in education.[11-13] ChatGPT is well-versed in a wide range of undergraduate subjects, including biology, engineering, and health sciences. It could provide answers to a variety of queries that kids might have both within and outside of the classroom. He is especially skilled at writing code and is proficient in a number of programming languages. In a successful example, the relevant code can be created in the desired programming language. It can help with teaching by drawing students' attention with entertaining techniques like question-answer games in all subject areas. However, ChatGPT may not be sufficient for visual outputs. Furthermore, it has been noted that ChatGPT does a good job at answering exam questions. It can precisely respond to both verbal and numeric questions without omitting any information or cautions. Given that it is a text-based model, its ability to answer test questions can be enhanced by addressing any limitations in graphical representation or interpretation. It is evident that ChatGPT is utilized in undergraduate courses, particularly for tests administered remotely.[14] Questions about undergraduate education and questions about specialist education were found to have a very high positive association in this study. Chatbots can be reliable in answering various questions regarding both undergraduate and specialty education in dentistry. However, it should not be forgotten that visual outputs are also of great importance, especially in dentistry education, where manual dexterity is of great importance. We believe that in the future, developing chatbots with visual outputs may increase their use in the field of dentistry. However, these responses must be verified by professional and expert dentists.

The ability of large language models (LLMs) like ChatGPT[15], Gemini[16], and Microsoft Copilot[17] to comprehend and generate text that is human-like has completely changed the field of artificial intelligence. These chatbot models learn a great deal of knowledge and linguistic subtleties through training on a variety of online datasets. LLMs are useful instruments for information distribution and decision assistance in a number of fields since they can handle a wide range of activities, from responding to inquiries to creating logical and contextually relevant answers.[18] In this study, it was observed that the chatbot that gave the most valid responses to the questions was ChatGPT. The fact that ChatGPT provides a conversational interactive platform for users and answers questions directly instead of sending them to different websites is one of its biggest advantages over conventional online search engines. ChatGPT allows users to analyze and query the veracity and sources of information. ChatGPT has the ability to create new content during real-time conversations. To answer user questions, ChatGPT employs a variety of AI models that have been trained on a vast quantity of textual data. Provides conversational responses to user queries. It is capable of remembering user input and its own response to the conversation thread, and builds its previous outputs with subsequent queries. With the use of a conversational interface, ChatGPT enables users to carry out a wide range of text-based operations at a never-

before-seen scale, including answering queries, producing codes, translating text, and producing original content. ChatGPT provides an intelligent learning platform that enhances student learning by allowing learning materials to be tailored to each student's needs. It seems like a good idea to use ChatGPT as a learning platform, and there are no issues with it. In each web search, there is always a chance of finding false information. This is also true of ChatGPT, which ensures that users double-check information when unsure.[8]

The number of studies evaluating the accuracy, consistency and reliability of chatbots, a new and popular source of information, is quite limited.[1,19] It appears that issues related to restorative dentistry have not yet been researched. Therefore, this study aimed to evaluate the accuracy, consistency, validity, and recommendability of the responses given to frequently asked questions in the field of restorative dentistry by chat robots, which have become very popular recently. When the internal consistency and reliability of the chatbots were analyzed, Copilot and Gemini were found to show good internal consistency and good reliability. ChatGPT, on the other hand, was found to show good internal consistency and good-moderate reliability. Therefore, the second null hypothesis of the study was rejected. However, all three chatbots gave incorrect responses to questions about undergraduate education and questions about specialty education. On the contrary, it was determined that all three chatbots produced correct responses on controversial issues. The accuracy of the responses given to controversial questions is very important. We think that this study may provide ideas for further studies in terms of evaluating the responses to various questions, including especially controversial questions regarding different areas of dentistry.

Limitations of this study include the constant updating and improvement of existing versions of the chatbots used. It is very important to keep up with artificial intelligence technology, which is renewed and developed every day. This study focused on questions related to the field of restorative dentistry, one of the fields of dentistry. We think that evaluating questions related to other fields of dentistry with further studies and possibly the most up-to-date chat robots will make valuable contributions to the literature.

Conclusions

Within the limitations of this current study, it was concluded that: There is a strong statistically significant consistency in 13 out of 35 total questions (37%) within each chatbot. There is a very strong positive statistically significant correlation between the questions about undergraduate education and the questions about specialist education. Out of a total of 105 responses, Copilot produced 48 valid responses (45%) and 57 invalid responses (54%). While Gemini produced 54 valid (51%) and 51 invalid (48%) responses, ChatGPT produced 58 valid (55%) and 47 invalid (44%) responses. However, these responses must be verified by professional and expert dentists. We think that this study may provide ideas for further studies in terms of evaluating the responses to various questions, including especially controversial questions regarding different areas of dentistry.

Acknowledgements

Not applicable.

Conflicts of Interest Statement

Not applicable.

Authorship Contributions

Conceptualization, C.Y.Ç., M.N., and S.Ö.; methodology, C.Y.Ç., M.N., and S.Ö.; validation, C.Y.Ç., M.N., and S.Ö.; formal analysis, C.Y.Ç., M.N., and S.Ö.; investigation, C.Y.Ç., M.N., and S.Ö.; writing—original draft preparation, C.Y.Ç., M.N., and S.Ö.; writing—review and editing, C.Y.Ç., M.N., and S.Ö.; supervision, C.Y.Ç., M.N., and S.Ö.; project administration, C.Y.Ç., M.N., and S.Ö. All authors have read and agreed to the published version of the manuscript.

References

1. Mohammad-Rahimi H, Ourang SA, Pourhoseingholi MA, Dianat O, Dummer PMH, Nosrat A. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. Int Endod J 2024;57:305-314.
2. Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. Eur J Orthod 2024;cjae017.
3. Alhaidry HM, Fatani B, Alrayes JO, Almana AM, Alfhaed NK. ChatGPT in dentistry: a comprehensive review. Cureus 2023;15:e38317.
4. Kaftan AN, Hussain MK, Naser FH. Response accuracy of ChatGPT 3.5 Copilot and Gemini in interpreting biochemical laboratory data a pilot study. Sci Rep 2024;14:8233.
5. Tiwari A, Kumar A, Jain S, Dhull KS, Sajjanar A, Puthenkandathil R, et al. Implications of ChatGPT in public health dentistry: A systematic review. Cureus 2023;15:e40367.
6. Lee Y, Shin T, Tessier L, Javidan A, Jung J, Hong D, et al. Harnessing artificial intelligence in bariatric surgery: comparative analysis of ChatGPT-4, Bing, and Bard in generating clinician-level bariatric surgery recommenda tions. Surg Obes Relat Dis 2024;20:603-608.
7. Mohammad-Rahimi H, Khoury ZH, Alamdari MI, Rokhshad R, Motie P, Parsa A, et al. Performance of AI chatbots on controversial topics in oral medicine, pathology, and radiology. Oral Surg Oral Med Oral Pathol Oral Radiol 2024;137:508-514.
8. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT-A double-edged sword for healthcare education? Implications for assessments of dental students. Eur J Dent Educ 2024;28:206-211.
9. Buldur B, Teke F, Kurt MA, Sagtas K. Perceptions of dentists towards artificial intelligence: validation of a new scale. Cumhuriyet Dent J 2024;27:109-117.
10. Fang Q, Reynaldi R, Araminta AS, Kamal I, Saini P, Afshari FS, et al. Artificial Intelligence (AI)-driven dental education: Exploring the

role of chatbots in a clinical learning environment. J Prosthet Dent 2024.
11. Su N-Y, Yu C-H. Survey of dental students' perception of chatbots as learning companions. J Dent Sci 2024;19:1222-1223.
12. Elnagar MH, Yadav S, Venugopalan SR, Lee MK, Oubaidin M, Rampa S, et al., editors. ChatGPT and dental education: Opportunities and challenges. Seminars in Orthodontics; 2024: Elsevier.
13. Aminoshariae A, Nosrat A, Nagendrababu V, Dianat O, Mohammad-Rahimi H, O'Keefe AW, et al. Artificial intelligence in endodontic education. J Endod 2024;50:562-578.
14. Oguz FE, Ekersular MN, Sunnetci KM, Alkan A. Can Chat GPT be Utilized in Scientific and Undergraduate Studies? Ann Biomed Eng 2024;52:1128-1130.
15. ChatGPT. https://chat.openai.com Accessed March 22, 2024
16. Gemini. https://gemini.google.com/app Accessed March 22, 2024
17. Copilot. https://copilot.microsoft.com Accessed March 22, 2024
18. Tepe M, Emekli E. Assessing the Responses of Large Language Models (ChatGPT-4, Gemini, and Microsoft Copilot) to Frequently Asked Questions in Breast Imaging: A Study on Readability and Accuracy. Cureus 2024;16:e59960.
19. Rokhshad R, Zhang P, Mohammad-Rahimi H, Pitchika V, Entezari N, Schwendicke F. Accuracy and consistency of chatbots versus clinicians for answering pediatric dentistry questions: A pilot study. J Dent 2024;144:104938.