Exploring AI-Driven Insights on Oral Implant Surgery: A Study of Four Different AI Applications

Oral İmplant Cerrahisinde Yapay Zeka Destekli Hasta Bilgilendirmesi: Dört Farklı Sistem Üzerine Karşılaştırmalı Bir İnceleme

Meltem Özden YÜCE[1] https://orcid.org/0000-0002-7088-9701
Emine ADALI[2] https://orcid.org/0000-0002-0623-5746
İrem YAMAN[1] https://orcid.org/0000-0003-4657-2525
Betül ALPAGUTER[1] https://orcid.org/0009-0005-8505-9101
Berk KARADENİZ[1] https://orcid.org/0009-0006-2715-8918

[1]Ege University Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, Izmir
[2]Private Clinic

Citation: Yüce M Ö, Adalı E, Yaman İ, Alpaguter B, Karadeniz B. Exploring AI-Driven Insights on Oral Implant Surgery: A Study of Four Different AI Applications. *Int Arc Dent Sci*. 2025;46(1):19-25.

ABSTRACT

INTRODUCTION: Artificial intelligence (AI) chatbots are increasingly influential in healthcare, including in dental procedures like implants. However, their accuracy and reliability of the information they provide have not been comprehensively evaluated.This study aimed to assess the responses of four AI chatbots—ChatGPT-4, Gemini, Claude, and Microsoft Copilot—by comparing them with those provided by oral surgeons in response to common patient queries about dental implants.

METHODS: This study aimed to assess the responses of four AI chatbots—ChatGPT-4, Gemini, Claude, and Microsoft Copilot—by comparing them with those provided by oral surgeons in response to common patient queries about dental implants. Fifteen frequently asked questions were posed to the chatbots, and five oral surgeons scored their responses using the Global Quality Scale (GQS).

RESULTS: Statistical analysis revealed that ChatGPT received a significantly higher median rating than both Gemini and Copilot. Notably, Copilot exhibited negative Cronbach's α values, suggesting a lack of response consistency and raising concerns about reliability.

CONCLUSION: While all four AI chatbots provided responses that were at least satisfactory, the risk of patient misunderstanding remains. Patients are advised to validate AI-provided information obtained from these platforms with healthcare professionals and trusted sources, highlighting the importance of professional guidance in patient education

Keywords: *Artificial intelligence, patient information, implant surgery, chatbot*

ÖZ

GİRİŞ ve AMAÇ: Yapay zeka (AI) sohbet robotları, implantlar gibi dental prosedürler de dahil olmak üzere sağlık hizmetlerinde giderek daha etkili hale gelmektedir. Ancak, sağladıkları bilgilerin doğruluğu ve güvenilirliği kapsamlı bir şekilde değerlendirilmemiştir. Bu çalışmanın amacı, dört yapay zeka sohbet robotunun yanıtlarını, dental implantlarla ilgili yaygın hasta sorularına yanıt olarak ağız cerrahları tarafından verilen yanıtlarla karşılaştırarak değerlendirmektir.

YÖNTEM ve GEREÇLER: Sık sorulan on beş hasta sorusu oluşturulmuş ve dört YZ sohbet robotuna (Chat GPT-4, Gemini, Claude ve Microsoft Copilot) sunulmuş ve yanıtlar Orijinal Global Kalite Ölçeği (GQS) kullanılarak beş ağız cerrahı tarafından puanlanmıştır.

BULGULAR: İstatistiksel analiz ChatGPT'nin hem Gemini hem de Copilot'tan önemli ölçüde daha yüksek bir medyan derecelendirme aldığını ortaya koymuştur. Özellikle Copilot'un negatif Cronbach α değerleri sergilemesi, yanıt tutarlılığının eksik olduğunu göstermekte ve güvenilirlikle ilgili endişeleri artırmaktadır.

SONUÇ: Dört YZ sohbet robotu da en azından tatmin edici yanıtlar vermiş olsa da, hastaların yanlış anlama riski devam etmektedir. Hastalara, bu platformlardan elde edilen YZ tarafından sağlanan bilgileri sağlık uzmanları ve güvenilir kaynaklarla doğrulamaları tavsiye edilmekte ve hasta eğitiminde profesyonel rehberliğin önemi vurgulanmaktadır

Anahtar Kelimeler: *Yapay zeka, hasta bilgilendirme, implant cerrahisi, chatbot*

Corresponding author: dtiremyaman@gmail.com
Received Date: 01.11.2024
Accepted Date: 22.11.2024

INTRODUCTION

Artificial Intelligence (AI), encompassing diverse technologies that emulate human cognition, has brought about transformative changes across multiple sectors, including health care.[1] Among AI applications; chatbots—utilizing natural language processing (NLP) to simulate human conversation—are increasingly prevalent. Enabled by NLP, these chatbots interpret and respond to user queries in a conversational manner.[2]

With rapid advancements in AI, AI-powered chatbots have become increasingly common. These chatbots engage with users and enhance their capabilities through a variety of AI techniques.[3] They are widely used across numerous sectors including but not limited to finance, customer service, and education and are now making significant inroads in the healthcare industry as well[4,5]

In dentistry, however, the application of chatbots remains largely under-researched. AI-powered chatbots have substantial potential to provide patients with valuable information. Patients frequently turn to the internet for insights into their health concerns and possible treatment options, especially when faced with barriers in reaching a healthcare provider or seeking second opinions.[6] NLP platforms offer numerous advantages, including 24/7 availability, enabling patients to access information they need at any time. Nevertheless, chatbot responses can be inconsistent and may even mislead patients[7]

While NLP platforms provide advantages like 24/7 accessibility for patients, they also pose significant limitations. Chatbots may lack the empathy and understanding, which is crucial in health care,[8] and their responses might lead to miscommunication or misinformation, which could affect patients' health decisions. Often, these AI systems cannot fully grasp the nuances and emotional context of human language, making them a less reliable source of information compared to direct communication with healthcare professionals.[9]

This study aimed to assess the accuracy, quality and reliability of responses generated by four AI chatbots—ChatGPT, Gemini, Claude, and Microsoft Copilot—compared with responses from oral surgeons regarding common dental implant-related queries. 15 questions frequently asked by patients at our clinic were developed and posed to each chatbot as queries. Five oral surgeons rated these responses based on the Original Global Quality Scale (GQS).[10]

MATERIAL and METHODS

In this study, 15 frequently asked questions about dental implant procedures—covering topics such as risks, recovery times, and implant types—were developed to capture a comprehensive range of patient concerns.

In order to conduct this study, new accounts for the chatbots ChatGPT-4 (https://chat.openai.com), Google Gemini (https://gemini.google.com), Claude (Anthropic) (https://claude.ai) and Microsoft Copilot (https://copilot.microsoft.com) were created. Each question was posed to each chatbot three times on the same day each session began with a new chat to reduce potential biases and each question was asked three times in a row. Five oral surgeons who were blinded to each other's responses, evaluated the chatbot replies. Each question was carefully crafted to prevent any grammatical or syntactical mistakes.

Responses were rated on a five-point Likert-type GQS, providing scores based on specific quality criteria:[10]

Table 1. Global Quality Scale (GQS) Classification

1. Poor quality; poor flow of the video; most information missing; not at all useful.
2. Generally poor quality and poor flow; some information listed, but many important topics missing; of very limited use
3. Moderate quality; suboptimal flow; some important information adequately discussed, but other information poorly discussed; somewhat useful
4. Good quality and generally good flow; most of the relevant information listed, but some topics not covered; useful
5. Excellent quality and flow; very useful.

Table 2. The queries that were asked to ChatGPT-4, Google Gemini, Claude and Microsoft Copilot

Queries
1. What is the lifetime of a dental implant?
2. What is the duration of dental implant surgery?
3. Can a tooth be fitted immediately after the dental implant has been placed?
4. Will dental implants look like my natural teeth?
5. How many days of rest do I need after dental implant surgery?
6. Is it difficult to clean dental implants?
7. Is dental implant treatment painful?
8. How long does it take for a dental implant to heal?
9. What will happen if there is not enough bone in the jaw for dental implant treatment?
10. Is dental implant treatment expensive?
11. Is there a possibility that the body might reject the dental implant?
12. When can I start eating and drinking normally after dental implant operation?
13. Can a dental implant fall out?
14. Can my jawbone be damaged during a dental implant procedure?
15. Could a dental implant cause allergy?

Since the study was based on publicly available information, approval from the Institutional Review Board was not necessary.

Statistical analysis

The main analysis metric was the experts' ratings for each question across different AI systems. Cronbach's α score was used to assess reliability, while descriptive statistics (median, mean, and standard deviation) summarized the data. The Shapiro–Wilk test checked data normality, and the Kruskal–Wallis test compared ratings among AIs, with Bonferroni correction for multiple comparisons. The Friedman test was used to assess response consistency within each AI system over time. All statistical analyses were performed using SPSS 15.0 software for Windows (SPSS, Inc., Chicago, Illinois).

RESULTS

Three responses per question were collected, and we analyzed potential variations within each AI's responses per question (Table 3). The Friedman test results revealed significant differences for Claude in ratings of Q4, Q8, Q9, Q10, and Q12. For Q4, there was a statistically significant difference between the median rating at T0 (5) and T1 and T2 (3; $p < 0.05$). For Q8 and Q12, the median rating changed from 4 at T0 and T1 to 3 at T2, indicating significant differences ($p < 0.05$). For Q9, the median rating was 5 at T0 and 4 at T1 and T2, and this difference was also statistically significant (p=0.05). Finally, for Q10, the median rating was 4 at T0 and 3 at both T1 and T2, a difference that is borderline significant ($p = 0.050$). Overall, these results suggest that while ChatGPT, Gemini, and Copilot were consistent in their responses and received similar ratings repeatedly per question, Claude's responses somehow changed for 5 questions out of 10, and these changes were reflected in its quality ratings.

Table 1. Descriptive Statistics for Ratings (per Question and AI) and comparison of mean ratings among AIs per question.

	Chat GPT		Gemini		Claude		Copilot		
	Median	Mean± Std dev.	Median	Mean± Std dev.	Median	Mean± Std dev.	Median	Mean± Std dev.	p
Q1	4.00	3.87±0.69	3.00	3.40±0.55	3.00	3.40±0.72	3.67	3.40±0.72	0.546
Q2	4.00	3.87±0.51	4.00	3.60±0.55	3.00	3.53±1.10	3.33	3.27±0.49	0.566
Q3	4.33	4.40±0.37	4.00	4.33±0.47	3.33	3.67±0.67	3.00[b]	3.13±0.87	0.031*
Q4	4.00	4.20±0.45	4.00	3.80±0.45	3.33	3.80±0.84	3.00	3.27±0.43	0.115
Q5	3.00	3.20±0.45	3.00	3.27±0.43	3.67	3.60±0.43	3.00	3.00±0.00	0.086
Q6	4.00	4.07±0.15	3.67	3.53±0.51	3.67	3.87±0.69	4.00	4.00±0.00	0.124
Q7	4.00	3.93±0.15	3.00	3.47±0.65	4.00	4.20±0.65	3.00	3.40±0.55	0.154
Q8	4.00	4.27±0.43	3.00	3.40±0.55	3.67	3.53±0.38	3.00	3.20±0.45	0.025*
Q9	4.33	4.47±0.38	3.67	3.53±0.51	4.33	4.40±0.28	3.67	3.67±0.33	0.005*
Q10	4.00	4.00±0.24	4.00	3.67±0.62	3.33	3.27±0.49	3.67	3.80±0.18	0.136
Q11	4.00	3.73±0.55	4.00	4.00±1.00	3.67	3.87±0.69	3.33	3.47±0.51	0.755
Q12	4.00	3.80±0.45	4.00	3.47±0.96	3.67	3.73±0.49	4.00	3.93±0.15	0.922
Q13	4.00	3.80±0.84	4.00	4.33±0.62	3.67	3.80±0.18	3.67	3.80±0.87	0.528
Q14	3.00	3.40±0.55	4.00	3.80±1.30	3.33	3.53±0.61	3.00	3.20±0.84	0.748
Q15	4.00	3.87±0.51	4.00	3.60±0.55	4.00	3.87±0.77	4.00	4.07±0.28	0.529

*$p<0.05$

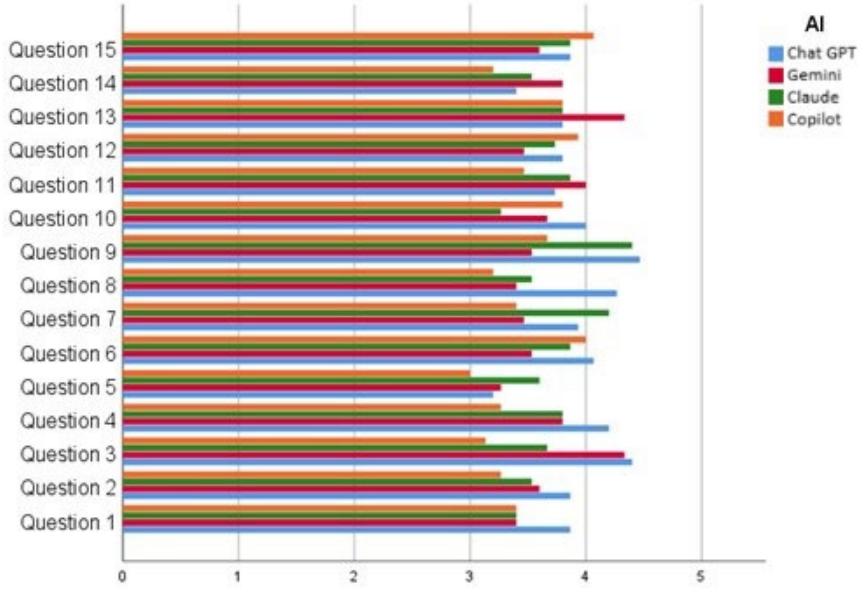

Figure 1. Mean scores for AI responses to 15 frequently asked questions, each asked three times

In addition to the question-by-question analysis (Table 1), we calculated overall scores for each AI system and conducted a comparative analysis using the Kruskal–Wallis test (Table 2). Results indicated statistically significant differences among the AI systems ($p < 0.001$), with ChatGPT' median rating notably higher than those of Gemini and Copilot. However, no significant differences were found between Claude's ratings and those of the other AIs.

Table 2. Overall evaluation of AI performances based on expert ratings

	Median	Mean± Std dev.	p
Chat GPT[a]	4.00	3.92±0.54	0.000*
Gemini[b]	4.00	3.68±0.70	
Claude[a,b]	3.67	3.74±0.64	
Copilot[b]	3.67	3.51±0.58	

*$p<0.05$

Table 3. Cronbach's Alpha Values for Rating Consistency of Different AIs for Each Question

Questions	ChatGPT	Gemini	Claude	Copilot
1	0.837	1.000	0.830	0.894
2	0.913	1.000	0.944	0.682
3	0.500	0.900	0.488	0.971
4	1.000	1.000	0.833	0.882
5	1.000	0.882	0.441	-
6	-	0.913	0.558	-
7	-	0.947	0.711	1.000
8	0.882	1.000	0.462	1.000
9	0.692	0.913	0.429	0.600
10	0.000	0.943	0.682	-1.000
11	0.889	1.000	0.558	0.913
12	1.000	0.976	0.341	-
13	1.000	0.943	-	0.949
14	1.000	1.000	0.591	1.000
15	0.913	1.000	0.849	-

In order to investigate the differences among AIs Kruskal Wallis tests were run per question, and significant differences were found among the AI systems (ChatGPT, Gemini, Claude, and Copilot). More specifically, median ratings of Q3 (4.33) and Q8 (4.00) for ChatGPT, were significantly greater than the median ratings of Q3 (3) and Q8 (3) for Copilot ($p<0.05$). This result indicates ChatGPT performed better compared to Copilot on these questions. On the other hand, ChatGPT and Claude both received median ratings of 4.33, but Gemini and Copilot had lower median ratings of 3.67 each ($p<0.05$), suggesting that ChatGPT and Claude responses for Q9 were received higher ratings than those of Gemini and Copilot (see Table 1 for detailed descriptives and test results).

Table 3 represents Cronbach's Alpha values for the expert ratings of different AI (ChatGPT, Gemini, Claude, and Copilot) responses for 15 questions. The Cronbach's Alpha values provide insights into the internal consistency of responses generated by each AI system for each question. The Cronbach's Alpha values for different AI methods range between 0.341 and 1.000, indicating generally consistent responses. Specifically, values between 0.70 and 1 suggest high reliability, values between 0.30 and 0.70 suggest moderate reliability, and values between 0 and 0.30 indicate low reliability. An inspection of Table 3 shows that $\alpha = 1.000$ under Copilot, this negative value for Cronbach's Alpha suggests that there was a potential issue with response reliability.

Table 4. Comparison of Repeated Ratings Within Each AI Per Question.

		T0		T1		T2		
		Median	Mean± Std dev.	Median	Mean± Std dev.	Median	Mean± Std dev.	p
Question 1	Chat GPT	4	4±1.22	4	3.8±0.45	4	3.8±0.45	0.717
	Gemini	3	3.4±0.55	3	3.4±0.55	3	3.4±0.55	-
	Claude	4	3.6±0.55	3	3.6±0.89	3	3±1	0.150
	Copilot	3	3.4±0.55	4	3.4±0.89	4	3.4±0.89	1.000
Question 2	Chat GPT	4	4±0.71	4	3.8±0.45	4	3.8±0.45	0.368
	Gemini	4	3.6±0.55	4	3.6±0.55	4	3.6±0.55	-
	Claude	4	4±1	3	3.4±1.14	3	3.2±1.3	0.061
	Copilot	3	3.2±0.45	3	3.2±0.45	4	3.4±0.89	0.717
Question 3	Chat GPT	5	4.6±0.55	4	4.2±0.45	4	4.4±0.55	0.368
	Gemini	4	4.2±0.45	4	4.4±0.55	4	4.4±0.55	0.368
	Claude	4	4.2±0.45	3	3.6±0.89	3	3.2±1.3	0.319
	Copilot	3	3.2±0.84	3	3.2±0.84	3	3±1	0.368
Question 4	Chat GPT	4	4.2±0.45	4	4.2±0.45	4	4.2±0.45	-
	Gemini	4	3.8±0.45	4	3.8±0.45	4	3.8±0.45	-
	Claude	5[a]	4.6±0.55	3[b]	3.6±0.89	3[b]	3.2±1.3	0.024*
	Copilot	3	3.2±0.45	3	3.2±0.45	3	3.4±0.55	0.368
Question 5	Chat GPT	3	3.2±0.45	3	3.2±0.45	3	3.2±0.45	-
	Gemini	3	3.4±0.55	3	3.2±0.45	3	3.2±0.45	0.368
	Claude	4	4±0	4	3.8±0.45	3	3±1	0.061
	Copilot	3	3±0	3	3±0	3	3±0	
Question 6	Chat GPT	4	4±0	4	4±0	4	4.2±0.45	0.368
	Gemini	3	3.4±0.55	4	3.6±0.55	4	3.6±0.55	0.368
	Claude	5	4.8±0.45	3	3.6±0.89	3	3.2±1.3	0.060
	Copilot	4	4±0	4	4±0	4	4±0	-
Question 7	Chat GPT	4	3.8±0.45	4	4±0	4	4±0	0.368
	Gemini	3	3.4±0.55	3	3.6±0.89	3	3.4±0.55	0.368
	Claude	4	4.2±0.45	4	4.4±0.55	4	4±1.22	0.607
	Copilot	3	3.4±0.55	3	3.4±0.55	3	3.4±0.55	-
Question 8	Chat GPT	4	4.4±0.55	4	4.2±0.45	4	4.2±0.45	0.368
	Gemini	3	3.4±0.55	3	3.4±0.55	3	3.4±0.55	-
	Claude	4[a]	3.8±0.45	4[a]	3.8±0.45	3[b]	3±0.71	0.050*
	Copilot	3	3.2±0.45	3	3.2±0.45	3	3.2±0.45	-
Question 9	Chat GPT	5	4.8±0.45	4	4.4±0.55	4	4.2±0.45	0.097
	Gemini	4	3.6±0.55	4	3.6±0.55	3	3.4±0.55	0.368
	Claude	5[a]	4.8±0.45	4[b]	4.4±0.55	4[b]	4±0	0.050*
	Copilot	4	4±0	4	3.6±0.55	3	3.4±0.55	0.097
Question 10	Chat GPT	4	4±0	4	4±0.71	4	4±0	1.000
	Gemini	4	3.6±0.55	4	3.6±0.55	4	3.8±0.84	0.368
	Claude	4[a]	3.8±0.45	3[b]	3±0.71	3[b]	3±0.71	0.050*
	Copilot	4	3.6±0.55	4	4±0	4	3.8±0.45	0.368
Question 11	Chat GPT	4	4±0.71	4	3.6±0.55	4	3.6±0.55	0.135
	Gemini	4	4±1	4	4±1	4	4±1	-
	Claude	5	4.8±0.45	3	3.6±0.89	3	3.2±1.3	0.060
	Copilot	4	3.6±0.55	3	3.4±0.55	3	3.4±0.55	0.368
Question 12	Chat GPT	4	3.8±0.45	4	3.8±0.45	4	3.8±0.45	-
	Gemini	4	3.6±1.14	4	3.4±0.89	4	3.4±0.89	0.368
	Claude	4[a]	4.2±0.45	4[a]	4±0	3[b]	3±1.22	0.039*
	Copilot	4	4±0	4	4±0	4	3.8±0.45	0.368
Question 13	Chat GPT	4	3.8±0.84	4	3.8±0.84	4	3.8±0.84	-
	Gemini	4	4.4±0.55	4	4.4±0.55	4	4.2±0.84	0.368
	Claude	4	4±0	4	4±0	3	3.4±0.55	0.050
	Copilot	4	4±1	4	3.8±0.84	3	3.6±0.89	0.223
Question 14	Chat GPT	3	3.4±0.55	3	3.4±0.55	3	3.4±0.55	-
	Gemini	4	3.8±1.3	4	3.8±1.3	4	3.8±1.3	-
	Claude	4	4±0.71	3	3.4±0.55	3	3.2±1.1	0.257
	Copilot	3	3.2±0.84	3	3.2±0.84	3	3.2±0.84	-
Question 15	Chat GPT	4	4±0.71	4	3.8±0.45	4	3.8±0.45	0.368
	Gemini	4	3.6±0.55	4	3.6±0.55	4	3.6±0.55	-
	Claude	4	4.4±0.55	4	3.8±0.84	3	3.4±1.14	0.060
	Copilot	4	4±0	4	4±0	4	4.2±0.84	0.717

$p<0.05$

The Friedman test results show significant time-based differences in responses for Q4, Q8, Q9, Q10, and Q12 for only the Claude system. For Q4, Claude's median response decreased from 5 at T0 to 3 at T1 and T2, showing a significant reduction over time ($p < 0.05$). For Q8 and Q12, the median decreased from 4 at T0 and T1 to 3 at T2, indicating significant changes at T2 ($p < 0.05$). For Q9, the median changed from 5 at T0 to 4 at T1 and T2, with the reduction over time being statistically significant (p=0.05). For Q10, the median response dropped from 4 at T0 to 3 at both T1 and T2, with a borderline significance ($p = 0.050$). These results show a reduction in Claude's response scores over time for these specific questions.

DISCUSSION

With the widespread accessibility of online sources, individuals frequently seek answers to their health-related questions on major online platforms like YouTube, Google, and AI-driven chatbots, where access is simple and mostly free of charge. YouTube provides a vast array of professional discussions and visual content on health issues, enhancing user comprehension. However, as an ever-evolving media site, YouTube sees new content uploaded at a speed of roughly one video every minute. As a result, the outcomes of a search may differ based on when a query is made.[11]

Google offers a broad array of articles, research papers, and credible sources for exploring health concerns.[12] While it refrains from adding any commentary to the search results, which may reduce bias, it's important to recognize that Google as a search engine does not authenticate the information, which includes both academic sources but also advertisements. As such, individuals with limited knowledge may be at risk of encountering misinformation.[13]

In recent years, AI chatbots have also emerged as primary information sources for patients seeking accurate health data. However, as with many other fields, the quality of AI-generated data requires further examination, particularly regarding patient health.[14] This study evaluated responses from various AI chatbots, including ChatGPT-4, Google Gemini, Claude, and Microsoft Copilot, to patient-centered questions about dental implant procedures.

Our findings suggest that ChatGPT exhibits a relatively higher degree of internal consistency in its responses when the same questions are repeated. In contrast, Copilot's negative Cronbach's α value indicates low reliability, potentially due to a summarization approach across responses. However, this may result in patients receiving incomplete information.

As evidenced in Table 3, the greatest discrepancy was observed in the responses provided by the AI systems for questions 3, 8, and 9. Although some deviation was observed across other questions, these differences were not statistically significant. This demonstrates that the information accessible to patients online varies considerably depending on the specific circumstances and the quality of the questions. For example, Question 9, "What will happen if there is not enough bone in the jaw for dental implant treatment?" elicited responses ranging from general statements about bone supplementation to specific procedural details like "sinus lift, short implantation, all-on-four concepts, etc." This variability likely contributed to the observed statistical significance.

These results underscore the importance of healthcare professionals critically evaluating AI-generated content before sharing it with patients, as inconsistencies could lead to misunderstandings or misinformed treatment decisions.[15] Although AI has potential as a supplementary tool for patient education, it is essential to recognize its limitations. Ongoing updates and training for these AI systems are essential to align them with current medical guidelines and research, thereby enhancing their clinical utility.[16] Collaboration between healthcare professionals and AI developers could further enhance patient care by ensuring that AI complements, rather than replaces, human expertise.

This study concluded that ChatGPT, Google Gemini, Claude, and Microsoft Copilot generally provide satisfactory responses to patient inquiries and may be appropriate for patient use. However, it remains crucial for patients to interpret the source of the information accurately. Even with accurate data, misunderstandings are possible. Thus, patients should verify AI-provided information with healthcare professionals and the relevant channels.

In conclusion, the potential for chatbots to be trained and become more useful in the future is significant, with ongoing developments in various domains. Lifelong learning dialogue systems allow chatbots to learn from user interactions and external sources, enhancing their language understanding and conversational skills over time. (Kaynak) In healthcare, structured training methods for AI chatbots could be developed to ensure accuracy and safety, particularly in sensitive contexts like postpartum care.[17]

Limitations

This study has certain limitations, including the number of questions evaluated and potential changes in AI algorithms. While power analysis determined the sample size of expert surgeons, increasing the sample size could strengthen the study's conclusions.

REFERENCES

1. Rathore FA, Rathore MA**.** The emerging role of artificial intelligence in healthcare. *J Pak Med Assoc.* 2023;73(9):1368-1369. doi: 10.47391/JPMA.23-48.
2. Matic R, Kabiljo M, Zivkovic M, Cabarkapa M. Extensible Chatbot Architecture Using Metamodels of Natural Language Understanding. *Electronics*. 2021; 10(18):2300. https://doi.org/10.3390/electronics10182300
3. Perez-Pino A, Yadav S, Upadhyay M, Cardarelli L, Tadinada A. The accuracy of artificial intelligence-based virtual assistants in responding to routinely asked questions about orthodontics. *Angle Orthod*. 2023;93(4):427-432. doi:10.2319/100922-691.1
4. Sidlauskiene J, Joye Y, Auruskeviciene V. AI-based chatbots in conversational commerce and their effects on product and price perceptions. *Electron Mark*. 2023;33(1):24. doi:10.1007/s12525-023-00633-8
5. Acar AH. Can natural language processing serve as a consultant in oral surgery?. *J Stomatol Oral Maxillofac Surg*. 2024;125(3):101724. doi:10.1016/j.jormas.2023.101724
6. Fatani B. ChatGPT for Future Medical and Dental Research. *Cureus*. 2023;15(4):e37285. Published 2023 Apr 8. doi:10.7759/cureus.37285
7. Shali A, Prashanth PH, G DR, Shriram S, Assel M, Naskath J. Bots using natural language processing in medical sector. In: *2022 1st International Conference on Computational Science and Technology (ICCST)*. IEEE; 2022:250-254. doi: 10.1109/ICCST55948.2022.10040432
8. Mendapara H, Digole S, Thakur M, Dange A. AI based healthcare chatbot system by using natural language processing Int J Sci Res Eng Dev. 2021;4(2):89-96.
9. Akinwande M, Adeliyi O, Yussuph T. Decoding AI and Human Authorship: Nuances Revealed through NLP and Statistical Analysis. Int J Cyber Info 2024;13(4):85-103. doi:10.5121/ijci.2024.130408
10. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol*. 2007;102(9):2070-2077. doi:10.1111/j.1572-0241.2007.01325.x
11. Yüce MÖ, Adalı E, Kanmaz B. An analysis of YouTube videos as educational resources for dental practitioners to prevent the spread of COVID-19. *Ir J Med Sci*. 2021;190(1):19-26. doi:10.1007/s11845-020-02312-5
12. Van Riel N, Auwerx K, Debbaut P, Van Hees S, Schoenmakers B. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open*. 2017;1(2):bjgpopen17X100833. Published 2017 May 17. doi:10.3399/bjgpopen17X100833
13. Metaxa, D, Torres-Echeverry, N. Google's Role in Spreading Fake News and Misinformation. Social Science Research Network. 2017.
14. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery?. *J Stomatol Oral Maxillofac Surg*. 2023;124(5):101471. doi:10.1016/j.jormas.2023.101471
15. Monteith S, Glenn T, Geddes JR, Whybrow PC, Achtyes E, Bauer M. Artificial intelligence and increasing misinformation. *Br J Psychiatry*. 2024;224(2):33-35. doi:10.1192/bjp.2023.136
16. Soleas EK, Dittmer D, Waddington A, van Wylick R. Demystifying artificial intelligence for health care professionals: continuing professional development as an agent of transformation leading to artificial intelligence-augmented practice. *J Contin Educ Health Prof.* Published online August 13, 2024. doi: 10.1186/s13643-024-02646-6
17. Lin J, Joseph T, Parga-Belinkie JJ, et al. Development of a practical training method for a healthcare artificial intelligence (AI) chatbot. *BMJ Innov.* 2021;7(4):441-444. doi: 10.1136/bmjinnov-2020-000530