Araştırma Makalesi – Original Article

Yapay zeka sohbet robotlarının ‘diş hekimliğinde flor’ konusu ile ilgili sorulara verdikleri yanıtların değerlendirilmesi*

Evaluation of Responses Given by Artificial Intelligence Chatbots to Questions on the ‘Fluoride in Dentistry’

Seçkin Aksu[i], Fatma Ayça Bakır[ii]

[i]Dr. Öğr. Üyesi, Mersin Üniversitesi, Diş Hekimliği Fakültesi, Çocuk Diş Hekimliği A.D., https://orcid.org/0000-0002-5196-215X

[ii]Araş. Gör. Dt., Mersin Üniversitesi, Diş Hekimliği Fakültesi, Çocuk Diş Hekimliği A.D., https://orcid.org/0009-0008-2382-7373

ÖZ

Amaç: Bu çalışmanın amacı, ‘diş hekimliğindeki flor’ konusuyla ilgili sıkça sorulan sorulara farklı yapay zeka sohbet robotları tarafından sağlanan bilgilerin doğruluk, eksiksizlik, güvenilirlik-kalite ve okunabilirlik yönünden değerlendirilmesidir.

Yöntem: Diş hekimliğinde flor uygulamaları hakkında sık sorulan açık uçlu 15 soru ChatGPT 4.0, ChatGPT 4.10, Copilot, Gemini ve Perplexity platformlarına yöneltildi. Tüm sorular yalnızca bir kez, aynı bilgisayar IP adresinden ve aynı sabit fiber internet ağı kullanılarak soruldu. Doğruluk ve eksiksizlik için Likert Ölçekleri, kalite ve güvenilirlik için mDISCERN ölçeği ve okunabilirlik için Ateşman Okunabilirlik İndeksi değerlendirme ölçütü olarak kullanıldı. İstatistiksel analizde Fisher’s Exact Testi, Kruskal Wallis Testi, Dunn Testi ve Mann Whitney U Testi kullanıldı.

Bulgular: Likert Doğruluk ve Eksiksizlik Ölçekleri ortanca değerlerine göre; Gemini> Perplexity> ChatGPT 4.10> ChatGPT 4.0> Copilot şeklindedir (p=0.016), (p=0.002). mDISCERN indeksine göre sıralama; Perplexity= Gemini> ChatGPT 4.10> ChatGPT 4.0> Copilot şeklindedir (p=0.008). Tüm yapay zeka sohbet motorlarının verdiği cevaplar ortalama orta zorlukta okunabilirlikte iken; Gemini ile Perplexity arasında istatistiksel anlamlı farklılık vardır (p=0.010).

Sonuç: ‘Diş hekimliğinde flor’ konusuyla ilgili cevapları tüm yapay zeka sohbet robotları için çoğunlukla doğru ve geçerli bilgiler olmakla birlikte eğitilmeleri için kullanılan veri kümelerinin kaliteli ve en güncel bilimsel yönergeleri içeren kaynaklardan yararlanması ve daha kolay okunabilirliğin sağlanması gerekmektedir. Konuyla ilgili daha fazla araştırma ve geliştirmeye ihtiyaç vardır.

Anahtar Kelimeler: Yapay zeka sohbet robotu, diş hekimliği, flor, kalite ve güvenilirlik, okunabilirlik

ABSTRACT

Aim: This study aimed to evaluate the accuracy, completeness, reliability-quality and readability of the information provided by different artificial intelligence chatbots to frequently asked questions on the topic of ‘fluoride in dentistry’.

Methods: 15 frequently asked open-ended questions about fluoride applications in dentistry were directed to ChatGPT 4.0, ChatGPT 4.10, Copilot, Gemini, and Perplexity platforms. All questions were asked only once, from the same computer IP address and using the same fixed fiber internet network. Based on current literature and clinical practices, all responses were evaluated by a consensus of three expert pediatric dentists with at least ten years of clinical experience. Likert Scales were used for accuracy and completeness, the mDISCERN scale for quality and reliability, and the Ateşman Readability Index for readability. Fisher’s Exact, Kruskal-Wallis, Dunn, and Mann-Whitney U Test were used for statistical analysis.

Results: According to the median values of Likert Accuracy and Completeness Scales, Gemini> Perplexity> ChatGPT 4.10> ChatGPT 4.0> Copilot (p=0.016), (p=0.002). The ranking according to mDISCERN index: Perplexity Gemini> ChatGPT 4.10> ChatGPT 4.0> Copilot (p=0.008). While the answers given by all AI chat engines had an average readability of medium difficulty, there was a statistically significant difference between Gemini and Perplexity (p=0.010).

Conclusion: Although the answers on the subject of ‘Fluoride in dentistry’ were accurate primarily and valid for all AI chatbots, the datasets used for their training should benefit from sources containing quality and the most up-to-date scientific guidelines, and easier readability should be provided. More research and development are needed on the subject.

Keywords: Artificial intelligence chatbot, dentistry, fluoride, quality and reliability, readability

Mersin Üniversitesi Tıp Fakültesi Lokman Hekim Tıp Tarihi ve Folklorik Tıp Dergisi 2025;15 (3):1069-1078

DOI: 10.31020/mutftd.1746766

e-ISSN: 1309-8004

Geliş Tarihi – Received: 20. Temmuz.2025; Kabul Tarihi- Accepted:10.Eylül.2025

İletişim- Correspondence Author: Seçkin Aksu < seckinaksu@mersin.edu.tr >

Etik Kurul Onayı: Mersin Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurul Başkanlığı (Tarih: 09.07.25,Sayı: 779)

Giriş

Günümüzde birçok farklı alanda bilgiye ulaşmak için kullanılan internet, sağlık hizmetleri konusunda da yararlanılan oldukça yaygın bir uygulamadır.[1] Hasta ve hasta yakınları, genellikle sağlık çalışanlarının kendilerine söylediği bilgileri yeterince anlamadığı için veya sadece kişisel endişelerini gidermek amacıyla mevcut bilgilerinin sağlamasını yapma isteği nedeniyle internet bilgilerine başvurmaktadır. Bu eğilim, 2000'li yılların başında hastaların sağlık bilgisi edinmek için interneti kullanma alışkanlıklarını ifade eden 'Dr. Google' teriminin ortaya çıkmasına yol açmıştır.[2]

Zaman ilerledikçe günlük yaşamımızda bir dizi önemli değişime sebep olan popüler gelişmelerden biri ise yapay zekâ teknolojisidir.[3] Bilgisayarlı sentetik insan bilişsel işlev olarak tanımlanan "yapay zeka" terimi, aslında ilk olarak 1956'da ortaya çıkmıştır.[4] Karar alma, sorun çözme ve deneyimden öğrenme gibi genellikle insan zekası gerektiren görevleri yerine getirmek üzere tasarlanmış makineleri veya yazılım sistemlerini tanımlamaktadır.[5] Yapay zekanın bir alt kümesi olan sohbet robotları, doğal dil işleme ve makine öğrenimi algoritmalarını kullanarak insan kullanıcı sorgularını otomatik olarak anlayan ve yanıtlayan dil modelleridir. Günümüzde farklı özelliklere sahip farklı dil modelleri bulunmaktadır.[6]

OpenAI (San Francisco, CA, ABD) tarafından geliştirilen ChatGPT, metni yorumlayabilen ve üretebilen güçlü bir dil modelidir. İlerleyen teknoloji ile birlikte çok çeşitli konularda fazla miktarda bilgi sağlayabilir ve bu da onu dil çevirisi, özetleme ve metin tamamlama gibi çeşitli uygulamalar için kullanışlı hale getirir.[7,8]

Microsoft (Microsoft Corporation, Redmond, WA, ABD), Copilot[4] olarak yeniden başlatılan GPT-4 dil modelini kullanan Bing Chat AI sohbet robotunu tanıtmıştır. Copilot, kodlama ve işleme için sırasıyla GPT-4 teknolojisinin yanı sıra Kod Yorumlayıcı ve DALL-E 3'ü kullanır. Microsoft 365 uygulamalarıyla da etkili bir şekilde çalışarak canlı internet erişimi aracılığıyla güncel olaylardan haberdar olmak ve alınan bilgiler için kaynaklara bağlantılar içeren dipnotlar sağlamak gibi avantajları barındırmaktadır.[7-9]

Google (Google Ireland Limited, Dublin, İrlanda), başlangıçta LaMDA ve daha sonra PaLM 2 tarafından desteklenen Gemini olarak yeniden başlatılan Google Bard dil modelini tanıtmıştır. ChatGPT'nin yanıtları önceden var olan eğitim verilerine dayanırken, Gemini yanıtları oluştururken güncel bilgileri dahil etmek için internete gerçek zamanlı erişim sağlamaktadır.[10,11]

Perplexity, Perplexity AI tarafından kullanılan web sitelerine hiper bağlantılar gerçekleştirerek çıktıdaki kaynakların sunumunu kolaylaştırır. Ayrıca, önceden oluşturulmuş "ilgili" girdiler çıktının altında sağlanır ve bu da kullanıcının daha fazla yorum yapmasına veya orijinal çıktıyı daha fazla keşfetmesine yardımcı olur.[12,13]

Teknolojik gelişmeler sayesinde yapay zeka sohbet robotlarının kullanımı hastalar ve hatta sağlık profesyonelleri arasında tıp ve diş hekimliği bilgisi için uygun bir kaynak olarak giderek daha popüler hale gelmektedir.[14] Günün her saati erişilebilirlik, sorulara hızlı yanıtlar sunma ve randevu ihtiyacını en aza indirerek her an bilgiye erişimi sağlar.[15] Bu özelliklere ek olarak, "halüsinasyonlar" olarak bilinen gerçek dışı yanıtlar verme potansiyeli, farklı zamanlarda farklı yanıtlar verme olasılığı ve eğitim süreçlerinin güvenilirliğiyle ilgili endişeler, sağlık gibi kritik alanlarda ciddi kaygılara yol açmaktadır.[15,16] Yapay zeka tarafından oluşturulan sağlık ile ilgili içeriğin nesnel olarak değerlendirilmesi esastır. Sohbet robotlarındaki bilgi kaynağının güvenilirliği ve kalitesi, hastaların tedavide iş birliği ve uyumunu ve ayrıca doktor-hasta iletişimini ve güvenini etkileyebileceği için çok önemlidir.[5]

Flor, özellikle çocuk diş hekimliği alanında, ülkemizde ve dünyada koruyucu uygulamalar arasında en sık tercih edilen ajanlardan biridir. Diş hekimleri tarafından hem terapötik hem de profilaktik ajan olarak topikal ve sistemik uygulamaları bulunmaktadır.[17] Günümüzde çoğunlukla topikal kullanımını tavsiye eden klinisyenler doğru dozda terapötik etkiye sahip olan flor uygulamalarının aşırı dozlarda yaratabileceği toksik etkiler

hakkında detaylı bilgiye sahiptir ve hastalarını bu doğrultuda bilgilendirmektedir.[18] Ancak, son çalışmalar toplumda bu uygulamalara karşı önyargının devam ettiğini göstermektedir.[19,20]

Bu çalışmanın amacı, diş hekimliğindeki flor konusu ile ilgili sıkça sorulan sorulara farklı yapay zeka sohbet robotları tarafından sağlanan bilgilerin doğruluk, eksiksizlik, güvenilirlik-kalite ve okunabilirlik yönünden değerlendirilmesidir.

Gereç ve Yöntem

Bu çalışmada "Diş hekimliğinde flor uygulamaları hakkında sık sorulan sorular" arama terimine yanıt veren web sitelerini belirlemek için Google Arama Aracı kullanılarak bir arama yapıldı.[3] Literatürde, arama robotu kullanıcılarının %90'ının yalnızca arama sonuçlarının ilk üç sayfasını görüntülediği bildirilmektedir.[14] Bu doğrultuda yazarlar tarafından hazırlanan açık uçlu sorulardan 15 adeti belirlenip üç uzman diş hekiminden görüş alınarak soruların netleştirilmesi ve sadeleşmesi sağlandı.

Belirlenen sorular ***(Tablo 1)***, 17 Temmuz 2025 tarihinde dahil edilme ve dışlanma kriterlerine uyan beş yapay zeka sohbet robotuna [ChatGPT 4.0 (OpenAI), ChatGPT 4.10 (OpenAI), Copilot (Bing, Microsoft), Gemini (Google) ve Perplexity (Perplexity.AI)] platformlarına yöneltildi. Çalışmaya dahil edilme kriterleri; yapay zeka sohbet robotlarının genel kullanıcı erişimine açık olan web tabanlı veya ücretsiz sürüme sahip olması, aktif olarak güncellenmiş ve çok dilli desteği bulunması, akademik, eğitsel veya genel bilgi verme amaçlı kullanılabilecek şekilde tasarlanmış sistemler olması, kullanıcı verilerinin güvenliğini temel alan ve açık etik politikaları bulunması ve metin tabanlı doğal dil işleme yeteneklerine sahip olması koşullarını sağlamasıdır. Güncellenmemiş, bilgi kesim tarihi eski olan ya da bakım desteği sona ermiş sistemler, sınırlı dil desteğine sahip olup Türkçe dilinde anlamlı etkileşim sağlayamayan ya da yanıt üretim kapasitesi düşük modeller, doğrudan bilgi üretmek yerine yalnızca yönlendirme yapan veya yalnızca tek işlev (örneğin sohbet, randevu hatırlatma) üzerine odaklanmış dar kapsamlı diyalog sistemleri, genel kullanıcı erişimine kapalı olan, yalnızca kurumsal lisans ile erişilebilen veya deneysel aşamada olan yapay zeka sohbet robotları, gizlilik, veri güvenliği ve etik politika açısından şeffaflık sağlamayan, kullanıcı verilerinin nasıl işlendiğine dair açık beyan sunmayan sistemler çalışmadan dışlandı.

Tüm sorular yalnızca bir kez, aynı bilgisayar IP adresinden ve aynı sabit fiber internet ağı kullanılarak soruldu. Yanıtların farklı versiyonunun seçenek olarak sunulduğu durumlar olduğunda birinci versiyon değerlendirme için seçildi. Farklı konuşma stili modu sunulması durumunda ise standart mod seçildi. Yanıtların korelasyonunu önlemek için her sohbet robotu için yeni bir kullanıcı kaydı oluşturuldu ve her soru için ayrı bir sohbet penceresi açıldı. Tüm yanıtlar, güncel literatür ve klinik uygulamalara dayalı olarak değerlendirilmeye tabi tutuldu. Doğruluk ve eksiksizlik, güvenilirlik ve kalite ile okunabilirlik puanları belirlendi.

Açık uçlu yanıtların ve klinik senaryoların doğruluğu ve eksiksizliği, önceden tanımlanmış olan Likert Doğruluk ve Eksiksizlik Ölçekleri kullanılarak değerlendirildi.[21] Likert Doğruluk Ölçeği'nde, 1 puan tamamen yanlışı, 2 puan doğru öğelerden daha fazla yanlışı, 3 puan doğru ve yanlış öğelerin eşit dengesini, 4 puan yanlış öğelerden daha fazla doğru öğeyi, 5 puan ise tamamen doğruyu temsil etmektedir. Likert Eksiksizlik Ölçeği'nde ise, 1 puan sorunun yalnızca bazı yönlerini ele alan ve önemli kısımları eksik veya tamamlanmamış olan eksik yanıtı, 2 puan sorunun tüm yönlerini ele alan ve eksiksizlik için gereken asgari bilgiyi sağlayan yeterli yanıtı ve 3 puan ise sorunun tüm yönlerini kapsayan ve beklentilerin ötesinde ek bilgi veya bağlam sunan kapsamlı bir yanıtı ifade etmektedir.

DISCERN ölçeği, çevrimiçi sağlık bilgilerinin güvenilirliğini ve kalitesini değerlendirmek için önceki çalışmalarda kullanılan üç bölümlü bir ölçektir.[22] İlk bölüm, bilgilerin güvenilirliğini değerlendirmek için sekiz sorudan oluşmakta iken; ikinci bölüm, tedavi seçenekleri hakkındaki bilgilerin kalitesini değerlendirmek için yedi sorudan oluşmaktadır. Son bölüm ise, tedavi seçenekleri hakkında bilgi kaynağı olarak verinin genel kalitesine

odaklanmaktadır. Ancak bu çalışmada, sorularımızın hepsi tedavi ile ilgili olmadığından önceki çalışmaların ışığında yapay zeka sohbet robotlarının yanıtlarının güvenilirliğini değerlendirmek için ilk sekiz soruluk bölümden oluşan modifiye edilmiş versiyonu (mDISCERN Ölçeği) kullanıldı.[7] Ölçekteki her soru için toplam puan, hayır cevabı 1, kısmi yanıt 2-3-4 ve evet cevabı 5 olarak puanlanarak hesaplanarak %40' ın altındaki puanlar (8-15) kötü, %40-79 (16-31) orta ve %80'in üzerindeki puanlar (32-40) iyi olarak derecelendirildi.

Son olarak, yapay zeka sohbet robotlarının sunduğu yanıtların okunabilirliği, Türkçe metinler için sıklıkla kullanılan Ateşman Okunabilirlik İndeksi ile değerlendirildi.[23] Ateşman Okunabilirlik İndeksi, kelime ve cümle uzunluğu analizine dayanarak tasarlandı. Okunabilirlik puanını hesaplama formülü şu şekildedir: 198.825– 40.175 × (toplam hece/toplam kelime) – 2.610 × (toplam kelime/toplam cümle). Buna göre, bir metnin okunabilirliği, sayısal değeriyle ters orantılıdır; daha yüksek değerler daha kolay okunabilirliği (100'e yakın) ve daha düşük değerler daha zor okunabilirliği (0'a yakın) göstermektedir. 1-29 puan çok zor, 30-49 puan zor, 50-69 puan orta, 70-89 puan kolay ve 90-100 puan çok kolay okunabilir olarak sınıflandırıldı.

İstatistiksel Yöntem

Veriler bilgisayar ortamında analiz edildi. Normal dağılıma uygunluk Shapiro-Wilk Testi ile incelendi. Sohbet robotları ile kategorik değişkenler arasındaki bağlantının incelenmesinde Monte Carlo Düzeltmeli Fisher's Exact Testi kullanıldı ve çoklu karşılaştırmalar Bonferroni Düzeltmeli Z Testi ile yapıldı. Sohbet robotlarına göre normal dağılıma uymayan değişkenlerin karşılaştırılmasında Kruskal Wallis Testi kullanıldı ve çoklu karşılaştırmalar Dunn Testi ile yapıldı. Okunabilirlik durumuna göre normal dağılıma uymayan değişkenlerin karşılaştırılmasında Mann Whitney U Testi kullanıldı. Analiz sonuçları kategorik değişkenler için frekans (yüzde) şeklinde, nicel değişkenler için ortalama ± standart sapma ve ortanca (minimum – maksimum) şeklinde sunuldu. Önem düzeyi $p<0.05$ olarak alındı.

Araştırmanın Etik Yönü

Çalışmanın etik onayı Mersin Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurul Başkanlığı'ndan (Karar Sayısı:779, 09.07.25) alınmıştır.

Tablo 1. Yapay zeka sohbet robotlarına 'diş hekimliğinde flor kullanımı' hakkında sorulan sorular

	Sorular
1	Flor nedir?
2	Flor diş çürüklerine nasıl etki eder?
3	Flor dişlere nasıl uygulanır?
4	Flor uygulaması dişlere zarar verir mi?
5	Okullarda flor uygulaması yapılması doğru mu?
6	Flor zeka geriliği yapar mı?
7	Florlu diş macunları zararlı mı?
8	Flor fazlalığında insan vücudunda ne olur?
9	Florlu macun kaç yaşında kullanılır?
10	Diş macunları ne kadar flor içermelidir?
11	Dişlere flor uygulaması sırasında ve sonrasında nelere dikkat edilmelidir?
12	Flor jelleri ve vernikleri arasında fark var mıdır?
13	Dişlere kaç yaşına kadar flor uygulanır?
14	Dişlere ne kadar sıklıkla flor uygulanmalıdır?
15	Flor dişte renk değişikliği yapar mı?

Bulgular

Likert Doğruluk Ölçeği, Likert Eksiksizlik Ölçeği, mDISCERN Ölçeği ve Ateşman Okunabilirlik İndeksi'ne göre yapay zeka sohbet robotlarının yanıtlarına verilen puanlamaların istatistiksel karşılaştırılması Tablo 2'de sunulmuştur. Likert Doğruluk Ölçeği ortanca değerlerine göre gruplar arasındaki puanlama sıralaması Gemini> Perplexity> ChatGPT 4.10> ChatGPT 4.0> Copilot şeklinde olup istatistiksel olarak anlamlı bir farklılık bulunmuştur ($p=0.016$). Likert Eksiksizlik Ölçeği ortanca değerlerine göre ise; Gemini> Perplexity> ChatGPT

4.10> ChatGPT 4.0> Copilot şeklindedir (p=0.002). Düşük güvenilirlik ve kalitede olan yanıtların sayısı 3 iken orta olanların sayısı 72'dir. mDISCERN indeksine göre sıralama; Perplexity= Gemini> ChatGPT 4.10> ChatGPT 4.0> Copilot şeklindedir (p=0.008). ***(Tablo 2)***

Tablo 2. Yapay zeka sohbet robotlarının doğruluk, eksiksizlik ve kalite-güvenirlilik değerlerinin karşılaştırılması

Yapay Zeka Sohbet Robotları	Likert Doğruluk Ölçeği		Likert Eksiksizlik Ölçeği		mDISCERN Ölçeği	
	Ortalama±ss	Ortanca (min-maks)	Ortalama±ss	Ortanca (min-maks)	Ortalama±ss	Ortanca (min-maks)
ChatGPT 4.0	4.07 ± 0.88	4 (2 - 5)^{bc}	2.07 ± 0.46	2 (1 - 3)^{abc}	24.93 ± 2.37	25 (20 - 28)^{b}
ChatGPT 4.10	4.20 ± 0.94	4 (2 - 5)^{bc}	2.47 ± 0.64	3 (1 - 3)^{c}	24.47 ± 4.49	25 (12 - 31)^{b}
Copilot	3.20 ± 1.01	3 (2 - 5)^{b}	1.60 ± 0.63	2 (1 - 3)^{b}	21.00 ± 3.85	21 (14 - 27)^{b}
Gemini	4.83 ± 0.52	5 (4 - 5)^{ac}	2.87 ± 0.35	3 (2 - 3)^{a}	27.47 ± 1.30	28 (25 - 30)^{a}
Perplexity	4.73 ± 0.46	5 (4 - 5)^{a}	2.60 ± 0.51	3 (2 - 3)^{ac}	28.33 ± 2.97	30 (21 - 31)^{ab}
p*	**0.016**		**0.002**		**0.008**	

*Kruskal Wallis Testi; a-c: Aynı harfe sahip yapay zeka sohbet robotları arasında istatistiksel bir farklılık yoktur.

Okunabilirlik açısından sohbet robotları arasında istatistiksel olarak anlamlı bir farklılık bulunmuştur (p=0.010). Burada farklılık okunabilirliğin orta ve zor olması açısından Gemini ile Perplexity arasında görülmüştür. ***(Tablo 3)***

Tablo 3. Yapay zeka sohbet robotlarının okunabilirlik karşılaştırılması

Okunabilirlik	ChatGPT 4.0	ChatGPT 4.10	Copilot	Gemini	Perplexity	p
Kolay	0 (0.00)	1 (6.70)	1 (6.70)	0 (0.00)	0 (0.00)	
Orta	10 (66.7)ab	7 (46.7)ab	10 (66.7)ab	14 (93.3)b	5 (33.3)a	**0.010**
Zor	5 (33.3)ab	7 (46.7)ab	4 (26.7)ab	1 (6.70)b	10 (66.7)a	

Tartışma

Günümüzde hastaların ve ebeveynlerin önemli bir kısmı, sağlık sorunları hakkında daha fazla bilgi edinmek için interneti ve sosyal medya platformlarını kullanmaktadır. Yapay zeka ve insan etkileşimlerindeki son gelişmeler ile kullanımı giderek yaygınlaşan yapay zeka sohbet robotları hastalara tedaviler, olası yan etkiler, komplikasyonlar, prognozlar ve sonuçlar hakkında bilgi sağlayabilir.[7,24] Tüm yapay zeka sohbet robotu modelleri dil işlemeye dayalı olsa da sinir ağı mimarilerindeki farklılıklar, eğitim verilerinin kalitesi, miktarı, çeşitliliği, kullanılan optimizasyon stratejileri ve teknikler nedeniyle aynı sorulara farklı yanıtlar üretmektedir. Her modelin tipik olarak farklı güçlü yönleri, zayıflıkları, yetenekleri ve sınırlamaları vardır.[13] Bu nedenlerden dolayı, uygulanabilirliklerini belirlemeden önce performanslarının ve potansiyel hatalarının karşılaştırılmalı değerlendirilmesi esastır.[25] ChatGPT farklı sürümlerinin doğruluk ve güvenilirliği ile ilgili nispeten daha fazla çalışma[7,13,16,24,26] olduğu tespit edilmekle birlikte bu çalışma, özellikle Copilot, Gemini ve Perplexity gibi bugüne kadar kapsamlı bir şekilde araştırılmamış yeni teknolojiler hakkında önemli veriler ortaya koymayı amaçlamıştır.

Diş hekimliğinde flor kullanımı, toplum temelli uygulamalarla birlikte, diş çürüklerinin önlenmesinde son derece önemli koruyucu yaklaşımlardır.[27] Bireysel ve profesyonel olarak uygulanabilen topikal florür, demineralizasyonu önler ve remineralizasyona katkıda bulunur. Öte yandan, hem akut hem de kronik olarak yüksek dozda florüre sistemik maruziyet çeşitli toksisitelere neden olabilir.[28-30] Mineral yataklarında yüksek flor içeriği bulunan ülkelerde gerçekleştirilen bazı çalışmalar, flor maruziyetinin bilişsel ve nöral gelişim üzerinde olumsuz etkileri olduğunu gösterse de, bu çalışmaların metodolojisi doğada aşırı miktarda florüre maruz kalmış kişiler üzerinde yapılması ve bilişsel ve sinirsel gelişim üzerinde etkisi olabilecek birçok olası faktörün etkilerinin gözardı edilmesi yönünden yetersizdir.[31,32] Ancak, bu çalışmalar yayınlandıkları günden bu yana dünya çapında florür karşıtı lobinin güçlenmesine katkıda bulunmuştur.[27] Dünya çapında bilimsel geçerliliği kanıtlanmış kuruluşlar tarafından uygun doz ve zamanda alınan florürün bireyin genel sağlığı üzerinde olumsuz bir etki yaratmadan dental sağlığı açısından önemli katkılar sağladığı bilgisi giderek daha

fazla vurgulanmaktaysa da birçok kaynaktan basına yansıyan çelişkili bilgiler nedeniyle özellikle çocuk diş hekimliği alanında en çok tartışılan konulardan biridir. Aşılama protokolü yaklaşımına benzer şekilde ticari şirketler ve reklam kaynakları tarafından, özellikle sosyal medyada ve internette, flor hakkında yapılan iddiaların nesnelliği aileleri şüpheye sürüklemektedir.[33,34] Bu nedenle, konu hakkında bilgiye erişim yönünde potansiyel bir kaynak olan yapay zeka sohbet robotlarının sunduğu sağlık bilgilerinin kalitesi önemlidir. Bu çalışma, diş hekimliğinde flor kullanımı ile ilgili en çok merak edilenlere odaklanarak, beş farklı sohbet robotu yanıtlarının doğruluğunu, eksiksizliğini, güvenilirliğini ve okunabilirliğini değerlendiren ilk çalışmalardan biridir.

Bildiğimiz kadarıyla, yapay zeka sohbet robotunun florür kullanımına ve etkilerine verdiği tepkilerin değerlendirilmesine yönelik bilimsel literatürdeki ilk çalışma Buldur ve ark.'na aittir.[27] Klinik senaryoların yalnızca ChatGPT'ye yönlendirildiği çalışmada alınan yanıtlar Amerikan Diş Hekimleri Birliği'nin web sitesindeki bilgilerle benzerlik sergilemiştir. Öte yandan COVİD-19 pandemisinin ardından gerçekleşen bilimsel gelişmeler doğrultusunda rehberlerde Eylül 2021'den sonra gerçekleşen bazı değişiklikler hakkında ChatGPT'nin bilgisinin sınırlı olduğu tespit edilmiştir.[35] Çalışmamızdaki yanıtlarının 4.0 ve 4.10 versiyonu için benzer şekilde yeterli bulunmasına rağmen özellikle florlu macunlar ile ilgili olan 9 ve 10. sorularda örnek çalışmada belirtildiği gibi güncel değişimler konusunda Gemini dışındaki sohbet robotlarının yeterli bilgisinin olmadığı görülmektedir. Bulgularımız, ChatGPT'nin 2021'e kadar eğitilen GPT3'e dayanması nedeniyle en son verileri dikkate almadığı bilgisiyle örtüşmektedir.[2]

Yapay zeka sohbet robotlarının yaygın erişilebilirliği göz önüne alındığında, performanslarının titizlikle değerlendirilmesi ve sağladıkları bilgilerin araştırmacılar ve ilgili değerlendirme kuruluşları tarafından kapsamlı bir şekilde incelenmesi zorunludur. Sağlık profesyonelleri halkı, sohbet robotlarından elde edilen bilgileri değerlendirmek ve kullanmak konusunda eğitmelidir.[13] Doğası gereği olasılıkçı olan sohbet robotları, eğitim verilerindeki istatistiksel örüntülere dayanarak bir sonraki kelimenin olasılığını tahmin ederek yanıtlar oluştururlar. Sonuç olarak, aynı soruyu iki kez sormak her zaman aynı yanıtları vermeyebilir.[36] Dolayısıyla sorular, tek seferde ve aynı bilgisayar kullanıcısı tarafından yöneltilmiştir.

Literatürde, yapay zeka sohbet robotu yanıtlarının doğruluğunu değerlendirmek için farklı indeksler kullanılmıştır, bu çalışmada doğruluk ve eksiksizlik için iki tip Likert ölçeği kullanılmıştır.[7-10] Tüm yapay zeka sohbet robotlarının genellikle tüm sorulara makul ölçüde doğru yanıtlar verdiği tespit edilmiştir. Johnson ve ark.'nın[37] diş travmasıyla ilgili sorulara farklı yapay zeka sohbet robotlarının verdiği yanıtları değerlendirdikleri çalışmanın bulguları, Copilot'un ChatGPT ve Gemini'den daha düşük güvenilirlik sergilediğini göstererek çalışmamızın bulgularını desteklemektedir. Çalışmamızda çoğu soruya genel geçerli nitelikte olsa da kimi zaman ebeveynleri yanlış yönlendirebilecek ve diş hekimliği ile örtüşmeyen yanıtlar sunması Copilot'u diğer sohbet robotlarının gerisine taşımıştır. Öte yandan, genel ortodonti ile ilgili klinik sorulara verilen yanıtları değerlendiren bir çalışmada, Makrygiannakis ve ark., en iyi yanıtların sırasıyla Copilot ve ChatGPT-4 tarafından verildiğini bulmuştur.[11] Bu farklılık soruların ayrıntı ve teknik içeriğiyle ilgili olabilir. Balel, ChatGPT'nin potansiyel hasta sorularına iyi ila mükemmel güvenilir yanıtlar verdiğini, potansiyel hekim sorularına verilen yanıtların ise orta kalitede olduğunu bildirmiştir.[26] Bu durum, yapay zeka sohbet robotlarının hasta bakış açısından bilgilendirme aracı olarak önemli bir potansiyele sahip olabileceğini, ancak teknik sorularda ve profesyonellerin bakış açısından eğitimlerde kullanımında tamamen güvenli olmayabileceğini düşündürmektedir.

Sohbet robotlarının yanlış veya yanıltıcı bilgiler sunması mümkündür ve hastalar bu bilgilere inanmaya meyilli olabilir.[3] Bilgi kaynakları, atıflar ve referanslar sağlık bilgilerinin güvenilirliğini değerlendirmede önemlidir.[22] Dolayısıyla mDISCERN indeksine göre puanlama sıralamasının Likert ölçeklerinden farklı olmasının nedeni sohbet robotlarının yanıtlara kaynak gösterip göstermediği veya kaynakların bilimsel yönleri ile

ilişkilendirilebilir. Soruların genelinde ChatGPT'nin iki sürümünün de birbirine yakın ve tatmin edici yanıtlar verdikleri fakat referans sunmamaları kalite ve güvenirlilik puanlarını sınırlandırmaktadır. En tatmin edici yanıtlar Gemini tarafından verilmiş görülse de birkaç soru dışında referans sunmaması, kendisinden daha az yeterli yanıtlar sunmasına rağmen tüm sorularda referans sunmuş olan Perplexity'i öne geçirmiştir.

DISCERN aracını kullanarak ChatGPT-4 tarafından periodontal hastalık hakkında sağlanan bilgilerin kalitesini değerlendiren literatürdeki bir çalışma, bazı sınırlamalara rağmen yapay zekanın çoğu yanıt için doğru rehberlik sağladığını bulmuş olsa da;[38] Bhattacharyya ve ark. ise, tıbbi içerikte sağlanan referansların çoğunun yanlış olduğu tespit edildiğinden, ChatGPT'de tıbbi bilgi ararken dikkatli olunması gerektiğini vurgulamaktadır.[39] Çalışmamızın bulguları bu durumu destekleyici sonuçlar ortaya koymakla birlikte kanıt sunan Copilot, Gemini, ve Perplexity sohbet robotlarının referanslarının özel diş sağlığı kliniklerinin reklam içerikli sayfalarına veya bilimsel yönden zayıf ve güncel olmayan web sayfalarına yönlendirdiğini gözlemledik. Dolayısıyla yanıtlar bilimsel olarak temellendirilmediğinde ve hakemli referanslar sağlanamadığında sohbet robotlarının birincil kaynak olarak kullanılmaya henüz uygun olmadığı, eğitilmeleri için kullanılan veri kümelerinin güncel yönergeleri içermesi gerektiği anlaşılmıştır.

Sağlık okuryazarlığı, sağlıkla ilgili bilgileri okuma ve anlama yeteneği olarak tanımlanır ve bireylerin bilinçli kararlar almasını ve tedavi süreçlerini yönetmesini sağlar. Literatürde okunabilirlik, sağlık okuryazarlığının temel bir unsuru olarak tanımlanır ve belgelerin anlaşılmasını sağlar.[22] Okunabilirlik, cümle uzunluğu ve zor kelimelerin kullanımı gibi farklı indeksler kullanılarak değerlendirilir.[40] Literatürden taradığımız kadarıyla yapay zeka robotlarının diş hekimliği alanındaki yanıtlarının okunabilirliğini kendi dilimizde değerlendiren ilk çalışma olduğunu gözlemledik. Dolayısıyla İngilizce metinler üzerinde yapılan önceki çalışmaların çoğunun okunabilirliğini değerlendirmek için kullanılan indekslere göre bir kıyaslama yapmanın uygun olmayacağını düşünmekteyiz. Fakat yine de birçok yabancı kaynakta sohbet robotlarının yanıtlarının okunabilirlik seviyelerinin çoğunlukla 'zor' olarak tespit edilmesi çalışmamızın bulgularıyla benzerlik göstermektedir.[7,41,42] ChatGPT 4.0'da okunabilirlik zor olanların oranı %33,3, ChatGPT 4.10'da %46,7, Copilot'da %26,7, Gemini'de %6,7 ve Perplexity'de bu değer %66,7 olarak elde edilmiştir. Sohbet robotlarının okunabilirliğindeki zorluğun halk tarafından kolay kullanımlarını sınırlayacağı düşüncesindeyiz.

Sonuç ve Öneriler

ChatGPT 4.0, ChatGPT 4.10, Copilot, Gemini ve Perplexity yapay sohbet robotları 'diş hekimliğinde flor kullanımı' konusunda genel geçerli bilgiler sunsa da performanslarında belirgin farklılıklar gözlemlenmektedir. Bu çalışma kapsamında Gemini ve Perplexity sohbet robotları, diğerlerine kıyasla daha nitelikli bilgiler sunmuş olsa da, farklı çalışmalar ve soru türlerinde güvenilirliğinin değerlendirilmesi amacıyla tekrarlanmasına ihtiyaç duyulmaktadır.

Öte yandan, okunabilirlik puanlamaları açısından değerlendirildiğinde, çalışmada yer alan tüm sohbet robotlarının genel olarak orta düzeyde bir okunabilirlik değerine sahip olduğu görülmüştür. Ancak Perplexity, diğerlerine kıyasla daha düşük bir okunabilirlik düzeyi sergilemiştir. Bu durum, sohbet robotlarının Türkçe dilindeki metin okunabilirliğini geliştirmelerinin, daha geniş kitlelere ulaşmaları açısından faydalı olabileceğini göstermektedir.

Ayrıca yapay zeka sohbet robotlarının potansiyel faydalarını tam olarak gerçekleştirmek adına eğitilmeleri için kullanılan veri kümelerinin kaliteli bilimsel yönergeleri içeren kaynaklardan yararlanması gerekmektedir. Konuyla ilgili daha fazla araştırma ve geliştirmeye ihtiyaç vardır.

Araştırmanın Kısıtlılıkları

Yanıtların yalnızca Türkçe olarak analiz edilmesi nedeniyle sonuçlar tüm dillere genelleştirilemez. Tüm soruların yalnızca bir kez sorulduğu ve ilk yanıtların değerlendirildiği unutulmamalıdır. Derin öğrenmeye dayalı sohbet robotları her saniye internete yüklenen yeni verileri toplar ve yeni yanıtlar oluşturur. Sonuç olarak, yanıtları değiştirmek kontrolü zorlaştırır. Aynı soru tekrar sorulursa veya farklı zaman noktalarında sorulursa, farklı bir yanıtla karşılaşılması son derece olasıdır. Sorunun ifadesi veya tonu biraz değiştirildiğinde farklı bir yanıt verilmesi, sağlıkla ilgili bilgi arayan hastalar için çok yanıltıcı olabilir. Bu nedenle, verilen yanıtlar güvenilir ve değerli görünse de, bu sınırlamalar ışığında yorumlanmalıdır.

Değişen deneyim seviyelerine sahip daha fazla hekimle daha büyük bir çalışma yürütmek, sohbet robotlarının sağlık hizmetleri için doğruluğunu ve değerini daha fazla araştırmak için bir sonraki adım olabilir. Benzer çalışmaların daha geniş örneklem üzerinde, daha fazla kullanıcıyla ve daha geniş soru yelpazesiyle yürütülmesi gerekmektedir.

Yapay zeka platformları, çocuk diş hekimliği veya diş hekimliğinde flor kullanımı konusunda özel olarak eğitilmemiştir ve bu çalışma yalnızca hasta sorularına odaklanmıştır. Sağlık alanında pek çok konu, ekol farklılıkları nedeniyle farklı yaklaşımları benimseyebilen sağlık profesyonelleri tarafından anket çalışmaları ile ele alınmaya devam ederken yapay zeka sohbet robotlarının kullanıcıya tek tip standart yanıt stili sunmaları beklentisi gerçekçi olmamaktadır.[43-45]

Referans sunan yapay zeka sohbet robotlarının kaynak gösterdiği sayfaların kalitesinin ve güvenirliliğinin detaylıca değerlendirilmesi de ayrı bir çalışma konusu olabilir. Çeşitli yanıtların sağlanması kullanıcılara belirli bir konu hakkında daha kapsamlı bir bakış açısı sunabilse de, sohbet robotlarının yanıt üretmek için kullandığı kaynakların halk sağlığı örgütleri veya yüksek etkili bilimsel dergiler gibi güvenilir olması çok önemlidir. Bunun sağlanamaması gerçekçi olmayan bilgilerin yayılmasına neden olabilir. Bu amaçla kanıta dayalı bilimsel verilerle desteklenen web sitelerinin geliştirilmiş içerik kaynaklarının uzmanlar tarafından hazırlanması teşvik edilebilir.

Bilgi

Yazarlar bu makale için gerçek, potansiyel veya algılanan çıkar çatışması olmadığını beyan ederler.

Bu çalışmada soruların hazırlanması ve değerlendirilmesine katkıda bulunan Mersin Üniversitesi Diş Hekimliği Fakültesi Çocuk Diş Hekimliği Anabilim Dalı'nın değerli öğretim üyelerine kıymetli destekleri için teşekkür ederiz.

Etik Onay

Mersin Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurul Başkanlığı'ndan (Karar Sayısı:779, Karar Tarihi: 09.07.25) etik kurul izni alınmıştır.

Araştırmacı Katkı Oranı Beyanı

Seçkin Aksu: Fikir, tasarım, veri toplama ve işleme, analiz ve yorum, kaynak taraması, makale yazımı, eleştirel inceleme, fon sağlama.

Fatma Ayça Bakır: Veri toplama ve işleme, analiz ve yorum, kaynak taraması, fon sağlama.

Kaynaklar

1. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. Annu Rev Public Health. 2020; 4:433-51.

2. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. Eur J Cardiovasc Nurs. 2024;23(1):95-8.

3. Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. Int J Environ Res Public Health. 2018;15(12).
4. Thurzo A, et al. Where Is the Artificial Intelligence Applied in Dentistry? Systematic Review and Literature Analysis. Healthcare (Basel). 2022;10(7).
5. Acar AH. Can natural language processing serve as a consultant in oral surgery? J Stomatol Oral Maxillofac Surg. 2024;125(3):101724.
6. Chakraborty CS, et al. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front Artif Intell. 2023; 6:1237704.
7. Dursun D, Bilici Gecer R. Can artificial intelligence models serve as patient information consultants in orthodontics? BMC Med Inform Decis Mak. 2024;24(1):211.
8. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. Eur J Cardiovasc Nurs. 2023;22(7): e55-9.
9. Cascella M, et al. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. J Med Syst. 2024;48(1):22.
10. Daraqel B, et al. The performance of artificial intelligence models in generating responses to general orthodontic questions: ChatGPT vs Google Bard. Am J Orthod Dentofacial Orthop. 2024;165(6):652-62.
11. 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;46.
12. Gravina AG, et al. Charting new AI education in gastroenterology: Cross-sectional evaluation of ChatGPT and perplexity AI in medical residency exam. Dig Liver Dis. 2024;56(8):1304-11.
13. Mustuloglu S, Deniz BP. Evaluation of Chatbots in the Emergency Management of Avulsion Injuries. Dent Traumatol. 2025; 0:1-8.
14. Abu Arqub S, et al. Content analysis of AI-generated (ChatGPT) responses concerning orthodontic clear aligners. Angle Orthod. 2024;94(3):263-72.
15. Chen S, et al. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. JAMA Oncol. 2023;9(10):1459-62.
16. Suarez A, et al. Unveiling the ChatGPT phenomenon: Evaluating the consistency and accuracy of endodontic question answers. Int Endod J, 2024;57(1):108-13.
17. Veneri F, Vinceti SR, Filippini T. Fluoride and caries prevention: a scoping review of public health policies. Ann Ig. 2024;36(3):270-80.
18. Manchanda S, et al. Topical fluoride to prevent early childhood caries: Systematic review with network meta-analysis. J Dent. 2022; 116:103885.
19. Chi DL. Parent Refusal of Topical Fluoride for Their Children: Clinical Strategies and Future Research Priorities to Improve Evidence-Based Pediatric Dental Practice. Dent Clin North Am. 2017;61(3):607-17.
20. Chi DL, et al. A conceptual model on caregivers' hesitancy of topical fluoride for their children. PLoS One. 2023;18(3):e0282834.
21. Lyons RJ, et al. Artificial intelligence chatbot performance in triage of ophthalmic conditions. Can J Ophthalmol. 2024;59(4): e301-8.
22. Kilinc DD, Mansiz D. Examination of the reliability and readability of Chatbot Generative Pretrained Transformer's (ChatGPT) responses to questions about orthodontics and the evolution of these responses in an updated version. Am J Orthod Dentofacial Orthop. 2024;165(5):546-55.
23. Varli B, et al. Evaluation of readability levels of online patient education materials for female pelvic floor disorders. Medicine (Baltimore). 2023;102(52):e36636.
24. Ghanem YK, et al. Dr. Google to Dr. ChatGPT: assessing the content and quality of artificial intelligence-generated medical information on appendicitis. Surg Endosc. 2024;38(5):2887-93.
25. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Med Educ Online. 2023;28(1):2181052.
26. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? J Stomatol Oral Maxillofac Surg. 2023;124(5):101471.
27. Buldur M, Sezer B. Can Artificial Intelligence Effectively Respond to Frequently Asked Questions About Fluoride Usage and Effects? A Qualitative Study on Chatgpt. Fluoride. 2023;56(3):201-16.
28. Amadeu de Oliveira F, et al. The effect of fluoride on the structure, function, and proteome of intestinal epithelia. Environ Toxico. 2018;33(1):63-71.
29. Melo CGS, et al. Enteric innervation combined with proteomics for the evaluation of the effects of chronic fluoride exposure on the duodenum of rats. Sci Rep. 2017;7(1):1070.
30. Ullah R, Zafar MS, Shahani N. Potential fluoride toxicity from oral medicaments: A review. Iran J Basic Med Sci. 2017;20(8):841-8.
31. Do LG, et al. Early Childhood Exposures to Fluorides and Child Behavioral Development and Executive Function: A Population-Based Longitudinal Study. J Dent Res. 2023;102(1):28-36.
32. Johnston NR, Strobel SA. Principles of fluoride toxicity and the cellular response: a review. Arch Toxicol. 2020;94(4):1051-69.
33. Basch CH, Milano N, Hillyer GC. An assessment of fluoride related posts on Instagram. Health Promot Perspect. 2019;9(1):85-8.

34. Carpiano RM, Chi DL. Parents' attitudes towards topical fluoride and vaccines for children: Are these distinct or overlapping phenomena? Prev Med Rep. 2018; 10:123-8.
35. Association AD Fluoridation FAQs; https://www.ada.org/resources/community-initiatives/fluoride-in-water/fluoridation-faqs
36. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452-9.
37. Johnson AJ, et al. Evaluation of validity and reliability of AI Chatbots as public sources of information on dental trauma. Dent Traumatol. 2025;41(2):187-93.
38. Alan R, Alan BM. Utilizing ChatGPT-4 for Providing Information on Periodontal Disease to Patients: A DISCERN Quality Analysis. Cureus. 2023;15(9):e46213.
39. Bhattacharyya M, et al. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. Cureus. 2023;15(5): e39238.
40. McInnes N, Haglund BJ. Readability of online health information: implications for health literacy. Inform Health Soc Care. 2011;36(4):173-89.
41. Onder CE, et al. Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. Sci Rep. 2024;14(1):243.
42. Seth I, et al. Comparing the Efficacy of Large Language Models ChatGPT, BARD, and Bing AI in Providing Information on Rhinoplasty: An Observational Study. Aesthet Surg J Open Forum. 2023;5: ojad084.
43. Delikan E, Ertürk-Avunduk T, Aksu S. Approaches of general and specialist dentists to deep caries management: a cross-sectional study from Turkey. J Dent Indonesia. 2021;28(2):94–104.
44. Mishra A, et al. Knowledge, Awareness, Attitudes of Radiation Hazards and Protection Among Dentist: A Questionnaire Survey. J Pharm Bioallied Sci. 2025;17(1):281-3.
45. Dămăşaru MS, et al. Implications of type 1 diabetes mellitus in the etiology and clinic of dento-maxillary anomalies-questionnaire-based evaluation of the dentists' opinion. Med Pharm Rep. 2025;98(1):135.