

ORJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/dentalsci.2025-108696

Diş Hekimliğinde Uzmanlık Sınavında Sorulan Çocuk Diş Hekimliği Sorularını Cevaplamada ChatGPT-4o Plus ve Gemini'nin Karşılaştırmalı Analizi: Kesitsel Çalışma

Comparative Analysis of ChatGPT-4o Plus and Gemini in Answering Pediatric Dentistry Questions for the Dentistry Specialization Exam: A Cross-Sectional Study

^{1b} Ezgi MERİÇ^a

^aAdana Yüreğir Karşıyaka Ağız ve Diş Sağlığı Hastanesi, Çocuk Diş Hekimliği Kliniği, Adana, Türkiye

ÖZET Amaç: Yapay zekâ destekli sohbet robotlarının diş hekimliği alanında kullanımı giderek artmaktadır. Ancak sunulan bu teknolojinin diş hekimliği alanındaki bilgi düzeyi belirsizdir. Çalışmanın amacı, diş hekimliği uzmanlık sınavında çıkmış olan çocuk diş hekimliği soruları kullanılarak yapay zekâ destekli sohbet robotlarının eğitim destek araçları olarak potansiyelini araştırmaktır. **Gereç ve Yöntemler:** 2012-2021 yılları arasındaki Diş Hekimliğinde Uzmanlık Eğitimi Giriş Sınavı'nda sorulan çocuk diş hekimliği soruları konulara göre 13 grupta sınıflandırılmıştır [Diş embriyolojisi ve diş sürmesi; diş gelişim anomalileri; çürük, beslenme ve risk değerlendirmesi; çürüklü dişlerde pulpal problemler; florid; restoratif materyaller; okluzyon ve ortodontik aparatlar; periodontal problemler ve sistemik hastalıklar; davranış yönlendirmesi (farmakolojik ve non-farmakolojik); lokal anestezi; dental travma ve tedavileri; pedodontide ilaç kullanımı; diş dokularının histolojik ve morfolojik özellikleri]. Toplam 127 adet çoktan seçmeli soru Chat Generative Pre-trained Transformer-4 Omni (ChatGPT-4o Plus) ve Gemini sohbet robotlarına aynı gün yöneltilmiştir. Sohbet robotlarının verdiği cevaplara yanlışa 0 puan, doğru ise 1 puan verilmiştir. Anlamlılık değeri $p < 0,05$ olarak kabul edilmiştir. **Bulgular:** Araştırmanın sonuçlarına göre ChatGPT-4o Plus %70,7 (n=89); Gemini ise %76 (n=96) doğruluk oranı vermiştir. Ancak, ChatGPT-4o Plus ve Gemini'nin çalışmaya dâhil edilen sorulara verdikleri yanıtlara karşılık yıllara göre aldıkları puanlar arasında anlamlı fark görülmemiştir ($p=0,502$). Sohbet robotlarının konu başlıklarına göre bilgi düzeyleri arasında da anlamlı fark bulunmamıştır ($p=0,397$). **Sonuç:** ChatGPT-4o Plus ve Gemini'nin çocuk diş hekimliği alanındaki bilgi düzeyleri benzer yükseklikte olsa da, soruları doğru yanıtlama yüzdeleri karşılaştırıldığında performanslarının sınırlı olduğu görülmektedir.

ABSTRACT Objective: The use of artificial intelligence powered chatbots in dentistry is increasing. However, the knowledge level of these technologies in the dental field remains uncertain. This study aimed to investigate the potential of artificial intelligence-supported chatbots as educational support tools using pediatric dentistry questions from the dentistry specialization exam. **Material and Methods:** Pediatric dentistry questions from the Dentistry Specialization Exam between 2012 and 2021 were classified into 13 groups according to the topics [Dental embryology and tooth eruption; dental anomalies; dental caries, nutrition, and caries risk assessment; pulpal problems in decayed teeth; dental fluoride; restorative materials; dental occlusion and orthodontic appliances; periodontal issues and systemic diseases; behavioral guidance (pharmacological and non-pharmacological); dental local anesthesia; dental trauma and treatments; medication use in pedodontics; histological and morphological features of dental tissues]. 127 multiple-choice questions were posed to Chat Generative Pre-trained Transformer-4 Omni (ChatGPT-4o Plus) and Gemini chatbots on the same day. The chatbot responses were scored 1 point for correct answers and 0 for wrong answers. Significance value was accepted as $p < 0.05$. **Results:** According to the results of the research, ChatGPT-4o Plus has an accuracy rate of 70.7% and Gemini 76%. No significant difference was found between their scores on the questions included in the study ($p=0.502$). There was also no significant difference in the chatbots' knowledge levels regarding the topics ($p=0.397$). **Conclusion:** Although ChatGPT-4o Plus and Gemini have similarly high knowledge levels in pediatric dentistry, their performance is limited when comparing percentages of correct answers.

Anahtar Kelimeler: Diş hekimliği; pedodonti; büyük dil modelleri; yapay zekâ

Keywords: Dentistry; pediatric dentistry; large language models; artificial intelligence

Correspondence: Ezgi MERİÇ

Adana Yüreğir Karşıyaka Ağız ve Diş Sağlığı Hastanesi, Çocuk Diş Hekimliği Kliniği, Adana, Türkiye

E-mail: ezgimeric@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Dental Sciences.

Received: 20 Jan 2025

Received in revised form: 01 Apr 2025

Accepted: 07 Apr 2025

Available online: 13 Jun 2025

2146-8966 / Copyright © 2025 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Yapay zekâ (YZ) destekli sohbet robotları her geçen gün hızla yenilenmekte ve geliştirilmektedir. En yaygın olarak bilinen YZ destekli sohbet robotları arasında OpenAI (San Francisco, ABD) ve Google (Mountain View, ABD) tarafından geliştirilen ChatGPT (Chat Generative Pretrained Transformer) (OpenAI; San Francisco, ABD) ve Gemini (Google; Mountain View, ABD) yer almaktadır.¹ Kasım 2022’de, ChatGPT’nin prototip varyantı piyasaya sürülmüştür. Sık kullanılan diğer bir YZ destekli sohbet robotu olan Bard (Google; Mountain View, ABD), Google tarafından Mayıs 2023’te kullanıma sunulmuştur. Yakın zamanda Google, Bard’ın güncel sürümü olan Gemini’yi en büyük ve en donanımlı YZ teknolojisi olarak kullanıma sunmuştur.² ChatGPT-3,5’un geliştirilmiş ve abonelik sürümü olan ChatGPT-4o, rakipleriyle karşılaştırıldığında kullanımı kolay ve OpenAI’nin web sitesinde herkesin erişimine açıktır. Bu yaygın erişilebilirlik, bu sohbet robotunu birçok kullanıcı için önde gelen bir tercih haline getirmektedir.

YZ teknolojisi, insan benzeri etkileşimleri yakından taklit eden gelişmiş iletişim yeteneklerine sahiptir. Bu yetenek, sanal asistanlar, sohbet robotları ve çevrimiçi öğrenme destek sistemleri gibi eğitim alanlarında mükemmel bir donanımla desteklenmiştir. ChatGPT, şu ana kadar eğitim, araştırma ve sağlık hizmetlerinde en çok incelenen YZ teknolojisiidir.¹ Sağlık hizmetlerinde klinik uygulamalarda sağlayabileceği faydalar arasında maliyet tasarrufu, belge yönetimi, kişiselleştirilmiş tıp, sağlık okuryazarlığı ve iş akışının düzenlenmesi yer alabilirken, diş hekimliği ve ağız sağlığı hizmetlerinde, ChatGPT görüntü analizini, radyografilerin yorumlanmasını tanı koyabilmeyi, veri sentezini, daha iyi sonuçlar elde etmek için malzeme ve klinik teknikler hakkında bilgi sağlama, hasta kayıt yönetimi, adli diş hekimliği, ortodonti, periodontoloji ve endodonti alanlarındaki diğer uygulamaları, diş çürüğü tanısını, tedavi planlamasını ve hasta iletişimi sayılabilmektedir.³ YZ teknolojisi kullanılarak, klinik sorular kullanıcıların cep telefonlarında saniyeler içinde cevaplanabilir ve devam eden eğitim güncellemeleri sürekli hale getirilebilir.^{3,4} Bu da diş hekimlerinin klinik uzmanlığı ve hastanın tedavi ihtiyaçları ile tercihleriyle dikkatli bir şekilde entegre edildiğinde, yoğun çalışan klinik uz-

manların kanıta dayalı diş hekimliği yaklaşımını ağız sağlığı hizmetlerine uygularken karşılaştıkları zorlukları aşmalarına yardımcı olabilir.⁵

YZ destekli sohbet robotları kullanılarak yapılan doğru cevap oranlarının değerlendirmeleri, tıp doktorları, hemşireler ve eczacılar için ulusal sınavlarda raporlanmış ve bu robotların eğitim destek aracı olarak potansiyelini göstermiştir. ChatGPT-4o’nin, ChatGPT-3,5’e kıyasla istatistiksel olarak anlamlı bir şekilde daha fazla diş hekimliği bilgisine sahip olduğu bildirilmiştir.¹ ChatGPT-4o Plus ise abonelerine yazılı metin içeren soruların yanı sıra, görseller ve dosyalar yükleyerek soru sorabilme imkânı sunabilmektedir.

ChatGPT-4o ve diğer YZ destekli sohbet robotlarının performansı, tıp ve diş hekimliği alanlarında değerlendirilmiştir. Ancak, Türkiye’de diş hekimliğinde uzmanlık eğitimi giriş sınavında (DUS) sorulan çocuk diş hekimliği soruları henüz herhangi bir YZ destekli sohbet robotu tarafından değerlendirilmemiştir. Bu nedenle, bu çalışma, DUS’ta çıkmış olan çocuk diş hekimliği soruları kullanılarak YZ destekli sohbet robotlarının eğitim destek araçları olarak potansiyelini araştırmak amacıyla yapılmıştır. Çalışmanın 0 hipotezi ise sohbet robotlarının çocuk diş hekimliği sınav sorularını doğru cevaplamada bilgi düzeylerinin benzer olduğudur.

GEREÇ VE YÖNTEMLER

Bu çalışmada incelenen sorular kamuya açık bir platform olan Öğrenci Seçme ve Yerleştirme Merkezi’nin (ÖSYM) resmi sayfasından edinildiği için etik kurul onayına gerek duyulmamıştır. DUS, 40 adet temel bilimler sorusu ve 80 adet klinik bilimler sorusu olmak üzere toplam 120 sorudan oluşan merkezi bir sınavdır. Soruların tümü 5 şıktan oluşan çoktan seçmeli sorulardır. Klinik bilimler testinin 10 soruluk kısmı çocuk diş hekimliği sorularından oluşmaktadır. DUS, ÖSYM tarafından 2015-2022 arasında yılda 1 kez gerçekleşmiş olup; 2012-2014 arasında ve 2023’ten bu yana ilkbahar ve sonbahar dönemi sınavları olarak yılda 2 kez yapılmaktadır. Çalışmamızda, 2022’ye kadar gerçekleştirilmiş 13 adet sınava ait DUS soruları konularına göre sınıflandırılmıştır. 2022 yılından itibaren ÖSYM’nin DUS sorularını yayınlamaması sebebiyle 2022, 2023 ve 2024 yıllarına

ait DUS soruları çalışma dışı bırakılmıştır. Çocuk diş hekimliği, referans kitaplar baz alınarak belirlenmiş 30 konu başlığı ile çok geniş konu yelpazesine sahiptir. DUS'ta 7 adet konu başlığıyla alakalı hiç soru sorulmamıştır. Bu nedenle kalan 23 konu (benzer kategorilerde değerlendirilebilecekse) daha kapsamlı ana başlıklar olacak şekilde sadeleştirilerek, 13 başlık altında toplanmıştır. Konu başlıkları: diş embriyolojisi ve diş sürmesi; diş gelişim anomalileri; çürük, beslenme ve risk değerlendirmesi; çürüklü dişlerde pulpal problemler; florid; restoratif materyaller; okluzyon ve ortodontik apareyler; periodontal problemler ve sistemik hastalıklar; davranış yönlendirmesi (farmakolojik ve non-farmakolojik); lokal anestezi; dental travma ve tedavileri; pedodontide ilaç kullanımı; diş dokularının histolojik ve morfolojik özellikleri olarak 13 grupta sınıflandırılmıştır. 2013 ilkbahar ve sonbahar dönemlerinde 1'er soru ve 2021 yılında 1 soru olmak üzere toplam 3 soru iptal edildiği için 127 soru çalışmaya dâhil edilmiştir. YZ'nin görsel tanı koyabilme becerisi sebebiyle çalışma yalnızca metinsel değil görsel materyal içeren soruları da kapsamaktadır.

Sorulara ÖSYM'nin resmî internet sitesinden ulaşılmıştır. Bu çalışmada, 127 adet çoktan seçmeli soru, ChatGPT-4o Plus (ücretli) ve Gemini (ücretsiz) sohbet

robotlarına tek operatör tarafından, eş zamanlı olarak 3 Ocak 2025 tarihinde sorulmuştur. Sorular, konulara göre sınıflandırıldıktan sonra sırasıyla sohbet robotlarına sorulmuştur. Yazılımların verdiği cevaplar 0 ve 1 puan olacak şekilde değerlendirilmiştir: Cevap şıkkı hatalı ise "0" puan; cevap şıkkı doğru ise "1" puan verilmiştir. Elde edilen puanlar konulara ve sınav yıllarına göre ayrı ayrı hesaplanmıştır. Bu çalışmanın istatistiksel analizinde IBM SPSS 26 (IBM, SPSS, Chicago, IL, ABD) yazılım programı kullanılmıştır. Normal dağılım Shapiro Wilk testi ile sağlandığı için farklı konu başlıkları dağılımları göz önüne alınarak doğru bilinen soru sayıları arasında farklılık olup olmadığı ise bağımsız örneklem t-testiyle incelenmiştir. Anlamlı düzey $p < 0,05$ olarak belirlenmiştir.

BULGULAR

Her iki sohbet robotuna 127 adet çoktan seçmeli çocuk diş hekimliği sorusu sorulmuştur. Başarı kavramı sohbet robotlarının soruları doğru cevaplama yüzdeleri göz önüne alınarak kullanılmıştır. Sohbet robotlarının soruları cevaplama başarıları yüzde olarak ifade edilmiştir. Çalışmaya dâhil edilen soruların konu başlıklarına ve yıllara göre dağılımı Tablo 1'de, sınav yıllarına aldıkları toplam puanlar ise Şekil 1'de gösterilmektedir. Araştırmaya dâhil edilen sorular-

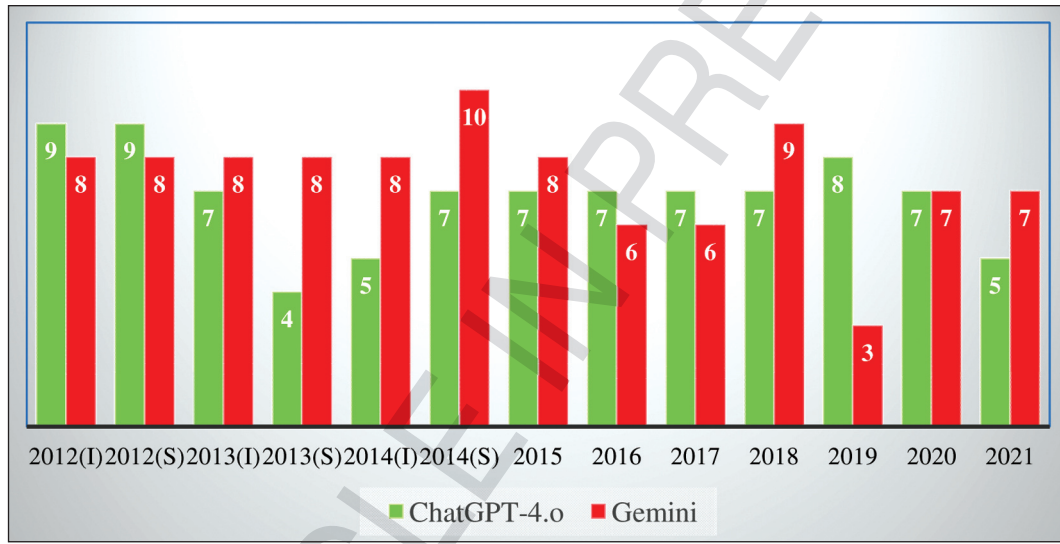
TABLO 1: Konu başlıklarına ait soru sayılarının sınav yıllarına göre dağılımı

Sınav yılı	1	2	3	4	5	6	7	8	9	10	11	12	13	Geçerli soru sayısı
2012(i)			1		2		1		2	2	1		1	10
2012(s)			1		2	2	1		1		2		1	10
2013(i)	1	1			1	1	1	1	2				1	9
2013(s)	1				1	2	1		1	2		1		9
2014(i)	2				1	1	1	2		1		2		10
2014(s)	1		1	1	1	1		1	1	1	1	1		10
2015	1			1	1		1	1	1	2	1	1		10
2016				1	1	1	3	1	1	1		1		10
2017	1				1	3	1	1	1			1	1	10
2018					1	1	2	2	2	1		1		10
2019	2					1	1	1	1	3	1			10
2020	2					2	1		2	1	1	1		10
2021	3					1	1	1	2	1				9
Toplam soru sayısı	14	1	3	3	12	16	15	11	17	15	7	9	4	127

1: Davranış yönlendirme (farmakolojik ve nonfarmakolojik); 2: Diş dokularının histolojik ve morfolojik özellikleri; 3: Pedodontide ilaç kullanımı; 4: Lokal anestezi; 5: Okluzyon ve ortodontik apareyler; 6: Periodontal problemler ve sistemik hastalıklar; 7: Restoratif materyaller; 8: Diş embriyolojisi ve diş sürmesi; 9: Dental travma ve tedavileri; 10: Diş gelişim anomalileri; 11: Çürük, beslenme ve risk değerlendirmesi; 12: Çürüklü dişlerde pulpal problemler; 13: Florid; i: ilkbahar dönemi; s: sonbahar dönemi

daki konu dağılımına bakıldığında en çok sorunun “dental travmalar” konu başlığından sorulduğu görülmüştür. Sohbet robotlarının başarı oranları konulara göre farklılık göstermektedir (Tablo 2). Sohbet robotlarının 1 soru için alabileceği en yüksek puan “1” olarak değerlendirildiği için, 127 soru için alınabilecek en yüksek puan 127 olarak hesaplanmıştır. Tüm sorular değerlendirildiğinde, ChatGPT-4o Plus toplam 89 puan alırken, Gemini’nin 96 puan aldığı görülmüştür. Sınav yıllarına göre sohbet robotları tarafından alınan puanların analizi Şekil 1’de gösterilmektedir. Araştırmaya dâhil edilen sorular içinde “diş

dokularının histolojik ve morfolojik özellikleri” başlığından sadece 1 adet soru sorulmuştur, bu soru her iki sohbet robotu tarafından doğru yanıtlandığı için başarı oranı %100 olarak ifade edilmiştir. Sohbet robotlarının soruları doğru yanıtlanma oranının en düşük olduğu konular “floridler” ve “lokal anestezi” başlıkları olmuştur. Tablo 2’de konulara göre sınavda çıkmış soru sayıları, sohbet robotlarının başarı oranları ve totalde elde edilmiş başarı oranları ifade edilmektedir. Konu başlıkları dikkate alınarak sohbet robotlarının başarı yüzdelerinin ortalama ve standart sapma (SS) değerleri ve t-test sonuçları Tablo 3’te



ŞEKİL 1: Sınav yıllarına göre ChatGPT-4o Plus ve Gemini'nin doğru cevapladığı sorulara aldıkları puanların gösterimi

TABLO 2: Konulara göre ve toplamda ChatGPT-4o Plus ve Gemini başarı yüzdeleri

	Çıkmış Soru Sayısı (n)	Doğru cevapların yüzdesi (%)	
		ChatGPT-4o Plus	Gemini
Davranış yönlendirme (farmakolojik ve nonfarmakolojik)	14	%92,85	%92,85
Diş dokularının histolojik ve morfolojik özellikleri	1	%100	%100
Pedodontide ilaç kullanımı	3	%67	%100
Lokal anestezi	3	%67	%33
Okluzyon ve ortodontik apareyler	12	%58,33	%58,33
Periodontal problemler ve sistemik hastalıklar	16	%81,25	%93,75
Restoratif materyaller	15	%67	%67
Diş embriyolojisi ve diş sürmesi	11	%63,63	%81,81
Dental travma ve tedavileri	17	%82,35	%76,47
Diş gelişim anomalileri	15	%60	%67
Çürük, beslenme ve risk değerlendirmesi	7	%67	%67
Çürüklü dişlerde pulpal problemler	9	%67	%78
Florid	4	%25	%50
Toplam	127	%70,7	%76

TABLO 3: Konulara göre doğru cevaplanan soruların SS değerleri ve bağımsız örneklem t test sonuçları

Yazılım	N	SS	p değeri
ChatGPT-4o Plus	13	1,46	0,397*
Gemini	13	1,71	

p<0,05; N: Konu sayısı; SS: Standart sapma

TABLO 4: Çalışmaya dâhil edilen sınav yıllarında ChatGPT-4o Plus ve Gemini tarafından alınan puanların, ortalama ve SS değerleri ve bağımsız örneklem t-testi sonuçları

Yazılım	Ortalama	N	SS	p değeri
ChatGPT-4o Plus	69,11	13	18,34	0,502*
Gemini	74,25	13	20,06	

p<0,05; N: Konu sayısı; SS: Standart sapma

gösterilmektedir. ChatGPT-4o Plus ve Gemini arasında farklı konu başlıkları dağılımları değerlendirilip doğru bilinen soru sayısı incelendiğinde istatistiksel açıdan anlamlı bir fark bulunmamıştır (p=0,397). Çalışmaya dâhil edilen sınav yıllarında ChatGPT-4o ve Gemini tarafından alınan puanların, ortalama ve SS değerleri Tablo 4'te gösterilmektedir. Araştırmaya dâhil edilen 13 sınava ait sorulara sohbet robotları tarafından verilen yanıtlar incelendiğinde, ChatGPT-4o Plus ve Gemini'nin aldıkları puanlar arasında anlamlı bir fark bulunmamıştır (p=0,502).

TARTIŞMA

Özellikle son yıllarda hızlı bir popülarite yakalayan YZ teknolojilerinin halen çocuk diş hekimliği alanında kullanımı sınırlıdır.⁶ Khanagar ve ark.nın yaptıkları bir çalışmada çocuk diş hekimliği alanındaki araştırmalar sırasıyla diş çürüğü tespiti, çocukların ağız sağlığı durumunun ve tedavi ihtiyaçlarının tahmini, kronolojik yaş tespiti, fissür örtücülerin tespiti ve sınıflandırılması, süt dişlerinde plak tespiti, dişlerin otomatik tespiti ve çizelgelenmesi gibi alanlara yönelmiştir.⁷ ChatGPT-4'ün ChatGPT-3,5'tan daha yüksek doğruluk oranları sergilediği araştırmalarla gösterilmiştir.^{1,8-10} Jin ve ark. sohbet robotlarını kullanarak tıp, eczacılık, diş hekimliği ve hemşirelik branşlarına ait 23 sınavın sonuçlarını derlemiştir. Bu çalışmanın sonuçlarına göre ChatGPT-3,5 için %36 ila %77 ve ChatGPT-4o için %64,4-100 ara-

sında değişen doğruluk oranları bildirmişlerdir.⁸ Yamaguchi ve ark.nın farklı sohbet robotlarının Japon diş hijyenistliği sınavındaki performanslarını araştırdıkları çalışmalarında, ChatGPT-4o'nun en yüksek doğruluğu (%75,3) elde ettiğini vurgulamışlardır. ChatGPT-3,5 için bu oranın %63 olduğu ve çoktan seçmeli sorularda ChatGPT-4o'nun genel olarak diğer modellerden daha iyi performans gösterdiğini bildirmişlerdir.¹ Uehara ve ark.nın Japon Ulusal Diş Hekimliği Sınavı sorularında sohbet robotlarının performansını araştırdıkları çalışmalarında doğruluk oranını ChatGPT-4o için %84,63, ChatGPT-3,5'in ise %45,46 olarak bulmuşlardır. ChatGPT-4o'nun, diş hekimliği konularında ChatGPT-3,5'i geride bıraktığını bildirmişlerdir.⁹

Dashti ve ark. sohbet robotlarının Amerika Birleşik Devletleri'nde diş hekimliği bilgisini ölçen 3 ayrı sınavdaki doğruluk oranlarını değerlendirmişlerdir. Teorik bilgi ve vaka raporu olarak ikiye ayırdıkları çoktan seçmeli sorularda ChatGPT-4o, ChatGPT-3,5'tan daha yüksek doğruluk oranları elde etmiştir.¹⁰ Aynı çalışma, klinik diş hekimliği sorularını cevaplama sohbet robotlarının doğruluk performansını değerlendiren Giannakopoulos ve ark. çalışmasıyla benzer sonuçlar vermiştir.³ Sohbet robotlarının doğruluk oranlarının kıyaslandığı başka bir çalışmada ChatGPT-3,5'in yanıltıcı ve yanlış cevap verme oranı yüksek olmasına karşılık ChatGPT-4o ve Gemini performansların birbirine yakın ve yüksek doğruluk sağladığı bildirilmiştir.¹¹ Çocuk diş hekimliği özelinde yapılmış olan benzer çalışmalar değerlendirildiğinde de sorulan soruları doğru cevaplayabilme açısından ChatGPT-4o'nun ChatGPT-3,5'e üstünlük sağladığı bildirilmiştir.^{12,13} Bu sonuç, yazılım sürümünün sohbet robotlarının doğruluk performansında etkili olduğunu ve YZ teknolojisinin geliştirilebilir ve iyileştirilebilir olduğunu göstermektedir. Çalışmamıza benzer bir araştırmada Jung ve ark. ChatGPT-3,5 (%35,3) ile Gemini'nin (%33,0) doğruluk performanslarını kıyaslanmış ve aralarında anlamlı fark bulunmadığını bildirmişlerdir.¹³ Çalışmamızda ChatGPT-4o Plus (%70,7) ile Gemini (%76) doğruluk performansı benzer yükseklikte bulunmuştur. Ayrıca elde edilen sonuçlara göre çalışmamızın 0 hipotezi geçerli kabul edilmiştir. Sohbet robotlarının farklı zamanlarda sorulan aynı soru-

lara farklı şekilde cevaplar verebildiği bilinmektedir. Bu durumu standardize etmek için araştırmamızda tüm sorular eş zamanlı sorulmuştur. Çalışmamızda tüm sorular sohbet robotlarına bir defa sorulmuş olup cevap tutarlılığı incelenmemiştir. Jung ve ark.nın yapmış oldukları aynı çalışmada sınav soruları her iki sohbet robotuna 7 kez sorulup cevap tutarlılığı da incelenmiştir.¹³ Bu çalışmanın sonuçlarına göre ChatGPT-3,5 iyi tutarlılık Gemini ise mükemmel tutarlılık göstermiştir. Çalışmamızda cevap tutarlılığının incelenmemiş olması ChatGPT 4o Plus ile Gemini'in benzer tutarlılık göstermesine neden olmuş olabilir. Jung ve ark. tarafından yapılan aynı araştırmada, sohbet robotlarının konu bazında farklı doğruluk oranları da araştırılmıştır.¹³ Buna göre en yüksek doğruluk oranı ChatGPT-3,5 için "pulpa tedavileri" ve "travma" konularında iken; Gemini için "diş gelişimi ve morfolojisi" ve "oklüzyon/ortodonti" konularında olmuştur.¹³ Çalışmamızda en yüksek doğruluk oranı ChatGPT-4o Plus için "diş dokularının histolojik ve morfolojik özellikleri" ile "davranış yönlendirmesi" iken Gemini için "diş dokularının histolojik ve morfolojik özellikleri" ile "pedodontide ilaç kullanımı" konularında olmuştur. ChatGPT-4o Plus'ın en başarısız olduğu konu %25 başarı oranı ile "floridler" olurken, Gemini'nin en başarısız olduğu konu %33 ile "lokal anestezi" başlığı olmuştur.

Sohbet robotlarının oluşturduğu cevaplar, sorulan soruya ve önceki konuşmalara göre değişebilmektedir.² Araştırmacılar bunu bir çeşit "halüsinasyon fenomeni" olarak adlandırmışlardır. Bir çalışmada bu fenomenin, sohbet robotlarının ikna edici, tekrarlayan ama yanlış cevaplar verebildiği üzerinde durulmuştur. Bu durum sağlık alanında YZ'nin yanlış tıbbi tavsiyeler verebilme riski konusunda endişe yaratmaktadır.¹³ Ayrıca soru formatının değişmesi ve soruların farklı dilde sorulması durumunda verilebilecek cevapların değişkenlik gösterip göstermeyeceği de araştırılmalıdır. Sohbet robotlarının eğitim ve klinik destek araçları olarak etkin şekilde kullanılabilmesi, yanlış teşhis ve yanıtların önlenmesi için tıbbi terminolojinin doğru bir şekilde çevrilmesi çok önemlidir.¹⁴ Bu çalışmada, orijinal sınav sorularının anlamını ve bağlamını korumak amacıyla soru-

lar Türkçe olarak sorulmuştur. Jung ve ark. yaptıkları çalışmanın sonuçlarıyla kıyaslandığında, farklı konularda farklı doğruluk yüzdelerinin gözlenmesine soruların formatı ve Türkçe olarak sorulmasının da etken olabileceğini düşünmekteyiz. Sohbet robotlarının verdiği cevapların tutarlılığının araştırılmaması, soru formatları değiştirilerek yapılan yanıt modellerinin analiz edilmemesi ve Gemini Advanced gibi daha üst sürümün kullanılmamış olması çalışmanın sınırlılıklarındandır.

Sohbet robotları, sağlık alanında uzman seviyesinde bilgiye sahip olamama, doğrulanmamış kaynaklar, referans güvenilirliği gibi sınırlamalara sahiptir.² Diş hekimliğinin tüm branşlarında bilgi edinmek için daha güvenilir kaynaklara başvurma önemi unutulmamalıdır. Çalışmamız çocuk diş hekimliği alanında sorulmuş sınav soruları kullanılarak sohbet robotlarının doğruluk oranlarını ölçmeyi hedefleyen ilk araştırmalardan birisidir. Ülkemizde en sık kullanılan sohbet robotlarının son sürümleri seçilerek sonraki araştırmalara referans olabilmesi amaçlanmıştır. Sohbet robotlarının özellikle cevap tutarlılıklarının araştırılması için daha fazla sayıda çalışmaya ihtiyaç vardır.

SONUÇ

Bu çalışma, YZ'nin çocuk diş hekimliğini ilgilendiren konularda ne kadar sağlıklı değerlendirmeler yapabildiğini araştırmayı amaçlamıştır. Hem ChatGPT-4o Plus hem de Gemini, çocuk diş hekimliği sınav sorularını doğru yanıtlama konusunda genel olarak iyi başarı gösterse de, halen bir takım eksiklikleri bulunmaktadır. Çocukların ağız diş sağlığı durumunu ve tedavi ihtiyaçlarını tespit edebilecek algoritmaların geliştirilmesi ile YZ'nin hem bireysel hem de toplum düzeyinde önleyici stratejilerin geliştirilmesine yardımcı olması hedeflenmelidir. Çocuk diş hekimleri, bu yeni teknolojinin avantajlarının ve sınırlamalarının farkında olmalı ve YZ'yi hasta yararını gözeterek dikkatli bir şekilde kullanmalıdır.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya

herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Çıkar Çatışması

Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite

üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.

Yazar Katkıları

Bu çalışma tamamen yazarın kendi eseri olup başka hiçbir yazar katkısı alınmamıştır.

KAYNAKLAR

1. Yamaguchi S, Morishita M, Fukuda H, Muraoka K, Nakamura T, Yoshioka I, et al. Evaluating the efficacy of leading large language models in the Japanese national dental hygienist examination: a comparative analysis of ChatGPT, Bard, and Bing Chat. J Dent Sci. 2024;19(4):2262-7. PMID: 39347065; PMCID: PMC11437298.
2. Bilgin Avcı D, Ertan AA. Diş hekimliğinde uzmanlık sınavında sorulan protetik diş tedavisi sorularının ChatGPT-3.5 ve Gemini tarafından cevaplanma performanslarının karşılaştırmalı olarak incelenmesi: kesitsel araştırma [A comparative study of ChatGPT-3.5 and Gemini's performance of answering the prosthetic dentistry questions in dentistry specialty exam: cross-sectional study]. Türkiye Klinikleri J Dental Sci. 2024;30(4):668-73. doi:10.5336/dentalsci.2024-104610
3. Giannakopoulos K, Kavaddella A, Aaqel Salim A, Stamatopoulos V, Kaklamano EG. Evaluation of the performance of generative ai large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. J Med Internet Res. 2023;25:e51580. PMID: 38009003; PMCID: PMC10784979.
4. Schwendicke F, Blatz M, Uribe SE, Cheung W, Verma M, Linton J, et al. Artificial Intelligence for Dentistry: FDI Artificial Intelligence Working Group. Geneva: FDI World Dental Federation (FDI); 2023. p.1-20. https://www.fdi-worlddental.org/sites/default/files/2023-01/FDI%20ARTIFICIAL%20INTELLIGENCE%20WORKING%20GROUP%20WHITE%20PAPER_0.pdf
5. FDI World Dental Federation [Internet]. Evidence-based dentistry (EBD). ©2016 FDI [Cited: January 12, 2025]. Available from: <https://www.fdiworld-dental.org/evidence-based-dentistry-ebd>
6. Yüce F, Taşşöker M. Diş hekimliğinde yapay zeka uygulamaları [The applications of artificial intelligence in dentistry]. Yeditepe J Dent. 2023;19(2):141-9. doi:10.5505/yeditepe.2023.05668
7. Khanagar SB, Alfouzan K, Alkadi L, Albalawi F, Iyer K, Awawdeh M. Performance of artificial intelligence (ai) models designed for application in pediatric dentistry-a systematic review. Appl Sci. 2022;12(19):9819. <https://doi.org/10.3390/app12199819>
8. Jin HK, Lee HE, Kim E. Performance of ChatGPT-3.5 and GPT-4 in national licensing examinations for medicine, pharmacy, dentistry, and nursing: a systematic review and meta-analysis. BMC Med Educ. 2024;24(1):1013. PMID: 39285377; PMCID: PMC11406751.
9. Uehara O, Morikawa T, Harada F, Sugiyama N, Matsuki Y, Hiraki D, et al. Performance of ChatGPT-3.5 and ChatGPT-4o in the Japanese national dental examination. J Dent Educ. 2025;89(4):459-66. PMID: 39538434.
10. Dashti M, Ghasemi S, Ghadimi N, Hefzi D, Karimian A, Zare N, et al. Performance of ChatGPT 3.5 and 4 on U.S. dental examinations: the INBDE, ADAT, and DAT. Imaging Sci Dent. 2024;54(3):271-5. PMID: 39371301; PMCID: PMC11450412.
11. Guven Y, Ozdemir OT, Kavan MY. Performance of artificial intelligence chatbots in responding to patient queries related to traumatic dental injuries: a comparative study. Dent Traumatol. 2025;41(3):338-47. PMID: 39578674.
12. 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. PMID: 38499280.
13. Jung YS, Chae YK, Kim MS, Lee HS, Choi SC, Nam OH. Evaluating the accuracy of artificial intelligence-based chatbots on pediatric dentistry questions in the Korean national dental board exam. J Korean Acad Pediatr Dent. 2024;51(3):299-309. <https://doi.org/10.5933/JKAPD.2024.51.3.299>
14. Liu Z, Zhang L, Wu Z, Yu X, Cao C, Dai H, et al. Surviving ChatGPT in healthcare. Front Radiol. 2024;3:1224682. PMID: 38464946; PMCID: PMC10920216.