

ÖZGÜN ARAŞTIRMA

Ortodontik Tedavi Sürecinde Karşılaşılan Sorunlarının Çözümünde Yapay Zekâ Tabanlı Büyük Dil Modellerinin Etkinliği

Effectiveness of AI-Based Large Language Models in Addressing Patient Problems During Orthodontic Treatment

Dr. Öğr. Üyesi Ahmet YILDIRIM

Zonguldak Bülent Ecevit Üniversitesi,
Diş Hekimliği Fakültesi, Ortodonti Anabilim Dalı,
Zonguldak

ORCID ID: 0009-0005-6804-1276

Dr. Öğr. Üyesi Aybüke Asena ATASEVER İŞLER

Bolu Abant İzzet Baysal Üniversitesi,
Diş Hekimliği Fakültesi, Ortodonti Anabilim Dalı, Bolu

ORCID ID: 0000-0003-0738-6797

Dr. Öğr. Üyesi Yavuz Selim GENÇ

Giresun Üniversitesi, Diş Hekimliği Fakültesi,
Ortodonti Anabilim Dalı, Giresun

ORCID ID: 0000-0003-0556-2830

Dr. Öğr. Üyesi Ruşen ERDEM

Kafkas Üniversitesi, Diş Hekimliği Fakültesi,
Ortodonti Anabilim Dalı, Kars

ORCID ID: 0000-0002-5298-7949

Dr. Öğr. Üyesi Şule USANMAZ

Kafkas Üniversitesi, Diş Hekimliği Fakültesi,
Ortodonti Anabilim Dalı, Kars

ORCID ID: 0000-0001-6857-0382

Dr. Vildan UKA

Ortodonti Uzmanı, Bağımsız Araştırmacı, Adana

ORCID ID: 0000-0003-3506-4787

Geliş tarihi: 18.03.2025

Kabul tarihi: 17.07.2025

doi: 10.5505/yeditepe.2026.78309

Yazışma adresi:

Dr. Öğr. Üyesi Ahmet YILDIRIM

Adres: Zonguldak Bülent Ecevit Üniversitesi, Diş Hekimliği Fakültesi, Ortodonti Anabilim Dalı, Kozlu, 67000, Zonguldak, Türkiye

Tel: 0507 361 44 48

E-posta: drahmetyyildirim@gmail.com

ÖZET

Amaç: Bu çalışma, ortodontik tedavi gören hastaların acil durumlarda ortodontistlerine ulaşamadığı senaryoları simüle ederek, yapay zeka modellerinin bilgi sağlama yetkinliğini incelemeyi amaçlamaktadır.

Gereç ve Yöntem: Ortodontik tedavi gören hastaların acil durumlarla ilgili olarak en sık sorduğu sorular, ChatGPT-3.5, ChatGPT-4o, Copilot, Gemini, Claude 3.7 Sonnet ve Medisearch yapay zeka platformlarına tek bir araştırmacı tarafından yöneltilmiştir. Elde edilen yanıtlar anonimleştirilmiş ve iki bağımsız değerlendirici tarafından GQS indeksi ile körleme yöntemi kullanılarak skorlanmıştır. Ayrıca, yanıtların bilimsel doğruluk ve alaka düzeyi nitel analiz yöntemiyle değerlendirilmiş, intra-değerlendirici güvenilirliği dört hafta sonra tekrar puanlama yapılarak test edilmiştir.

Bulgular: Kruskal-Wallis testi, modeller arasında anlamlı fark olduğunu göstermiştir ($H=37,819$; $p<0,0001$). En yüksek GQS ortalaması ChatGPT-4o'ya ($3,803\pm 0,4008$) aitken, en düşük GQS ortalaması ise Medisearch'e ($3,2576\pm 0,7084$) ait bulunmuştur. ChatGP-4o, Copilot ve ChatGPT-3.5'in yanıt kalitesi Medisearch'ten istatistiksel anlamda daha yüksek skorlar almış, Gemini ise Copilot ve ChatGPT-4o'ya kıyasla istatistiksel anlamda daha düşük skorlar almıştır ($p<0,05$).

Sonuçlar: Çalışmamızda ortodonti kapsamlı hasta sorularına yapay zeka tabanlı sohbet robotlarının verdikleri yanıtlar değerlendirilmiş ve olumlu sonuçlar gösterdikleri tespit edilmiştir. Hastaların hekimlerine ulaşamadıkları durumlarda tedavilerine ilişkin acil durum sorularına sohbet modellerinin yüksek kalitede yanıtlar üretebilmesi, hasta-hekim ilişkisini geliştirebilecek ve hastaların bilinçlenerek tedaviye daha uyumlu olmalarını sağlayabilecektir.

Anahtar Kelimeler: Yapay zeka, büyük dil modelleri, ortodonti, ChatGPT

ABSTRACT

Aim: This study aims to evaluate the information-providing competency of artificial intelligence models by simulating scenarios in which patients undergoing orthodontic treatment are unable to reach their orthodontists during urgent situations.

Materials and Method: The most frequently asked questions by patients undergoing orthodontic treatment regarding urgent situations were submitted by a single researcher to various artificial intelligence platforms, including ChatGPT-3.5, ChatGPT-4o, Copilot, Gemini, Claude 3.7 Sonnet, and Medisearch. The responses were anonymized and independently evaluated by two blinded raters using the Global Quality Scale (GQS). In addition, the scientific accuracy and relevance of

the responses were assessed through qualitative analysis, and intra-rater reliability was tested via repeated scoring after a four-week interval.

Results: The Kruskal-Wallis test revealed a statistically significant difference among the models ($H=37,819$; $p<0,0001$). The highest mean GQS score was observed for ChatGPT-4o ($3,803\pm 0,4008$), while the lowest was recorded for Medisearch ($3,2576\pm 0,7084$). The response quality of ChatGPT-4o, Copilot, and ChatGPT-3.5 was significantly higher than that of Medisearch, whereas Gemini received significantly lower scores compared to Copilot and ChatGPT-4o ($p<0,05$).

Conclusions: In our study, the responses provided by artificial intelligence-based chatbot models to orthodontics-related patient questions were evaluated, and it was found that they demonstrated positive performance. The ability of chatbot models to generate high-quality answers to urgent treatment-related questions in situations where patients cannot reach their orthodontists may strengthen the patient-clinician relationship and help patients become more informed and compliant with their treatment.

Keywords: Artificial intelligence, large language models, orthodontics, ChatGPT

GİRİŞ

Yapay zeka teknolojileri ve bunların uygulamaları son birkaç yıl içinde artış göstermiştir. Yapay zeka teknolojilerinin büyük bir kısmını, makine öğrenimi ve büyük dil modelleri oluşturmaktadır ve bu teknolojiler, diş hekimliği de dahil olmak üzere çeşitli alanlarda kullanılmaktadır.¹ Günümüzde çeşitli büyük dil modelleri (Large Language Model) (LLM) ortaya çıkmıştır. Ücretsiz erişilebilen ChatGPT (OpenAI Global, San Francisco, California, U.S) sürümü, GPT-3.5 modeline dayalıdır. Daha yeni olan GPT-4 sürümü ise ChatGPT-4o olarak sunulmaktadır. Şubat 2023'te Microsoft (Microsoft Corporation, Redmond, WA, U.S), GPT-4 modelini kullanan ve günümüzde Copilot olarak lanse edilen Bing Chat yapay zeka modelini tanıtmıştır. Mart 2023'te ise Google (Google Ireland Limited, Dublin, Ireland), başlangıçta LaMDA (kendi LLM ailesi) ile çalışan ve daha sonra PaLM 2 LLM tarafından desteklenen ve şu anda Gemini olarak adlandırıldıkları Bard modelini piyasaya sürmüştür.^{2,3} Anthropic ise Claude 3.7 Sonnet (Anthropic, USA) modelini geliştirmiştir. MediSearch (Weierstrass Limit Inc, 2023) ise özel olarak sağlık alanı için geliştirilmiştir. LLM programları, yapay zeka destekli sohbet botu (chatbot) modelleri gibi, akıllı insan-bilgisayar etkileşimlerine dayanmaktadır. Bu sohbet botu modelleri, internet üzerinden insan kullanıcılarla yapılan konuşmaları simüle etmek amacıyla tasarlanmıştır.³ Bu modeller, akıcı ve tutar-

lı metinler oluşturabilmekte, diyaloglarda soruları yanıtlayabilmekte, dilleri çevirebilmekte ve dil ile ilgili birçok başka görevi yerine getirebilmektedir.¹ Genel olarak LLM'ler, internet üzerindeki geniş metin veri setleri (Wikipedia, dijital kitaplar, makaleler gibi) üzerinde eğitilmiş sinir ağlarıdır. Amaçları, verilen bir girdiye dayalı olarak tutarlı ve insan benzeri yanıtlar üretmektir. Bu süreç, derin öğrenme algoritmaları ve gelişmiş modelleme yöntemleriyle gerçekleştirilir.^{4,5}

Hastaların tedaviler hakkında bilgilendirilmesi, tedavi sürecini kolaylaştırmalarına yardımcı olabilir. Ancak, hastaların endişeleri hakkında bilgi almak için her zaman doktorlarına veya sağlık personeline ulaşmaları mümkün olmayabilir. Yapay zeka tabanlı LLM'ler, hastaların ilgilendikleri konularda bilgi alabilmeleri için büyük bir potansiyele sahiptir.⁶ Bu teknolojiler yalnızca hastalar ve genel halk için değil, aynı zamanda diş hekimleri ve öğrenciler için de faydalı araçlar olarak kullanılmaktadır.¹ Ortodontistlerin bilgi sağlamak ve hastaların destek aramak için internet tabanlı sosyal medya sitelerini kullanabilecekleri bir fırsat olduğunu göstermektedir. Hastalar, genellikle, aparey kullanımına uyumlarını artırmak için kullanılabilecek bir uygulamanın geliştirilmesine ilgi duyduklarını ifade etmişlerdir.⁷ Bununla birlikte, LLM'lerin tıbbın olumlu ve olumsuz yönlerini tam anlamıyla ele almaması, yanlış cevaplar verebilmesi, anlamsız içerikler üretmesi ve yanlış bilgileri ya da dezenformasyonu gerçekmiş gibi sunabilmesi gibi durumlar, sağlık gibi kritik alanlarda ciddi endişeler yaratmaktadır.¹

Web tabanlı bilginin güvenilirliği ve kalitesi büyük önem taşımaktadır, çünkü bu durum hastaların tedavi sürecindeki iş birliğini, uyumlu ilerlemeyi, doktorlarıyla olan iletişimlerini ve hekim-hasta güvenini etkileyebilir.⁸ Milenyum kuşağının %90'ı aktif sosyal medya kullanıcısıdır ve bu grup günümüzde ortodonti hasta kitlesinin çoğunluğunu oluşturmaktadır.⁹ Bu nedenle, bu kuşağın ortodontik tedavi sürecinde karşılaşılabilecekleri acil durumlar hakkında bilgi arayışlarını karşılamak üzere LLM'lerin sunduğu cevapların incelenmesi, hem hasta deneyimi hem de dijital sağlık hizmetlerinin geliştirilmesi açısından önem taşımaktadır. Bu çalışma, ortodontik tedavi gören hastaların ve hasta velilerinin acil durumlarda ortodontistlerine ulaşamadığı senaryoları simüle ederek, yapay zeka modellerinin hızlı, talep üzerine bilgi verme ve açıklama yapma kapasitelerini değerlendirmeyi amaçlamaktadır. Ayrıca, etkileşimlerin tek sorularla sınırlı tutulması, yapay zeka modellerinin karmaşık sorulara kısa ve ilgili yanıtlar verme yeteneğini daha odaklanmış bir şekilde incelemeye olanak sağlamıştır.

GEREÇ VE YÖNTEM

Çalışmada insan veya hayvan materyalleri kullanılmadığı için etik onay gerekli olmamıştır. Ortodontik tedavi gör-

mekte olan hastaların acil durumlarla ilgili en sık sorduğu soruların belirlenmesi amacıyla, alanında en az 5 yıl deneyimli dört ortodontist, kendilerine en çok yöneltilen 25 soruyu listelemiştir. Toplamda 100 soru elde edilmiş, ortak ve benzer sorular elimine edilerek 66 soruya indirgenmiştir (Tablo 1). Ortodontik tedavi gören hastaların acil durumlarda sıkça sorduğu sorular, tek bir araştırmacı tarafından 2024 yılı aralık ayı içinde ChatGPT-3.5, ChatGPT-4o, Copilot, Gemini, Claude 3.7 Sonnet ve Medisearch yapay zeka platformlarına iletilmiştir. Tüm sorular aynı dizüstü bilgisayar (MacBook Air M3, 16GB RAM; Apple, Cupertino, Calif), 4,5G internet bağlantısı ve sanal özel ağ (Virtual Private Network) (VPN) sunucusu (sürüm 3,9; Astrill Systems Corp, Santa Clara, Calif) kullanılarak sorulmuştur. Sorular uygun terminoloji ile yazılmış olup, metin tabanlı yanıt gerektiren açık uçlu sorulardan oluşmuştur.¹¹ Her bir soru, her bir yapay zeka modeline yalnızca bir kez sorulmuş ve yanıtların yeniden ifade edilmesine veya modelin cevaplayamaması durumunda ek açıklama yapılmamıştır. Ayrıca, başka bir araştırmacı tarafından sorular tekrar sorulmamıştır. Sorgulamanın başarılı bir şekilde tamamlanmasının ardından, her yapay zeka modelinden elde edilen yanıtlar toplanmış ve elektronik randomizasyon aracı (<https://www.random.org/lists/>) kullanılarak rastgele dağıtılmıştır. Her modelden alınan yanıtlar, değerlendiricilerin hangi yapay zeka modeline ait yanıtı değerlendirdiklerini bilmemeleri için, her bir yapay zeka modeline bir harf atanarak (A, B, C, D, E, F) körlleme yöntemi ile sunulmuştur. Rastgele dağıtılmış bu 396 yanıt, değerlendirilmek üzere iki bağımsız değerlendiriciye gönderilmiştir. GQS (Global Kalite İndeksi) aracı ile yanıtların kalitesini değerlendirmek amacıyla kullanılmış,^{6,11,12} ayrıca yanıtlar nitel analiz ile değerlendirilmiştir.¹³ GQS indeksinde yanıtlar beş ana madde üzerinden değerlendirilmektedir.¹² Çalışmamızda sorulara verilen yanıtların daha etkili bir şekilde değerlendirilmesi amacıyla modifiye GQS indeksi⁶ kullanılmıştır (Tablo 2). Her madde, 1 ile 5 arasında puanlanmış ve maksimum 5 puanlık GQS skoru elde edilmiştir. Yanıtları değerlendiren araştırmacılar, her soru için "altın standart"² belirleyerek bu yanıtı 5/5 puan olarak kabul edilmiştir. Altın standart cevaplar soru havuzunu oluşturan dört ortodontist tarafından ilgili literatür, kitabi bilgiler ve klinik tecrübeleri göz önünde bulundurularak oluşturulmuştur. İlk değerlendirmeden dört hafta sonra, intra-değerlendirici güvenilirliği değerlendirmek amacıyla yanıtlar tekrar puanlanmıştır.

Tablo 1. Ortodontik tedavi sürecine ilişkin acil durum soruları

Sorular
1 Braketim koptu, ne yapabilirim?
2 Braketimi ve kovan telimi yuttum ne yapmalıyım?
3 Ortodontik tedavi görüyorum tel yanağıma batıyor, ne yapmalıyım?
4 Çene genişletme aparatı başımı çok ağrıtıyor, çevirmeyi bırakmalı mıyım?
5 Ortodontik tedavi sırasında dişlerimin çok sallandığını fark ettim, bu normal mi?
6 Ortodontik tedavi sonrası doktorumun verdiği retansiyon plağını kırdım, ne yapmalıyım?
7 Ortognatik cerrahi ameliyatı geçirdim, ne zaman yemek yiyebilirim?
8 Ortodontik şeffaf plak tedavisi görüyorum ve plağımın biri kayboldu, bir sonraki plağa geçmeli miyim?
9 Vidalı hareketli ortodontik aparatım ağzıma oturmuyor, ne yapmalıyım?
10 Ortodontik mini vidam sallanıyor, ne yapmalıyım?
11 Ortodontik tedavi sırasında diş etlerim şiş ve kanıyor ne yapmalıyım?
12 Dişlerimde çok fazla ağrı oluyor, geceleri retansiyon plaklarımı kullanmasam olur mu?
13 Tatile çıkıyorum, retansiyon plaklarımı kullanmasam olur mu?
14 Ağız içi lastiklerimi çiftler çiftler taksam tedavim hızlanır mı?
15 Şeffaf plaklarım ağzımda iken sıcak içecekler içebilir miyim?
16 Şeffaf plaklarım takıldıktan sonra konuşmamı değiştirdi, ne yapmalıyım?
17 Ağız içi elastiklerim çok ağrı yapıyor, sadece geceleri takabilir miyim?
18 Çene genişletme aparatım kaydı ama yerinden çıkmadı, doktoruma haber vermeme gerek var mı?
19 Ortodontik tedavi görüyorum, tel takıldıktan sonra ağzımda yara oluştu, ne yapmalıyım?
20 Ortodontik tedavi görüyorum, tel takıldıktan sonra dişlerim sızlıyor, ne yapmalıyım?
21 Diş tellerim baş ağrısına neden oldu, ağrı kesici kullanmam olur mu?
22 Ortognatik cerrahi oldum, yüzümde oluşan şişlikler ne zaman geçer? Yüzümdeki şişlerin geçmesi için ne yapabilirim?
23 Çene genişletme aparatım ağzımdan düştü, evde nasıl yapıştırabilirim?
24 Doktorumun üst dişlerimin alttaki braketlere çarpmaması için yaptığı yükseltici dolgu düştü, problem olur mu?
25 Şeffaf plak atışmanlarımdan 2 tanesi koptu, şeffaf plakları takmaya devam mı etmeliyim?
26 İnvizalığın şeffaf plaklarım ağızda koku yapıyor, bunu düzeltmek için nasıl bir ölçüm alabilirim?
27 İnvizalığın şeffaf plaklarımı taktım ancak ağzımdan çıkaramıyorum, ne yapmam gerekir?
28 Tellerim takıldıktan sonra dişlerimde ağrı oldu, hangi ağrı kesicileri içmeliyim?
29 Çeneler arası kullandığım paket lastiğim bitti, bunun yerine başka bir paket lastik kullanırsam problem olur mu?
30 Ağzımda sadece retainer tellerim var, pekiştirme plağımı kullanmıyorum, dişlerim eskiye döner mi?
31 Gömülü dişimin üzerine yapıştırılan butonun teli koptu, yeniden ameliyat olmam gerekir mi?
32 Ortodontik yüz maskesi kullanıyorum, çenemde yaralar oluştu, ne yapmalıyım?
33 Ortognatik cerrahi oldum, ağzımı açamıyorum, bu normal mi? Ağzımın daha fazla açılması için ne yapabilirim?
34 Ortognatik cerrahi oldum, dişlerim üst üste kapanmıyor, tamamen kapanması için ne yapabilirim?
35 Ortodontik tedavimde yapılan vidanın etrafında kızamıklık oldu ve ilgili bölgede şişlik ve ağrı var, doktoruma ulaşmıyorum, antibiyotik içeyim mi?
36 Bugün ortodontik tedaviye başladım ve metal braketler takıldı, ancak ağzımın etrafında kızamıklıklar oldu, bu neden olmuş olabilir ve ne yapmalıyım?
37 Dudak damak yarığına sahip bebeğimin tedavisi için yapılan aparatımdan dolayı bebeğimin ağzında yara oluştu, ne yapmalıyım?
38 Dudak damak yarığına sahip bebeğimin tedavisi için yapılan aparatımdan dolayı bebeğimin burnunda yara oluştu, ne yapmalıyım?
39 Dudak damak yarığına sahip bebeğimin tedavisi için yapılan aparatımdan dolayı bebek sürekli ağlıyor, ne yapmalıyım?
40 Dudak damak yarığına sahip bebeğimin tedavisi için yapılan aparatımdan dolayı bebeğimin yanakları tahriş oldu, ne yapmalıyım?
41 Ortodontik hareketli aparatımı kırdım, ne yapmalıyım?
42 Ortodontistimin bağladığı uzun tel çıktı, ne yapmalıyım?
43 Ortodontistimin bağladığı lastik çıktı, ne yapmalıyım?
44 Ortodontistimin bağladığı yay çıktı, ne yapmalıyım?
45 Ortodontik tedavi için sabit fonksiyonel aparatı takıldı, ama eve gidince yerinden çıktı, ne yapmalıyım?
46 Ortodontik tedavi için çocuğuma aktivatör aparatı yapıldı, ama geceleri çocuğumun ağzından çıkıyor, ne yapmalıyım?
47 Çocuğuma çene ilerletmesi için aparatı takıldı, ama geceleri ağzından çıkıyor, sadece gündüz takabilir miyiz?
48 Ortodontik tedavi sonrası takılan retansiyon plağım kırdım, ne yapmalıyım?
49 Hareketli aparatım ağzımı kesiyor, ne yapmalıyım?
50 Ortognatik cerrahi ameliyatı sonrası atılan dikişlerden birkaçı koptu, ama kanamam yok, doktoruma gitmeli miyim?
51 Çocuğuma ortodontik yüz maskesi takıldı, öndeki parçası oynuyor, ben evde sıkıştırabilir miyim?
52 Doktorum bebeğime yaptığı dudak-damak aparatının burna uzanan teli eğildi, evde kendimiz düzeltebilir miyiz?
53 Şeffaf plağımın ucu batıyor, bıçakla kesip düzeltebilir miyim?
54 Ağız içi lastiklerim doktorumun belirttiği şekilde taktığımda çok ağrı yapıyor, başka şekilde takabilir miyim?
55 Ortodontik mini vida dudagımda yaraya neden oldu, ne yapmalıyım?
56 Tel takıldıktan sonra dudagımda uçuk çıktı, ne yapmalıyım?
57 Tel takıldıktan sonra dudagımda aft çıktı, ne yapmalıyım?
58 Hareketli aparat sürekli ağzımdan düşüyor, ne yapmalıyım?
59 Tel takıldıktan sonra dişlerimde aralıklar oluşmaya başladı ve endişeleniyorum, bu normal mi?
60 Teller takılıyken dişlerimi fırçalayamıyorum, ne yapmalıyım?
61 Braketlerin etrafında dişlerim çürümeye başladı ve endişeleniyorum, ne yapmalıyım?
62 Teller takıldıktan sonra dişlerimde renklemeler oluştu, bu durum düzelir mi?
63 Ortodontik tedavi görüyorum, takılan tel koptu, ne yapmalıyım?
64 Ağzımda bulunan ortodontik distalizasyon aparatı dilimde kesikler oluşturdu, gargara yapsam geçer mi?
65 Hareketli aparatımı kırdım, tedavim geriler mi?
66 Dudak damak yarıklı 2 aylık bir bebeğim var, beslenme plağımı kaybettik, doktorumuza da çok uzadı ne yapabiliriz?

Değerlendiriciler, LLM yanıtlarının bilimsel doğruluğu ve alaka düzeyine ilişkin nitel bir değerlendirmeyi serbest metin biçiminde sunmuştur. Özellikle, seçtikleri LLM yanıtlarını "altın standart" ile eleştirel biçimde karşılaştırarak açıklayıcı yorumlar yapmaları istenmiştir. Bu yorumlarda, yanıtların olumlu yönleri olarak ayrıntılı, doğru ve iyi ifade edilmiş olması ile konuyu yeterince ele alması gibi özel-

likler belirtilebilmektedir. Ayrıca olumsuz yönler olarak da yanıtların yanlış, belirsiz ya da eksik olması ve "altın standart" ile uyuşmaması nedeniyle, kanıta dayalı uygulamalar için ilgili ve bilimsel açıdan doğru rehberlik sağlayamaması gibi noktalar vurgulanabilmektedir.¹

Tablo 2. Orijinal ve modifiye global kalite ölçekleri

Orijinal global kalite ölçüğü (GQS)	
Skor 1	Düşük kalite, sitenin akışı zayıf, bilgilerin çoğu eksik, hastalar için hiç faydalı değil.
Skor 2	Genellikle düşük kalite ve zayıf akış, bazı bilgiler listelenmiş ancak birçok önemli konu eksik, hastalar için çok sınırlı fayda sağlıyor.
Skor 3	Orta kalite, yetersiz akış, bazı önemli bilgiler yeterince ele alınmış ancak diğerleri zayıf bir şekilde tartışılmış, hastalar için kısmen faydalı.
Skor 4	İyi kalite ve genel olarak iyi akış, ilgili bilgilerin çoğu listelenmiş ancak bazı konular eksik, hastalar için faydalı.
Skor 5	Mükemmel kalite ve mükemmel akış, hastalar için çok faydalı.
Modifiye global kalite ölçüğü (MGQS)	
Skor 1	Düşük kalite, bilgilerin akışı zayıf, çoğu bilgi eksik, hastalar veya eğitim için hiç faydalı değil.
Skor 2	Genellikle düşük kalite ve zayıf akış, bazı bilgiler listelenmiş ancak birçok önemli konu eksik, hastalar veya eğitim için çok sınırlı fayda sağlıyor.
Skor 3	Orta kalite, yetersiz akış, bazı önemli bilgiler yeterince ele alınmış ancak diğerleri zayıf bir şekilde tartışılmış, hastalar veya eğitim için kısmen faydalı.
Skor 4	İyi kalite ve genel olarak iyi akış. İlgili bilgilerin çoğu listelenmiş ancak bazı konular eksik. Hastalar veya eğitim için faydalı.
Skor 5	Mükemmel kalite ve akış, hastalar veya eğitim için çok faydalı.

İstatistiksel Analiz

İstatistiksel analizler IBM SPSS (Statistical Package For Social Sciences, Chicago, IL, USA), Versiyon 26.0, yazılımı kullanılarak gerçekleştirilmiş, $p < 0,05$ değeri istatistiksel anlamlılık sınırı olarak kabul edilmiştir. İki bağımsız değerlendirici arasındaki tutarlılığı değerlendirmek için iki yönlü rastgele etkiler modeli kullanılarak Intraclass Correlation Coefficient (ICC) analizi uygulanmıştır. Değerlendirici içi güvenilirliği test etmek amacıyla, aynı değerlendirici tarafından dört hafta arayla yapılan tekrar puanlamalarla ICC değeri hesaplanmıştır. Verilerin normal dağılıma uygunluğu Kolmogorov-Smirnov testi ile değerlendirilmiş, verilerin normal dağılım göstermemesi nedeniyle gruplar arası karşılaştırmalarda Kruskal-Wallis H testi kullanılmıştır. Anlamlı fark saptanan durumlarda, ikili grup karşılaştırmaları Bonferroni düzeltmeli post-hoc test ile yapılmıştır.

BULGULAR

İki bağımsız değerlendirici arasındaki güvenilirliği değerlendirmek için Intraclass Correlation Coefficient (ICC) analizi uygulanmış ve ICC değerleri 0,820 ile 0,968 arasında değişmiştir. Bu sonuç, değerlendiriciler arasındaki tutarlılığın yüksek olduğunu göstermektedir. Ayrıca, değer-

lendirici içi güvenilirliğini test etmek amacıyla skorlamalar 4 hafta sonra tekrar yapılmış ve ICC değeri 0,89 olarak hesaplanmıştır. Bu değer, aynı değerlendiricinin farklı zamanlardaki değerlendirmelerinin yüksek düzeyde tutarlı olduğunu göstermektedir.

Tanımlayıcı istatistik sonuçlarında GQS skorlarına göre en yüksek ortalama puanı ChatGPT-4o modeli ($3,803 \pm 0,4008$) alırken, en düşük ortalama puan Medisearch modeline ($3,2576 \pm 0,7084$) aittir. Kruskal-Wallis testi sonuçlarına göre, yapay zeka modelleri arasında GQS skorları arasında istatistiksel olarak anlamlı bir fark olduğu saptanmıştır ($H=37,819$; $p < 0,0001$). Değerlendiricilerin körleme yöntemiyle gerçekleştirdiği puanlamalar sonucunda, en yüksek sıralama ortalaması ChatGPT-4o modeline (240,37), en düşük sıralama ortalaması ise Medisearch modeline (154,3) ait bulunmuştur. GQS skorlarında en yüksek medyan puan ChatGPT-4o, Copilot, Claude 3.7, ChatGPT-3.5 modellerinde 4 olarak belirlenmiştir. Bu modellerin çeyrekler açıklığı (IQR) değeri 1 olup, ChatGPT-4o modelinin IQR değeri 0 olarak hesaplanmıştır. Bu durum, ChatGPT-4o modeline ait skorların oldukça homojen olduğunu göstermektedir. Öte yandan Gemini ve Medisearch modellerinin medyan değerleri 3 olarak saptanmış ve her iki modelin IQR değeri 1 olmuştur (Tablo 3).

Tablo 3. Büyük dil modelinin verdiği yanıtlara ilişkin puanların karşılaştırılması.

GQS	Maks	Min	Medyan	Çeyrekler Açıklığı (IQR)	SS	Ort	Sıralama Ort	H	p
ChatGPT-4o	4	3	4	0,00	0,4008	3,803	240,37		
Gemini	4	2	3	1,00	0,5513	3,393	166,61		
							9		
Copilot	5	1	4	1,00	0,5755	3,712	226,57		
							1		
Medisearch	4	2	3	1,00	0,7084	3,257	154,3	37,819	<0,001
							6		
Claude 3.7 Sonnet	4	1	4	1,00	0,6378	3,530	196,7		
							3		
ChatGPT-3.5	4	1	4	1,00	0,6071	3,590	206,45		
							9		

İstatistiksel anlamlılık düzeyi $p < 0,05$ olarak kabul edilmiştir. H: Kruskal-Wallis analizi. (Maks: Maksimum, Min: Minimum, SS: Standart Sapma, Ort: Ortalama)

Tablo 4. İkili karşılaştırma sonuçları

Anlamlı Farklılıklar	p değeri	Düzeltilmiş p değeri	Etki Büyüklüğü
Medisearch - ChatGPT-3.5	0,002	0,036	0,26
Medisearch - Copilot	<0,001	<0,001	0,36
Medisearch - ChatGpt-4o	<0,001	<0,001	0,43
Gemini - Copilot	<0,001	0,007	0,30
Gemini - ChatGpt-4o	<0,001	<0,001	0,37

Bonferroni düzeltmeli $p < 0,05$ değeri istatistiksel anlamlılık düzeyidir.

İkili karşılaştırmalar sonucunda ChatGP-4o, Copilot ve ChatGPT-3.5 modellerinin yanıt kalitesi Medisearch modeline kıyasla anlamlı düzeyde daha yüksek bulunmuştur ($p < 0,05$) (Tablo 4). Gemini modelinin, Copilot ve ChatGPT-4o ile karşılaştırıldığında istatistiksel olarak anlamlı düzeyde daha düşük GQS skorlarına sahip olduğu bulunmuştur ($p < 0,05$).

TARTIŞMA

Yapay zeka tabanlı LLM'lerin sağlık alanındaki kullanımı gün geçtikçe yaygınlaşmakta ve bu bağlamda yapılan çalışmalar yaygınlaşmaktadır. Rao ve ark.¹⁴ tarafından gerçekleştirilen meme kanseri taramasıyla ilgili bir çalışmada, ChatGPT, meme ağrısı ve meme kanseri taraması konusunda Amerikan Radyoloji Koleji Uygunluk Kriterleri temel alınarak yöneltilen açık uçlu sorularda yüksek puanlar almıştır. Diş hekimliğine ilişkin olarak gömülü diş tedavileri, önleyici ortodontik tedavi ve çene ameliyatları hakkında genellikle doğru ve kapsamlı bilgiler sunabilmiş,¹⁵ endodonti alanında nispeten yüksek güvenilirlik ortaya koymuş,¹⁶ ortodontik sorulara verdiği yanıtların yüksek doğruluk seviyesine sahip olduğu ifade edilmiştir.¹⁷ Ayrıca bu LLM'lerden el-bilek radyografilerinden kemik yaşı tespiti gibi görüntü analiz desteği de alınmıştır.¹⁸

Yapay zeka modelleri sürekli güncellenmekte ve öğrenmektedir. Nitekim bir çalışmada, ChatGPT'nin Mart 2023 sürümündeki yanıtları ile Mayıs 2023'te güncellenen sürümün yanıtları arasında farklılıklar olduğu ve modelin tutarlılığında gelişim gözlemlenmiştir. Bu da zaman içinde elde edilecek sonuçların değişebileceğine işaret etmekte ve sürekli gelişen yapay zeka tabanlı LLM'lerin yeniden değerlendirilmeleri gerekliliğini ortaya koymaktadır.¹⁹ Bu sonuçlar dikkate alınarak, her geçen gün daha da gelişen LLM'lerin ilerleyen performanslarının araştırılması amacıyla çalışmamızda, altı farklı yapay zeka tabanlı LLM'nin ortodontik tedavi gören hastaların acil destek kapsamlı sorularına verdiği yanıtlar değerlendirilmiştir.

Literatürdeki benzer çalışmalara paralel olarak çalışmamızda da ChatGPT-4o en doğru yanıtları veren model olarak değerlendirilmiş,²⁰⁻²³ Copilot ise en başarılı 2. LLM olarak tespit edilmiştir.²¹ En yüksek performansı gösteren bu iki model arasındaki fark ise istatistiksel anlamlılık düzeyine ulaşmamıştır. Çalışmamızın bulgularıyla uyumlu olarak başka bir araştırmada da ChatGPT-4o'nun Gemini'ye kıyasla daha yüksek kaliteli yanıtlar ürettiği gösterilmiştir.²²

Nguyen ve ark.'nın²⁴ yaptığı çalışmada Copilot'un ücretli versiyonu olan CopilotPro en başarılı LLM olarak tespit edilmiştir. ChatGPT-4o en başarılı 2. model olarak bulunmuş ve Gemini'nin ücretli versiyonu olan Gemini Advanced'dan daha iyi performans gösterdiği ortaya konulmuştur. Çalışmamızın sonuçları ile uyumlu olarak Claude 3.7 Sonnet görece düşük performans göstermiştir. Bu çalışmada CopilotPro'nun ChatGPT-4o'dan daha başarılı bulunmuş olması, iki ücretli platform arasındaki farkı göstermektedir. Ancak ChatGPT'nin en yeni versiyonu olan ChatGPT-Pro'nun da gelecek çalışmalarda test edilmesi bu kıyaslamayı daha güvenilir kılacaktır.

Makrygiannakis ve ark.'nın² genel ortodonti ile ilgili dört farklı LLM'nin yanıtlarını değerlendirdiği bir çalışmada, en iyi yanıtların sırasıyla Bing (Copilot), ChatGPT-4o, Google

Bard (Gemini) ve ChatGPT-3.5 tarafından sağlandığını bildirmiştir. Çalışmamızın sonuçlarından farklı olarak Copilot, ChatGPT-4o'dan daha iyi performans göstermiştir. Genel olarak, bu bulgular çalışmamızla örtüşmektedir ve LLM'lerin belirli bir doğruluk düzeyi sergileyebilmesine rağmen aynı soruya tekrar tekrar farklı yanıtlar verebileceği de gösterilmektedir.

Çalışmamızın sonuçlarına göre en düşük yanıt performansına sahip olduğu görülen MediSearch modeli, bir başka çalışmanın sonuçlarına göre GQS puanlarında yüksek skorlar almıştır.²⁵ Ancak bu çalışmada MediSearch'ün sunduğu "standart" ve "detaylı" olmak üzere iki farklı yanıt seçeneğinden "detaylı" seçeneği tercih edilmiştir. Diğer LLM'ler ile yapılacak kıyaslamaların tarafsızlığının etkilenmemesi adına çalışmamızda bu opsiyon seçilmemiştir. Nitekim bu seçim yapıldığında ChatGPT-4o modelinde mevcut olan "derin araştırma" seçeneğinin de seçilmesi gerekecek ve bunun da klasik anlamdaki değerlendirmenin güvenilirliğini zedelemesi muhtemel görülecektir.

Yapmış olduğumuz çalışmada en düşük yanıt performansını gösteren Medisearch modelinin çoğu soruya yüzeysel önerilerle karşılık verdiği, daha çok "ortodontistimize başvurun" yanıtını sunduğu dikkat çekmiştir. Bunun yanında ortodontik tedavi sırasında meydana gelen ağrının nasıl azaltılabileceğine yönelik iletilen soruya, Claude 3.7 Sonnet modeli buz torbası uygulanması gibi pek de geçerliliği olmayan bir öneride bulunurken, Medisearch modeli sakız çiğnenmesi gibi pratik bir önermede bulunmuştur. Bu soruya başka bir platform sakız çiğnenmesi önermesini sunmamıştır. Buradan da anlaşılacağı üzere Medisearch gibi genel toplamda başarısız gibi görünen bazı LLM'ler de farklı öğrenimlere sahip olarak fark yaratabilmektedir. ChatGPT-3.5'in cevapları arasındaki retansiyon plağı kırılan bir hastaya, plağını kullanmadığı süre zarfında dişlerini aynada kontrol ederek, özellikle ön dişlerin pozisyonunu takip etmesi gerektiği önerisi değerlendirici tarafından kayda değer bir yanıt performansı olarak değerlendirilmiştir (Tablo 5). Aynı düzey yanıtın, ücretli versiyon olan ChatGPT-4o tarafından sunulmaması, versiyonlar arasındaki veri tabanlarının farklılık gösterebileceğini, ücretsiz versiyon veri tabanındaki yapay zeka öğreniminin de bazı noktalarda daha iyi performans sunabileceğini ortaya koymaktadır.

Dil modellerinin ortak noktaları olmasına rağmen, bu LLM'ler farklı mimariler üzerine inşa edilmiştir. Örneğin, ChatGPT, büyük veri kümeleri üzerinde ön eğitim aldıktan sonra belirli görevler için ince ayarlama yapılan GPT (Generative Pre-trained Transformer) mimarisini kullanmaktadır. Buna karşılık, Gemini, daha doğru yanıt üretimi için bağlamı daha iyi anlamayı amaçlayan Google'ın LaMDA (Language Model for Dialogue Application) sinir ağı

mimarisine dayanmaktadır. Microsoft Copilot ise belirli bir görev veya uygulamaya bağlı olarak GPT-4 gibi çeşitli öğrenme modellerini kullanmaktadır. Ağ mimarilerindeki farklılıklar ve eğitim verilerinin miktar ve çeşitliliğindeki değişkenlikler, dil modellerinin aynı sorulara farklı yanıtlar üretmesine neden olmakta ve her modelin farklı güçlü ve zayıf yönleri sahip olmasına neden olmaktadır.² Yanıtların farklılık göstermesi veya doğruluk açısından eksiklikler barındırmasının bir diğer nedeni, kesinliğin sağlanması için gerekli olan istem (prompt) ayrıntı seviyesidir. LLM'ler, soruların ayrıntı seviyesine duyarlı olduğundan, bazı sorular yeterli doğrulukta ifade edilmemiş olabilir ve bu durum, LLM'lerin soruları uygun şekilde anlamasını engelleyebilir.^{2,26}

Diğer taraftan, yapay zekâ tarafından oluşturulan yanıtların değerlendirilmesinde, değerlendirmeyi yapan kişilerin uzmanlığı ve yetkinliği de kritik bir öneme sahiptir. Literatürde,^{6,27} farklı grupların algısında farklılık olabileceği gösterilmiştir. Yapılan bir çalışmada ortodontistlerin, diş hekimliği öğrencilerinin ve hastaların LLM'lerin cevaplarına verdikleri puanlar karşılaştırılmış, ortodonti hastalarının model yanıtlarını kalite açısından daha yüksek değerlendirdiği, ortodontistlerin ise daha eleştirel yaklaştığı rapor edilmiştir.²⁷ Balel, ChatGPT'nin hastalardan gelen genel sorulara verdiği yanıtların güvenilir olduğunu ancak doktorların sorduğu sorulara verilen yanıtların orta kalitede olduğunu bildirmiştir.⁶ Bu sonuçlar, LLM'lerin hastalar için bir bilgi aracı olarak önemli bir potansiyel taşıdığını ancak teknik sorular ve eğitim amaçlı kullanım açısından tamamen güvenli olmayabileceğini göstermekte ve bu aracın kesinlikle profesyonel görüşün yerine geçmemesi gerektiği vurgulanmaktadır.¹⁵ Ortodonti uzmanları, hastaların bu araçları nasıl kullandığını fark etmeli ve uygun yönlendirmelerde bulunmalıdır.²⁸

Hastalar, LLM'lerin bu akıl yürütme süreçlerini sorgulama yeteneğine sahip olmayabileceğinden, belirli güvenlik önlemleri gerekebilir. Rokhshad ve ark.'nın¹⁰ çalışmalarında değerlendirilen LLM'lerden Google Bard (Gemini), belirli bir doğruluk olasılığı ile yanıt sağlayamadığında tıbbi sorulara cevap vermeyi reddederek "Ben metin tabanlı bir yapay zeka modeliyim ve bu konuda yardımcı olamıyorum" gibi bir uyarı ile bu tür bir önlem almıştır. Model geliştiricileri, belirsizliği halüsinatif içerik üretmek yerine doğrudan işaretleme stratejisini benimsemeyi düşünebilir.

Tablo 5. Sohbet robotlarının yanıtlarının bilimsel doğruluk ve alaka düzeyine ilişkin değerlendirici analizlerinden bazıları

Sohbet robotlarının yanıtlarının bilimsel doğruluk ve alaka düzeyine ilişkin değerlendirici analizi	
•Kopan braket veya tel parçasının yutulması halinde hastanın derhal acil müdahale için başvurusu, hatta ambulans çağırması gerektiği gibi hastanın kaygı düzeyini artıran pek de geçerli olmayan bir cevap verilmiş.	Gemini 2. Soru Yanıtı
•Çocukların aparey kullanımının artırılmasına yönelik bir ödül sistemi tavsiye edilmiş. Bu noktada dikkat çekici bir öneri.	Gemini 47. Soru Yanıtı
•Dudak damak yarıklı bebeklerin annelerine deneyimlerinden faydalanmak adına, başka dudak damak yarıklı bebeklerin anneleriyle iletişim kurmalarına yönelik güzel bir öneride bulunulmuş.	Gemini 66. Soru Yanıtı
•Elastikleri düzenli olarak kullanmanın, elastiklere bağlı oluşan ağrının azaltılmasını sağlayabileceği önerisi hem elastik kullanımı teşvik ediyor hem de ağrıyı kontrol edici bir tavsiye sunuyor.	Copilot 17. Soru Yanıtı
•Relapsın "10 yıl içinde veya daha kısa sürede" oluşacağı gibi hastayı rahvete yönlenebilecek başarısız bir yanıt.	Medisearch 13. Soru Yanıtı
•Farklı markalardaki lastiklerin kuvvetlerinde meydana gelen azalma oranlarının farklılık gösterebileceğinden, farklı bir paket lastik kullanmadan önce bu durumun göz önünde bulundurulması gerektiğinden bahsedilerek iyi bir noktaya değinilmiş.	Medisearch 29. Soru Yanıtı
•Ortodontik tedaviye bağlı dişlerinin sızladığını belirten hastaya yüzüne buz torbası uygulayabileceği yönünde pratik karşılığı zayıf bir yanıt verilmiş.	Claude 3.7 Sonnet 20. Soru Yanıtı

Yapay zeka tabanlı LLM'ler "halüsinasyon" olarak adlandırılan riski taşımalarına rağmen bazen konuya yeterince uygun olmayan yanlış veya eksik bilgiler üretebilir.²⁹ Mevcut modellerde görülen halüsinasyon yani "gerçekliği olmayan bilgi" sorununu engellemek için, model mimarisine kaynak doğrulama mekanizmaları eklenmesi yararlı olacaktır. LLM'ler yanıt verirken tıbbi bir iddiayı mutlaka güvenilir bir literatür kaynağı ile desteklemeli ve kullanıcıya referans sunmalıdır. Klasik ChatGPT yanıtları akademik referanslar içermediği için bilimsel olarak güvenilir kabul edilmemektedir. Ancak derin araştırma seçeneği ile yanıtlara bilimsel atıf ekleyebilmektedir. Bunun da araştırmaya değer olduğu görülen bir başka çalışmanın konusu olabileceği düşünülmektedir.

SONUÇLAR

Çalışmamızda ortodonti kapsamlı hasta sorularına yapay zeka tabanlı LLM'lerin verdikleri yanıtlar değerlendirilmiş ve iyi sonuçlar gösterdikleri tespit edilmiştir. Hastaların hekimlerine ulaşamadıkları durumlarda tedavilerine ilişkin acil durum sorularına LLM'lerin kaliteli yanıtlar üretebilmesi, hasta-hekim ilişkisini geliştirebilecek ve hastaların bilinçlenerek tedaviye daha uyumlu olmalarını sağlayabilecektir. Gelecek çalışmalar, bu tür LLM'lerin gerçek hasta sonuçlarına etkisini de değerlendirmelidir. Örneğin, LLM desteği alan hastaların tedavi sürecindeki kaygı düzeyleri veya uyum oranlarının, destek almaya kıyasla farklılık gösterip göstermediği araştırılması gereken bir konudur. Bu araştırmalar sonucunda elde edilecek veriler, LLM'lerin hangi alanlarda en fazla fayda sağladığını ve nerelerde sınırlı kaldığını ortaya koyacaktır.

KAYNAKLAR

1. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. *J Esthet Restor Dent* 2023; 35: 1098-1102. doi: 10.1111/jerd.13046.
2. Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. *Eur J Orthod* 2024; cjae017. doi: 10.1093/ejo/cjae017.
3. Adamopoulou E, Moussiades L. An overview of chatbot technology. In *IFIP international conference on artificial intelligence applications and innovations*. Springer 2020; 373-383. doi: 10.1007/978-3-030-49186-4_31.
4. Vaishya R, Misra A, Vaish A. ChatGPT: Is this version good for healthcare and research? *Diabetes Metab Syndr* 2023; 17: 102744. doi: 10.1016/j.dsx.2023.102744.
5. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*. MDPI 2023; 11: 887. doi: 10.3390/healthcare11060887.
6. Balel Y. Can ChatGPT be used in oral and maxillofacial surgery? *J Stomatol Oral Maxillofac Surg* 2023; 124: 101471. doi: 10.1016/j.jormas.2023.101471.
7. Henzell M, Knight A, Antoun JS, Farella M. Social media use by orthodontic patients. *NZ Dent J* 2013; 109: 130-133.
8. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. *J Med Internet Res* 2018; 20: 253. doi: 10.2196/jmir.9364.
9. Siddiqui N, Chia M, Sharif MO. Social media and orthodontics: Are our patients scrolling? *J Orthod* 2022; 49: 179-184. doi: 10.1177/14653125211042025.
10. Rokhshad R, Zhang P, Mohammad-Rahimi H, Pitchika V, Entezari N, et al. Accuracy and consistency of chatbots versus clinicians for answering pediatric dentistry questions: A pilot study. *J Dent* 2024; 144: 104938. doi: 10.1016/j.jdent.2024.104938.
11. Acar AH. Can natural language processing serve as a consultant in oral surgery? *J Stomatol Oral Maxillofac Surg* 2024; 125: 101724. doi: 10.1016/j.jormas.2023.101724.
12. Bernard A, Langille M, Hughes S, Rose C, Leddin D, et al. A systematic review of patient inflammatory bowel disease information resources on the World Wide. *Am J Gastroenterol* 2007; 102: 2070-2077.
13. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos 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. doi: 10.2196/51580.
14. Rao A, Kim J, Kamineni M, Pang M, Lie W, et al. Evaluating ChatGPT as an adjunct for radiologic decision-making. *MedRxiv* 2023; 2023-02. doi: 10.1101/2023.02.02.23285399.
15. Alkhamees A. Evaluation of Artificial Intelligence as a Search Tool for Patients: Can ChatGPT-4 Provide Accurate Evidence-Based Orthodontic-Related Information? *Cureus* 2024; 16. doi: 10.7759/cureus.65820.
16. Mohammad-Rahimi H, Ourang SA, Pourhoseingholi MA, Dianat O, Dummer PMH, et al. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. *Int Endod J* 2024; 57: 305-314. doi: 10.1111/iej.14014.
17. Daraqel B, Wafaie K, Mohammed H, Cao L, Mheissen S, 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: 652-662. doi: 10.1016/j.ajodo.2024.01.012.
18. Yıldırım, A., Cicek, O., Genç, Y. S. Can AI-Based ChatGPT Models Accurately Analyze Hand-Wrist Radiographs? A Comparative Study. *Diagnostics* 2025; 15(12), 1513. doi: 10.3390/diagnostics15121513.
19. Kiliç DD, Mansız 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: 546-555. doi: 10.1016/j.ajodo.2023.11.012.
20. Rudolph J, Tan S, Tan S. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *J. Appl. Learn. Teach.* 2023; 6: 364-389. doi: 10.37074/jalt.2023.6.1.23.
21. Jeong H, Han SS, Yu Y, Kim S, Jeon KJ. How well do large language model-based chatbots perform in oral and maxillofacial radiology? *Dentomaxillofac Radiol* 2024; 53: 390-395. doi: 10.1093/dmfr/twae021.
22. Arslan C, Kahya K, Cesur E, Cakan DG. An evaluation of orthodontic information quality regarding artificial intelligence (AI) chatbot technologies: A comparison of ChatGPT and google BARD. *Aust Orthod J* 2024; 40:149-157. doi: 10.2478/aoj-2024-0012.
23. Dursun D, Bilici Geçer R. Can artificial intelligence models serve as patient information consultants in orthodontics?. *BMC Medical Informatics and Decision Making* 2024; 24(1), 211. doi: 10.1186/s12911-024-02619-8.
24. Nguyen HC, Dang HP, Nguyen TL, Hoang V, Nguyen VA. Accuracy of latest large language models in answering multiple choice questions in dentistry: A comparative study. *PloS one* 2025; 20: e0317423. doi: 10.1371/journal.pone.0317423.
25. Yurdakurban E, Topsakal KG, Duran GS. A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient informa-

tion. *J Stomatol Oral Maxillofac Surg* 2024; 125: 101757. doi: 10.1016/j.jormas.2023.101757.

26. Mago J, Sharma M. The potential usefulness of ChatGPT in oral and maxillofacial radiology. *Cureus* 2023; 15. doi: 10.7759/cureus.42133.

27. Kurt Demirsoy K, Buyuk SK, Bicer T. How reliable is the artificial intelligence product large language model ChatGPT in orthodontics? *Angle Orthod* 2024; 94: 602-607. doi: 10.2319/031224-207.1.

28. Abu Arqub S, Al-Moghrabi D, Allareddy V, Upadhyay M, Vaid N, et al. Content analysis of AI-generated (ChatGPT) responses concerning orthodontic clear aligners. *Angle Orthod* 2024; 94: 263-272. doi: 10.2319/071123-484.1.

29. Korngiebel DM, Mooney SD. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *NPJ Digital Medicine* 2021; 4: 93. doi: 10.1038/s41746-021-00464-x.