Dil farklılıklarının ChatGPT-3.5, Copilot ve Gemini'nin pediatrik oftalmoloji ve şaşılık çoktan seçmeli sorulardaki etkinliğinin değerlendirilmesi

Evaluation of language differences on the effectiveness of ChatGPT-3.5, Copilot and Gemini in pediatric ophthalmology and strabismus multiple choice questions

Eyüpcan Şensoy[1], Melike Şensoy[1], Mehmet Çıtırık[1]

[1] Ankara Etlik Şehir Hastanesi, Göz Hastalıkları Anabilim Dalı

Öz

Amaç: Bu çalışmada yapay zeka programlarının pediatrik oftalmoloji ve şaşılık ile ilişkili çoktan seçmeli soruları cevaplamadaki başarı düzeylerine dil farklılıklarının etkilerinin incelenmesi amaçlandı.

Yöntemler: Pediatrik oftalmoloji ve şaşılık ile ilişkili 44 soru çalışmaya dâhil edildi. Soruların Türkçe çevirileri sertifikasyonlu çevirmen (native speaker) tarafından gerçekleştirildikten sonra hem İngilizce hem Türkçe versiyonları ChatGPT-3,5, Copilot ve Gemini yapay zeka sohbet botlarına uygulandı. Sorulara verilen cevaplar cevap anahtarı ile karşılaştırılarak doğru ve yanlış olarak gruplandırıldı.

Bulgular: İngilizce sorulara ChatGPT-3,5, Copilot ve Gemini sırası ile %56,8, %72,7 ve %56,8 oranında doğru cevap verdi (p=0,206). Türkçe sorulara ChatGPT-3,5, Copilot ve Gemini sırası ile %45,5, %68,2 ve %56,8 oranında doğru cevap verdi (p=0,099). Yapay zeka programları soruların İngilizce ve Türkçe versiyonlarını cevaplamada benzer başarı düzeylerine sahipti (p>0,05).

Sonuç: Sohbet botları her ne kadar soruları cevaplamada benzer performans göstermiş olsa bile sorular ayrı ayrı incelendiğinde aynı sorulara farklı cevaplar üretebilmişlerdir. Bu durum kullanıcıların sohbet botlarının doğruluğuna olan güvenini zedeleyebilir. Sohbet botlarının dil performanslarının geliştirilmeye ihtiyacı vardır.

Anahtar Sözcükler: ChatGPT-3,5; Copilot; Gemini; oftalmoloji; pediatri; şaşılık

Abstract

Aim: This study aimed to investigate the effects of language differences on the success levels of artificial intelligence programs in answering multiple-choice questions related to pediatric ophthalmology and strabismus.

Methods: Forty-four questions related to pediatric ophthalmology and strabismus were included in the study. After the questions were translated into Turkish by a certified native speaker, both English and Turkish versions were applied to ChatGPT-3.5, Copilot, and Gemini artificial intelligence chatbots. The answers given to the questions were compared with the answer key and grouped as correct and incorrect.

Results: ChatGPT-3.5, Copilot, and Gemini answered the English questions correctly at a rate of 56.8%, 72.7%, and 56.8%, respectively (p = 0.206). ChatGPT-3.5, Copilot, and Gemini answered the Turkish questions correctly at a rate of 45.5%, 68.2%, and 56.8%, respectively (p = 0.099). Artificial intelligence programs had similar levels of success in answering the English and Turkish versions of the questions (p>0.05).

Conclusion: Although chatbots performed similarly in answering questions, they could produce different answers to the same questions when examined separately. This situation may undermine users' trust in the chatbots' accuracy. The language performance of chatbots needs to be improved.

Keywords: ChatGPT-3.5; Copilot; Gemini; ophthalmology; pediatrics; strabismus.

Geliş/*Received* : 30.11.2024
Kabul/*Accepted*: 28.12.2024

DOI: 10.21673/anadoluklin.1593858

Yazışma yazarı/*Corresponding author*

Eyüpcan Şensoy

Ankara Etlik Şehir Hastanesi, Göz Hastalıkları Anabilim Dalı, Ankara, Türkiye.
E-posta: dreyupcansensoy@yahoo.com

ORCID

Eyüpcan Şensoy: 0000-0002-4401-8435
Melike Şensoy: 0000-0002-8273-3851
Mehmet Çıtırık: 0000-0002-0558-5576

GİRİŞ

Teknolojideki hızlı gelişim ile birlikte bilgisayar biliminin bir alt dalı olarak ortaya çıkan yapay zeka uygulamaları tıbbın her alanında yaygın bir şekilde etkisini göstermeye başlamıştır. (1). Bu konudaki ilk çalışmalar her ne kadar yapay zeka fikrinin ortaya çıkışı daha eskilere dayansa da 1970'li yıllara kadar gecikmiş, daha sonra hızlı bir ilerleme katetmiştir (2). Yapay zeka uygulamalarının oftalmoloji alanında kullanımları ise 2015 yılından sonra hızla artmış, çok çeşitli hastalıkların tanı ve tedavi izlemlerinde etkin bir rol üstlenmeyi başarmıştır (3–5). Pediatrik oftalmoloji ve şaşılık da yapay zeka uygulamalarının kendisine yer bulduğu oftalmolojinin önemli alanlarından biridir (6,7).

Yapay zeka teknolojilerinin önemli bir kolu olan Büyük Dil Modeli (BDM) temelli sohbet botları ise son zamanlarda hayatımıza girmiş, çok geniş veriler ile eğitilmiş, kavramlar arasındaki örüntüyü anlayabilen ve daha sonra gelişebilecek olasılıkları tahmin edebilen yani insan düşünce yapısını taklit etmeyi amaçlayan yeni bir teknolojidir (8). Üç farklı büyük üretici tarafından ücretsiz olarak piyasaya sürülmüş olan ChatGPT-3,5 (Open AI), Copilot (Microsoft) ve Gemini (Google AI) BDM temelli bu programların önemli örneklerindendir (8–10).

Çalışmamızın amacı son zamanlarda yaygın olarak kullanıma girmiş olan ChatGPT-3,5, Copilot ve Gemini yapay zeka sohbet botlarının pediatrik oftalmoloji ve şaşılık ile ilgili İngilizce ve Türkçe aynı sorularda gösterdikleri performanslara dil farklılığının etkilerini araştırmaktır.

GEREÇ VE YÖNTEM

Araştırmanın tasarımı

Amerikan Akademi ve Oftalmoloji 2023-2024 Basic and Clinical Science Course (BCSC) Pediatrik Oftalmoloji ve Şaşılık kitabı çalışma soruları kısmında yer alan 44 sorunun tamamı çalışmaya dâhil edildi (11). İngilizce olan soruların sertifikasyonlu çevirmen (native speaker) tarafından Türkçe çevirileri gerçekleştirildi. İngilizce ve Türkçe aynı sorular ücretsiz olarak erişim sağlanabilen ChatGPT-3,5 (OpenAI; San Francisco, CA), Copilot (Dengeli mod) (Microsoft, Redmond, WA) ve Gemini (Google, Mountain View, California, United States) yapay zeka sohbet botlarına 19 Temmuz 2024 tarihinde uygulandı. Sorular sohbet botlarına uygulanmadan önce 'Sana çoktan seçmeli sorular soracağım. Lütfen bana doğru cevap seçeneğini ver.' komutu verildi ve her soru sonrasında oturum sonlandırılarak tekrar oturum başlatıldı. Sohbet botlarının her soruya verdiği cevap seçenekleri kitap arkasında yer alan cevap anahtarı ile karşılaştırılarak sorular doğru ve yanlış olarak gruplandırıldı. Ayrıca yapay zeka programlarının İngilizce ve Türkçe aynı sorulardaki performansları, her soru düzeyinde ayrı ayrı incelenerek kaydedildi.

Yapay zeka programlarının her birinin İngilizce ve Türkçe sorulardaki doğru cevaplama performansları araştırmamızın birincil sonuçlarını oluştururken, bu programların her sorunun İngilizce ve Türkçe versiyonlarına verdikleri doğru cevap düzeyleri araştırmamızın ikincil sonuçlarını oluşturmaktadır.

Kullanılan yapay zeka sohbet botları ve özellikleri

ChatGPT-3,5 yaklaşık olarak 175 milyar parametre veri grubuyla eğitilmiş, çok geniş bir bilgi birikimine sahip olan fakat Eylül 2021 tarihinden sonra güncelleme almamış bir yapay zeka programıdır. Bu programın güncel internet erişimi mevcut değildir. Girilen verilerle insan düşünce yapısını taklit etmek ve bu doğrultuda cevaplar üretebilmek amaçlanmaktadır (12).

Copilot; Şubat 2023 tarihi ile birlikte GPT-4 entegrasyonu sağlanarak daha da geliştirilmiş, güncel internet erişimine sahip sürekli yenilenmeleri devam eden bir yapay zeka programıdır. Çeşitli bağlamsal bilgileri analiz edip cevaplar üretebilmekte ve bu bilgilerin kaynaklarını araştırıcılara sunarak daha ayrıntılı bilgilere ulaşılabilmede kolaylık sağlayabilmektedir (10).

Gemini; çok sayıda bilgileri algılayabilen, geniş bir bağlamda inceleyebilen ve bunun sonucunda kesin yanıtlar üretebilen bir yapay zeka programıdır. Soruları cevaplamadaki tatminkar eminliği ve güncel internet erişiminin mevcut olması bu programın öne çıkan özelliklerinden bazılarıdır (9).

Etik kurul

Çalışmamız etik kurul izni gerektiren araştırmalar kapsamında olmayıp, çalışmada insan/insana ait veriler ve hayvan deneyleri verileri bulunmadığından etik kurul onayına ihtiyaç yoktur.

İstatistiksel analiz

Verilerin istatistiksel analizinde Statistical Package for the Social Sciences sürüm 23 (SPSS Inc., Chicago, IL, USA) programı kullanıldı. Yüzdelik değerler hesaplandı. Nominal birbirinden bağımsız gruplardaki verilerin istatistiksel karşılaştırılmasında Pearson ki-kare testi, nominal bağımlı grup verilerinin karşılaştırılmasında McNemar testi kullanıldı. P değerinin 0,05'in altında olması anlamlılık düzeyi olarak kabul edildi.

BULGULAR

Pediatrik oftalmoloji ve şaşılık ile ilişkili 44 soru yapay zeka sohbet botlarına İngilizce olarak soruldu. ChatGPT-3,5 sorulan soruların 25'ine (%56,8) doğru cevap, 19'una (%43,2) yanlış cevap verdi. Copilot sorulan soruların 32'sine (%72,7) doğru cevap, 12'sine (%27,3) yanlış cevap verdi. Gemini sorulan soruların 25'ine (%56,8) doğru cevap, 19'una (%43,2) yanlış cevap verdi (Şekil 1). Her üç sohbet botunun İngilizce sorulardaki başarıları arasında istatistiksel olarak anlamlı düzeyde bir fark tespit edilmedi (p=0,206).

Pediatrik oftalmoloji ve şaşılık ile ilişkili 44 sorunun Türkçe versiyonları yapay zeka sohbet botlarına uygulandı. ChatGPT-3,5 sorulan soruların 20'sine (%45,5) doğru cevap, 24'üne (%54,5) yanlış cevap verdi. Copilot sorulan soruların 30'una (%68,2) doğru cevap, 14'üne (%31,8) yanlış cevap verdi. Gemini sorulan soruların 25'ine (%56,8) doğru cevap, 19'una (%43,2) yanlış cevap verdi (Şekil 1). Her üç sohbet botunun Türkçe sorulardaki başarıları arasında istatistiksel olarak anlamlı düzeyde bir fark tespit edilmedi (p=0,099).

ChatGPT-3,5; İngilizce ve Türkçe soruların 29'una (%65,9) aynı cevap, 15'ine (%34,1) ise farklı cevap verdi. Farklı cevap verdiği soruların 5'inde (%33,3) sorular Türkçe sorulduğunda doğru cevaplanırken, 10'unda (%66,7) sorular Türkçe sorulduğunda yanlış olarak cevaplandı. ChatGPT-3,5'in, İngilizce ve Türkçe sorulardaki performans düzeyleri arasında istatistiksel olarak anlamlı düzeyde bir fark tespit edilmedi (p=0,302) (Tablo 1).

Copilot; İngilizce ve Türkçe soruların 30'una (%68,2) aynı cevap, 14'üne (%31,8) ise farklı cevap verdi. Farklı cevap verdiği soruların 6'sında (%42,9) sorular Türkçe sorulduğunda doğru cevaplanırken, 8'inde (%57,1) sorular Türkçe sorulduğunda yanlış olarak cevaplandı. Copilot'un, İngilizce ve Türkçe sorulardaki performans düzeyleri arasında istatistiksel olarak anlamlı düzeyde bir fark tespit edilmedi (p=0,791) (Tablo 1).

Gemini; İngilizce ve Türkçe soruların 32'sine (%72,7) aynı cevap, 12'sine (%27,3) ise farklı cevap verdi. Farklı cevap verdiği soruların 5'inde (%41,7) sorular Türkçe sorulduğunda doğru cevaplanırken, 7'sinde (%58,3) sorular Türkçe sorulduğunda yanlış olarak cevaplandı. Gemini'nin, İngilizce ve Türkçe sorulardaki performans düzeyleri arasında istatistiksel olarak anlamlı düzeyde bir fark tespit edilmedi (p=0,1) (Tablo 1).

TARTIŞMA VE SONUÇ

Yapay zeka sürekli olarak kendini yenilemekte ve gerçek hayatta yerini günden güne güçlendirmeye devam etmektedir. Sıklıkla geçmişte üretilmiş yapay zeka programları kalıpları öğrenerek sonuçlar üretebilen derin öğrenme temelli yapay zeka uygulamaları iken, günümüzdeki gelişmelerle birlikte BDM temelli yapay zeka sohbet botlarının kullanımı yaygınlaşmış ve her alanda olduğu gibi tıbbi alanlarda da kendisine çok çeşitli kullanım yerleri bulmuş, çeşitli faydaları sıklıkla araştırılan bir konu olmuştur (13,14). BDM temelli yapay zeka sohbet botlarının bu kullanım yerlerine; çok çeşitli hastalıklarla ilgili bilgi edinme, ayırıcı tanı yapabilme, tedavi modaliteler ile ilgili bilgi verebilme hatta tıbbi eğitimde literatür tarayabilme, özet çıkarabilme ve dil çevirisi yapabilme örnek olarak gösterilebilir (13,15). Çok çeşitli bu faydaları sayesinde tıp eğitimi alan öğrencilerden tıp profesyonellerine kadar çok geniş aralıkta bir etki alanına sahip olmuştur (15).

BDM temelli yapay zeka programlarının kullanımlarının araştırıldığı bir diğer önemli konu da hasta eğitiminde bu programlarının kullanılabilirliğidir. Bu programlar hastaların hastalıkları ile ilgili gerekli bilgileri elde edebilmesinde, bunların iyileştirilmesinde veya bu hastalıklardan korunabilmede yapılması gereken durumlar gibi çok çeşitli konularda hastalara öneriler sunabilmekle birlikte çok çeşitli karmaşaları da peşinde getirebilmektedir (16-18). Her ne kadar amaç hasta bakımının kolaylaştırılması ve sağlığa kavuşulmanın hızlandırılması olsa da çok çeşitli etik sorunlar beraberinde gelebilecek, hastaların kendileri-

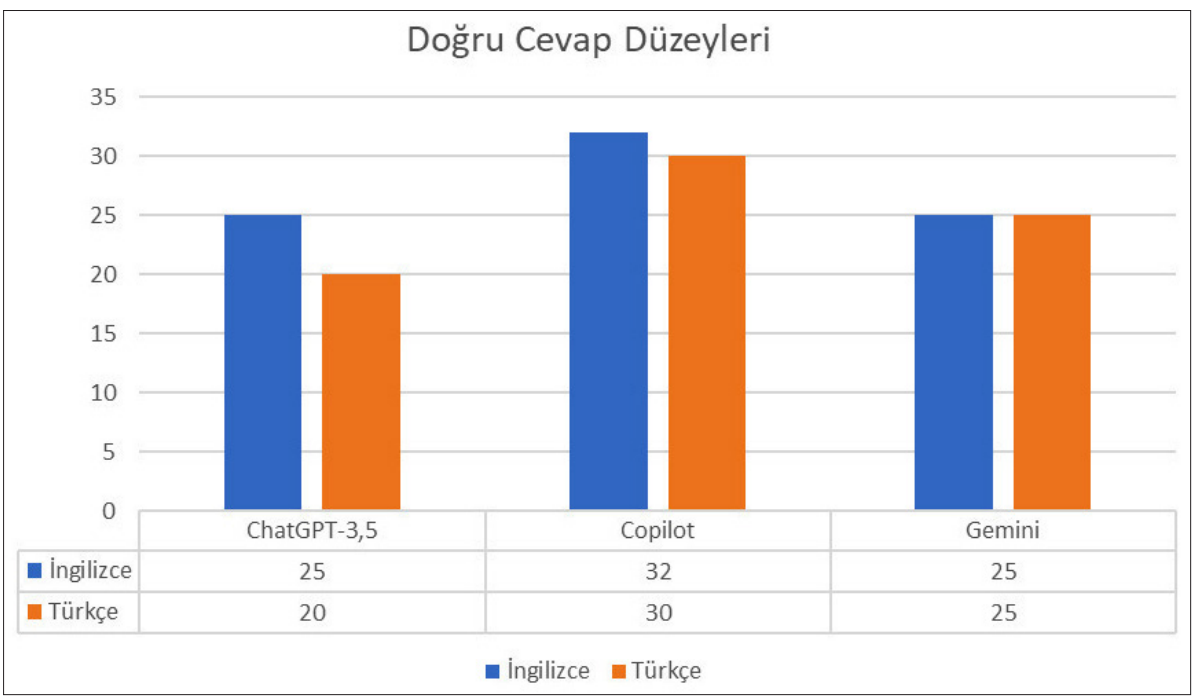
Tablo 1. Yapay zeka sohbet botlarının aynı sorulara verdikleri cevaplar ve değişimleri

Cevaplar	ChatGPT-3,5 (İngilizce)	ChatGPT-3,5 (Türkçe)	Copilot (İngilizce)	Copilot (Türkçe)	Gemini (İngilizce)	Gemini (Türkçe)
Doğru	25 (%56,8)	20 (%45,5)	32 (%72,7)	30 (%68,2)	25 (%56,8)	25 (%56,8)
Yanlış	19 (%43,2)	24 (%54,5)	12 (%27,3)	14 (%31,8)	19 (%43,2)	19 (%43,2)
p değeri	0,302*		0,791*		0,1*	
Aynı cevabı üretme	29 (%68,2)		30 (%68,2)		32 (%72,7)	
Farklı cevabı üretme	15 (%34,1)		14 (%31,8)		12 (%27,3)	
Doğru-yanlış değişimi	10 (%66,7)		8 (%57,1)		7 (%58,3)	
Yanlış-doğru değişimi	5 (%33,3)		6 (%42,9)		5 (%41,7)	

***McNemar testi**

ni yanlış tanılara inandırabilmesine, bu araçların başvuruda tek kaynak olarak kullanılabilmesine ve hatta hasta veri gizliliğinin kaybolabilmesi gibi çok çeşitli olaylara sebebiyet verebilecektir. Bu programlar her ne kadar potansiyel olarak faydalı gibi görülmekle beraber peşinden çok çeşitli ikilemleri sürüklemekte ve bu durumların da ayrı ayrı dikkate alınarak düzeltilmesi ve incelenmesini zorunlu hale getirmektedir. Bu eksikliklerin düzenlenip geliştirilmesi sağlık alanındaki profesyonellerin yüklerini azaltmakla kalmayıp hastaların doğru şekilde daha ayrıntılı bilgiler alabilmeleri için gerekli zaman ve bilgi aktarımı sağlayabilecek bir ortam hazırlamasına yardımcı olacaktır (16-19).

Son gelişmelerle birlikte yapay zeka sohbet botlarının tüm bu faydalı özellikleri göz önüne alınarak oftalmoloji alanında da performansları araştırılan bir konu olmuş, pediatrik oftalmoloji ve şaşılık konuları da rağbet gören bu oftalmoloji konularından biri haline gelmiştir. Ambliyopi ile ilişkili sorularda ChatGPT-3,5, Bard ve Bing yapay zeka sohbet botlarının uygulana-



Şekil 1. ChatGPT-3,5, Copilot ve Gemini'nin pediatrik oftalmoloji ve şaşılık ile ilişkili İngilizce ve Türkçe soruları doğru cevaplama düzeyleri

bilirlik, anlaşılabilirlik ve okunabilirlik performansları incelenen bir çalışmada Bard'ın her üç alanda da en yüksek performansı gösterdiği fakat yine de sohbet botlarının hepsinin kabul edilebilir düzeyde ayrıntılı ve uygun yanıtlar elde etme potansiyeline sahip oldukları belirtilmiştir (20). Oftalmoloji ile ilişkili çoktan seçmeli sorularda sohbet botlarının performanslarını inceleyen bir çalışmada 125 soru ChatGPT'ye uygulanmış ve bu sorularda %46 oranında başarı gösterdiği, ayrıca pediatrik oftalmoloji sorularında ise %33 düzeyinde başarı gösterdiği ifade edilmiştir (21). ChatGPT-3,5 ve Bing'in 913 oftalmoloji sorusundaki başarısının incelendiği bir başka çalışmada ise sohbet botları sırası ile %59,69 ve %73,6 oranında soruları doğru cevaplamış, pediatrik oftalmoloji sorularında ise sırası ile %57,45 ve %76 oranında başarı göstermişlerdir (22). Farklı bir konunun ele alındığı başka bir çalışmada bu sohbet botlarının oftalmoloji sorularındaki başarısının internet erişiminin sağlandığı ülkeye göre değişip değişmediği araştırılmıştır. Bunun sonucunda Gemini'nin erişilen ülkeye göre soruları cevaplamadaki başarılarının değişiklik gösterebildiği, bu değişikliklerin pediatrik oftalmoloji ve şaşılık ile ilişkili soruları cevaplamada da izlendiği belirtilmiştir. Bulunduğu ülkeye göre Gemini'nin pediatrik oftalmoloji ve şaşılık ile ilişkili sorularda %50 ile %80 arasında değişen oranlarda başarılı oldukları ifade edilmiştir (23). ChatGPT-3,5, Bing ve Bard'ın 200 Türkçe oftalmoloji sorularındaki başarısının incelendiği bir çalışmada ise bu programların sırası ile %51, %63 ve %45,5 oranında başarı gösterdiği belirtilmiş, pediatrik oftalmoloji, klinik refraksiyon ve genetik ile ilişkili sorularda sırası ile %57,1, %57,1 ve %71,4 oranında başarı gösterdiği belirtilmiştir. Yazarlar çalışmanın sonucunda ise bu programların potansiyel faydasının varlığına fakat şu an için çok yetkin olmayıp gelişimlerine devam etmesinin gerektiği düşüncesine vurgu yapmışlardır (24). Biz de çalışmamızda ChatGPT-3,5, Copilot ve Gemini'nin İngilizce sorulardaki performanslarını sırası ile %56,8, %72,7 ve %56,8 düzeylerinde; Türkçe sorulardaki performanslarını ise sırası ile %45,5, %68,2 ve %56,8 düzeylerinde olduğunu tespit ettik. Ayrıca bu çalışma, yaptığımız literatür taraması kadarı ile aynı İngilizce ve Türkçe sorularda bu sohbet botlarının pediatrik oftalmoloji ve şaşılık alanındaki başarısını değerlendiren ilk araştırmadır. Her ne kadar uygulamaların performansları arasında ve aynı zamanda programların kendi içlerinde İngilizce ve Türkçe soruları cevaplamadaki başarı düzeyleri arasında istatistiksel olarak fark tespit edememiş olsak da bu programların farklı dilde aynı ve tek doğru cevabı verdiği düşüncesini taşımıyoruz. Cevapları incelediğimizde, doğru cevaplanan soruların Türkçe olarak uygulandığında yanlış cevap oranını artırdığı, ancak bazı sorular açısından da doğru yanıta ulaşmayı kolaylaştırdığı belirledik. Bu farklılık sohbet botlarının farklı dillerde anlama, yorumlama ve fikir üretebilme kabiliyetlerindeki farklılıklarından kaynaklanıyor olabilir. Ayrıca bu sonuç bizde soruların hangi dillerde sorulduğunun verilen bilgilerin doğruluk düzeylerine etki ettiği ve doğru bilgiye ulaşmada karşımıza hız kısıtlayıcı bir basamak olarak çıktığı düşüncesini ortaya çıkarmaktadır.

Önemli bilgileri barındıran ve temel kitaplar arasında yer alan Amerikan Akademi ve Oftalmoloji 2023-2024 Basic and Clinical Science Course (BCSC) Pediatrik Oftalmoloji ve Şaşılık kitabının çalışma soruları 44 soruyu içeriyordu ve sohbet robotlarına bu soruların tamamını sorduk. Soru sayısının az olmasının istatistiksel sonucun anlamlı çıkıp çıkmamasına etki edebilecek önemli bir faktör olduğu kanaatindeyiz. Fakat temel bilgileri ölçen bu kitabın sorularına ilave soru eklemeyi uygun bulmadık. Daha fazla soru içeren testlerde farklı değerlerin çıkıp çıkmayacağının ileri araştırmalar ile desteklenmesi gerekmektedir.

Çalışmamızın bazı kısıtlılıkları mevcuttur. Çalışmadaki soru sayımızın azlığı, tek merkez çalışması olması, soruların sadece doğru ve yanlış olarak değerlendirilip bu programların anlama ve yorumlama özelliklerinin incelenmemesi, yapay zekadaki sürekli gelişimler nedeniyle gösterilen başarıların sadece kesitsel olarak değerlendirilebilmesi ve gerçek hasta hekim ilişkisini ve bu bağlamda ortaya çıkabilen çeşitli yanılsamaların incelenememesi çalışmamızın kısıtlı yönlerini oluşturmaktadır.

Sonuç olarak yapay zeka sohbet botları gelecekte tıp eğitimi içerisinde kendisine güçlü bir yer bulma potansiyeline sahip olsa da geliştirilmesi gereken önemli problemleri içerisinde barındırmaktadır. Farklı dillerdeki performansların aynı düzeye getirilmesi araştırıcılar için doğru bilgi erişiminde daha güvenilir bir kaynak olarak seçilmesine aracılık edecektir.

Çıkar çatışması ve finansman bildirimi

Yazarlar bildirecek bir çıkar çatışmaları olmadığını beyan eder. Yazarlar bu çalışma için hiçbir finansal destek almadıklarını da beyan eder.

KAYNAKLAR

1. Rahimy E. Deep learning applications in ophthalmology. Curr Opin Ophthalmol. 2018;29(3):254-60.
2. Patel VL, Shortliffe EH, Stefanelli M, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med. 2009;46(1):5-17.
3. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. Prog Retin Eye Res. 2018;67:1-29.
4. Antaki F, Coussa RG, Kahwati G, Hammamji K, Sebag M, Duval R. Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. Br J Ophthalmol. 2023;107(1):90-5.
5. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167-75.
6. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. Surv Ophthalmol. 2019;64(2):233-40.
7. de Figueiredo LA, Dias JVP, Polati M, Carricondo PC, Debert I. Strabismus and Artificial Intelligence App: Optimizing Diagnostic and Accuracy. Transl Vis Sci Technol. 2021;10(7):22.
8. Mikolov T, Deoras A, Povey D, Burget L, Černocký J. Strategies for training large scale neural network language models. 2011 IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE. 2011;196-201.
9. Google AI updates: Bard and new AI features in search. Erişim Tarihi: 04.07.2024, https://blog.google/technology/ai/bard-google-ai-search-updates/.
10. Bing Chat | Microsoft Edge. Erişim Tarihi: 04.07.2024, https://www.microsoft.com/en-us/edge/features/bing-chat?form=MT00D8.
11. Khan AO, Chang TCP, El-Dairi MA, et al. (2023), Pediatric ophthalmology and strabismus. San Francisco: American Academy of Ophthalmology.
12. Wen J, Wang W. The future of ChatGPT in academic research and publishing: A commentary for clinical and translational medicine. Clin Transl Med. 2023;13(3):e1207.
13. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci. 2023;39(2):605-7.
14. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.
15. Jeblick K, Schachtner B, Dexl J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. Eur Radiol. 2024;34(5):2817-25.
16. Yılmaz İE. The Promise and the Challenge: Large Language Models for Patient Education - Are We There Yet?. Exp Appl Med Sci. 2024;5(3):137-49.
17. Yılmaz IBE, Doğan L. Talking technology: exploring chatbots as a tool for cataract patient education. Clin Exp Optom. 2025;108(1):56-64.
18. Edhem Yılmaz İ, Berhuni M, Özer Özcan Z, Doğan L. Chatbots talk Strabismus: Can AI become the new patient Educator?. Int J Med Inform. 2024;191:105592.
19. Tan Yip Ming C, Rojas-Carabali W, Cifuentes-González C, et al. The Potential Role of Large Language Models in Uveitis Care: Perspectives After ChatGPT and Bard Launch. Ocul Immunol Inflamm. 2024;32(7):1435-9.
20. Doğan L, Özçakmakcı GB, Yılmaz İE. The Performance of Chatbots and the AAPOS Website as a Tool for Amblyopia Education. J Pediatr Ophthalmol Strabismus. 2024;61(5):325-31.
21. Mihalache A, Popovic MM, Muni RH. Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment. JAMA Ophthalmol. 2023;141(6):589-97.
22. Tao BK, Hua N, Milkovich J, Micieli JA. ChatGPT-3.5 and Bing Chat in ophthalmology: an updated evaluation of performance, readability, and informative sources. Eye (Lond). 2024;38(10):1897-902.
23. Mihalache A, Grad J, Patil NS, et al. Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment. Eye (Lond). 2024;38(13):2530-5.
24. Canleblebici M, Dal A, Erdağ M. Evaluation of the Performance of Large Language Models (ChatGPT-3.5, ChatGPT-4, Bing and Bard) in Turkish Ophthalmology Chief-Assistant Exams: A Comparative Study. Turkiye Klinikleri J Ophthalmol. 2024;33(3):163-70.