

ChatGPT-3.5, Copilot ve Gemini'nin Oküler İnflamasyon ve Üveit Konusundaki Çoktan Seçmeli Sorularda Performans Analizi: Dil Farklılıklarının Etkisi: Kesitsel Araştırma

Performance Analysis of ChatGPT-3.5, Copilot and Gemini on Multiple Choice Questions on Ocular Inflammation and Uveitis: The Effect of Language Differences: a Cross-Sectional Research

^{1b} Eyüpcan ŞENSOY^a, ^{1b} Mehmet ÇITIRIK^a

^aAnkara Etlik Şehir Hastanesi, Göz Hastalıkları Kliniği, Ankara, Türkiye

ÖZET Amaç: Ücretsiz erişim sağlanabilen ChatGPT-3.5, Copilot ve Gemini yapay zekâ sohbet botlarının oküler inflamasyon ve üveit alanındaki çoktan seçmeli sorulardaki başarısının soruların dil farklılığına bağlı olarak değişimlerini değerlendirmektir. **Gereç ve Yöntemler:** Oküler inflamasyon ve üveit ile ilgili 36 soru çalışmaya alındı. Her soru sertifikasyonlu çevirmen tarafından Türkçeye çevrildikten sonra hem İngilizce hem Türkçe versiyonları yapay zekâ programlarına uygulandı. Sorular cevap anahtarları ile karşılaştırılarak doğru ve yanlış olarak gruplandırıldı. Doğru cevaplama düzeyleri istatistiksel olarak karşılaştırıldı. **Bulgular:** İngilizce sorulara ChatGPT-3.5, Copilot ve Gemini sırası ile %63,9, %63,9 ve %50 oranında doğru cevap verdi. Türkçe sorulara ChatGPT-3.5, Copilot ve Gemini sırası ile %52,8, %52,8 ve %66,7 oranında doğru cevap verdi. ChatGPT-3.5, Copilot ve Gemini İngilizce ve Türkçe soruların sırası ile %22,2'sine, %30,6'sına ve %25'ine farklı cevaplar üretti. ChatGPT-3.5'in farklı cevaplar ürettiği soruların %75'i; Copilot'un farklı cevaplar ürettiği soruların %63,6'sı; Gemini'nin farklı cevaplar ürettiği soruların %22,2'si İngilizce sorulduğunda doğru cevaplanmışken Türkçe sorulduğunda yanlış cevaplandı. Yapay zekâ programları İngilizce ve Türkçe soruları cevaplama farklı doğru cevap oranına sahip olsa da başarıları arasında istatistiksel olarak anlamlı düzeyde bir fark gözlenmedi ($p>0,05$). **Sonuç:** Yapay zekâ programları her ne kadar üvea alanında umut vadedse de bilgi düzeyleri ve dil çeviri, algılama ve cevap verebilme kabiliyetlerinin geliştirilmeye ihtiyacı vardır.

ABSTRACT Objective: To evaluate the changes in the success of free-accessible ChatGPT-3.5, Copilot, and Gemini artificial intelligence chatbots in multiple choice questions in the field of ocular inflammation and uveitis depending on the language difference of the questions. **Material and Methods:** Thirty-six questions regarding ocular inflammation and uveitis were included in the study. After each question was translated into Turkish by a native speaker, both English and Turkish versions were applied to artificial intelligence programs. The questions were grouped as correct and incorrect by comparing them with the answer key. Correct answer levels were compared statistically. **Results:** ChatGPT-3.5, Copilot and Gemini correctly answered the questions asked in English at a rate of 63.9%, 63.9%, and 50%, respectively. ChatGPT-3.5, Copilot and Gemini answered the Turkish questions correctly at 52.8%, 52.8%, and 66.7%, respectively. ChatGPT-3.5, Copilot and Gemini produced different answers to 22.2%, 30.6%, and 25% of questions asked in English and Turkish, respectively. Seventy-five percent of the questions for which ChatGPT-3.5 produced different answers; 63.6% of the questions for which Copilot produced different answers; and 22.2% of the questions for which Gemini produced different answers were answered correctly when asked in English but incorrectly when asked in Turkish. Although the artificial intelligence programs had different correct answer rates in answering English and Turkish questions, there was no statistically significant difference between their success ($p>0.05$). **Conclusion:** Although artificial intelligence programs are promising in uveitis, their knowledge levels and language translation, perception, and response capabilities need to be improved.

Anahtar Kelimeler: ChatGPT-3.5; Copilot; Gemini; İngilizce ve Türkçe; oküler inflamasyon ve üveit

Keywords: ChatGPT-3.5; Copilot; Gemini; English and Turkish; ocular inflammation and uveitis

KAYNAK GÖSTERMEK İÇİN:

Şensoy E, Çitirik M. ChatGPT-3.5, Copilot ve Gemini'nin oküler inflamasyon ve üveit konusundaki çoktan seçmeli sorularda performans analizi: Dil farklılıklarının etkisi: Kesitsel araştırma. Türkiye Klinikleri J Ophthalmol. 2025;34(1):12-6.

Correspondence: Eyüpcan ŞENSOY
Ankara Etlik Şehir Hastanesi, Göz Hastalıkları Kliniği, Ankara, Türkiye
E-mail: dreyupcansensoy@yahoo.com



Peer review under responsibility of Türkiye Klinikleri Journal of Ophthalmology.

Received: 21 Jul 2024

Received in revised form: 22 Oct 2024

Accepted: 30 Oct 2024

Available online: 21 Jan 2025

2146-9008 / 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/>).

Üveit, hâlen dünya çapında tüm yaş grubunu etkileyebilen önemli bir körlük nedenidir. Çok çeşitli nedenlerinin olması, teşhisinin zor olması ayrıca doğru tanı ve tedavi için ileri düzeyde uzmanlık istemesi hastaların doğru tanı ve tedaviyi almasında önemli bir engel oluşturmakta ve hastalığın prognozunu olumsuz yönde etkilemektedir.¹⁻³ Ayrıca üveit alanında uzmanlaşmış oftalmologların arasında bile üveit olgularının tanılarında kesin bir görüş birliği mevcut değildir.⁴

Son yıllarda teknolojiadaki gelişmelerle birlikte üveit hastalıklarının tanı ve tedavisinde doğru bilgiler verebilecek yapay zekâ programlarının geliştirilme denemeleri gerçekleştirilmiş, ayrıca mevcut yapay zekâ sohbet botlarının bu alanla ilgili soruları cevaplamaları da sıklıkla araştırılan bir başka konu olmuştur.^{3,5} Tüm bu araştırmalar neticesinde yapay zekâ sohbet botlarının çok çeşitli faydalarının yanında çok sayıda eksiklikleri de ortaya çıkmıştır.^{6,7}

Çalışmamızın amacı, ChatGPT-3.5 (OpenAI; San Francisco, ABD), Copilot (Microsoft; Redmond, ABD) ve Gemini (Google; Mountain View, ABD) yapay zekâ sohbet botlarının üveit ve oküler inflamasyon ile ilişkili olan farklı dillerdeki aynı sorulara verdikleri cevaplarda dil farklılığının etkisinin olup olmadığını araştırmaktır.

GEREÇ VE YÖNTEMLER

Amerikan Akademi ve Oftalmoloji 2023-2024 Temel ve Klinik Bilim Kursu [Basic and Clinical Science Course (BCSC)] Üveit ve Oküler İnflamasyon Kitabı çalışma soruları kısmında yer alan 36 sorunun tamamı çalışmaya dâhil edildi ve bu soruların Türkçe çevirileri sertifikasyonlu çevirmen (native speaker) tarafından gerçekleştirildi.⁸ Soruların İngilizce ve Türkçe versiyonları ChatGPT-3.5, Copilot ve Gemini yapay zekâ sohbet botlarına 6 Temmuz 2024 tarihinde soruldu. Sorular sohbet botlarına sorulmadan önce “Sana çoktan seçmeli sorular soracağım. Bana doğru cevap şikkını bildir.” komutu verildi ve her soru sonrasında oturum sonlandırıldı. Sohbet botlarının verdikleri cevaplar kitap arkasında yer alan cevap anahtarları ile karşılaştırılarak doğru ve yanlış olarak gruplandırıldı. Çalışmamızda insan ve deney hayvanı kaynaklı veriler kullanılmadığından etik kurul onayı gerekmemektedir.

ChatGPT-3.5, Eylül 2021 tarihinden sonra güncelleme almayan, çok geniş bir veri ağı ile eğitilmiş ve insan zekâsına benzer yanıtlar üretebilen Büyük Dil Modeli (BDM) temelli yapay zekâ sohbet botudur.⁹ Copilot, Şubat 2023 tarihinden itibaren GPT-4 entegre edilerek daha da geliştirilmiş, güncel internet erişimi olan BDM temelli bir yapay zekâ sohbet botudur.¹⁰ Gemini karmaşık dil modellerini kullanarak kapsamlı cevaplar üretebilen, karşılaştığı sorulara kesin yanıtlar üretebilen BDM tabanlı bir yapay zekâ sohbet botudur.^{11,12}

İSTATİSTİKSEL ANALİZ

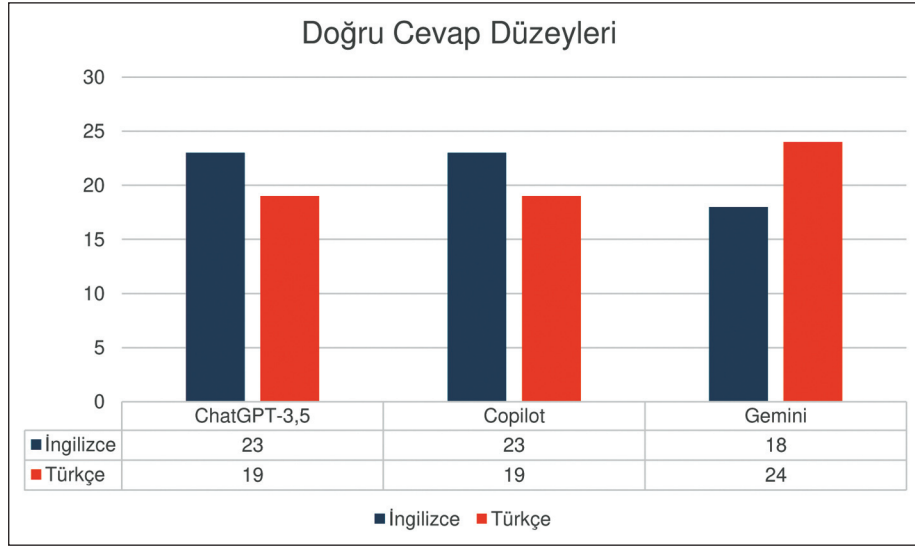
İstatistiksel analizde SPSS sürüm 23 (SPSS Inc., Chicago, IL, ABD) programı kullanıldı. Yüzde değerler hesaplandı. Nominal birbirinden bağımsız verilerin karşılaştırılmasında Pearson ki-kare testi, nominal bağımlı verilerin karşılaştırılmasında McNemar testi kullanıldı. İstatistiksel olarak anlamlılık düzeyi $p < 0,05$ olarak kabul edildi.

BULGULAR

Üveit ve oküler inflamasyonla ilişkili olan 36 soru yapay zekâ programlarına İngilizce olarak uygulandı. ChatGPT-3.5 sorulan soruların 23'üne (%63,9) doğru cevap, 13'üne (%36,1) yanlış cevap verdi. Copilot sorulan soruların 23'üne (%63,9) doğru cevap, 13'üne (%36,1) yanlış cevap verdi. Gemini sorulan soruların 18'ine (%50) doğru cevap, 18'ine (%50) yanlış cevap verdi (Şekil 1). Her üç yapay zekâ sohbet botlarının İngilizce soruları doğru cevaplamada birbirlerine üstünlükleri tespit edilmedi ($p=0,38$, Pearson ki-kare test).

Aynı soruların Türkçe versiyonları yapay zekâ sohbet botlarına uygulandı. ChatGPT-3.5 sorulan soruların 19'una (%52,8) doğru, 17'sine (%47,2) yanlış cevap verdi. Copilot sorulan soruların 19'una (%52,8) doğru, 17'sine (%47,2) yanlış cevap verdi. Gemini sorulan soruların 24'üne (%66,7) doğru, 12'sine (%33,3) yanlış cevap verdi (Şekil 1). Her üç yapay zekâ sohbet botlarının İngilizce soruları doğru cevaplamada birbirlerine üstünlükleri tespit edilmedi ($p=0,39$, Pearson ki-kare test).

ChatGPT-3.5; İngilizce ve Türkçe sorulan aynı soruların 28'ine (%77,8) aynı cevap, 8'ine (%22,2)



ŞEKİL 1: ChatGPT-3.5, Copilot ve Gemini'nin oküler inflamasyon ve üveit ile ilişkili İngilizce ve Türkçe soruları doğru cevaplama düzeyleri.

TABLO 1: Yapay zekâ sohbet botlarının oküler inflamasyon ve üveit ile ilişkili aynı sorulara verdikleri cevaplar ve değişimleri.

Cevaplar	ChatGPT-3.5 (İngilizce)	ChatGPT-3.5 (Türkçe)	Copilot (İngilizce)	Copilot (Türkçe)	Gemini (İngilizce)	Gemini (Türkçe)
Doğru	23 (%63,9)	19 (%52,8)	23 (%63,9)	19 (%52,8)	18 (%50)	24 (%66,7)
Yanlış	13 (%36,1)	17 (%47,2)	13 (%36,1)	17 (%47,2)	18 (%50)	12 (%33,3)
p değeri	0,29*		0,29*		0,11*	
Aynı cevabı verme	28 (%77,8)		25 (%69,4)		27 (%75)	
Farklı cevabı verme	8 (%22,2)		11 (%30,6)		9 (%25)	
Doğru-yanlış değişimi	6 (%75)		7 (%63,6)		2 (%22,2)	
Yanlış-doğru değişimi	2 (%25)		4 (%36,4)		7 (%77,8)	

*McNemar testi.

farklı cevaplar verdi. Farklı cevap verilen soruların; 6'sında (%75) sorular Türkçe sorulduğunda yanlış cevaplanırken, 2'sinde (%25) sorular Türkçe sorulduğunda doğru cevaplandı. ChatGPT-3.5'in İngilizce ve Türkçe soruları cevaplamadaki başarı oranları arasında istatistiksel olarak anlamlı fark tespit edilmedi ($p=0,29$, McNemar testi) (Tablo 1).

Copilot; İngilizce ve Türkçe soruların aynı soruların 25'ine (%69,4) aynı cevap, 11'ine (%30,6) farklı cevaplar verdi. Farklı cevap verilen soruların; 7'sinde (%63,6) sorular Türkçe sorulduğunda yanlış cevaplanırken, 4'ünde (%36,4) sorular Türkçe sorulduğunda doğru cevaplandı. Copilot'un İngilizce ve Türkçe soruları cevaplamadaki başarı oranları arasında istatistiksel olarak anlamlı fark tespit edilmedi ($p=0,29$, McNemar testi) (Tablo 1).

Gemini; İngilizce ve Türkçe soruların aynı soruların 27'sine (%75) aynı cevap, 9'una (%25) farklı cevaplar verdi. Farklı cevap verilen soruların; 2'sinde (%22,2) sorular Türkçe sorulduğunda yanlış cevaplanırken, 7'sinde (%77,8) sorular Türkçe sorulduğunda doğru cevaplandı. Gemini'nin İngilizce ve Türkçe soruları cevaplamadaki başarı oranları arasında istatistiksel olarak anlamlı fark tespit edilmedi ($p=0,11$, McNemar testi) (Tablo 1).

TARTIŞMA

Teknolojinin gelişimi ile birlikte yapay zekâ programları hayatımızın her alanında karşımıza çıkar hâle gelmiş tıbbi bilimlerde bu durumdan etkilenmişlerdir.¹³ Yapay zekâ uygulamalarının bir alt dalı olan BDM'ler kullanıcı tarafından verilen bilgileri anla-

yabilen, özetleyebilen ve meydana gelebilecek cümle bağlamlarını çeşitli olasılıklarla tahmin edebilen güncel ve sürekli gelişimi devam eden bir yan koldur.¹⁴ Dil çeviri kabiliyetleri, kişisel öğrenme planı yapabilmemesi, bilgiye hızlı ulaşım gibi çok sayıdaki yararlı özelliği sayesinde her aşamadaki sağlık personeline yardımcı olabilecek potansiyel bir güç hâline gelmiştir.^{15,16} Yapay zekâ sohbet botlarının tüm bu avantajları göz önüne alındığında oftalmoloji alanına da yapabileceği katkılar araştırılmaya başlanmış, üvea alanı da bu konu ile ilgili araştırmaların ilgi odaklarından biri hâline gelmiştir. Güncel bir çalışmada yapay zekâ sohbet botlarının üvea hastalıklarında tanı ve tıbbi karar almadaki başarısı incelenmiş ve bu başarılar üvea eğitilmiş uzmanlar ile karşılaştırılmıştır. Araştırmacılar her ne kadar uzman oftalmologların daha başarılı olduklarını belirtmiş olsa da sohbet botlarının üveit tanı ve tedavisinde potansiyel faydalarının olabileceği ve bu konudaki geliştirme ve araştırma faaliyetlerinin devam etmesi gerekliliğini vurgulamıştır.³ Bir başka çalışmada, 6 vaka ChatGPT ve Glass adlı iki yapay zekâ sohbet botuna sorulmuş ve burada gösterdikleri başarıları üveit uzmanları ile karşılaştırılmıştır. Çalışmanın sonucunda ChatGPT %66'lık ve Glass %33'lük bir başarı oranına sahipken, üveit uzmanları %100 başarı göstermişlerdir. Yazarlar bu verilerden yola çıkarak ChatGPT'nin üveit vakalarını çözümlenmede ümit vadettiğini, ayrıca göz doktorlarının da yapay zekânın kliniğe entegre edilme konusunda heyecanlı olduklarını ifade etmişlerdir.¹⁷ Araştırılan başka bir konu ise çoktan seçmeli sorularda yapay zekâ sohbet botlarının başarılarıdır. ChatGPT'nin 125 oftalmoloji sorusundaki başarısını test eden çalışmada soruların %46'sının doğru cevaplandığını, üvea ile ilişkili soruların ise %73'üne doğru cevap verdiği belirtilmiştir.¹⁸ Ama üvea soru sayısındaki azlık (11 soru) doğru cevaplamadaki bu farkın sebebi olabilir. ChatGPT-3.5 ve Bing'in çoktan seçmeli sorulardaki etkinliğini değerlendiren ve 913 oftalmoloji sorusunun incelendiği bir çalışmada; ChatGPT'nin %59,7, Bing'in %73,6 oranında başarı gösterdiği, bu başarılarının oküler inflamasyon ve üveit alanında sırası ile %55,38 ve %72,9 oranında olduğu belirtilmiştir. Araştırmacılar sonuç olarak Bing sohbet botunun oftalmoloji öğrencilerine ek fayda sağlayabileceği yönünde bir çıkarımda bulunmuşlardır.¹⁹ Türkçe oftalmoloji sorularının ChatGPT,

Bing ve Bard yapay zekâ sohbet botlarına uygulandığı ve bunların karşılaştırıldığı bir çalışmada; ChatGPT-3.5 %51, Bing %63 ve Bard %45 başarı göstermiş, bu programlar üvea ve oküloplasti ile ilişki sorulara ise sırası ile %62,1, %69 ve %58,6 oranında doğru cevaplar üretmiştir. Bu araştırma sonucunda araştırmacılar sonuçların umut vadetmekle birlikte yeterli düzeyde olmadığını ve geliştirilmesi gerektiğini vurgulamışlardır.²⁰ Yapay zekâ sohbet botlarının farklı bir performanslarını inceleyen ve buldukları ülkelere göre sorulara verilen yanıtlardaki farklılıkları inceleyen bir çalışmada, Gemini 150 tane çoktan seçmeli oftalmoloji sorusunda Amerika versiyonunda ortalama %71 oranında başarı gösterdiği ve üvea ilişkili sorularda %80 oranında başarı gösterdiği belirtilmiştir. Daha sonra bu sorular Vietnam, Hollanda ve Brezilya internet erişiminden uygulanmış ve programların doğruluklarının bulunduğu bölgeye göre değişebildiği belirtilmiştir. Sonuç olarak araştırmacılar bu konunun programları iyileştirmek için geliştiricilere faydalı olabileceği kanaatine varmışlardır.²¹

Bu çalışmada yapay zekâ sohbet botlarının tüm bu araştırılan konulardaki performansına ek olarak aynı soruların İngilizce ve Türkçe versiyonunda başarı farklılığının varlığını değerlendirdik. Çalışmamız sonucunda her ne kadar istatistiksel olarak fark tespit edememiş olsak da ChatGPT-3.5 ve Copilot'un İngilizce sorularda daha başarılı olduğunu, Gemini'nin ise Türkçe sorularda daha başarılı olduğunu gözlemledik. Bu farklılıklar programların anlama yorumla ve bildiği bilgileri sorulan dillere göre yorumlayabilme kabiliyetlerindeki farklılıktan meydana gelmiş olabilir. Ayrıca üveit prevalansının görece yüksek olarak izlendiği ülkemizde Gemini'nin Türkçe sorulardaki başarısı, bu alanda bizlere daha doğru bilgi vermede aracılık yapabileceği düşüncesini akıllara getirmiştir.

Önemli bilgileri barındıran ve temel kitaplar arasında yer alan Amerikan Akademi ve Oftalmoloji 2023-2024 BCSC Üveit ve Oküler İnflamasyon Kitabı'nın çalışma soruları 36 soruyu içeriyordu ve sohbet robotlarına bu soruların tamamını uyguladık. Soru sayısının az olmasının istatistiksel sonucun anlamlı çıkıp çıkmamasına etki edebileceğini ön görmekteyiz. Fakat temel bilgileri ölçen bu kitabın sorularına ilave soru eklemeyi uygun bulmadık. Daha fazla soru

içeren testlerde farklı değerlerin çıkıp çıkmayacağı-
nın araştırılması gerektiğini düşünmekteyiz.

Soru sayımızın azlığı, buna bağlı olarak soruları
alt konu dallarına ayıramamamız ve birbirini üstün-
lüklerinin varlığını araştıramamamız çalışmamızın en
önemli kısıtlı yönlerini oluşturmaktadır.

SONUÇ

Bildiğimiz kadarı ile oküler inflamasyon ve üveit ala-
nındaki İngilizce ve Türkçe aynı sorularda yapay
zekâ sohbet botlarının performansını inceleyen ilk ça-
lışmadır. Yapay zekâ sohbet botlarının bilgi düzey-
lerinin iyileştirilmesine ek olarak bu programların
farklı dillerdeki anlama yorumlama ve cevaplama ka-
biliyetlerinin de geliştirilmeye ihtiyacı vardır.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğru-
dan 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, her-
hangi bir firmada çalışma durumu, hissedarlık ve benzer du-
rumları yoktur.

Yazar Katkıları

Fikir/Kavram: Eyüpcan Şensoy; **Tasarım:** Eyüpcan Şensoy; **De-
netleme/Danışmanlık:** Eyüpcan Şensoy, Mehmet Çıtırık; **Veri
Toplama ve/veya İşleme:** Eyüpcan Şensoy; **Analiz ve/veya
Yorum:** Eyüpcan Şensoy, Mehmet Çıtırık; **Kaynak Taraması:**
Eyüpcan Şensoy; **Makalenin Yazımı:** Eyüpcan Şensoy; **Eleştirel
İnceleme:** Eyüpcan Şensoy, Mehmet Çıtırık; **Kaynaklar ve Fon
Sağlama:** Eyüpcan Şensoy; **Malzemeler:** Eyüpcan Şensoy.

KAYNAKLAR

- Standardization of Uveitis Nomenclature (SUN) Working Group. Development of Classification Criteria for the Uveitides. Am J Ophthalmol. 2021;228:96-105. [Crossref] [PubMed] [PMC]
- Solomon L, Tsegaw A. Pattern of uveitis at a tertiary eye care and training center, North-West Ethiopia. Ocul Immunol Inflamm. 2022;30(7-8):1848-52. [Crossref] [PubMed]
- Rojas-Carabali W, Cifuentes-González C, Wei X, Putera I, Sen A, Thng ZX, et al. Evaluating the diagnostic accuracy and management recommendations of ChatGPT in uveitis. Ocul Immunol Inflamm. 2024;32(8):1526-31. [Crossref] [PubMed]
- Jabs DA, Dick A, Doucette JT, Gupta A, Lightman S, McCluskey P, et al; Standardization of Uveitis Nomenclature Working Group. Interobserver agreement among uveitis experts on uveitis diagnoses: the standardization of uveitis nomenclature experience. Am J Ophthalmol. 2018;186:19-24. [Crossref] [PubMed] [PMC]
- Agrawal R, Ludi Z, Betzler BK, Testi I, Mahajan S, Rousset A, et al. The Collaborative Ocular Tuberculosis Study (COTS) calculator—a consensus-based decision tool for initiating antitubercular therapy in ocular tuberculosis. Eye (Lond). 2023;37(7):1416-23. [Crossref] [PubMed] [PMC]
- Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med. 2023;6(1):75. [Crossref] [PubMed] [PMC]
- Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. JAMA. 2023;329(8):637-9. [Crossref] [PubMed]
- Smith WM, Armbrust Karen R, Dahr Sam S, Dodds EM, Gangaputra Sapna, Knickelbein Jared E, et al. Uveitis and Ocular Inflammation. San Francisco: American Academy of Ophthalmology; 2023.
- 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. [Crossref] [PubMed] [PMC]
- Bing Chat [Internet] Microsoft Edge. © Microsoft 2024 [Cited: July 4, 2024]. Available from: [Link]
- Waisberg E, Ong J, Masalkhi M, Zaman N, Sarker P, Lee AG, et al. Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology. Eye (Lond). 2024;38(4):642-5. [Crossref] [PubMed] [PMC]
- Google AI updates: Bard and new AI features in Search.(accessed July 4, 2024) [Link]
- Rahimy E. Deep learning applications in ophthalmology. Curr Opin Ophthalmol. 2018;29(3):254-60. [Crossref] [PubMed]
- Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. 2019;1(8):9. [Link]
- Jeblick K, Schachtner B, Dextl J, Mittermeier A, Stüber AT, Topalis J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. Eur Radiol. 2024;34(5):2817-25. [Crossref] [PubMed] [PMC]
- Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci. 2023;39(2):605-7. [Crossref] [PubMed] [PMC]
- Rojas-Carabali W, Sen A, Agarwal A, Tan G, Cheung CY, Rousselot A, et al. Chatbots vs. human experts: evaluating diagnostic performance of chatbots in uveitis and the perspectives on AI adoption in ophthalmology. Ocul Immunol Inflamm. 2024;32(8):1591-8. [Crossref] [PubMed]
- Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. JAMA Ophthalmol. 2023;141(6):589-97. [Crossref] [PubMed] [PMC]
- 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. [Crossref] [PubMed]
- 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. Türkiye Klinikleri J Ophthalmol. 2024;33(3):163-70. [Crossref]
- Mihalache A, Grad J, Patil NS, Huang RS, Popovic MM, Mallipatna A, Kertes PJ, Muni RH. Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment. Eye (Lond). 2024;38(13):2530-5. [Crossref] [PubMed] [PMC]