**Özgün Araştırma / Original Article**

# Gelişen Zihinler: Oftalmoloji Eğitiminde Doğal Öğrenme, Yapay Öğrenmeye Karşı

## Evolving Minds: Natural Learning vs. Artificial Learning in Ophthalmology Training

Ali Safa Balcı[1], Zeliha Yazar[2], Banu Turgut Öztürk[3], Çiğdem Altan[4]

[1]Sağlık Bilimleri Üniversitesi, Sancaktepe Şehit Prof. Dr. İlhan Varank Eğitim ve Araştırma Hastanesi, Göz Hastalıkları Kliniği, İstanbul, Türkiye
[2]Ankara Bilkent Şehir Hastanesi, Göz Hastalıkları Kliniği, Ankara, Türkiye
[3]Selçuk Üniversitesi Tıp Fakültesi, Göz Hastalıkları Anabilim Dalı, Konya, Türkiye
[4]Sağlık Bilimleri Üniversitesi, Beyoğlu Göz Eğitim ve Araştırma Hastanesi, Göz Hastalıkları Kliniği, İstanbul, Türkiye

## Öz

**Amaç:** Bu çalışma, Türkiye genelinde yapılan oftalmoloji asistan eğitim sınavlarında ChatGPT'nin yıl bazında gösterdiği başarı değişimini, aynı dönemde asistanların gösterdiği başarı değişimiyle karşılaştırmayı amaçlamaktadır.

**Gereç ve Yöntem:** Bu gözlemsel çalışmaya, Türk Oftalmoloji Yeterlilik Kurulu tarafından 2023 ve 2024 yıllarında düzenlenen Uzmanlık Eğitim Gelişim Sınavı'na her iki yılda da katılan oftalmoloji asistanları dâhil edilmiştir. 2023 sınavı 69 tek doğru yanıtlı çoktan seçmeli sorudan oluşmakta olup ChatGPT-3.5'e sorulmuştur. 2024 sınavı ise 72 sorudan oluşmakta olup ChatGPT-4o'ya sorulmuştur. Her iki sınava katılan asistanlar ile ChatGPT'nin başarı yüzdeleri karşılaştırılmıştır.

**Bulgular:** ChatGPT'nin doğruluk oranı 2023'te %53,6 iken 2024'te %84,7'ye yükselmiştir. Her iki yılda da sınava katılan toplam 501 asistanın ortalama puanı %48,2'den %53,1'e çıkmıştır. ChatGPT, 2023 yılında 292. sırada yer alırken 2024'te birinci olmuştur. Başarı yüzdesindeki artışa göre sıralandığında ChatGPT-4o genel sıralamada 8. olmuştur. ChatGPT açısından en büyük gelişim şaşılık (%75 artış), nöro-oftalmoloji (%40 artış) ve optik (%40 artış) alanlarında görülmüştür. Asistanlar arasında en fazla gelişim okuloplasti alanında (%33,5 artış) olurken, kornea ve oküler yüzey alanında %4,1 oranında bir düşüş gözlenmiştir.

**Sonuç:** ChatGPT-4o, selefiyle karşılaştırıldığında oftalmoloji sorularında belirgin bir başarı artışı göstermiştir; buna karşın, asistanların öğrenme süreci daha kademeli ilerlemiştir. ChatGPT'deki bu hızlı gelişim, yapay öğrenmenin belirli sınırlar içinde çok hızlı ilerleyebileceğini ortaya koymaktadır. Buna karşılık insan öğrenmesi daha derinlikli ve zaman gerektiren bir süreçtir. Elde edilen sonuçlar, gelişen büyük dil modellerinin tıp eğitimi ve klinik karar destek sistemlerinde giderek daha önemli roller üstleneceğini göstermektedir.

**Anahtar Kelimeler:** Eğitim, üretken yapay zekâ, asistan eğitimi

## Abstract

**Objectives:** This study aimed to compare year-over-year change in ChatGPT's performance on nationwide ophthalmology exams with the performance change among residents over the same period.

**Materials and Methods:** This observational study included ophthalmology residents in Türkiye who participated in both the 2023 and 2024 Resident Training Development Exams organized by the Turkish Ophthalmological Association Qualifications Committee. The 2023 examination consisted of 69 single-best-answer multiple-choice questions and was administered to ChatGPT-3.5. The 2024 version, containing 72 questions, was administered to ChatGPT-4o. The success rates of ChatGPT and the residents who participated in both exams were compared.

**Results:** ChatGPT's accuracy improved from 53.6% in 2023 to 84.7% in 2024. Among the 501 residents who participated in both years, the average score increased from 48.2% to 53.1%. ChatGPT ranked 292nd among residents in 2023 but achieved the top score in 2024. Based on percentage improvement in scores, ChatGPT-4o ranked 8th overall. The most notable performance gains for ChatGPT were seen in the areas of strabismus (+75%), neuro-ophthalmology (+40%), and optics (+40%). Among residents, the largest improvement occurred in oculoplastics (+33.5%), while a decrease was observed in cornea and ocular surface (-4.1%).

**Conclusion:** ChatGPT-4o showed a marked improvement in answering ophthalmology questions compared to its predecessor,

**Cite this article as:** Balcı AS, Yazar Z, Turgut Öztürk B, Altan Ç. Evolving Minds: Natural Learning vs. Artificial Learning in Ophthalmology Training. Turk J Ophthalmol. 2026;56:1-7

**Yazışma Adresi/Address for Correspondence:** Ali Safa Balcı, Sağlık Bilimleri Üniversitesi, Sancaktepe Şehit Prof. Dr. İlhan Varank Eğitim ve Araştırma Hastanesi, Göz Hastalıkları Kliniği, İstanbul, Türkiye
**E-posta:** alisafabalci@gmail.com
**ORCID-ID:** orcid.org/0000-0001-9161-7839

**Geliş Tarihi/Received:** 02.08.2025
**Revizyon Talebi/Revision Requested:** 25.09.2025
**Son Revizyon Alınma/Last Revision Received:** 27.10.2025
**Kabul Tarihi/Accepted:** 11.12.2025
**Yayın Tarihi/Publication Date:** 18.02.2026

**DOI:** 10.4274/tjo.galenos.2025.16517

## Abstract

whereas resident learning progressed more gradually. This rapid advancement in ChatGPT highlights the potential speed with which artificial learning can progress within defined boundaries. In contrast, human learning remains a deeper and more time-intensive process. Results suggest that evolving large language models will play an increasingly significant role in medical education and clinical support.

**Keywords:** Education, generative artificial intelligence, resident training

## Giriş

OpenAI tarafından geliştirilen ChatGPT gibi büyük dil modelleri, insan benzeri yanıtlar üretebilme yeteneğine sahip, ileri düzey yapay zekâ sistemleridir ve doğal dil işleme tekniklerini temel alarak çalışmaktadır. Bu modeller, üretici ön eğitimli dönüştürücü ("*Generative Pre-trained Transformer*"; GPT) mimarisi üzerine inşa edilmiş olup, oldukça geniş ve çeşitli metin veri kümeleri üzerinde eğitilerek bağlamsal olarak tutarlı ve anlamlı yanıtlar üretebilme kapasitesine ulaşmışlardır. ChatGPT-3.5 ve ChatGPT-4.0 gibi önceki sürümler, doğal dil anlama alanında önemli ilerlemeler sağlamış olsa da, 13 Mayıs 2024 tarihinde kullanıma sunulan ChatGPT-4o, dilsel doğruluk ve etkileşimsel performans açısından önceki sürümlere kıyasla kayda değer bir gelişim göstermektedir.[1] Tıbbi, eğitsel ve akademik bağlamlardaki yetkinliklerinin artmasıyla birlikte ChatGPT, yüksek doğruluk ve yanıt verebilirlik düzeyi sergilemektedir.[2] Bununla birlikte, özellikle tıp alanındaki yanıtlarının, hem sağlık profesyonelleri hem de hastalar açısından klinik güvenilirliğinin sağlanabilmesi için sürekli olarak değerlendirilmesi gerekmektedir.

Tıp alanında ChatGPT'nin hızla artan popüleritesi, sağlıkla ilişkili çeşitli görevlerdeki işlevselliğini değerlendirme yönünde artan bir ilgiye yol açmıştır. Oftalmoloji alanında yapılan araştırmalar, ChatGPT'nin alt uzmanlık alanlarına yönelik sorulara verdiği yanıtların doğruluğunu inceleyerek, bu yapay zekâ modelinin tamamlayıcı bir eğitim aracı olarak kullanılabileceğine dair önemli bulgular sunmuştur.[3,4,5] Ayrıca, kornea ülseri, katarakt yönetimi ve retinal patolojiler gibi klinik senaryolarda oftalmolojik durumları yorumlama ve yönetme yeteneği de araştırılmıştır.[6,7,8] ChatGPT, yalnızca tanı koymada değil, aynı zamanda tıbbi raporların hazırlanması gibi belgeleme süreçlerinde ve karmaşık oftalmolojik bilgilerin daha anlaşılır ve erişilebilir eğitim içeriklerine dönüştürülmesinde de yardımcı olabilecek bir araç olarak dikkat çekmektedir.[9]

Bu çalışmada, art arda iki yıl benzer formatta uygulanan asistanlara yönelik oftalmoloji sınavında hem iki yıl üst üste sınava giren asistanların hem de ChatGPT'nin bilgi düzeyi ve performansını değerlendirmeyi; ayrıca iki grubun başarı değişimlerini birbirleriyle karşılaştırmayı amaçladık.

## Gereç ve Yöntem

26 Mayıs 2023 tarihinde, Türk Oftalmoloji Derneği Yeterlik Kurulu tarafından "Uzmanlık Eğitimi Gelişim Sınavı"nın üçüncüsü gerçekleştirildi. Önceki çalışmamızda, bu sınavın sorularını ChatGPT'nin önceki sürümü olan ChatGPT-3.5'e yanıtlatmış ve performansını sınava ülke çapında katılan oftalmoloji asistanlarının sonuçlarıyla karşılaştırmıştık.[10] Bir yıl sonra, 31 Mayıs 2024 tarihinde "Asistan Eğitimi Gelişim Sınavı"nın dördüncüsü gerçekleştirildi. Türkiye genelindeki 80 eğitim merkezinden toplam 1.013 oftalmoloji asistanı bu sınava katıldı. Sınav soruları, ChatGPT'nin en güncel sürümü olan ChatGPT-4o'ya soruldu. Önceki yıl yapılan sınavla karşılaştırma yapılabilmesini sağlamak amacıyla, çalışmaya yalnızca her iki yıl da sınava katılmış olan asistanlar dahil edildi. Asistanlar, 2024 sınav tarihi itibarıyla eğitim sürelerine göre yıllara ayrılarak gruplandırıldı.

Her iki sınav, Türk Oftalmoloji Derneği Yeterlik Kurulu tarafından aynı alt uzmanlık alanlarını kapsayacak şekilde ve benzer bilişsel zorluk düzeyinde hazırlanmıştır. Her sınavda toplam 75 soru yer almış; her iki yılda da tek bir doğru cevabı çoktan seçmeli sorular dışındaki sorular çalışma dışında bırakılmıştır. Böylece değerlendirme 2023 yılı için 69, 2024 yılı için 72 geçerli soru üzerinden yapılmıştır. 2023 ve 2024 yıllarında sorulan soruların branşlara göre dağılımı sırasıyla Tablo 1 ve Tablo 2'de verilmiştir.

Sorular İngilizce'ye çevrildikten sonra, ChatGPT-4o'nun (model tanımlayıcısı: gpt-4o-2024-05-13) internet sitesindeki (https://chat.openai.com) resmî web arayüzü kullanılarak 21 Mart 2025 tarihinde bireysel oturumlarda yöneltildi. Her soru için sistem geçmişi sıfırlandı. Görsel veya grafiksel içerik içeren soru bulunmadığından, ek bir metinleştirme ya da tanımlama süreci uygulanmamıştır. Büyük dil modellerinin sürekli güncellenen yapısının sonuçlar üzerindeki etkisini önlemek amacıyla, her soruya şu yönerge eşlik etti: "Answer the following question using the knowledge available as of May 31, 2024."

ChatGPT-4o'nun her bir soru için verdiği yanıtlar ve açıklamalar kaydedildi. Her bir yanıt, önceden belirlenmiş cevap anahtarına göre doğru ya da yanlış olarak değerlendirildi. Hem asistanlar hem de ChatGPT-4o, doğru yanıt sayısına göre 100 üzerinden puanlandırıldı. Ayrıca, doğru yanıt sayısına göre bir sıralama oluşturuldu. Hem her iki yıl da sınava giren asistanlar hem de ChatGPT için genel ve yan dal bazında performans değişimleri analiz edildi. ChatGPT ve asistanlar arasında sıralama yapılırken

**Tablo 1. Hem ChatGPT-3.5'in hem de her iki yıl da sınava giren asistanların 2023 sınavında yan dal kategorilerine göre ortalama doğru cevap sayısı**

| Yan dallar (soru sayısı) | Tüm asistanlar (n=501) | 1. yıl asistanları (n=249) | 2. yıl asistanları (n=132) | 3. yıl asistanları (n=120) | ChatGPT-3.5 |
|---|---|---|---|---|---|
| Lens ve katarakt (n=9) | 3,82±1,64 | 3,16±1,36 | 4,21±1,63 | 4,77±1,61 | 7 |
| Kornea/oküler yüzey/ön segment (n=9) | 4,55±1,24 | 4,4±1,29 | 4,57±1,24 | 4,83±1,09 | 4 |
| Glokom (n=8) | 3,34±1,41 | 2,92±1,34 | 3,54±1,4 | 4,01±1,29 | 4 |
| Nöro-oftalmoloji (n=5) | 2,21±0,946 | 2,12±0,93 | 2,2±0,94 | 2,41±0,97 | 3 |
| Oküloplasti (n=4) | 1,30±0,83 | 1,11±0,78 | 1,35±0,76 | 1,63±0,89 | 2 |
| Pediatrik oftalmoloji ve şaşılık (n=8) | 3,96±1,51 | 3,62±1,46 | 4,27±1,49 | 4,32±1,52 | 0 |
| Optik (n=5) | 2,07±1,13 | 1,91±1,06 | 2,12±1,16 | 2,36±1,19 | 3 |
| Retina (n=16) | 8,98±2,53 | 7,85±2,27 | 9,68±2,33 | 10,56±2,15 | 11 |
| Üveit (n=5) | 3,0±1,14 | 2,79±1,11 | 3,01±1,16 | 3,45±1,08 | 3 |
| Toplam (n=69) | 33,24±7,32 | 29,88±6,31 | 34,95±6,34 | 38,33±6,74 | 37 |

**Tablo 2. Hem ChatGPT-4o'nun hem de her iki yıl da sınava giren asistanların 2024 sınavında yan dal kategorilerine göre ortalama doğru cevap sayısı**

| Yan dallar (soru sayısı) | Tüm asistanlar (n=501) | 2. yıl asistanları (n=249) | 3. yıl asistanları (n=132) | ≥4. yıl asistanları (n=120) | ChatGPT-4o |
|---|---|---|---|---|---|
| Lens ve katarakt (n=9) | 4,14±1,59 | 3,81±1,50 | 4,21±1,46 | 4,73±1,71 | 8 |
| Kornea/oküler yüzey/ön segment (n=11) | 5,11±1,90 | 4,59±1,81 | 5,39±1,89 | 5,87±1,76 | 9 |
| Glokom (n=7) | 3,34±1,28 | 3,11±1,21 | 3,47±1,25 | 3,66±1,36 | 5 |
| Nöro-oftalmoloji (n=4) | 1,84±1,05 | 1,78±1,05 | 1,87±1,06 | 1,95±1,03 | 4 |
| Oküloplasti (n=7) | 4,61±1,32 | 4,20±1,28 | 4,74±1,30 | 5,32±1,07 | 6 |
| Pediatrik oftalmoloji ve şaşılık (n=8) | 4,34±1,59 | 3,96±1,49 | 4,52±1,64 | 4,93±1,52 | 6 |
| Optik (n=5) | 2,21±0,98 | 2,04±0,92 | 2,26±0,97 | 2,49±1,06 | 5 |
| Retina (n=16) | 9,30±2,46 | 8,62±2,33 | 9,52±2,52 | 10,45±2,20 | 14 |
| Üveit (n=5) | 3,20±1,17 | 2,96±1,07 | 3,24±1,21 | 3,66±1,16 | 4 |
| Toplam (n=72) | 38,20±8,47 | 35,12±7,07 | 39,42±8,99 | 43,25±7,82 | 61 |

her yılın kendi katılımcı popülasyonu temelinde hesaplama yapıldı. Yıllar arası performans farkları, her iki sınava da katılan 501 asistanın ortalama doğruluk oranları üzerinden değerlendirilmiştir.

Katılımcı bilgileri anonimleştirildiği ve herhangi bir kişisel veri kullanılmadığı için etik kurul onayı alınmadı.

### İstatistiksel Analiz

İstatistiksel analizler SPSS Version 26 (IBM, Armonk, NY, ABD) programı kullanılarak gerçekleştirildi. Verilerin dağılımını ve normalliğini değerlendirmek için Kolmogorov-Smirnov testi uygulandı. Asistanlara ait veriler bireysel bazda değil, 501 asistanın ortalama doğruluk oranı üzerinden değerlendirilmiştir. ChatGPT ile asistan grubu arasındaki karşılaştırmalar istatistiksel test düzeyinde değil, tanımlayıcı olarak yapılmıştır. ChatGPT tek bir model çıktısı sunduğundan varyasyon hesaplanmamış, farklar yalnızca yüzdelik doğruluk oranları üzerinden karşılaştırılmıştır. Sürekli değişkenler, ortalama ± standart sapma ve minimum-maksimum değerler şeklinde sunuldu. Asistanların 2023 ve 2024 sınavları arasındaki genel ve yan dal bazlı doğru cevap yüzdelerindeki değişimi analiz etmek için Wilcoxon işaretli sıralar testi kullanıldı. Asistan katılımcı grubuna ve ChatGPT modellerine ait doğruluk oranları için %95 güven aralıkları (GA) hesaplandı. Yan dallara ait karşılaştırmalarda tip I hata olasılığını azaltmak amacıyla Bonferroni düzeltmesi yapıldı ve anlamlılık düzeyi $p<0,005$ olarak belirlendi.

## Bulgular

Her iki yıl da sınava katılan toplam asistan sayısı 501 idi. 2024 yılı itibarıyla eğitim süresine göre dağılım

incelendiğinde, 249 asistanın 12 ila 23 ay, 132 asistanın 24 ila 35 ay ve 120 asistanın 36 ay ve üzeri deneyime sahip olduğu görüldü. Asistanların ortalama eğitim süresi 28,4±10,6 ay (13-64 ay aralığında) olarak hesaplandı. 2024 sınavında her iki yıl da sınava katılan asistanlar, 72 sorudan ortalama 38,2±8,5'ini doğru yanıtlayarak %53,1'lik (%95 GA: %52,2-%54,0) bir başarı oranı elde etti. İkinci yıl asistanları ortalama 35,1±7,1 doğru cevapla %48,8 (95% GA: %47,5-%50,1), üçüncü yıl asistanları 39,4±8,9 doğru cevapla %54,8 (%95 GA: %53,3-%56,3) ve dördüncü yıl asistanları ise 43,3±7,8 doğru cevapla %60,1 (%95 GA: %58,7-%61,5) başarı oranına ulaştı. Buna karşılık, ChatGPT-4o 72 sorunun 61'ine doğru yanıt vererek %84,7 (%95 GA: %74,7-%91,3) doğruluk oranı elde etti. 2023 sınavında ChatGPT-3.5, doğru cevap sayısına göre 292. sırada yer alırken; 2024 sınavında ChatGPT-4o, tüm asistanlardan daha yüksek bir puan elde etti. Her iki yıl da sınava katılan asistanların ve ChatGPT-3.5'in 2023 yılına ait alt uzmanlık alanlarına göre ortalama doğru cevap sayıları Tablo 1'de, aynı asistanların ve ChatGPT-4o'nun 2024 yılına ait ortalamaları ise Tablo 2'de sunulmuştur.

Asistanlar, önceki yıla kıyasla çoğu yan dalda doğru cevap yüzdelerinde genel bir artış göstermiş olsa da bu artış nöro-oftalmoloji alanında istatistiksel olarak anlamlı bulunmamıştır (p=0,655). Asistanların performansının azaldığı tek yan dal kornea ve oküler yüzey hastalıkları olup, bu düşüş istatistiksel olarak anlamlıydı (p<0,001). Buna karşılık, ChatGPT tüm yan dallarda doğru cevap yüzdesini artırmıştır. Önceki yıla göre ChatGPT'nin doğru cevap yüzdesi %30,4 oranında artış göstermiştir. İki yıl da sınava giren tüm asistanlar ve ChatGPT, doğru cevap yüzdesindeki artışa göre sıralandığında, ChatGPT listenin 8. sırasında yer almıştır. İki sınav arasındaki doğru cevap yüzdesindeki değişimler Tablo 3'te özetlenmiştir.

## Tartışma

Bu çalışma, Türkiye'de art arda iki yıl ulusal düzeyde gerçekleştirilen asistanlık eğitim sınavları temelinde, oftalmoloji asistanları ile bir büyük dil modelinin yıl bazında performans değişimini ve başarı düzeylerindeki ilerlemeyi değerlendirmeyi amaçlamıştır. Bu yönüyle, doğal öğrenme ile yapay öğrenme arasındaki karşılaştırmayı ortaya koymayı hedeflemiştir. Bulgularımız, 2023 ile 2024 yılları arasında asistanların ortalama performansının çoğu yan dalda artış gösterdiğini; buna karşın ChatGPT-4o'nun, selefi ChatGPT-3.5'e kıyasla tüm alanlarda tutarlı bir gelişim sergilediğini ve 2024 sınavında tüm insan katılımcıların üzerinde bir başarı elde ettiğini ortaya koymaktadır.

**Tablo 3. Her iki yıl da sınava giren asistanlar ve ChatGPT için iki sınav arasındaki doğru yanıt yüzdesindeki değişim (minimum–maksimum)**

| Yan dallar | Tüm asistanların yüzdelik değişimi (%) (n=501) | ChatGPT'nin yüzdelik değişimi (%) | p* |
|---|---|---|---|
| Lens ve katarakt | 3,48±21,38 (-66,67-66,67 arası) | 11,11 | **<0,001** |
| Kornea/oküler yüzey/ön segment | -4,06±20,78 (-57,58-72,73 arası) | 37,37 | **<0,001** |
| Glokom | 5,85±23,36 (-50,0-73,21 arası) | 21,43 | **<0,001** |
| Nöro-oftalmoloji | 1,82±29,73 (-80,0-80,0 arası) | 40,0 | 0,655 |
| Oküloplasti | 33,46±26,34 (-50,0-100,0 arası) | 35,71 | **<0,001** |
| Pediatrik oftalmoloji ve şaşılık | 4,74±22,63 (-62,50-75,0 arası) | 75,0 | **<0,001** |
| Optik | 2,67±27,91 (-80,0-80,0 arası) | 40,0 | 0,029 |
| Retina | 1,95±17,17 (-50,0-62,50 arası) | 18,75 | 0,011 |
| Üveit | 3,91±28,18 (-80,0-80,0 arası) | 20,0 | **0,002** |
| Toplam | 4,31±9,97 (-39,81-56,75 arası) | 30,38 | **<0,001** |

*Her iki yıl da sınava giren asistanların iki sınav arasındaki doğru yanıt yüzdelerindeki değişim: Wilcoxon işaretli sıralar testi

Sağlık alanında yapay zekânın yaygın şekilde benimsenmesi hem hastalar hem de sağlık profesyonelleri arasında bu araçlara tıbbi bilgi edinme ve eğitim desteği sağlama amacıyla artan bir yönelime yol açmıştır.[11,12] Kullanımları özellikle ChatGPT-4o gibi ileri düzey büyük dil modelleri aracılığıyla giderek yaygınlaştıkça, bu sistemlerin ürettiği yanıtların güvenilirliğini ve bilimsel doğruluğunu değerlendirmek her zamankinden daha önemli hale gelmektedir. Hızlı ve erişilebilir bilgi sunma avantajına sahip olsalar da, klinik karar verme süreçleri ve tıp eğitimi üzerindeki potansiyel etkileri, bu sistemlerin alanına özgü, kanıta dayalı sorulara verdikleri yanıtların titizlikle değerlendirilmesini gerekli kılmaktadır.

Yapay zekâ, sürekli gelişen ve öğrenen bir yapıya sahiptir. Göz içi lenslerle ilgili aynı sorular ChatGPT-4.0'a altı ay arayla sorulduğunda, yanıtlarının doğruluk oranının arttığı bildirilmiştir.[13] Tıbbi soruların sorulduğu bir başka çalışmada ise, ChatGPT tarafından başlangıçta yanlış yanıtlanan sorular belirli bir süre sonra tekrar sorulduğunda, modelin bu soruların çoğuna doğru yanıt verdiği bildirilmiştir.[14]

İnsan öğrenmesi deneyim, düşünme ve bağlamla şekillenen kademeli bir süreçken; ChatGPT gibi büyük dil modelleri, bilgiyi dönemsel olarak gerçekleştirilen geniş ölçekli yeniden eğitim döngüleri aracılığıyla edinir.[15] ChatGPT-4o gibi her yeni sürüm, giderek daha çeşitli, güncel ve alana özgü veri setlerinden elde edilen bilgilerle geliştirilen aşamalı bir ilerlemeyi yansıtır. Bu süreç, bilgi doğruluğu ve işlevsel başarı açısından hızlı ve etkili gelişmelere olanak tanır. Ancak bu gelişim, insan öğrenmesinde yer alan süreklilik, etik muhakeme ve deneyime dayalı derinlik gibi özellikleri taşımaz.[16] Buna karşılık, insanlar daha yavaş fakat çok daha bütüncül bir öğrenme sürecinden geçer. Bilgi yalnızca biçimsel eğitimle edinilmez; aynı zamanda deneme-yanılma, duygusal bağlam ve sosyal etkileşim yoluyla da şekillenir.[17] Özellikle tıp eğitiminde, bu tür öğrenme süreci klinik yargı, empati ve uyum yeteneği gibi nitelikleri geliştirir; bu özellikler ise günümüzdeki yapay zekâ sistemlerinin erişemediği alanlar arasında yer almaktadır.[18]

Büyük dil modelleri ile insanların oftalmolojiyle ilgili sorulardaki performansı, daha önce yapılan çalışmalarda da karşılaştırılmıştır.[19,20] 2020 ile 2023 yılları arasındaki oftalmoloji uzmanlık sınavı sorularının kullanıldığı başka bir çalışmada, büyük dil modellerinin doğruluk oranlarının dört yıl boyunca anlamlı bir değişim göstermediği bildirilmiştir.[21] Soruların büyük dil modellerine ne zaman yöneltildiği tam olarak belirtilmemiş olsa da, eğer tüm sorular yaklaşık olarak aynı dönemde sorulduysa, sınav yılları farklı olsa bile doğruluk oranlarının benzer kalması beklenir.[21] Taloni ve ark.[22] tarafından yapılan ve Amerikan Oftalmoloji Akademisi'ne ait BCSC soru setinden alınan 1.023 sorunun kullanıldığı çalışmada, ChatGPT-4.0 selefi ChatGPT-3.5'ten daha iyi performans göstermiştir. İnsan katılımcılar ise genel performans sıralamasında ikinci sırada yer almıştır. Benzer şekilde, Maino ve ark.[23] tarafından yapılan bir çalışmada, Avrupa Oftalmoloji Boardu Diploma Sınavı'nda daha önce uygulanmış 440 çoktan seçmeli soru değerlendirilmiş ve oftalmoloji asistanlarının ChatGPT-3.5'e kıyasla daha yüksek performans gösterdiği, ancak ChatGPT-4o'nun asistanlara kıyasla daha yüksek doğruluk oranı gösterdiği bildirilmiştir.

Bulgularımız genel olarak bu çalışmalarla uyumlu olmakla birlikte, önemli bir fark çalışma dizaynında ortaya çıkmaktadır. Önceki çalışmalar kesitsel bir değerlendirme yaklaşımı benimserken, bizim çalışmamız aynı asistan grubuna bir yıl arayla uygulanan benzer formatta iki ulusal sınav üzerinden yürütülmüş ve böylece zaman içindeki değişimi gözlemleme olanağı sağlamıştır. Ayrıca, yalnızca insan öğrenmesinin zaman içindeki gelişimini değil, aynı büyük dil modelinin ardışık sürümleri arasındaki performans değişimini de değerlendirdik. Bildiğimiz kadarıyla çalışmamız hem insan hem de yapay öğrenmenin zaman içindeki ilerleyişine paralel bir bakış sunan ilk çalışma olma özelliğini taşımaktadır.

Asistanların puanlarındaki genel artış, zaman içinde eğitimin verimliliğini gösteren olumlu bir göstergedir ve klinik deneyimle birlikte yapılandırılmış eğitim programlarının bilgi düzeyinin kalıcılığına katkı sağladığını düşündürmektedir. Dikkat çekici bir şekilde, istatistiksel olarak anlamlı bir gelişme gözlenmeyen tek yan dal nöro-oftalmoloji oldu. Bu alan, multidisipliner yapısı ve birçok eğitim merkezinde sınırlı klinik karşılaşma imkânı sunmasıyla bilinmektedir.[24] Ayrıca, asistanların performansının anlamlı düzeyde azaldığı tek alan kornea ve oküler yüzey hastalıkları oldu. Bu durum, eğitim müfredatında bu yan dala yeterince vurgu yapılmaması ya da klinik olgu sayısının yetersizliği gibi etkenlere işaret edebilir. Bu bulgular, özellikle güçlendirilmesi gereken alanları belirleyerek asistanlık eğitim programlarında gelecekte yapılacak düzenlemelere yön verebilir. Buna karşılık, ChatGPT-4o tüm yan dallarda güçlü bir performans sergilemiş ve ChatGPT-3.5'e kıyasla anlamlı düzeyde gelişim göstermiştir. Genel doğruluk oranı %84,7 olan ChatGPT-4o, asistan grubuna göre daha yüksek doğruluk ve tutarlılık düzeyi göstermiştir; ancak yıl bazındaki performans artışı açısından değerlendirildiğinde 8. sırada yer almıştır. Bu durum, büyük dil modellerinin tıp eğitiminde, özellikle sınav hazırlığı ve teorik bilgi desteği açısından, eğitim aracı olarak artan potansiyelini pekiştirmektedir. Ancak, bu modellerin tıbbi uygulamada önemli olan bağlamsal incelik, klinik yargı ve uygulamalı beceriler gibi unsurları içermediği göz önünde bulundurulmalıdır. Bu nedenle, bu tür yapay zekâ araçları geleneksel tıp eğitiminin yerini almaktan ziyade, onu destekleyici ve tamamlayıcı bir unsur olarak değerlendirilmelidir.

### Çalışmanın Kısıtlılıkları

Çalışmamızda belirtilmesi gereken bazı kısıtlılıklar bulunmaktadır. İlk olarak, her ne kadar zamana dayalı bir karşılaştırma yapılmış olsa da bireysel öğrenme ortamları, klinik deneyim düzeyi ve çalışma alışkanlıkları gibi değişkenlerin etkisi bilinememektedir. Her iki sınav içerik ve yapı açısından büyük ölçüde benzer olmakla birlikte, madde düzeyinde psikometrik eşitleme yapılmamıştır. Bu nedenle çalışma, yıllar arasındaki farkları mutlak değil, benzer koşullar altındaki göreceli başarı değişimi olarak değerlendirmektedir. Web tabanlı arayüz, uygulama sürümlerine kıyasla yanıt uzunluğu ve bağlam belleği üzerinde sınırlı kontrol sunmaktadır. Bu durum yanıt biçiminde küçük farklılıklara yol açabilmekte olup, metodolojik bir sınırlılık olarak dikkate alınmıştır. Her iki sınava da katılan asistanların seçilmesi, daha motive veya akademik eğilimli bireyleri içerebileceğinden olası bir seçilim yanlılığı oluşturabilir. Soru sayısının sınırlı olması ve çalışmanın tek bir ülkenin ulusal sınavına dayanması, bulguların farklı eğitim sistemlerine genellenebilirliğini kısıtlayabilir.

## Sonuç

Sonuç olarak, çalışmamız ChatGPT-4o'nun önceki sürümü olan ChatGPT-3.5'e kıyasla üstün bir performans sergilediğini ve 2024 ulusal oftalmoloji sınavında asistan grubuna kıyasla daha yüksek bir performans düzeyi gösterdiği ortaya konmuştur. Asistanlar zaman içinde kademeli bir gelişim göstermiş olsa da, ChatGPT-4o'nun tutarlı ilerleyişi büyük dil modellerinin gelişen yeteneklerini gözler önüne sermektedir. Bununla birlikte, doğruluk oranı yüksek olsa da bu modellerin zaman zaman hatalı veya yanıltıcı yanıtlar verebileceği unutulmamalıdır. Bu nedenle, tıp eğitimindeki rolleri tamamlayıcı nitelikte olmalı; insan eğitimiyle gelişen eleştirel düşünme ve deneyim temelli bilgilerin yerini almak yerine, onları destekleyici bir araç olarak değerlendirilmelidir.

### Etik

**Etik Kurul Onayı:** Gerekli değildir.

**Hasta Onayı:** Gerekli değildir.

### Beyan

### Yazarlık Katkıları

Konsept: A.S.B., Z.Y., B.T.Ö., Ç.A., Dizayn: A.S.B., Z.Y., B.T.Ö., Ç.A., Veri Toplama veya İşleme: A.S.B., Z.Y., Analiz veya Yorumlama: A.S.B., Literatür Arama: A.S.B., Yazan: A.S.B., Z.Y., B.T.Ö., Ç.A.

**Çıkar Çatışması:** Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

**Finansal Destek:** Çalışmamız için hiçbir kurum ya da kişiden finansal destek alınmamıştır.

## Kaynaklar

1. OpenAI. ChatGPT-4o: Advancements and new peatures. OpenAI Blog. Accessed June 11, 2025. Available from: https://openai.com/blog
2. Gumus Akgun G, Altan C, Balci AS, Alagoz N, Çakır I, Yaşar T. Using ChatGPT-4 in visual field test assessment. Clin Exp Optom. 2025;108:1031-1036.
3. Aydın FO, Aksoy BK, Ceylan A, Akbaş YB, Ermiş S, Kepez Yıldız B, Yıldırım Y. Readability and appropriateness of responses generated by ChatGPT 3.5, ChatGPT 4.0, Gemini, and Microsoft Copilot for FAQs in refractive surgery. Turk J Ophthalmol. 2024;54:313-317.
4. Durmaz Engin C, Karatas E, Ozturk T. Exploring the role of ChatGPT-4, BingAI, and Gemini as virtual consultants to educate families about retinopathy of prematurity. Children (Basel). 2024;11:750.
5. Bahar TS, Öcal O, Çetinkaya Yaprak A. Comparison of ChatGPT-4, Microsoft Copilot, and Google Gemini for pediatric ophthalmology questions. J Pediatr Ophthalmol Strabismus. 2025;62:428-434.
6. Gurnani B, Kaur K, Gireesh P, Balakrishnan L, Mishra C. Evaluating the novel role of ChatGPT-4 in addressing corneal ulcer queries: an AI-powered insight. Eur J Ophthalmol. 2025;35:1531-1541.
7. Sundaramoorthy S, Ratra V, Shankar V, Dorairajan R, Maskati Q, Fredrick TN, Ratra A, Ratra D. Conversational guide for cataract surgery complications: a comparative study of surgeons versus large language model-based chatbot generated instructions for patient interaction. Ophthalmic Epidemiol. 2025:1-8.
8. Balas M, Mandelcorn ED, Yan P, Ing EB, Crawford SA, Arjmand P. ChatGPT and retinal disease: a cross-sectional study on AI comprehension of clinical guidelines. Can J Ophthalmol. 2025;60:e117-e123.
9. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza Afrash M. Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. Int J Med Inform. 2024 Aug;188:105474.
10. Balci AS, Yazar Z, Ozturk BT, Altan C. Performance of ChatGPT in ophthalmology exam; human versus AI. Int Ophthalmol. 2024;44:413.
11. Sevgi M, Antaki F, Keane PA. Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. Br J Ophthalmol. 2024;108:1354-1361.
12. Balci AS, Çakmak S. Evaluating the Accuracy and Readability of ChatGPT-4o's responses to patient-based questions about keratoconus. Ophthalmic Epidemiol. 2025:1-6.
13. Aydin FO, Ermis S. Responses of ChatGPT-4 on intraocular lenses: an evolving artificial intelligence assessment. Clin Exp Optom. 2025:1-5.
14. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, Chang S, Berkowitz S, Finn A, Jahangir E, Scoville E, Reese T, Friedman D, Bastarache J, van der Heijden Y, Wright J, Carter N, Alexander M, Choe J, Chastain C, Zic J, Horst S, Turker I, Agarwal R, Osmundson E, Idrees K, Kieman C, Padmanabhan C, Bailey C, Schlegel C, Chambless L, Gibson M, Osterman T, Wheless L. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq [Preprint]. 2023:rs.3.rs-2566942.
15. Shi H, Xu Z, Wang H, Qin W, Wang W, Wang Y, Wang Z, Ebrahimi S, Wang H. Continual learning of large language models:

a comprehensive survey. arXiv preprint arXiv: 2404.16789. 2024. https://doi.org/10.48550/arXiv.2404.16789

16. Komasawa N, Yokohira M. Learner-centered experience-based medical education in an AI-driven society: a literature review. Cureus. 2023;15:e46883.
17. De Felice S, Hamilton AFC, Ponari M, Vigliocco G. Learning from others is good, with others is better: the role of social interaction in human acquisition of new knowledge. Philos Trans R Soc Lond B Biol Sci. 2023;378:20210357.
18. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. Med Teach. 2011;33:961-973.
19. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions. Br J Ophthalmol. 2024;108:1379-1383.
20. 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;38:1897-1902.
21. Bahir D, Zur O, Attal L, Nujeidat Z, Knaanie A, Pikkel J, Mimouni M, Plopsky G. Gemini AI vs. ChatGPT: a comprehensive examination alongside ophthalmology residents in medical knowledge. Graefes Arch Clin Exp Ophthalmol. 2025;263:527-536.
22. Taloni A, Borselli M, Scarsi V, Rossi C, Coco G, Scorcia V, Giannaccare G. Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. Sci Rep. 2023;13:18562.
23. Maino AP, Klikowski J, Strong B, Ghaffari W, Woźniak M, Bourcier T, Grzybowski A. Artificial Intelligence vs. Human Cognition: A Comparative Analysis of ChatGPT and candidates sitting the european board of ophthalmology diploma examination. Vision. 2025;9:31.
24. Mollan SP, Menon V, Cunningham A, Plant GT, Bennetto L, Wong SH, Dayan M. Neuro-ophthalmology in the United Kingdom: providing a sustainable, safe and high-quality service for the future. Eye (Lond). 2024;38:2235-2237.