

# Glokomun Medikal Tedavisi ile İlgili Sorulara ChatGPT Yanıtlarının Doğruluk, Kapsamlılık, Okunabilirlik Açısından Değerlendirilmesi

Berire Şeyma DURMUŞ ECE\*, Gözde HONDUR\*\*, Erdem YÜKSEL\*\*\*

## Öz

**Amaç:** Antiglomatöz tedavi kullanan hastaların, tedavi ile ilgili sıkça sorduğu sorulara ChatGPT tarafından verilen yanıtların doğruluğu, kapsamlılığı, okunabilirliğinin değerlendirilmesi.

**Gereç ve Yöntem:** ChatGPT 4o uygulamasına glokomun medikal tedavisi ile ilgili olarak internet kullanıcıları tarafından sıkça sorulan 20 soru yöneltildi. Soruların yanıtları birbirinin değerlendirmelerine kör 3 oftalmolog tarafından içeriğinin doğruluğu ve kapsamlılığı açısından 1-4 arasında puanlandı. Yanıtların okunabilirlik düzeyi Türkçe metinlerde okunabilirliğin ölçülmesi amacıyla Flesch formülü uyarlaması ile geliştirilmiş olan Ateşman Formülü kullanılarak skorlandı. Ek olarak yanıt metinleri doktora danışma tavsiyesi varlığı açısından değerlendirildi.

**Bulgular:** ChatGPT yanıtlarının toplam ortalama doğruluk puanı  $3,70 \pm 0,59$ , ortalama kapsamlılık puanı  $3,67 \pm 0,54$  olarak saptandı. Yanıtların 18'i (%90,0) hem doğruluk hem de kapsamlılık açısından 3 ve üzeri puan ortalamasına sahipti. Dokuz yanıtın (%45,0) hem doğruluk hem de kapsamlılık açısından üç değerlendireciden de tam puan (4 puan) aldığı saptandı. Ortalama okunabilirlik puanı  $43,79 \pm 6,91$  ile "zor" seviyesinde saptandı. Okunabilirlik zorluk düzeyleri yanıtların 4'ünde (%20,0) "orta güçlükte", 15'inde (%75,0) "zor", 1'inde (%5,0) "çok zor" seviyesinde idi. Yanıtların 12'sinde (%60,0) doktora başvurmak için teşvik ifadesinin yer aldığı saptandı.

**Sonuç:** ChatGPT'nin bazı sınırlamaları olmasına rağmen glokomun medikal tedavisi ile ilgili soruları yüksek doğruluk ve kapsamlılıkta yanıtladığını saptadık. Oftalmologların, tıbbi danışmanlık amacıyla ChatGPT kullanımının olası etkilerinin farkında olmaları, yanlışlıkları ve eksiklikleri düzeltmeleri açısından gereklidir.

**Anahtar Kelimeler:** Büyük dil modeli, ChatGPT, Glokom, Tıbbi tedavi, Yapay zeka

## Evaluation of Chatgpt Responses to Questions Regarding the Medical Treatment of Glaucoma in Terms of Accuracy, Comprehensiveness, and Readability

### Abstract

**Objectives:** To evaluate the accuracy, comprehensiveness, and readability of the responses provided by ChatGPT to frequently asked questions from patients undergoing antiglaucomatous treatment.

**Material and Method:** Twenty frequently asked questions from internet users regarding the medical treatment of glaucoma were posed to the ChatGPT 4o application. The responses were rated by three ophthalmologists, who were blinded to each other's evaluations, on a scale of 1 to 4 for accuracy and comprehensiveness. The readability level of the responses was assessed using the Ateşman Formula, an adaptation of the Flesch formula developed for measuring readability in Turkish texts. The results were classified according to a readability scale as very easy, easy, moderately difficult, difficult, and very difficult. Additionally, the response texts were evaluated for the presence of recommendations for consulting a doctor.

**Results:** The overall average accuracy score of the ChatGPT responses was  $3.70 \pm 0.59$ , while the average comprehensiveness score was  $3.67 \pm 0.54$ . Eighteen responses (90.0%) had an average score of 3 or higher in terms of both accuracy and comprehensiveness. Nine responses (45.0%) received a maximum score (4 points) from all three evaluators for both accuracy and comprehensiveness. The average readability score was  $43.79 \pm 6.91$ , indicating a "difficult" readability level. The readability difficulty levels were classified as follows: 4 responses (20.0%) were rated as "moderately difficult", 15 responses (75.0%) as "difficult", and 1 response (5.0%) as "very difficult". In 12 responses (60.0%), an encouragement to consult a doctor was included.

**Conclusion:** Despite some limitations of ChatGPT, we found that it answers questions regarding the medical treatment of glaucoma with high accuracy and comprehensiveness. It is essential for ophthalmologists to be aware of the potential impacts of using ChatGPT for medical consultation to correct inaccuracies and deficiencies.

**Keywords:** Large language model, ChatGPT, Glaucoma, Medical treatment, Artificial intelligence

\* Kastamonu Üniversitesi Tıp Fakültesi Göz Hastalıkları Ana Bilim Dalı, Kastamonu



\*\* Ulucanlar Göz Eğitim ve Araştırma Hastanesi Sağlık Bilimleri Üniversitesi, Ankara

\*\*\* Gazi Üniversitesi Tıp Fakültesi Göz Hastalıkları ABD, Ankara

**Yazışma Adresi:** Erdem Yüksel, Gazi Üniversitesi Tıp Fakültesi Göz Hastalıkları Ana Bilim Dalı, Ankara. e-posta: drerdemyuksel@gmail.com

**Geliş Tarihi:** 10.07.2025 **Revize Tarihi:** 18.07.2025 **Kabul Tarihi:** 26.09.2025

**ORCID No:** BŞDE: 0000-0003-3770-7520, GH: 0000-0002-3615-8884, EY:0000-0003-3713-8815

QR Kod	Bu makaleye online erişim
	<a href="https://medicalnetwork.com.tr">https://medicalnetwork.com.tr</a> • <a href="https://mnoftalmoloji.com.tr">https://mnoftalmoloji.com.tr</a> • <a href="https://mndijital.medicalnetwork.com.tr/category/mn-ofthalmoloji/">https://mndijital.medicalnetwork.com.tr/category/mn-ofthalmoloji/</a>
	<b>Bu çalışmanın kaynak olarak gösterimi:</b> Durmuş Ece BŞ. Hondur G. Yüksel E. Glokomun Medikal Tedavisi ile İlgili Sorulara ChatGPT Yanıtlarının Doğruluk, Kapsamlılık, Okunabilirlik Açısından Değerlendirilmesi. MN Oftalmoloji. 2025;32(4):239-245
	Copyright©: 2025 Durmuş Ece ve ark.. Bu eser; Creative Commons 4,0 Uluslararası lisansı ile lisanslanmıştır.

## Giriş

ChatGPT sorulara formüle edilmiş, inandırıcı ve insan benzeri yanıtlar verebilmesi ile dikkat çeken, makine öğrenme sistemine sahip büyük bir dil modelidir.<sup>1</sup> ChatGPT'nin giderek artan kullanımı ve farklı sağlık alanlarında kendine yer bulması, tıbbi uygulamalar ve hasta eğitim planlamalarında kullanım potansiyelinin bulunması, yanıtlarının içeriğini değerlendirmeyi gerekli kılmaktadır. Önceki bazı çalışmalar ChatGPT'nin tıbbi sınavlarda başarılı olabileceğini, hastaların tıbbi sorunlarına yanıt verebileceğini, tanı ve tedavi önerileri sağlayabileceğini göstermektedir.<sup>2-4</sup> Yapay zeka (YZ) teknolojisindeki ilerleme ile ChatGPT'nin farklı tıbbi amaçlar için kullanımı yaygınlaşmakla birlikte, tıp alanında kullanımının hasta güvenliği açısından tehlike oluşturup oluşturmayacağı, özellikle hastalar tarafından kullanıldığında doğru, kapsamlı ve anlaşılabilir bilgiler sunma yeteneği konusundaki araştırma ihtiyacı devam etmektedir.

Glokom, geri dönüşümsüz körlüğün önde gelen nedenlerinden biridir. Hastalığın tedavisinde temel hedef göz içi basıncının (GİB) düşürülmesidir. Bu amaçla medikal tedavi, lazer uygulamaları ve cerrahi yöntemler kullanılmaktadır. Açık açılı glokomda ilk tedavi olarak seçici lazer trabeküloplastisi kullanımı ile ilgili kanıtlar bulunmakla birlikte, tedavi yaklaşımında birinci basamak genellikle oküler hipotansif göz damlaları kullanımıdır.<sup>5</sup> Glokom progresyonu durdurmak veya yavaşlatmak için ilaç uyumu esastır ve glokomun medikal tedavisinde hasta uyumsuzluğu tedavi başarısızlığın en önemli nedenlerinden biridir.<sup>6</sup> Medikal tedaviye zayıf uyum, yetersiz ilaç etkinliği ile karıştırılabilir ve bu durum ek ilaç kullanımına veya cerrahi tedavi ihtiyacına neden olabilir.<sup>7</sup> Hasta uyumsuzluğu tedavi başarısının önündeki en önemli engellerden biridir. Bununla birlikte çalışmalar, hastaların medikal tedavinin önemi konusunda bilgi sahibi olmalarının ve hasta eğitim programları düzenlenmesinin tedaviye uyumu artırmada kritik rol oynadığını göstermektedir.<sup>8</sup> Glokom için medikal tedavi başlanmasını takiben hastalar genellikle tedaviyi başlatan hekim ya da eğitim hemşiresi tarafından bilgilendirilebilmektedir. Ancak bu eğitim bazen sınırlı zaman ve elverişsiz ortamlar nedeniyle kısıtlı kala-

bilir. Bir diğer sorun da hastaların glokom için medikal tedaviye başlamadan önce karşılaşılabilecekleri olası zorluklar hakkında henüz fikir sahibi olmamalarından kaynaklanan bilgi eksikliğinin, doktor viziti sırasında anlamlı sorular sormalarına engel oluşturmasıdır. Bu durum hastaların medikal tedavi ile ilgili yaşadıkları karışıklıklarda internete, hasta forumlarına ve online hasta eğitim materyallerine başvurması ile sonuçlanmaktadır.<sup>9</sup> YZ uygulamalarındaki gelişmelerle birlikte başvuru kaynakları arasına ChatGPT ve benzeri chatbotlar da eklenmiştir

Tıbbi uygulamalarda bilgi güvenilirliği hasta güvenliği için kritik olduğundan ChatGPT'nin sağlık alanındaki etkinliğini ve hasta tarafından kullanımının güvenilirliğinin sınırlarını tam olarak anlamak önemlidir. Bu sebeple çalışmamızın amacı, glokom sebebiyle medikal tedavi kullanan hastaların günlük hayatta karşılaştıkları durumlar ve damla kullanımını etkileyebilecek faktörle ilgili sıkça sorduğu sorulara ChatGPT uygulaması tarafından verilen yanıtların doğruluğu, kapsamlılığı, okunabilirliğinin değerlendirilmesidir.

## Gereç ve Yöntem

Bu çalışma için OpenAI tarafından geliştirilen, GPT-4 mimarisine sahip ChatGPT'nin Ekim 2024 sürümü kullanılmıştır. ChatGPT uygulamasına glokomun medikal tedavisi ile ilgili olarak glokom hastalarının sıkça sorulan 20 soru yöneltildi (Tablo 1). Sorular, glokom hastaları tarafından hasta forumları ve tıbbi danışmanlık sayfalarına yöneltilen sorular ile Google (Alphabet Inc.) uygulamasının "Diğer sorular" özelliği kullanılarak internet kullanıcıları tarafından sıkça sorulmuş sorular toplanarak oluşturuldu. Soruların hastalara ait demografik bilgileri içermemesi amacıyla hamilelik ve emzirme döneminde glokomun medikal tedavisi ile ilgili sorular çalışmadan hariç tutuldu. Çalışma için etik kurul onayı alındı (2024-KAEK-84/23.10.2024). Bu çalışma, Helsinki Bildirgesi'nin ilkelerine bağlı kalınarak yürütülmüştür ve çalışmada hasta verisi kullanılmadığı için bilgilendirilmiş onam alınmasına gerek duyulmamıştır.

**Tablo 1:** Glokomun medikal tedavisi ile ilgili sık sorulan sorular

1. Glokom damlalarının etkisini göstermesi ne kadar sürer?
2. Glokom tedavisinde verilen damlaları ömür boyu kullanmam gerekir mi?
3. Glokom damlaları, yaşadığım görme kaybını geri döndürebilir mi?
4. Glokom tedavisi için göz damlası kullanıyorum ancak herhangi bir iyileşme hissetmiyorum. Damlamı kullanmayı bırakmalı mıyım?
5. Glokom damlalarımı açtıktan sonra ne kadar süreyle kullanabilirim?
6. Glokom damlalarımı açtıktan sonra buzdolabında mı saklamalıyım?
7. Glokom tedavisi için günde iki kez göz damlası kullanıyorum. Sabah dozu almayı unuttum; akşam iki kez birden damla uygulamalı mıyım?
8. Glokom damlasını yanlışlıkla fazla uygularsam ne olur?
9. Glokom tedavisi için iki farklı göz damlası kullanıyorum. Bu damlaları arka arkaya kullanabilir miyim?
10. Glokom damlası uyguladıktan sonra gözümü kapatmam ya da ovuşturmam gerekir mi?
11. Glokom damlalarını damlattığımda gözümde yüzüme taşarsa ne olur?
12. Kontakt lens takarken glokom damlalarını kullanabilir miyim?
13. Glokom damlalarının daha etkili olması için yapabileceğim bir şey var mı?
14. Glokom tedavisinde kullanılan damlaların yan etkileri var mı?
15. Glokom damlaları göz rengimi değiştirir mi?
16. Glokom damlalarının uzun süreli kullanımı böbreklerime ya da karaciğerime zarar verir mi?
17. Gözlerim kurduğunda glokom damlalarımı daha fazla kullanabilir miyim?
18. Glokom tedavisi kullanırken düzenli göz muayenelerine gitmem gerekir mi? Gerekirse ne sıklıkla?
19. Muayene gününde glokom damlalarımı kullanmam gerekir mi?
20. Glokom randevuma bir ay var ancak damlalarım bitti. Yeni göz damlası almak için bir sonraki randevuma kadar bekleyebilir miyim?

Sorular, önceki konuşmaların yanıtları etkilememesini sağlamak amacıyla her bir soru için ayrı bir oturum kullanılarak 23 Ekim 2024 tarihinde ChatGPT'ye yöneltilmiş ve yanıtlar kaydedilmiştir (Bakınız Ek-1). Kaydedilen yanıtlar, farklı glokom birimlerinde görev yapan ve sırasıyla 19, 10 ve 10 yıllık deneyime sahip üç oftalmolog tarafından, birbirlerinin değerlendirmelerine kör olacak şekilde doğruluk ve kapsamlılık açısından değerlendirilmiş ve 1-4 arasında puanlanmıştır. Değerlendiriciler tarafından kullanılan puanlama ölçeği tablo 2'de gösterilmiştir. Ek olarak değerlendiricilerden hasta güvenliği açısından potansiyel risk taşıdığını düşündükleri yanıtları işaretlemeleri ve belirledikleri tehlikelerin gerekçelerini belirtmeleri talep edilmiştir.

Sorulara verilen yanıtların okunabilirlik düzeyini değerlendirmek için, Türkçe metinlerde okunabilirliğin ölçülmesi amacıyla Flesch formülü uyarlaması ile geliştirilmiş olan ve hece, kelime ve cümle sayısı üzerinden değerlendirme yapan Ateş-

man Formülü kullanıldı (Okunabilirlik Puanı=198,825 -40,175 x (toplam hece/toplam kelime) - 2.610 x (toplam kelime/toplam cümle)). Her bir yanıt metni, metne ait değişkenlerin denkleme yerleştirilmesi ile elde edilen sonucun okunabilirlik düzeyi ölçüğüne göre yorumlanmasıyla; 90-100 (çok kolay), 70-89 (kolay), 50-69 (orta güçlükte), 30-49 (zor), 1-29 (çok zor) olarak sınıflandı. <sup>10</sup> Ek olarak yanıt metinleri doktora danışmak için yapılan yönlendirmeler açısından değerlendirildi. Bunun için "doktorunuzla iletişime geçin, doktorunuza danışın veya doktorunuzla görüşün" ifadeleri ile açık bir şekilde yönlendirme içeren yanıtların sayısı belirlendi.

#### İstatistiksel Analiz

Veri analizi, tanımlayıcı istatistikler kullanılarak yapıldı. Tanımlayıcı analiz, SPSS Statistics yazılımı (sürüm 26.0, IBM Corp.) kullanılarak gerçekleştirildi. Doğruluk ve kapsamlılık puanları (değerlendirici ve yanıt ortalamaları), kelime sayısı ve okunabilirlik puanları ortalama ve standart sapma (SD) olarak

**Tablo 2:** ChatGPT-4o Yanıtlarının Değerlendirilmesine Yönelik Puanlama Ölçeği

	Değerlendirme Kriteri	Puanlama
Doğruluk	Yanıt tamamen doğru bilgiler içerir	4
	Yanıt çoğunlukla doğru bilgileri içermekler birlikte küçük hatalar içerir	3
	Yanıt bazı doğru bilgileri içermekle birlikte büyük oranda yanlış bilgi içerir	2
	Yanıt tamamen yanlış ve yanıltıcıdır	1
Kapsayıcılık	Yanıt tüm yönleri ile kapsayıcıdır	4
	Yanıt önemli bilgileri içermekle birlikte küçük detaylar eksiktir	3
	Yanıt soruyu bazı yönlerden açıklamış olmakla birlikte önemli bilgiler eksiktir	2
	Yanıtta gerekli bilgiler tamamen eksiktir	1

hesaplandı. Kategorik değişkenler ise sayı (n) ve yüzde (%) olarak gösterilmiştir. Üç değerlendirici arasındaki puanlama uyumunu değerlendirmek amacıyla sınıf içi korelasyon katsayısı (intraclass correlation coefficient, ICC) analizi yapıldı. Doğruluk ve kapsayıcılık puanlamaları için ICC değerleri ayrı ayrı hesaplandı ve %95 güven aralıkları ile raporlandı.

## Bulgular

ChatGPT yanıtlarının toplam ortalama doğruluk puanı  $3,70 \pm 0,59$ , ortalama kapsamlılık puanı  $3,67 \pm 0,54$  olarak saptandı. Değerlendiricilerin puan dağılımı Tablo 3'te gösterilmiştir. Yanıtlar için ayrı ayrı yapılan analizde hem doğruluk hem de kapsayıcılık açısından yanıtların 18'i (%90,0) 3 ve üzeri puan ortalamasına sahipti. Üç değerlendirici arasındaki uyum doğruluk puanlaması için orta düzeyde bulundu (ICC =0.56, %95 GA: 0,30-0,77), kapsayıcılık puanlaması için ise orta-iyi düzeyde bulundu (ICC =0.65, %95 GA: 0,41-0,82). Yanıtlardan 9'u (%45,0) tanesi her üç değerlendirici tarafından hem doğruluk hem de kapsayıcılık açısından tam puan (4 puan) ile puanlandı. Yanıtlardan hiçbirisi doğruluk ya da kapsayıcılık açısından 1 puan almadı. "Glokom damlaları göz rengimi değiştirir mi?" sorusunun yanıtı 2,33 puan ortalaması ile doğruluk açısından en düşük puan ortalamasına sahipti. Kapsamlılık açısından en

düşük puan ortalamasına (2,66) sahip yanıtlar "Glokom damlalarının etkisini göstermesi ne kadar sürer?" ve "Glokom damlalarını açtıktan sonra buzdolabında mı saklamalıyım?" sorularına aitti. Yanıtlardan hiçbirisi değerlendiriciler tarafından hasta güvenliği açısından potansiyel olarak tehlikeli bulunmadı. ChatGPT yanıtlarının "doktora danışma" tavsiyesi içermesi açısından yapılan değerlendirmede, yanıtların 12'sinde (%60,0) doktora başvurmak için teşvik ifadesinin yer aldığı saptandı.

ChatGPT yanıtlarının ortalama uzunluğu  $131,90 \pm 48,54$  kelime idi. Ortalama okunabilirlik puanı  $43,79 \pm 6,91$  ile "zor" düzeyde saptandı. Okunabilirlik zorluk düzeyi, yanıtların 4'ü (%20,0) için "orta güçlükte", 15'i (%75,0) için "zor", 1'i (%5,0) için "çok zor" olarak saptandı. Yanıtlardan hiçbirisi kolay veya çok kolay okunabilirlik sınıfında yer almadı (Tablo 4).

## Tartışma

Glokom son evreye kadar asemptomatik seyredebilmesi ve tedavinin sağladığı faydanın anlaşılmasının her zaman mümkün olmaması nedeniyle, tedavi motivasyonunun kritik olduğu oftalmolojik patolojilerdendir. Glokom hastaları medikal tedavinin kronik doğası nedeniyle tedaviye inancsızlık yaşayabilirler bile sağlık uzmanından yardım istemeyi zor bulabilirler.

**Tablo 3:** ChatGPT-4o Yanıtlarının Değerlendirme Sonuçları

Özellikler	Değerlendirici A	Değerlendirici B	Değerlendirici C
Doğruluk (Ortalama $\pm$ SS)	3,75 $\pm$ 0,55	3,80 $\pm$ 0,52	3,55 $\pm$ 0,68
Doğruluk Puan Dağılımı			
4	16/20	17/20	13/20
3	3/20	2/20	5/20
2	1/20	1/20	2/20
1	0/20	0/20	0/20
Kapsamlılık (Ortalama $\pm$ SS)	3,55 $\pm$ 0,60	3,85 $\pm$ 0,36	3,60 $\pm$ 0,59
Kapsamlılık Puan Dağılımı			
4	12/20	17/20	13/20
3	7/20	3/20	6/20
2	1/20	0/20	1/20
1	0/20	0/20	0/20

SS: Standart sapma

**Tablo 4:** ChatGPT-4o Yanıtlarının Okunabilirlik Özellikleri

Okunabilirlik Özellikleri	ChatGPT Yanıtları (n=20)
Kelime Sayısı (Ortalama $\pm$ SS)	131,90 $\pm$ 48,54
Okunabilirlik Puanı* (Ortalama $\pm$ SS)	43,79 $\pm$ 6,91
Okunabilirlik Zorluk Seviyesi* n (%)	
Orta Güçlükte (50-69)	4 (%20,0)
Zor (30-49)	15 (%75,0)
Çok Zor (1-29)	1 (%5,0)

SS: Standart sapma, \*Ateşman Formülü

Bilgi eksikliği, işitme engeli, hafıza sorunları gibi zorlukların varlığı da doktorlarından bilgi istemenin önündeki engellerdendir. Bununla birlikte farklı çalışmalarda hastaların teknoloji erişimlerine bağlı olarak, glokom hastalarının glokom hakkında internet ortamında sağlık bilgisi araştırmasının %43-73 oranlarında saptandığı bildirilmiştir.<sup>11,12</sup> Artan yapay zeka teknolojileri ile birlikte hastaların bilgi almak için eriştiği uygulamaların oftalmologlar tarafından tanınıp, doğruluğunun analizinin yapılması, yanlış bilginin yanlış teşhis ve uygunsuz tedaviye yol açabilmesi açısından önemlidir. Bu sebeple çalışmamız, glokomun medikal tedavisi ile ilgili sık sorulan sorulara ChatGPT yanıtlarının değerlendirilmesine odaklanmıştır.

Çalışmamızda ChatGPT'nin glokomun medikal tedavisi ile ilgili sağladığı yanıtların  $3,70 \pm 0,59$  doğruluk ortalaması ile yüksek bir doğruluğa sahip olduğunu saptadık. Bununla birlikte ChatGPT'nin yanlış içerikli yanıtlar sunduğu sorular mevcuttur. En düşük doğruluk puan ortalamasına sahip "Glokom damlaları göz rengimi değiştirir mi?" sorusunun yanıtında ChatGPT prostaglandin analoglarının iris rengi değişikliği yan etkisi ile ilgili yanlış bilgiler verilmiştir. Yanlış bilgiler içeren bazı yanıtlar olsa bile yanıtların %90,0'13 ve üzeri doğruluk puanına sahipti. Yalla ve ark.<sup>13</sup> Amerikan Oftalmoloji Akademisi'nin hasta eğitim broşüründen uyarlanan soruları kullandıkları çalışmalarında ChatGPT yanıtlarının %63'ünün "%76-%100 doğruluk" kategorisinde olduğunu saptamışlardır. Kerici ve ark.<sup>14</sup> glokom hakkında genel bilgi, tanı, tedavi ve takip ile ilgili soruların yanıtlarını değerlendirdikleri çalışmalarında ChatGPT tarafından verilen yanıtların %75'inin Glokom Terminolojisi ve Kılavuzu'na göre tamamen doğru olduğunu belirtmişlerdir. Wu ve ark.<sup>4</sup> glokomdaki bilgi eksiklikleri ile ilgili çalışmalarında ChatGPT yanıtlarının genel doğruluk oranını %67 olarak bildirmişlerdir. Doğruluğu saptamada ve puanlamadaki farklılıklara rağmen bu sonuçlar ChatGPT'nin hasta bilgilendirme konusundaki potansiyelinin umut vadettiğini göstermektedir.

Çalışmamızda ortalama kapsayıcılık puanını  $3,67 \pm 0,54$  saptadık. ChatGPT'nin sağlıkla ilgili sorulara kapsamlı yanıtlar verememesi, yanlış yanıtlar vermesiyle benzer olumsuz etkilere yol açabilir. Nitekim kapsayıcılık puanı en düşük sorulardan olan "Glokom damlalarının etkisini göstermesi ne kadar sürer?" sorusunun yanıtında hastalara tedavinin başlangıç etki süresi ile ilgili bilgi verilmekle birlikte maksimum GİB düşüş etkinliğinin sıklıkla 3-5 hafta sonra elde edileceği ile ilgili bilgi verilmemiştir. Hastaların bu bilgiden habersiz olmaları tedavi açısından potansiyel uyum sorunları oluşturabilir. Yanıtların kapsayıcılığı kabul edilebilir seviyede olmakla birlikte, mevcut eksiklikler YZ modelini geliştirmek için gerçekleştirilen eğitimin tıp alanı için spesifik olmaması ile ilişkili olabilir. ChatGPT yanıtlarının kapsamını geliştirmek için YZ modellerinde iyileştirmeler yapılması gerekli görünmektedir.

Hasta kullanımında ChatGPT tarafından sağlanan bilgilerin

doğru ve kapsamlı olması kadar önemli bir hususta açık ve anlaşılır olmasıdır. ChatGPT cevaplarının yeterince anlaşılır olmaması, hastaların yanlış çıkarımlarda bulunmasına sebep olabilir. Bu sebeple ChatGPT yanıtlarının okunabilirlik düzeyini araştırdık ve ortalama okunabilirlik puanının 43,79 ile "zor" seviyede olduğunu saptadık. ChatGPT'nin glokom ile ilgili İngilizce yanıtlarının okunabilirliği daha önce değerlendirilmiş olmakla birlikte, değerlendirdiğimiz yanıtların Türkçe metinler olması çalışmamızın önemli özelliklerindedir. Literatürde hem oftalmoloji alanında hem de oftalmoloji dışı sağlık alanlarında yapılmış ve sonuçlarımıza benzer şekilde ChatGPT yanıtlarının okunabilirlik düzeyinin zor olduğunu ve genel nüfus için karmaşık olabileceğini belirten başka çalışmalar da mevcuttur.<sup>15,16</sup> Bir çalışmada Wu ve ark.<sup>4</sup> ChatGPT yanıtlarını Flesch-Kincaid okunabilirlik testi ile değerlendirmişler ve ChatGPT yanıtlarının, 11-12. sınıf okuma düzeyinde okuduğunu anlama gerektirdiğine ancak ortalama bir Amerikalı'nın 7-8. sınıf okuma düzeyinde olduğuna dikkat çekmişlerdir. Bu sebeple okuyucunun daha basit açıklamalar elde etmesi için eğitim düzeyini paylaşarak bilgi talep etmesi önerisinde bulunmuşlardır.<sup>17</sup>

ChatGPT'nin internet ortamında glokom tedavisi ile ilgili bilgi arayan hastalara sağladığı bir avantaj, hasta bilgilendirme sitelerinin aksine sohbet botlarının hastaların sorularını etkileşimli bir şekilde yanıtlamasıdır. Bu olumlu katkıya rağmen ChatGPT'nin, hasta hekim ilişkisindeki empati ve etkileşimi sağlaması mümkün değildir. Yine de literatürdeki bazı çalışmalar ChatGPT yanıtlarını empati açısından değerlendirmiş ve ChatGPT'nin yüksek derecede empatiyle yanıtlar üretme yeteneği gösterebildiğini bildirmişlerdir.<sup>18,19</sup> Yanıtlarının soru sorma biçiminden etkilendiğini ve ChatGPT'nin empatiyle yanıtlar üretmesinin mümkün olabildiğini gösteren bu çalışmalara rağmen, ChatGPT her hasta için bireyselleşmiş öneriler sağlamak için gerekli olan deneyim ve klinik muhakeme yeteneğine sahip değildir. Bu açıdan hasta sorularının ChatGPT tarafından yanıtlanmasının etkisi, yıllar süren eğitim ve deneyime sahip oftalmologlar tarafından yanıtlanması ile benzer olamaz.

Çalışmamızda değerlendirdiğimiz yanıtların hiçbirinde hasta güvenliği açısından tehlike potansiyeli bulunan yanıt içeriği saptamadık. ChatGPT'nin doğruluk ve kapsamlılığının yeterli olmadığı yanıtlar olsa bile hasta güvenliği açısından açık tehlike potansiyeli taşıyan yanıtlar vermemiş olması sağlık danışmanlığında kullanımı açısından umut vadetmektedir. Bununla birlikte daha önceki çalışmalarda, ChatGPT'nin hastanın sağlığına zarar verebilecek tavsiyelerde bulunduğunu saptayan çalışmalar mevcuttur.<sup>13,20</sup> Bu farklılık yöneltilen sorunun içeriği ile ilgili olabilir. Nitekim ChatGPT'nin çalışmamızda olduğu gibi klinik yönetimler hakkında çoğunlukla uygun öneriler sunmasına rağmen, oftalmolojik acillerle ilgili soruların analiz edildiği bir çalışmada yanıtların %32'sinin hastaya zarar verme potansiyeline sahip olduğu saptanmıştır.<sup>21</sup>

Çalışmamızda ChatGPT'nin yanıtlarının %60,0'ının doktora başvurma için tavsiye ve teşvik ifadesi içerdiğini saptadık. ChatGPT'nin kullanımına ilişkin tehlikelerinden en önemlisi endişe veya sorununu ChatGPT'ye sorarak gidermeye çalışan hastaların, elde ettikleri yanıtların uygunluğunu değerlendirme potansiyellerinin olmaması ve edindikleri yanlış güven duygusu ile soruları hakkında doktorlarına danışmaktan vazgeçebilmeleridir. Bu olası tehlike ChatGPT'nin tıp alanı ile ilgili her sorunun yanıtına "tıbbi danışmanlık sağlayamayacağı, en doğru bilgi için sağlık uzmanlarına danışmanın gerekliliği" ile ilgili ifadelerin dahil edilmesi ile hafifletilebilir.

Çalışmamızda yanıtlarda dikkat çekici doğruluk oranları saptamamıza rağmen ChatGPT'nin tıp ve oftalmoloji alanında kullanımında bazı önemli sorunlar mevcuttur. Bunlardan ilki ChatGPT'nin "halüsinasyonlar" olarak adlandırılan gerçek dışı yanıtlar üretme potansiyelidir.<sup>22</sup> Bir diğer sorun da ChatGPT'nin "doğrulama yanlılığı" eğilimidir. Bu durum ChatGPT'nin sunduğu bilgiler doğru olsa bile kullanıcı tarafından düzeltilmediğinde, nesnel doğruluğunu korumak yerine, kullanıcı yanlılığından etkilenecek yanıtı değiştirme potansiyelinin bulunmasıdır.<sup>23</sup> Ayrıca ChatGPT'nin bir diğer doğal sınırlaması belirli tarihli veri kümeleri ile eğitilmiş olmasından kaynaklanan statik eğitim sürecinin bir sonucu olarak, güncel bilgiler ve yeni ilaçlarla ilgili eksik bilgi sunmasının mümkün olmasıdır. Bir başka sınırlama ise, ChatGPT'nin veri sunmak için kullandığı kaynakların ve kaynakların güvenilirliğinin kullanıcılar tarafından değerlendirilemiyor oluşudur.<sup>24</sup> ChatGPT'den paylaştığı bilgilerin referansları istendiğinde alakasız veya var olmayan referanslar sunması daha önce bildirilmiştir.<sup>25</sup> Gelecekte ChatGPT'ye sürekli öğrenme mekanizmaları entegre edilerek ve paylaştığı bilgiler için referans sunması sağlanarak, güncellik ve güvenilirlikle ilgili bu sınırlamanın aşılmasıyla kullanım potansiyelinin artması mümkün olabilir.

Çalışmanın kısıtlılıkları: Çalışmamızın bazı kısıtlılıkları vardır. Bunlardan ilki internet ortamında sıkça sorulan soruları derlemeye çalışsak da 20 soru ve yanıtta oluşan veri kümesinin nispeten küçük olmasıdır. Bu örneklem boyutu ChatGPT'nin karşılaşılabileceği soruların çeşitliliğini tam olarak yakalayamayabilir. Ayrıca çalışma tasarımı gereği soruların demografik

bilgileri içermemesi de çalışmamızın kısıtlılıklarından olabilir. Ancak bu durumu ChatGPT'nin verileri işleme sırasında cinsiyet ve ırk ilişkili olarak önyargı taşıyabilmesi sebebiyle özellikle tercih ettik.<sup>24</sup> Demografik bilgilerin ve hasta özgeçmiş bilgilerinin bulunduğu sorularda yanıtların doğruluk ve kapsamlılıklarında değişiklikler olabilir. Soruların yalnızca bir kez sorulması nedeniyle ChatGPT tutarlılığının ve kendini düzeltme kapasitesinin değerlendirilmemiş olması da çalışmamızın kısıtlılıklarındandır. Çalışmamızda değerlendiriciler, glokom birimlerinde deneyim sahibi olma ve farklı kliniklerde görev yapma ön kriterleri dikkate alınarak seçilmiştir. Ancak mesleki pozisyon, çalışılan kurumun özellikleri ve diğer bireysel/kurumsal farklılıklar açısından ek bir standardizasyon uygulanmamıştır. Bu durum, çalışmamızın kısıtlılıklarından biri olarak dikkate alınmalıdır.

## Sonuç

Çalışmamızda ChatGPT'nin bazı sınırlamaları olmakla birlikte glokomun medikal tedavisi ile ilgili soruları yüksek doğruluk ve kapsamlılıkta yanıtladığını saptadık. Bununla birlikte ChatGPT tavsiyelerinin glokomun medikal tedavisinin yönetiminde oftalmolog tavsiyelerinin yerine geçmesi mümkün görünmemektedir. ChatGPT'nin glokom tedavisinin yönetiminde destekleyici bir unsur olarak kullanılması ve oftalmoloji alanında genel kabul görmesi için daha fazla çalışmaya ve teknolojik iyileştirmeye ihtiyaç vardır. Oftalmologların, tıbbi danışmanlık amacıyla ChatGPT kullanımının hastalar üzerindeki etkilerinin farkında olması, yanlışlık ve eksikliklerin düzeltilmesi için gereklidir.

*Yazarlar arasında çıkar çatışması olmadığı ve çalışma için finansal destek alınmadığı bildirilmiştir.*

*Yazarların çalışmaya katkıları:*

*BŞDE: Fikir/kavram, tasarım, denetleme/danışmanlık, veri toplama ve/veya işleme, analiz ve/veya yorum, kaynak tarama, makale yazım. GH: Veri toplama ve/veya işleme, analiz ve/veya yorum, kaynak tarama, makalenin yazım. EY: Tasarım, denetleme/danışmanlık, veri toplama ve/veya işleme, analiz ve/veya yorum, kaynak tarama, makale yazım, eleştirel inceleme*

## Kaynaklar

1. Courtright P. Hutchinson AK. Lewallen S. Visual impairment in children in middle and lower income countries. Arch Dis Child. 2011;96:1129-34.
2. Gergely K. Gerinec A. Retinopathy of prematurity epidemics, incidence, prevalence, blindness. Bratisl Lek Listy. 2010;111:514-7.
3. Fevereiro-Martins M. Marques-Neves C. Guimarães H. Bicho M. Retinopathy of prematurity: A review of pathophysiology and signaling pathways. Surv Ophthalmol. 2023;68:175-210.
4. Cavallaro G. Filippi L. Bagnoli P. et al. The pathophysiology of retinopathy of prematurity: An update of previous and recent knowledge. Acta Ophthalmol. 2014;92:2-20.
5. Rasoulinejad SA. Maroufi F. Alizadeh A. Clinical parameters in different stages, zones, and remission/progression statuses of retinopathy of prematurity. Casp J Intern Med. 2022;13:385-92.
6. Sood BG. Madan A. Saha S. et al. Perinatal systemic inflammatory response syndrome and retinopathy of prematurity. Pediatr Res. 2010;67:394-400.

7. Hong HK, Lee HJ, Ko JH, et al. Neonatal systemic inflammation in rats alters retinal vessel development and simulates pathologic features of retinopathy of prematurity. *J Neuroinflammation*. 2014;11:87.
  8. Zeb A, Khurshid S, Bano S, et al. The Role of the Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio as Markers of Disease Activity in Ankylosing Spondylitis. *Cureus*. 2019;11:e6025.
  9. Howard R, Kanetsky PA, Egan KM. Exploring the prognostic value of the neutrophil-to-lymphocyte ratio in cancer. *Sci Rep*. 2019;9:19673.
  10. İlhan N, Daglıoğlu MC, İlhan O., et al. Assessment of Neutrophil/Lymphocyte Ratio in Patients with Age-related Macular Degeneration. *Ocul Immunol Inflamm*. 2015;23:287-90.
  11. Yue S, Zhang J, Wu J, Teng W, Liu L, Chen L. Use of the monocyte to lymphocyte ratio to predict diabetic retinopathy. *Int J Environ Res Public Health*. 2015;12:10009-19.
  12. Bas AY, Koc E, Dilmen U. ROP Neonatal Study Group. Incidence and severity of retinopathy of prematurity in Turkey. *Br J Ophthalmol*. 2015;99:1311-4.
  13. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol*. 2005;123:991-9.
  14. Early Treatment For Retinopathy Of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Arch Ophthalmol*. 2003;121:1684-94.
  15. Tremblay S, Miloudi K, Chaychi S, et al. Systemic inflammation perturbs developmental retinal angiogenesis and neuroretinal function. *Invest Ophthalmol Vis Sci*. 2013;54:8125-39.
  16. Lawrence SM, Eckert J, Makoni M, Pereira HA. Is the Use of Complete Blood Counts with Manual Differentials an Antiquated Method of Determining Neutrophil Composition in Newborns?. *Ann Clin Lab Sci*. 2015;45:403-13.
  17. Schmutz N, Henry E, Jopling J, Christensen RD. Expected ranges for blood neutrophil concentrations of neonates: the Manroe and Mouzinho charts revisited. *J Perinatol*. 2008;28:275-81.
  18. Turkmen K, Ozcicek F, Ozcicek A, Akbas EM, Erdur FM, Tonbul HZ. The relationship between neutrophil-to-lymphocyte ratio and vascular calcification in end-stage renal disease patients. *Hemodial Int*. 2014;18:47-53.
  19. Kurtul BE, Ozer PA. Neutrophil-to-lymphocyte ratio in ocular diseases: a systematic review. *Int J Ophthalmol*. 2019;12:1951-8.
  20. Hu YX, Xu XX, Shao Y, et al. The prognostic value of lymphocyte-to-monocyte ratio in retinopathy of prematurity. *Int J Ophthalmol*. 2017;10:1716-21.
  21. Kurtul BE, Kabatas EU, Zenciroglu A, et al. Serum neutrophil-to-lymphocyte ratio in retinopathy of prematurity. *J AAPOS*. 2015;19:327-31.
  22. Tong Q, Yin H, Zhao M, Li X, Yu W. Outcomes and prognostic factors for aggressive posterior retinopathy of prematurity following initial treatment with intravitreal ranibizumab. *BMC Ophthalmol*. 2018;18:150.
  23. Ozturk T, Durmaz Engin C, Kaya M, Yaman A. Complete blood count parameters to predict retinopathy of prematurity: when to evaluate and what do they tell us? *Int Ophthalmol*. 2021;41:2009-18.
  24. Akyüz Ünsal Aİ, Key Ö, Güler D, et al. Can complete blood count parameters predict retinopathy of prematurity? *Turk J Ophthalmol*. 2020;50:87-93.
  25. Tao Y, Dong Y, Lu CW, Yang W, Li Q. Relationship between mean platelet volume and retinopathy of prematurity. *Graefes Arch Clin Exp Ophthalmol*. 2015;253:1791-4.
-