

# ChatGPT'nin kadınlarda alt üriner sistem semptomları ile ilgili soruları yanıtlama performansının değerlendirilmesi

## Evaluation of ChatGPT's performance in answering questions about female lower urinary tract symptoms

Abuzer Öztürk<sup>1</sup>, İsmail Emre Ergin<sup>2</sup>

### ÖZ

**AMAÇ:** ChatGPT, OpenAI tarafından geliştirilen bir dil modelidir. Bir dil modeli olarak ChatGPT, soruları yanıtlama, bilgi sağlama, konuşmalara katılma ve daha fazlası gibi çeşitli doğal dil işleme görevleri için kullanılabilir. Kadınlarda alt üriner sistem semptomları (AÜSS) bilgilerinin güvenilirliğini ve doğruluğunu analiz ederek ChatGPT'nin hastaları ve tıp öğrencilerini doğru şekilde bilgilendirmeye uygunluğunu değerlendirmeyi amaçladık.

**GEREÇ ve YÖNTEMLER:** Kadınlarda AÜSS ile ilgili üroloji derneği web siteleri, hastaneler ve sosyal medya platformlarında sistematik olarak düzenlenen sık sorulan sorular ve Avrupa Üroloji Derneği (EAU) önerilerinden derlenen sorular analiz edildi. Tüm sorular sistematik olarak ChatGPT 3.5 ve 4.0 versiyonlarına gönderildi. İki uzman ürolog birbirinden bağımsız olarak verileri ve sorulara verilen cevapları değerlendirdi ve her cevaba 1'den 4'e kadar puan verdi.

**BULGULAR:** Kadınlarda AÜSS ile ilgili sık sorulan 103 soru ve EAU kılavuz önerilerinden derlenen 36 soru çalışmaya dahil edildi. Sık sorulan soruların %83,50'si (versiyon 4.0 için %85,44) ve EAU önerilerinden derlenen soruların %80,56'si (versiyon 4.0 için %80,56) 'tamamen doğru' (Grade 1) olarak derecelendirildi. Sorulara tamamen yanlış cevap verme oranı (Grade 4) sık sorulan sorular grubunda %0,97 (versiyon 4.0 için %0,97), rehber tabanlı sorular grubunda ise %2,78 (versiyon 4.0 için %2,78) olarak gerçekleşti. Her iki versiyonun da soruları yanıtlamadaki başarı oranı birbirine yakın bulunmuştur.

**SONUÇ:** ChatGPT, sorulan sorulara kabul edilebilir yanıtlar verdi. Kısıtlılıklarına rağmen sürekli gelişen bu platformun üroloji kliniklerinde ürologların rehberliğinde kullanılmasının, hastalara ve ailelerine tanı ve tedavi süreçlerinde katkı sağlayacağına inanıyoruz.

**Anahtar Kelimeler:** yapay zeka, ChatGPT, eğitim, kadın AÜSS

### ABSTRACT

**OBJECTIVE:** ChatGPT is a language model developed by OpenAI. As a language model, ChatGPT can be used for various natural language processing tasks, such as answering questions, providing information, engaging in conversations, and more. We aimed to evaluate ChatGPT's suitability to accurately inform patients and medical students by analyzing the reliability and accuracy of lower urinary tract symptoms (LUTS) information in women.

**MATERIAL and METHODS:** Systematically organized frequently asked questions from urology association websites, hospitals, and social media platforms regarding lower urinary tract symptoms (LUTS) in women, along with questions compiled from European Association of Urology (EAU) recommendations, were analyzed. All questions were systematically submitted to ChatGPT 3.5 and 4.0 versions. Two expert urologists independently evaluated the data and answers to the questions and assigned scores from 1 to 4 to each answer.

**RESULTS:** A hundred and three frequently asked questions and 36 questions compiled from EAU guideline recommendations regarding LUTS in women were included in the study. 83.50% of the frequently asked questions (85.44% for version 4.0) and 80.56% of the questions compiled from EAU recommendations (80.56% for version 4.0) were rated as 'completely correct' (Grade 1). The rate of completely incorrect answers to the questions (Grade 4) was 0.97% in the frequently asked questions group (0.97% for version 4.0) and 2.78% in the guide-based questions group (2.78% for version 4.0). The success rate of both versions in answering questions was found to be close to each other.

**CONCLUSION:** ChatGPT provided acceptable responses to the posed questions. Despite its limitations, we believe that utilizing this continuously evolving platform under the guidance of urologists in urology clinics will contribute to the diagnosis and treatment processes for patients and their families.

**Keywords:** artificial intelligence, ChatGPT, education, female LUTS

<sup>1</sup>Sivas Numune Hastanesi, Sivas, Türkiye

<sup>2</sup>Kızılcahamam Devlet Hastanesi, Ankara, Türkiye

#### Yazışma Adresi/ Correspondence:

Uzm. Dr. Abuzer Öztürk

Sivas Numune Hastanesi Üroloji Servisi Kat:5, Sivas, Türkiye

Tel: +90 346 258 05 21

E-mail: brusksidal@gmail.com

**Geliş/ Received:** 10.05.2024

**Kabul/ Accepted:** 10.06.2024

## GİRİŞ

Alt Üriner Sistem Semptomları (AÜSS), Uluslararası Kontinans Derneği (ICS) tarafından tanımlanan depolama, işeme ve işeme sonrası semptomları olmak üzere üç ana semptom grubundan oluşur.<sup>[1]</sup> Çok sayıda bilimsel ve deneysel çalışmanın bulgularına göre, kadın genitoüriner sisteminin anatomisi ve fizyolojisi ile ilişkili olarak kadınlar



AÜSS'ye erkeklerden daha duyarlıdır.<sup>[2]</sup> Benzer şekilde kadınların hamilelik ve doğum sırasındaki hormonal değişimleri, çevresel deneyimleri ve cinsiyet merkezli sosyal deneyimleri<sup>[3]</sup> bu sağlık sorunuyla ilişkilendirilmiştir.<sup>[4-6]</sup> AÜSS kadınların fiziksel, zihinsel ve cinsel sağlığını<sup>[7]</sup> büyük ölçüde ve olumsuz yönde etkilemiş ve aynı zamanda yaşam kalitelerini de düşürmüştür.<sup>[8]</sup> Bu nedenle, AÜSS yaşayan kadınlar tarafından fiziksel aktivitede azalma<sup>[9]</sup>, anksiyete, depresyon ve sosyal izolasyon<sup>[10]</sup> gibi olumsuz fiziksel, duygusal ve sosyal sonuçlar rapor edilmiştir.

Alt Üriner Sistem Semptomları yaşayan kadınlar, bir sağlık uzmanına danışmadan önce ön bilgi almayı ve meraklarını gidermeyi, internet ve yapay zekâ destekli bilgi sistemleri üzerinden karşılamayı tercih ederler. Bu durum, yapay zekâ destekli sohbet robotlarının sunduğu bilgilerin güvenilirliğinin değerlendirilmesine yönelik kritik ihtiyacın altını çiziyor. Yapay zekâ destekli sohbet robotları mantıksal ve bağlamsal olarak tutarlı yanıtlar oluşturma konusunda umut vaat ediyor olsa da sağladıkları tıbbi bilgilerin doğruluğunun tespit edilmesi hayati önem taşıyor.

ChatGPT, OpenAI tarafından oluşturulan ve internet veri tabanını kullanarak sorulara anında cevap veren bir yapay zekâ programıdır.<sup>[11]</sup> Bir akıllı telefon uygulaması veya bir internet tarayıcısı aracılığıyla erişilebilir. Çeşitli konulardaki verileri anında sentezleyerek kullanıcıların bilgiye ulaşmasını sağlayan bu program, hızla popülerlik kazanmaktadır. ChatGPT'nin tıp alanında doğru bilgi sağlama yeteneğini araştıran çeşitli makaleler yayımlanmıştır. Amerikan Tabipler Birliği Dergisi yakın geçmişte yayınları için sohbet robotlarının kullanımına ilişkin yönergeleri uygulamaya koymuştur.<sup>[12]</sup> Gilson ve ark.'nin yaptığı bir çalışmada ChatGPT'nin tıp fakültesi sınav sorularının %60'ına tatmin edici cevaplar verdiği belirlenmiştir.<sup>[13]</sup> Çağlar ve ark. ise çalışmanın sınırlılıklarına rağmen ChatGPT'nin pediatrik üroloji ile ilgili sorulara tatmin edici cevaplar sağladığını belirtmişlerdir.<sup>[14]</sup> ChatGPT'nin çeşitli tıp alanlarındaki cevaplarının doğruluğunu ve güvenilirliğini araştıran bugüne kadar yayımlanmış birçok makale olmasına rağmen<sup>[13-15]</sup>, AÜSS olan kadınlar hakkında böyle bir çalışmanın olmadığını fark ettik. Bu nedenle bu çalışmada kadınlarda AÜSS ile ilgili bir dizi bilinçli soruyla ChatGPT'nin mevcut kapasitesini değerlendirmeye çalıştık.

## GEREÇ ve YÖNTEMLER

ChatGPT ve Avrupa Üroloji Derneği (EAU) kılavuz bilgilerinin kullanımı kamuya açık olduğundan bu çalışma için etik onay alınmadı.<sup>[16]</sup>

Hastaların sağlık siteleri, üroloji derneği siteleri ve sosyal medya (Facebook, Instagram ve Twitter) üzerinden sıkça sorduğu tüm sorular kaydedildi. Sağlık web sitesi seçimi şu şekilde yapıldı: Değerlendirilecek alt konu Google'da arandığında ilk ortaya çıkan hasta bilgi sitelerinde sıkça sorulan sorular değerlendirildi. Avrupa Üroloji Derneği'nin kadınlarda AÜSS bölümünün her konusuna ilişkin güçlü öneri şeklinde tavsiye edilen öneriler, sorulara dönüştürüldü. Tekrarlayan sorular, dilbilgisi açısından uygun olmayan sorular, kişisel sağlıkla ilgili sorular ve yanıtları belirsiz olan sorular çalışmanın dışında bırakıldı. Araştırmada yer alan tüm sorular deneyimli iki ürolog tarafından oluşturuldu. İnternette elde edilen sorular ve EAU Kılavuzlarından elde edilen sorular şu şekilde kategorize edildi:

- > Aşırı Aktif Mesane
- > Stres İnkontinans
- > Az Aktif Mesane ve Mesane Çıkım Tıkanıklığı
- > Pelvik Organ Prolapsusu, Üriner Fistüller ve Üretral Divertikül
- > Fonksiyonel Üroloji Terminolojisi

Sorular, çerezleri ve geçmişi temizlenmiş internet tarayıcılarında ücretsiz ChatGPT 3.5 sürümü ve ücretli ChatGPT 4.0. sürümü kullanılarak yanıtlandı. ChatGPT'nin verdiği yanıtlar iki ürolog tarafından değerlendirilip puanlandı. Yazarlar her cevaba 1'den 4'e kadar bir puan verdi: Cevaplar, bir üroloğun hastaya sağlaması gereken tüm bilgileri doğru bir şekilde içeriyorsa, 1 (tamamen doğru) olarak derecelendirildi; doğru ancak yetersizse 2; doğru ve yanıltıcı bilgilerin bir karışımını içeriyorsa 3; tamamen yanlış cevaplar içeriyorsa 4 olarak puanlandı. Avrupa Üroloji Derneği kılavuzlarının güçlü öneri düzeyi temel alınarak oluşturulan soruların ise kılavuz bilgilerine uygunluğu değerlendirildi.

İki denetçi arasındaki puanlama farklılıkları aritmetik ortalamaya alınarak giderildi ve ardışık olarak puanlanan cevaplar denetçiler tarafından ortaklaşa incelenerek birincil kaynaklara göre değerlendirildi. ChatGPT'nin tüm sorulara verdiği yanıtlar, onaylandıktan sonra aynı soruların farklı IP adreslerinden sunulmasıyla oluşturuldu ve böylece yanıtların tekrarlanabilirliği onaylandı. ChatGPT'ye her sorunun zorluk seviyesini sorduk ve yanıtlarına göre sınıflandırdık. Yanıtlara göre tüm sorular kolay, orta ve zor olarak kategorize edildi.

## İstatistiksel Analiz

Cevaplardan toplanan veriler Excel'e (Versiyon 16, ABD) aktarıldı. Cevapların doğruluğu değerlendirildi ve cevapların doğruluk yüzdeleri hesaplandı. Tekrarlanabilirlik

açısından iki farklı bilgisayarda aynı soruya verilen anlamsal bütünlük açısından farklı yanıtlar olumsuz olarak değerlendirildi. Her iki gruptan gelen tüm veriler Windows için IBM Sosyal Bilimlerde İstatistik Paket Programı (SPSS) sürüm 25.0 (IBM Corp., 2018) kullanılarak karşılaştırıldı.

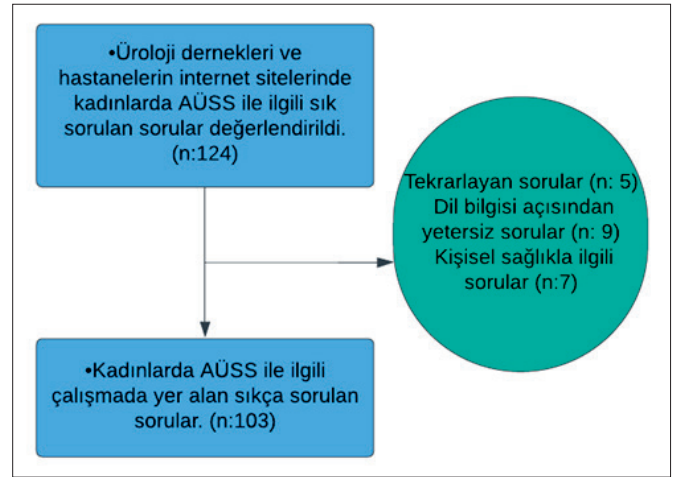
## BULGULAR

Çalışmaya başlangıçta toplam 124 soru dâhil edildi. Ancak dışlama kriterlerine göre 21 soru hariç tutuldu. Hariç tutulan bu sorulardan beşi tekrarlayan sorulardı, dokuzu dil bilgisi açısından yetersizdi ve yedisi kişisel sağlıkla ilgiliydi. Dışlama kriterleri uygulandıktan sonra geriye 103 soru kaldı ve çalışmaya dâhil edildi. Şekil 1’de araştırmaya dâhil edilen ve hariç tutulan sorular gösterilmektedir. Sıkça sorulan sorulara ek olarak EAU kılavuzu önerilerinden 36 soru derlenip çalışmaya dâhil edildi.

ChatGPT’nin sıkça sorulan sorular ve EAU kılavuzu önerilerinden derlenen sorular için sağladığı yanıt oranları Tablo 1’de sunulmaktadır. Sıkça sorulan soruların %83,50’si ‘tamamen doğru’ (Grade 1) olarak derecelendirilmiştir (4.0 sürümü için %85,44). Soruların yaklaşık %7,77’sinde eksik bilgi vardı (Grade 2) (4.0 sürümü için %6,80), %7,77’si doğru ve yanıltıcı bilgi karışımıydı (Grade 3) (4.0 sürümü için %6,80) ve %0,97’si tamamen yanlış yanıtlanmıştı (Grade 4) (4.0 sürümü için %0,97). Sıkça sorulan sorular aşırı aktif mesane, stres inkontinans, az aktif mesane ve mesane çıkım tıkanıklığı, pelvik organ prolapsusu, üriner fistüller ve üretral divertikül, fonksiyonel üroloji terminolojisi şeklinde sınıflandırıldı ve tamamen doğru yanıt oranları sırasıyla %93,10, %78,57, %80,95, %85,71, ve %72,22 oldu (4.0 sürümü için sırasıyla %93,10, %85,71, %80,95, %90,48, ve %72,22). Avrupa Üroloji Derneği önerilerinden derlenen soruların ise %80,56’sı ‘tamamen doğru’ (Grade 1) olarak derecelendirilmiştir (4.0 sürümü için %80,56). Soruların yaklaşık %5,56’sında eksik bilgi vardı (Grade 2) (4.0 sürümü için %8,83), %11,11’i doğru ve yanıltıcı bilgi karışımıydı (Grade 3) (4.0 sürümü için %8,83) ve %2,78’i tamamen yanlış yanıtlanmıştı (Grade 4) (4.0 sürümü için %2,78).

ChatGPT’ye her sorunun zorluk seviyesini sorduk ve yanıtlarına göre sınıflandırdık. Tüm sorular ChatGPT desteği ile kolay, orta ve zor olarak kategorize edildi. Sıkça sorulan sorulardan 56 soru kolay, 30 soru orta düzeyde ve 17 soru zor düzeyde kategorize edildi. Soru örnekleri Tablo 2’de gösterilmektedir. Kılavuzlardan elde edilen sorulardan 13’ü kolay, 17’si orta ve altısı zor olarak değerlendirildi.

Sorulara doğru cevap verme oranı (Grade 1 ve 2’nin birleşimi) sık sorulan sorularda %91,27 (4.0 sürümü için



Şekil 1. Çalışmada yer alan soruların akış şeması

Tablo 1. Kadınlarda AÜSS ile ilgili sorulara ChatGPT tarafından verilen yanıtların derecesi

Kadın AÜSS	ChatGPT Versiyon	Grade 1	Grade 2	Grade 3	Grade 4
Sıkça sorulan sorular (n: 103)	Versiyon 3,5	86 (%83,50)	8 (%7,77)	8 (%7,77)	1 (%0,97)
	Versiyon 4	88 (%85,44)	7 (%6,80)	7 (%6,80)	1 (%0,97)
Aşırı aktif mesane (n: 29)	Versiyon 3,5	27 (%93,10)	2 (%6,90)	—	—
	Versiyon 4	27 (%93,10)	2 (%6,90)	—	—
Stres inkontinans (n: 14)	Versiyon 3,5	11 (%78,57)	2 (%14,29)	1 (%7,14)	—
	Versiyon 4	12 (%85,72)	1 (%7,14)	1 (%7,14)	—
Az aktif mesane ve mesane çıkım tıkanıklığı (n: 21)	Versiyon 3,5	17 (%80,95)	—	3 (%14,29)	1 (%4,76)
	Versiyon 4	17 (%80,95)	1 (%4,76)	2 (%9,52)	1 (%4,76)
Pelvik organ prolapsusu, üriner fistüller ve üretral divertikül (n: 21)	Versiyon 3,5	18 (%85,71)	2 (%9,52)	1 (%4,76)	—
	Versiyon 4	19 (%90,48)	1 (%4,76)	1 (%4,76)	—
Fonksiyonel üroloji terminolojisi (n: 18)	Versiyon 3,5	13 (%72,22)	2 (%11,11)	3 (%16,67)	—
	Versiyon 4	13 (%72,22)	2 (%11,11)	3 (%16,67)	—
EAU kılavuzu önerileri (n: 36)	Versiyon 3,5	29 (%80,56)	2 (%5,56)	4 (%11,11)	1 (%2,78)
	Versiyon 4	29 (%80,56)	3 (%8,33)	3 (%8,33)	1 (%2,78)

**AÜSS:** Alt üriner sistem semptomları, **Grade 1:** Tamamen doğru, **Grade 2:** Doğru ancak yetersiz, **Grade 3:** Doğru ve yanıltıcı bilgilerin bir karışımı, **Grade 4:** Tamamen yanlış, **EAU:** Avrupa Üroloji Derneği.

**Tablo 2. Sıkça sorulan sorular**

- 1 Aşırı aktif mesane nedir?
- 2 Aşırı aktif mesane belirtileri nelerdir?
- 3 Aşırı aktif mesane için risk faktörleri nelerdir?
- 4 İdrar kaçırma tipleri nelerdir?
- 5 Sıkışma tipi idrar kaçırma nedir?
- 6 Kadınlarda idrar kaçırma tedavileri nelerdir?
- 7 Antikolinergik ilaçların yan etkileri nelerdir?
- 8 Mesane botoksu ne işe yarar?
- 9 Mesane botoksu yan etkileri nelerdir?
- 10 Mesane pili nedir?
- 11 Stres tipi idrar kaçırma nedir?
- 12 Stres tipi idrar kaçırma nedenleri nelerdir?
- 13 Pelvik taban kas egzersizleri nedir?
- 14 TOT ameliyatı nedir?
- 15 Az aktif mesane nedir?
- 16 Az aktif mesane belirtileri nelerdir?
- 17 Mesane çıkım tıkanıklığı nedir?
- 18 Noktüri nedir?
- 19 Noktüri sebepleri nelerdir?
- 20 Mesane sarkması nedir?
- 21 Rahim sarkması nedir?
- 22 Pelvik organ prolapsusu nedir? / AÜSS nedir?
- 23 Veziko-vajinal fistül nedir?
- 24 Veziko-vajinal fistül kendiliğinden kapanır mı?
- 25 İşeme testi nedir?
- 26 Sistoskopi/sistoüretroskopi nedir?
- 27 Ürodinami testi nedir?
- 28 Postmiksyonel rezidü idrar miktarı nedir?
- 29 Postmiksyonel rezidü idrar miktarı nasıl ölçülür?
- 30 Temiz aralıklı kateterizasyon nedir?

%92,24), kılavuzlardan alınan sorularda ise %86,12 olarak gerçekleşti (4.0 sürümü için %88,89). Sorulara tamamen yanlış cevap verme oranı (Grade 4) sık sorulan sorular grubunda %0,97 (4.0 sürümü için %0,97), kılavuza dayalı sorular grubunda ise %2,78 idi (4.0 sürümü için %2,78). Her iki sürümün soruları yanıtlamadaki başarı oranı oldukça benzer bulundu.

## TARTIŞMA

Yapay zekâ (AI) günümüzde giderek daha yaygın kullanılmaktadır. Hastaların kendi dillerinde rahatlıkla kullanabileceği ChatGPT programı OpenAI firmasının yaptığı bir değerlendirmedir. Etkinliği, özellikle bilgi teknolojisi, öğrenme, eğlence ve sağlık hizmetleri alanlarında büyük bir

potansiyel göstermektedir. ChatGPT, sağlık hizmetleri de dâhil olmak üzere çeşitli disiplinleri kapsayan geniş bir bilgi deposuna sahiptir.<sup>[17]</sup> Yapay zekânın özellikle hastalıkların teşhis ve tedavisinde kullanılması tıp alanında en popüler tartışmalı konulardan biridir. Yapay zekânın doğru ve etkili kullanımı, toplumun hastalıklar hakkındaki farkındalığını artırabilir ve hastaların tedavilere ve takip prosedürlerine daha uyumlu olmalarını sağlayabilir.<sup>[18]</sup>

YouTube, Facebook, Twitter, Reddit ve Pinterest gibi sosyal medya platformlarında paylaşılan tartışmalı ve yanlış sağlık bilgilerinin yaygınlığını gösteren çok sayıda çalışma daha önce yayımlanmıştır.<sup>[19]</sup> Örneğin Alsyouf ve ark. tarafından ürolojik malignitelerle ilgili sosyal medyada paylaşılan makaleler değerlendirildiğinde, yanlış ve yanıltıcı makale miktarının doğru makale sayısından belirgin şekilde fazla olduğu tespit edilmiştir.<sup>[20]</sup> AI kullanan sistemler için bu tür olumsuz bulgular bildirilmemiştir. Bulck ve Moons tarafından yapılan bir çalışmada seçilen hasta bilgilendirme yazılarının hastaların kolaylıkla anlayabileceği düzeyde özetlendiği ve sunulduğu görülmüştür.<sup>[21]</sup> Bu nedenle, kadınlarda oldukça yaygın olup insanların çektiği acılar ve ekonomik maliyet açısından önemi konusunda evrensel fikir birliği olan kadınlarda AÜSS hakkında ChatGPT'nin yanıtlarını analiz etmek için bu çalışmayı gerçekleştirdik.<sup>[22]</sup>

Makalemiz, ChatGPT'nin kadınlarda AÜSS ile ilgili sorulara verdiği yanıtların doğruluğunu ve güvenilirliğini değerlendiren ilk çalışmayı temsil etmektedir. Çalışmamızda ChatGPT'nin kadınlarda AÜSS ile ilgili sıkça sorulan soruların %91,27'sine (4.0 sürümü için %92,24) tamamen doğru veya doğru ancak yetersiz (Grade 1 ve Grade 2) olarak derecelendirilen yanıtlar verdiğini gözlemledik. Sosyal medya gibi diğer internet kaynaklarından farklı olarak ChatGPT, çok sayıda kaynaktan kapsamlı bilgi alır ve sunar. Kullandığı veri tabanının genişliğinin, sorulara doğru yanıt verme yüzdesinin daha yüksek olmasına katkıda bulunduğu açıktır.

Tıpta kılavuzlar uygulayıcılara kendi alanlarında spesifik ve evrensel olarak uygulanabilir kanıta dayalı öneriler sunmak için geliştirilmiştir. Kapsamlı tıbbi bilgi içeren kaynaklar hakkındaki soruların yanıtlanması ChatGPT için daha zor olabilir.<sup>[23]</sup> Ancak Çakır ve ark.'nın yaptığı çalışmada ChatGPT'nin EAU ürolityazis kılavuzlarına dayanarak soruların %80'ine tatmin edici cevaplar verdiğini keşfedilmiştir.<sup>[24]</sup> Çalışmamızda, ChatGPT'nin kapsamlı bilgi kaynaklarına ilişkin performansını değerlendirdik ve ChatGPT'nin kılavuzlardan derlenen soruların %86,12'sine (4.0 sürümü için %88,89) tamamen doğru veya doğru ancak yetersiz (Grade 1 ve Grade 2) olarak derecelendirilen yanıtlar

verdiğini gözlemledik. Bu, ChatGPT'nin, kılavuzlardaki güçlü tavsiyelerden elde edilenler gibi ayrıntılı bilgi gerektiren soruları yanıtlamadaki yüksek etkinliğini gösterir.

Bu çalışmada sıkça sorulan sorular ile kılavuza dayalı soruların cevap başarı oranlarının oldukça benzer olduğunu gözlemledik. Bu, ChatGPT'nin belirli bilgileri yüksek derecede başarıyla alma konusunda usta olduğunu gösterir. Ancak, sıkça sorulan soruların %8,74'ü (4.0 sürümü için %7,77) ve kılavuzlardan derlenen soruların %13,89'una (4.0 sürümü için %11,1) Grade 3 ve Grade 4 olarak derecelendirilen yanıtlar verdiğini belirtmekte fayda var. Bu veriler doğrultusunda ChatGPT'nin tıp eğitimi için tek başına yeterli olmayabileceğine ancak öğrenme sürecinde değerli bir tamamlayıcı araç olarak hizmet edebileceğine inanıyoruz.

Bilgiye bu kadar kolay ulaşmanın toplumun sağlık okuryazarlığını artırmasına rağmen hastalar semptomlarını internette araştırdıklarında sıklıkla malignitelerle ilgili konularla karşılaşılıyorlar ve bu da kaygı düzeylerini artırabiliyor. Yapay zekâ yöntemlerinin daha doğru bilgi sağlayabileceğini, dolayısıyla kaygı seviyelerini azaltabileceğini düşünmekteyiz. Ancak ChatGPT'nin kullanımının potansiyel dezavantajlarına karşı dikkatli olmak önemlidir. Örneğin Rafeh ve ark.'nın yaptığı bir çalışmada ChatGPT'nin sorgulanan vakaların %44'ünde hastalığın algılanan aciliyetini artırdığı bulunmuştur. Aciliyetteki bu artış, daha yüksek sağlık bakım maliyetlerine ve artan hasta kaygı düzeylerine yol açabilir.<sup>[25]</sup> Hoffman ve ark.'nın yaptığı başka bir çalışmada ise ChatGPT 3.5 ve 4.0 versiyonlarının ortopedik board sınav sorularına verdiği yanıtların başarı oranlarında istatistiksel olarak anlamlı farklılıklar gösterdiği belirlenmiştir. Çalışmalarında, 4.0 sürümünün yanıtlarının üç yıllık bir ortopedi uzmanının yanıtlarıyla aynı düzeyde olduğu, 3,5 sürümünün yanıtlarının ise bir ortopedi stajyerinin yanıtlarıyla karşılaştırılabilir olduğu bulunmuştur.<sup>[26]</sup> Ancak biz çalışmamızda her iki sürümün sorulara verdikleri yanıtların başarı oranlarını birbirine yakın bulduk. Bu farklı sonuçların yapay zekâ uygulamalarının bilgi kaynaklarının sürekli güncellenip yenilenmesine bağlı olduğunu düşünüyoruz.

Çalışmamızın bir avantajı ücretli ChatGPT 4.0 versiyonunun kullanılmış olmasıydı. Her iki versiyonu da değerlendirme fırsatı bulduk. ChatGPT 3.5 ve 4.0 versiyonlarının EAU 2023 rehberinden elde edilen soruları yakın başarı oranıyla cevaplama, her iki versiyonun da EAU 2023 rehberini veri tabanı olarak kullanması nedeniyle oldu. Çalışmamızın bir diğer avantajı ise daha önce İngilizce ağırlıklı yapılan ve diğer dillerde araştırma yapılmayan çalışmalarda görülen bir eksikliği gidermiş olmasıdır. İngilizce sorularla yapılan çalışmalara benzer sonuçlar elde etmemiz,

ChatGPT'nin yüksek çeviri başarısıyla bilgiye kolaylıkla ulaşabildiğimizi gösteriyor.

Çalışmamız, ChatGPT'nin kadınlarda AÜSS hakkındaki cevaplarının doğruluğunu ve güvenilirliğini değerlendiren ilk çalışma olmasına rağmen bazı sınırlamaları vardı. Öncelikle belli bir dönemi kapsıyor, oysa kadınlarda AÜSS ile ilgili internet kaynakları sürekli olarak yüklenip revize ediliyor. Çalışmamızı değerlendiren ürologların sayısı çalışmamızın bir diğer kısıtlılığıdır. Bu, gözlemcinin yorulma yanlılığına neden olabilir. Ayrıca, ChatGPT'nin mevcut sürümleri, sorulara doğru yanıtlar sağlama konusunda övgüye değer bir doğruluk sergilese de, bu başarının sonraki yıllarda sürdürülebilirliğinin belirsiz kaldığını kabul etmek zorunludur.

## SONUÇ

Sonuç olarak ChatGPT, sık sorulan sorulara ve EAU kılavuzlarından türetilen sorulara tatmin edici yanıtlar sundu. Ancak ChatGPT hemen her sorunun cevaplanmasında bir sağlık profesyoneli ile detaylı görüşmeyi önermekteydi. Bu nedenle, ChatGPT'nin üroloji kliniklerinde sağlık profesyonelleri gözetiminde uygulanmasını öneriyoruz. Bunun hastaların ve ailelerin kadınlarda AÜSS konularını anlamalarını geliştirebileceğini düşünüyoruz. Ayrıca tıp öğrencileri için değerli bir eğitim destek aracı olarak da hizmet verebileceğine inanıyoruz.

### Hakem Değerlendirmesi

Dış bağımsız.

### Çıkar Çatışması

Yazarlar çıkar ilişkisi olmadığını beyan etmişlerdir.

### Finansal Destek

Herhangi bir mali destek alınmamıştır.

### Peer-review

Externally peer-reviewed.

### Conflict of Interest

No conflict of interest was declared by the authors.

### Financial Disclosure

No financial support has been received.

## KAYNAKLAR

1. Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, Lee J, et al. An international Urogynecological association (IUGA)/international continence society (ICS) joint report on the terminology for female pelvic floor dysfunction. *Neurourol Urodyn.* 2010;29:4–20. [\[CrossRef\]](#)
2. Nasiri M, Sigaroudi AE, Moghadamnia MT, Leili EK. Lower urinary tract symptoms and related factors in Iranian female nurses. *Iran J Nurs Midwifery Res.* 2022;27:280–6. [\[CrossRef\]](#)
3. Brady SS, Bavendam TG, Berry A, Fok CS, Gahagan S, Goode PS, et al. The prevention of lower urinary tract symptoms (PLUS) in girls and women: developing a conceptual framework for a prevention research agenda. *Neurourol Urodyn.* 2018;37:2951–64. [\[CrossRef\]](#)

4. Maserejian NN, Chen S, Chiu GR, Wager CG, Kupelian V, Araujo AB, et al. Incidence of lower urinary tract symptoms in a population-based study of men and women. *Urology*. 2013;82:560–4. [CrossRef]
5. Brady SS, Berry A, Camenga DR, Fitzgerald CM, Gahagan S, Hardacker CT, et al. Applying concepts of life course theory and life course epidemiology to the study of bladder health and lower urinary tract symptoms among girls and women. *NeuroUrol Urodyn*. 2020;39:1185–2. [CrossRef]
6. Sutcliffe S, Bavendam T, Cain C, Epperson CN, Fitzgerald CM, Gahagan S, et al. The Spectrum of bladder health: the relationship between lower urinary tract symptoms and interference with activities. *J Womens Health (Larchmt)*. 2019;28:827–41. [CrossRef]
7. Coyne KS, Sexton CC, Kopp ZS, Ebel-Bitoun C, Milsom I, Chapple C. The impact of overactive bladder on mental health, work productivity and health-related quality of life in the UK and Sweden: results from EpiLUTS. *BJU Int*. 2011;108:1459–71. [CrossRef]
8. Pizzol D, Demurtas J, Celotto S, Maggi S, Smith L, Angiolelli G, et al. Urinary incontinence and quality of life: a systematic review and meta-analysis. *Aging Clin Exp Res*. 2021;33:25–35. [CrossRef]
9. Nygaard I, Girts T, Fultz NH, Kinchen K, Pohl G, Sternfeld B. Is urinary incontinence a barrier to exercise in women? *Obstet Gynecol*. 2005;106(2):307–314. [CrossRef]
10. Coyne KS, Wein AJ, Tubaro A, Sexton CC, Thompson CL, Kopp ZS, Aiyer LP. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. *BJU Int*. 2009;103(Suppl 3):4–11. [CrossRef]
11. OpenAI. ChatGPT: Optimizing Language Models for Dialogue. In: 2023. <https://chatgpt.r4wand.eu.org/>
12. Flanagan A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman “Authors” and implications for the integrity of scientific publication and medical knowledge. *JAMA* 2023;329:637–9. [CrossRef]
13. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How does ChatGPT perform on the United States Medical Licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. [Published 2023 Feb 8]. [CrossRef]
14. Caglar U, Yildiz O, Meric A, Ayranci A, Gelmis M, Sarilar O, Ozgor F. Evaluating the performance of ChatGPT in answering questions related to pediatric urology. *J Pediatr Urol*. 2024;20(1):26.e1–5. [CrossRef]
15. Yeo YH, Samaan JS, Ng WH, Ting P-S, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721–32. [CrossRef]
16. EAU Guidelines. Edn. presented at the EAU Annual Congress Milan 2023. ISBN 978 94-92671-19-6.
17. Li W, Fu M, Liu S, Yu H. Revolutionizing neurosurgery with GPT-4: a leap forward or ethical conundrum? [published online ahead of print, 2023 may 17]. *Ann Biomed Eng* 2023;51(10):2105–12. [CrossRef]
18. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13):1233–9. [CrossRef]
19. Ergul A. Quality and reliability of YouTube videos on surgical treatment of uterine leiomyomas. *Cureus*. 2021;13(11):e20044. [Published 2021 Nov 30]. [CrossRef]
20. Alsayouf M, Stokes P, Hur D, Amasyali A, Ruckle H, Hu B. ‘Fake News’ in urology: evaluating the accuracy of articles shared on social media in genitourinary malignancies. *BJU Int*. 2019;124(4):701–6. [CrossRef]
21. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: A proof-of-concept [Published online ahead of print, 2023 Aug 21]. *Eur J Cardiovasc Nurs*. 2024;23(2):122–6. [CrossRef]
22. Irwin DE, Mungapen L, Milsom I, Kopp Z, Reeves P, Kelleher C. The economic impact of overactive bladder syndrome in six Western countries. *BJU Int*. 2009;103(2):202–9. [CrossRef]
23. Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT knowledge evaluation in basic and clinical medical sciences: multiple choice question examination-based performance. *Healthcare (Basel)*. 2023;11(14):2046. [CrossRef]
24. Cakir H, Caglar U, Yildiz O, Meric A, Ayranci A, Ozgor F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *Int Urol Nephrol*. 2024;56(1):17–21. [CrossRef]
25. Abi-Rafeh J, Hanna S, Bassiri-Tehrani B, Kazan R, Nahai F. Complications following facelift and neck lift: implementation and assessment of large language model and artificial intelligence (ChatGPT) performance across 16 simulated patient presentations [Published online ahead of print, 2023 Aug 17]. *Aesthetic Plast Surg*. 2023;47(6):2407–14. [CrossRef]
26. Hofmann HL, Guerra GA, Le JL, Wong AM, Hofmann GH, Mayfield CK, et al. The rapid development of artificial intelligence: GPT-4’s performance on orthopedic surgery board questions [published online ahead of print, 2023 Sep 27]. *Orthopedics*. 2024;47(2):e85–9. [CrossRef]