

Yapay zekâ, erektil disfonksiyon ile ilgili sorulara nasıl yanıtlar veriyor: ChatGPT vs Google AI

How artificial intelligence responds to erectile dysfunction queries: ChatGPT vs Google AI

Mehmet Yılmaz¹, Gökhan Çil²

ÖZ

AMAÇ: Bu çalışma, Google AI ve ChatGPT'nin erektil disfonksiyon (ED) ile ilgili sorulara ne kadar doğru yanıt verdiğini araştırmayı amaçladı.

GEREÇ ve YÖNTEMLER: Ocak ve Şubat 2024 arasında, iki ürolog tarafından, erektil disfonksiyon (ED) ile ilgili 80 spesifik soru, üroloji kılavuzlarını (EAU kılavuzu 2023) temel alınarak oluşturuldu. Sorular, etioloji ve risk faktörleri, patofizyoloji, tanı ve tedavi gibi konuları kapsıyordu. Potansiyel önyargıları kontrol etmek için, tek bir araştırmacı tüm soruları ChatGPT ve Google AI'a sordu. Yapay zekâ modelleri tarafından verilen yanıtlar daha sonra soruları formüle eden doktorlar tarafından değerlendirildi. İstatistiksel analiz için IBM Sosyal Bilimlerde İstatistik Paket Programı (SPSS) sürüm 29.0.2.0 kullanıldı. Küçük örneklem büyüklüğü nedeniyle parametrik olmayan testler kullanıldı (n <30). Sonuçlar sayısal ve grafiksel olarak sunuldu ve her parametre için ortalama değerler hesaplandı. İki grup arasındaki karşılaştırma için Mann-Whitney U testi kullanıldı ve anlamlılık düzeyi p<0,05 olarak belirlendi.

BULGULAR: Etiyoloji ve risk faktörleri ile ilgili sorulara her iki dil modeli de yüksek doğruluk oranıyla cevaplar vermiş olsalar da ChatGPT biraz daha düşük oranda tamamen doğru yanıtlar verdi (sırasıyla 3,8±0,41, 3,65±0,58; p=0,56). Patofizyoloji açısından, her iki model de doğru cevaplar sağladı ve yanıltıcı bilgi bulunmadı (sırasıyla 3,55±0,51, 3,7±0,47; p=0,43). Tanı ve tedavi ile ilgili sorulara, her iki yapay zekâ modeli de benzer oranda doğru cevaplar verdiler (sırasıyla 3,6±0,59, 3,35±0,87; p=0,47) ve (sırasıyla 3,50±0,68, 3,4±0,94; p=0,70). Bununla birlikte tanı ve tedavi ile ilgili sorulara hem Google AI hem de ChatGPT daha fazla yanlış bilgiler de içeren cevaplar vermiştir.

SONUÇ: Google AI ve ChatGPT genellikle erektil disfonksiyon sorularına tatmin edici ve doğru yanıtlar vermiş olmalarına rağmen, performansları sorular arasında değişkenlik göstermektedir.

Anahtar Kelimeler: yapay zekâ, ChatGPT, erektil disfonksiyon, Google AI

ABSTRACT

OBJECTIVE: This study aimed to investigate how accurately Google AI and ChatGPT responded to questions regarding erectile dysfunction (ED).

MATERIAL and METHODS: In the study conducted by two andrologists between January and February 2024, 80 specific questions regarding erectile dysfunction (ED) were formulated based on established medical guidelines (EAU guidelines 2023). The questions covered topics including etiology, risk factors, pathophysiology, diagnosis, and treatment. To control potential biases, a single researcher inputted all questions into ChatGPT and Google AI. The responses provided by the artificial intelligence (AI) models were then evaluated by the physicians who formulated the questions. Statistical analysis was performed using IBM Statistical Package for Social Sciences (SPSS) program version 29.0.2.0, employing non-parametric tests due to the small sample size (n <30). Results were presented numerically and graphically, with mean values calculated for each parameter. The Mann-Whitney U test was utilized for comparison between the two groups, with a significance level set at p<0.05.

RESULTS: Both language models provided responses with high accuracy to questions related to etiology and risk factors; however, ChatGPT gave slightly lower rates of completely correct answers (3.8±0.41, 3.65±0.58; p=0.56). In terms of pathophysiology, both models provided correct answers without misleading information (3.55±0.51, 3.7±0.47; p=0.43). Regarding diagnosis and treatment-related questions, both artificial intelligence models provided similar rates of correct responses (3.6±0.59, 3.35±0.87; p=0.47) and (3.50±0.68, 3.4±0.94; p=0.70), respectively. However, both Google AI and ChatGPT provided responses containing more incorrect information concerning diagnosis and treatment.

CONCLUSION: Current AI models like Google AI and ChatGPT generally offer satisfactory and accurate responses to erectile dysfunction queries, yet performance fluctuates between questions.

Keywords: artificial intelligence, ChatGPT, erectile dysfunction, Google AI

Sağlık Bilimleri Üniversitesi, Bağcılar Eğitim ve Araştırma Hastanesi, Üroloji Kliniği, İstanbul, Türkiye

Yazışma Adresi/ Correspondence:

Uzm. Dr. Mehmet Yılmaz
Bağcılar eğitim ve araştırma hastanesi üroloji kliniği 34200 34 - Türkiye
Tel: +90 532 583 20 37
E-mail: cilgok@gmail.com

Geliş/ Received: 11.03.2024

Kabul/ Accepted: 15.04.2024

GİRİŞ

Günümüzde derin öğrenme ve yapay zekâ modelleri, insan benzeri bir dil formatında metin üreterek diyalog kurabilen ve büyük miktarda bilgi ağını işleyip analiz ederek kullanıcılara yüksek doğrulukla yanıt verebilir hale gelmişlerdir. OpenAI tarafından geliştirilen ChatGPT ve Google AI, geniş bir metin ve kod veri kümesinde eğitilmiş büyük

dil modelleridir ve en popüler olanlardır.^[1,2] Bu ürünler, kullanıma sunulduktan aylar sonra milyonlarca kullanıcıya ulaşmış ve bugün birçok alanda yerini almıştır. Tıbbi alanda, hem sağlık profesyonelleri hem de hastalar tarafından tercih edilirken, çeşitli endişeler de ortaya çıkmıştır.^[3,4]

Eretil disfonksiyon (ED), her yaş grubunda farklı şiddetlerde karşılaşılabilecek bir sorun olmasına rağmen, özellikle >40 yaş erkeklerde %40'ın üzerinde yaygınlığı olan küresel bir sorun olarak öne çıkmaktadır.^[5] Sağlık profesyonellerine ulaşmada yaşanan zorluklar en büyük etken gibi görünse de, utanç gibi nedenler erkek hastaların bazen üroloji uzmanlarına gitmeden önce androlojik sorunlarını internet üzerinden araştırmaya yöneliklerini göstermektedir.^[6] Hastalar, sağlıkla ilgili karar verme süreçlerinde bilgi aramaya yönelirken, sosyal medya %42 uygulama oranıyla en ilgi çekici alanlardan biridir.^[7] YouTube platformunun, sağlık profesyonelleri tarafından yüklenen akredite edilmiş videoların yanı sıra kalite ve içerik açısından farklılık gösteren ED videolarıyla dolu olması, birçok araştırmanın odak noktası haline gelmiştir.^[8-10] Bugünün popüler araçlarından biri olan yapay zekânın ED konusunda ne kadar güvenilir olduğunu araştıran çalışmalar bulunmamaktadır.

Bu çalışmanın amacı, günümüzde popüler iki ücretsiz yapay zekâ uygulamasının (ChatGPT, Google AI), ürologlar tarafından ED kılavuzunun içeriğinden üretilen sorulara ne kadar doğru yanıt verdiği araştırmak ve iki dil modeli arasında bir karşılaştırma yapmaktır.

GEREÇ ve YÖNTEMLER

Çalışma Tasarımı

Bu çalışma, Ocak ve Şubat 2024 tarihleri arasında ED ile ilgili sorular hazırlayan iki ürolog (GÇ, MY) tarafından yürütüldü. Temel endişe, soruların ve yanıtların açık, net ve tartışmasız olmasını sağlamaktı. Bu nedenle, tüm sorular EAU kılavuzları 2023'teki bilgilere uygun net cevapları olan sorulardan oluştu. Çalışma kapsamında ED'nin etiyolojisi ve risk faktörleri, patofizyolojisi, tanısı ve tedavis ile ilgili 80 spesifik soru hazırlandı. Tüm prosedürler, Helsinki Bildirgesi'nin etik kurallarına uygun olarak gerçekleştirildi.

Veri Toplama Süreci

Oluşabilecek önyargıları en aza indirmek amacıyla çalışma tasarımıımızın önemli bir parçası olarak, tüm soruların ChatGPT ve Google AI'a girilmesi işlemi tek bir araştırmacı tarafından gerçekleştirildi. Bu yöntem, soruların tutarlı bir şekilde sorulmasını sağlamak ve yapay zekâ

modellerinin soruları doğru bir şekilde anlamasına katkıda bulunmak amacıyla uygulandı. ChatGPT ve Google AI tarafından verilen yanıtların ardından, soruları hazırlayan doktorlara cevapları değerlendirebilmeleri için sunuldu.

Yapay Zekâ Modelleri

ChatGPT, çeşitli internet metinlerinde eğitilmiş olan OpenAI tarafından geliştirilmiş en son teknoloji dil modelidir. Bu yapay zekâ sistemi özellikle GPT olarak bilinen bir tür dönüştürücü model mimarisini kullanır ve makine öğrenme tekniklerinden yararlanır. Bu dil modelinin son güncellemesi Eylül 2021'de yapıldığından, model gerçek zamanlı olaylara, güncellemelere veya kaynaklara erişemez ve onlardan öğrenemez. Model, tıbbi kayıtları yönetme, hastalıklar hakkında bilgi sağlama ve hatta tedavi seçenekleri için destekleyici bir araç olarak hizmet etmek gibi tıbbi alanda potansiyel uygulamalara sahiptir. Google AI, Pathways Dil Modeli (PaDM) ailesine dayanarak geliştirilen geniş veri kümesi üzerinde eğitilmiş bir yapay zekâ modelidir ve ChatGPT gibi diğer dil modelleri ile karşılaştırılabilir bilgiler sunabilmektedir.

Yapay zekâ modelleri değerli yardımlar ve bilgiler sağlayabilirken, bir uzmanın yapay zekâ modelinin tıbbi değerlendirmesinin eksikliği nedeniyle tüm çıktıları gözden geçirmesi ve doğrulaması önemlidir. Bu çalışmanın ana odak noktası da budur.

Değerlendirme Prosedürü

Her bir alt başlık için her bir androlog tarafından 10 soru hazırlanarak toplamda 80 soru oluşturuldu. Bu sorular ED'nin etiyolojisi ve risk faktörleri, patofizyolojisi, tanısı ve tedavisiyle ilgili konuları kapsamaktadır. Her cevap, doğrulukları ve yeterlilikleri değerlendirebilmek amacıyla iki androlog tarafından incelendi. Yanıtlar bir ile dört arasında bir ölçekte puanlandı (4; tamamen doğru, 3; doğru ancak yetersiz, 2; doğru ve yanıltıcı bilgilerin karışımı, 1; tamamen yanlış). Yapay zekâ modellerine yöneltilen sorularda potansiyel önyargıları önlemek için daha önceden hiçbir yapay zekâ dil modellerine yönlendirilmedi ve yönetici hesabı olmayan bir bilgisayar kullanıldı.

İstatistiksel Analiz

Çalışmanın istatistiksel analizi için IBM Sosyal Bilimlerde İstatistik Paket Programı (SPSS) sürüm 29.0.2.0 (IBM Corp., Armonk, NY, USA) kullanıldı. Küçük örneklem büyüklüğü ve karşılaştırılan soru sayısı dikkate alınarak, parametrik olmayan testler kullanıldı. Sonuçlar, her yapay zekâ modeli için sayılar ve yüzdeler olarak hesaplandı

Erektıl Disfonksiyon ile ilgili Sorular

- Q1: Can you give information about the diseases that may play a role in the pathophysiology of erectile dysfunction?
- Q2: How many groups is erectile dysfunction generally divided into according to etiology?
- Q3: Do antihypertensive drugs play a role in the pathophysiology of ED? If so, which ones play a role?
- Q4: Erectile dysfunction is not observed much after external radiotherapy and brachytherapy, is it true?
- Q5: Which laboratory tests should be requested from a patient with erectile dysfunction complaints?
- Q6: Who are the patient groups with high cardiac risk when planning erectile dysfunction treatment?
- Q7: Who are the patient groups with low cardiac risk when planning erectile dysfunction treatment?
- Q8: What should be done first for the high cardiac risk patient group when planning erectile dysfunction treatment?
- Q9: What is Nocturnal penile tumescence and rigidity test?
- Q10: What is Intracavernous injection test?
- Q11: What does specific diagnostic tests for ED?
- Q12: What are the first-line treatments for erectile dysfunction?
- Q13: What are the oral pharmacotherapies for erectile dysfunction?
- Q14: What are common adverse events of the four PDE5Is ?
- Q15: What are the absolute contraindication to use PDE5Is for the ?
- Q16: What is alprostadil? Can it be used in the treatment of erectile dysfunction?
- Q17: What is in Trimix used in the treatment of erectile dysfunction?
- Q18: What are the drugs used for intracavernous therapy?
- Q19: What is LI-ESWT? Can it be used in the treatment of erectile dysfunction?
- Q20: Can Venous ligation surgery be used in the treatment of erectile dysfunction?
- Q21: What are the complications of penile prosthesis implantation?
- Q22: What is the Penile prostheses models available on the market?
- Q23: What is erectile dysfunction?
- Q24: How prevalent is erectile dysfunction, and does it vary by age?
- Q25: What role do psychological factors play in the development of ED?
- Q26: How does vascular health contribute to erectile function?
- Q27: Can hormonal imbalances lead to erectile dysfunction?
- Q28: What are some lifestyle factors that can contribute to ED?
- Q29: How is erectile dysfunction diagnosed?
- Q30: What are the available treatment options for erectile dysfunction?
- Q31: How do phosphodiesterase type 5 inhibitors (PDE5Is) work in treating ED?
- Q32: Are there alternative medications or therapies for ED?
- Q33: What is the role of testosterone replacement therapy in ED treatment?
- Q34: Can changes in diet and exercise improve erectile function?
- Q35: How effective are vacuum erection devices in treating ED?
- Q36: Is there a connection between smoking and erectile dysfunction?
- Q37: Can trauma or injury to the pelvic region lead to erectile dysfunction?
- Q38: What is the success rate of penile prosthesis implantation for ED?
- Q39: What are the accepted values for vascular insufficiency in penile Doppler ultrasound?
- Q40: What is the association between obesity and erectile dysfunction, and can weight loss help?

Erektıl Disfonksiyon ile ilgili Sorular devamı

- Q41: What are the primary risk factors for developing ED?
- Q42: How does age impact the risk of developing ED?
- Q43: Can genetic factors contribute to the risk of ED?
- Q44: What role does obesity play in increasing the risk of ED?
- Q45: Can excessive alcohol consumption lead to ED?
- Q46: Can psychological factors, such as depression, increase the risk of ED?
- Q47: Can certain medications increase the risk of ED?
- Q48: Is there a correlation between cardiovascular diseases and ED risk?
- Q49: Can hormonal imbalances, such as low testosterone, be a risk factor for ED?
- Q50: What role does metabolic syndrome play in increasing the risk of ED?
- Q51: Can an individual's mental health history impact the risk of ED?
- Q52: How does the vascular system contribute to the pathophysiology of ED?
- Q53: What is the role of nitric oxide in the physiology of penile erection?
- Q54: What is the impact of neurogenic disorders on the pathophysiology of ED?
- Q55: How do hormonal imbalances, such as hypogonadism, contribute to ED?
- Q56: Can trauma or injury to the pelvic region affect erectile function?
- Q57: What information is gathered during a detailed medical history for ED diagnosis?
- Q58: Can nocturnal penile tumescence tests aid in the diagnosis of ED?
- Q59: How does the International Index of Erectile Function (IIEF) help in diagnosis?
- Q60: Can imaging studies, such as MRI or CT scans, contribute to ED diagnosis?
- Q61: How is erectile function assessed through physical examination?
- Q62: What is the role of hormonal evaluation in diagnosing ED?
- Q63: What is the significance of evaluating psychological factors in ED diagnosis?
- Q64: Which PDE5 inhibitor can cause the most headache?
- Q65: How are nocturnal erections monitored during sleep studies?
- Q66: How is the prevalence of ED different in individuals with diabetes?
- Q67: Can a single test definitively diagnose ED?
- Q68: What are some lifestyle factors that can contribute to ED?
- Q69: Can excessive alcohol consumption contribute to ED?
- Q70: What is the role of testosterone in male sexual health?
- Q71: How does Peyronie's disease affect the pathophysiology of ED?
- Q72: Can chronic kidney disease contribute to the development of ED?
- Q73: How do changes in penile fibrosis contribute to the development of ED?
- Q74: How do hormonal changes during andropause influence ED in aging men?
- Q75: What is the importance of evaluating penile arterial inflow in ED diagnosis?
- Q76: How does the endocrinological evaluation contribute to ED diagnosis?
- Q77: How does the use of cavernosography contribute to ED diagnosis?
- Q78: Can a prostate-specific antigen (PSA) test aid in ED diagnosis?
- Q79: What is the role of a lipid profile in assessing cardiovascular risk for ED?
- Q80: Can genetic testing contribute to understanding the risk factors for ED?

Tablo 1. EAU ED kılavuzlarının etiyoloji ve risk faktörleri ile patofizyoloji alt başlıklarına göre ChatGPT ve Google AI'nın karşılaştırılması

	Etiyoloji ve Risk Faktörleri			Patofizyoloji		
	Google AI	ChatGPT	P değeri	Google AI	ChatGPT	P değeri
Ortalama± SD	3,8 ± 0,41	3,65 ± 0,58	0,56	3,55 ± 0,51	3,7 ± 0,47	0,43
Tamamen doğru; n (%)	16 (%80)	14 (%70)		11 (%55)	14 (%70)	
Doğru ancak yetersiz; n (%)	4 (%20)	5 (%25)		9 (%45)	6 (%30)	
Doğru ve yanıltıcı; n (%)	0 (0)	1 (%5)		0 (0)	0 (0)	
Tamamen yanlış; n (%)	0 (0)	0 (0)		0 (0)	0 (0)	

Tablo 2. EAU ED kılavuzlarının tanı ve tedavi alt başlıklarına göre ChatGPT ve Google AI'nın karşılaştırılması

	Tanı			Tedavi		
	Google AI	ChatGPT	P değeri	Google AI	ChatGPT	P değeri
Ortalama± SD	3,6 ± 0,59	3,35 ± 0,87	0,478	3,50 ± 0,68	3,40 ± 0,94	0,703
Tamamen doğru; n (%)	13 (%65)	11 (%55)		12 (60)	13 (%65)	
Doğru ancak yetersiz; n (%)	6 (%30)	6 (%30)		6 (30)	3 (%15)	
Doğru ve yanıltıcı; n (%)	1 (%5)	2 (%10)		2 (10)	3 (%15)	
Tamamen yanlış; n (%)	0 (0)	1 (%5)		0 (0)	1 (%5)	

ve grafiksel olarak sunuldu. Ayrıca, her parametre için ortalama değerler hesaplandı. Daha sonra, iki grup arasındaki karşılaştırma için Mann-Whitney U testi kullanıldı. $P < 0,05$ istatistiksel olarak anlamlı kabul edildi.

BULGULAR

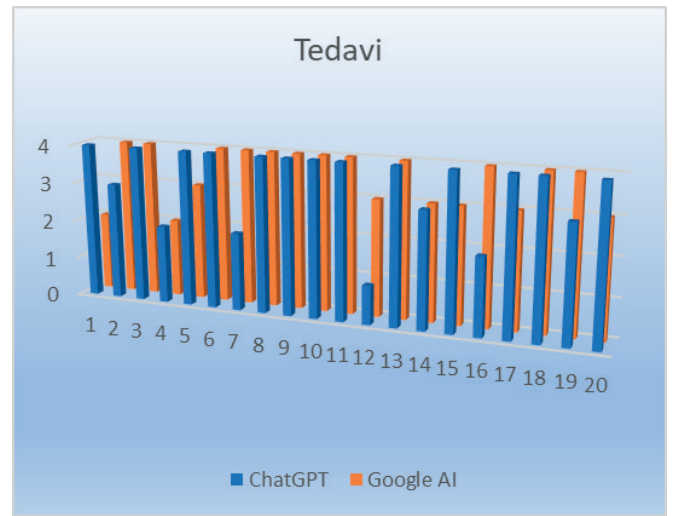
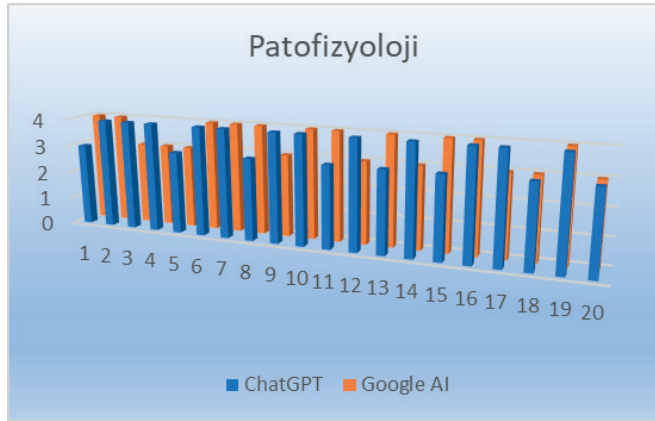
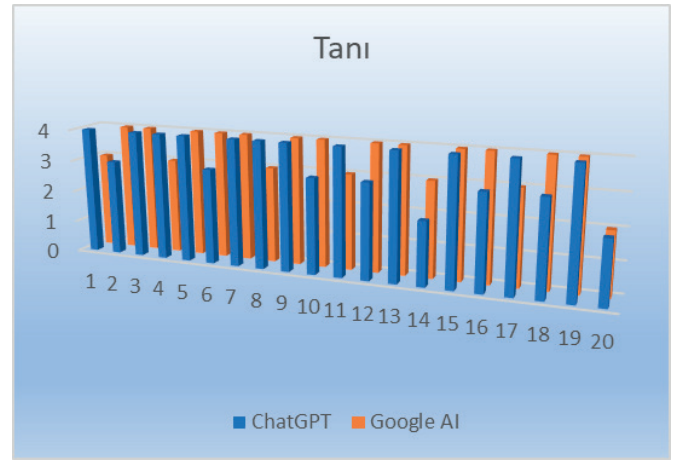
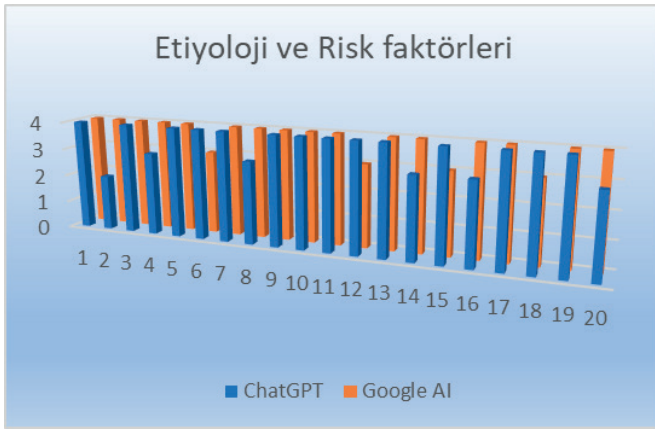
Etiyoloji ve risk faktörlerini içeren soruların yanıtları karşılaştırıldığında, Google AI ve ChatGPT'nin ortalama puanları sırasıyla $3,8 \pm 0,41$ ve $3,65 \pm 0,58$ 'dir ($p=0,56$). Google AI, 16 (%80) soruya tamamen doğru yanıt verirken, ChatGPT 14 (%70) soruya tamamen doğru yanıt verdi. Google AI, dört (%20), ChatGPT beş (%25) soruya doğru ancak yetersiz yanıt verdiler. ChatGPT bir (%5) soruya doğru ve yanıltıcı bilgilerin karışımı olarak yanıt verdi. Her iki grupta da tamamen yanlış yanıt saptanmadı (Tablo 1). Patofizyoloji ile ilgili soruların yanıtları karşılaştırıldığında, Google AI ve ChatGPT'nin ortalama puanları sırasıyla $3,55 \pm 0,51$ ve $3,7 \pm 0,47$ 'dir ($p=0,43$). Google AI, 11 (%55), ChatGPT 14 (%70) soruya tamamen doğru yanıt verdiler. Her iki dil modeli de patofizyoloji ile ilgili sorulara yanıltıcı veya yanlış yanıtlar vermediler. Tanı ile ilgili soruların yanıtları karşılaştırıldığında, Google AI ve ChatGPT'nin ortalama puanları sırasıyla $3,6 \pm 0,59$ ve $3,35 \pm 0,87$ 'dir ($p=0,478$). Google AI, 13 (%65) soruya tamamen doğru yanıtlar verirken, ChatGPT bunu 11 (%55) soruda gerçekleştirebildi. Google

AI, doğru ancak yanıltıcı bilgilerin de olduğu bir, ChatGPT ise iki yanıt verdi. Google AI hiç yanlış cevap vermedi ancak ChatGPT tamamen yanlış bir yanıt verdi. Tedavi ile ilgili yanıtlar karşılaştırıldığında, Google AI ve ChatGPT'nin ortalama puanları sırasıyla $3,50 \pm 0,68$ ve $3,40 \pm 0,94$ 'dir ($p=0,703$). Google AI 12 (%60) soruya tamamen doğru yanıt verirken, ChatGPT bunu 13 (%65) soruda gerçekleştirdi. Google AI hiç yanlış cevap vermedi, ChatGPT ise tamamen yanlış bir yanıt verdi. EAU ED kılavuzlarının ana başlıklarına dayanarak ChatGPT ve Google AI arasındaki detaylı karşılaştırma için Şekil 1–2'ye bakabilirsiniz.

TARTIŞMA

Yapay zekânın giderek artan kullanımı, özellikle sağlık gibi karmaşık alanlarda doğru ve kapsamlı cevaplar sağlama yeteneği konusunda soruları gündeme getirmektedir. Bu çalışmanın amacı, Google AI ve ChatGPT gibi popüler iki yapay zekâ modelinin, EAU ED kılavuzlarından elde edilen sorulara doğru ve yeterli cevaplar sağlama performansını karşılaştırmaktır.

Çalışma sonucunda Google AI ve ChatGPT arasında anlamlı bir fark saptanmadı. Ancak, patofizyoloji ile ilgili sorulara verilen yanıtlar dışında, Google AI'nın ChatGPT'ye kıyasla daha doğru yanıtlar sunduğunu ortaya koydu.



Şekil 1. EAU ED kılavuzlarına göre etiyoloji ve risk faktörleri (A), ve patofizyoloji (B) açısından ChatGPT ve Google AI'nın karşılaştırılması

Şekil 2. EAU ED kılavuzlarına göre tanı (A) ve tedavi (B) açısından ChatGPT ve Google AI'nın karşılaştırılması.

Google AI ve ChatGPT gibi yapay zekâ modelleri, üroloji gibi tıbbi alanlarda yaygın bir şekilde kullanılmakta ve giderek artan bir ilgiyle karşılanmaktadır. Bu modellerin akademik tıpta tıbbi rehberlik ve bilgi sağlama konusundaki avantajları, kalitesi ve sınırlamalarını araştıran çeşitli çalışmalar bulunmaktadır.^[11] Ancak, potansiyellerine rağmen, yapay zekâ modellerinin güvenilirliği, içerik kalitesi ve etik sonuçlarının kullanımıyla ilgili endişeler devam etmektedir. Szczeniewski ve ark.'nın yaptıkları çalışmada ürolojik hastalıklar ile ilgili sorulara ChatGPT, Google Bard ve Copilot (Microsoft) gibi yapay zekâ modellerinin verdikleri yanıtları değerlendirdi. Bu hastalıklar arasında prostat kanseri, böbrek kanseri, mesane kanseri, üriner sistem taşı hastalığı ve benign prostat hiperplazisi (BPH) bulunmaktadır.^[12] Çalışmada, Google Bard ve Copilot'un ürolojik hastalıklar hakkında daha detaylı açıklamalar sağladığı saptandı. Cerrahi tedavi için en yüksek puanları Google Bard aldı, onu ChatGPT ve ardından Copilot izledi.

Benzer şekilde, Connors ve ark.'ları endoüroloji ile ilgili 32 sorunun yanıtlarını ChatGPT ve Google Bard arasında karşılaştırdı.^[13] Her iki modelin de yeterli yanıtlar sağladığını ancak genellikle daha yüksek okuma seviyelerine ihtiyaç duyulabileceğini saptadılar. Bu durum,

kullanılabilirliklerini potansiyel olarak daha dar bir kitleye sınırlayabilir.

Çalışmamız hem Google AI'nın hem de ChatGPT'nin etiyoloji ve risk faktörleri ile ilgili sorulara tamamen doğru veya doğru ancak yetersiz yanıtlar sağladığını ortaya koydu. Ancak, patofizyoloji, tanı ve özellikle tedavi ile ilgili sorulara, tamamen doğru veya doğru ancak yetersiz yanıtların yanı sıra doğru ve yanıltıcı bilgiler veya tamamen yanlış yanıtlar da sundular. Bu durum, her iki yapay zekâ modelinin de, sorunun doğasıyla bağlantılı olarak yeterince doğru yanıtlar sağlayamayabileceğini ve dikkatli bir değerlendirme gerektirebileceğini göstermektedir.

Bu çalışmanın bazı sınırlamaları bulunmaktadır. ED ile ilgili soruların önemi vurgulanmış olup, bu soruların modellerin karşılaşılabileceği geniş tıbbi soru yelpazesini tam olarak yansıtmayabileceği kabul edilmiştir. Bu nedenle, sağlık alanındaki yapay zekâ modellerinin doğruluğunu ve güvenilirliğini değerlendiren gelecek araştırmalar, daha kapsamlı çalışmaları, daha büyük ve daha çeşitli veri kümelerini ve ek değerlendirme parametrelerini içermelidir.

SONUÇ

Google AI ve ChatGPT gibi yapay zekâ modelleri, ED ile ilgili sorulara genellikle yeterli ve doğru cevaplar verebilmiş olsalar da, cevapların doğruluğunun sorudan soruya değişebileceğini belirtmek önemlidir. Bu bulgular, gelişim ve öğrenme alanlarının önemini vurgularken, sürekli eğitim ve güncellenmenin kritik rolünü belirtmektedir. Teknolojinin ilerlemesiyle birlikte, büyük dil modellerine yapılacak güncellemeler, daha doğru ve güvenilir bilgilere erişimi artırabilir.

Etik Kurul Onayı

ChatGPT ve Google AI halka açık kullanılabilir yapay zeka modelleridir ve çalışmamızda hayvan veya insan araştırma katılımlarını bulunmamaktadır. Bu sebeplerle çalışmamıza etik komite onayı gerekmemiştir.

Hakem Değerlendirmesi

Dış bağımsız.

Çıkar Çatışması

Yazarlar çıkar ilişkisi olmadığını beyan etmişlerdir.

Finansal Destek

Herhangi bir mali destek alınmamıştır.

Ethics Committee Approval

ChatGPT and Google AI are publicly available artificial intelligence models, and our study does not involve any animal or human research participants. Therefore, ethical committee approval was not required for our study

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

- Deng J, Lin Y. The benefits and challenges of ChatGPT. An overview. *Front Comput Intelligent Syst.* 2022;2(2):81–3. <https://doi.org/10.54097/fcis.v2i2.4465>
- Mokrane S, SIAD. The promise and perils of Google's Bard for scientific research, 2023. <https://doi.org/10.17613/yb4n-mc79>
- King MR. The Future of AI in Medicine: A Perspective from a Chatbot. *Ann Biomed Eng.* 2023;51(2):291–5. <https://doi.org/10.1007/s10439-022-03121-w>
- Singh SK, Kumar S, Mehra PS; Chat GPT & Google Bard AI. A Review: International Conference on IoT, Communication and Automation Technology (ICICAT). Gorakhpur, India, 2023. pp. 1-6. <https://doi.org/10.1109/ICICAT57735.2023.10263706>
- EAU Guidelines. Edn. presented at the EAU Annual Congress Milan 2023. ISBN 978-94-92671-19-6.
- Fox S; Pew Research Center. The social life of health information; 2011. Available at: <https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/> Accessed: May 1, 2023.
- Warren C. 30 Facts & Statistics on social media and healthcare [internet]. Referral MD, 2018. <https://getreferralmd.com/2017/01/30-facts-statistics-on-social-media-and-healthcare/>
- Fode M, Nolsøe AB, Jacobsen FM, Russo GI, Østergren PB, Jensen CFS, et al; EAU YAU Men's Health Working Group. Quality of Information in YouTube Videos on Erectile Dysfunction. *Sex Med.* 2020;8(3):408–13. <https://doi.org/10.1016/j.esxm.2020.05.007>
- Alzahrani MA, Khan MA, Hakami BO, Alahmadi A, Alzahrani M, Alsaleh F, et al. Is Arabic Information on YouTube About Erectile Dysfunction Based on Scientific Evidence?. *Res Rep Urol.* 2023;15:261–72. <https://doi.org/10.2147/RRU.S410127>
- Hu M-Y, Ho DR, Weinberger JM, Osadchiy V, Mills JN, Eleswarapu SV. Guideline-nonconformant investigational treatments for erectile dysfunction: what are patients learning? *Urology.* 2023;173:111–118. <https://doi.org/10.1016/j.urology.2022.11.044>
- Whiles BB, Bird VG, Canales BK, DiBianco JM, Terry RS. Caution! AI Bot Has Entered the Patient Chat: ChatGPT Has Limitations in Providing Accurate Urologic Healthcare Advice. *Urology.* 2023;180:278–284. <https://doi.org/10.1016/j.urology.2023.07.010>
- Szczesniowski JJ, Ramos Alba A, Rodríguez Castro PM, Lorenzo Gómez MF, Sainz González J, Llanes González L. Quality of information about urologic pathology in English and Spanish from ChatGPT, BARD, and Copilot. *Actas Urol Esp (Engl Ed).* 2024;48(5):398–403. <https://doi.org/10.1016/j.acuroe.2024.02.009>
- Connors C, Gupta K, Khusid JA, Khargi R, Yaghoubian A, Levy M, et al. Evaluation of the current status of artificial intelligence for endourology patient education: a blind comparison of ChatGPT and Google Bard against traditional information resources. *J Endourol.* 2024. <https://doi.org/10.1089/end.2023.0696>