

Pediatric İlaç Dozu Hesaplamalarında Büyük Dil Modellerinin ve Hemşirelik Öğrencilerinin Verdikleri Yanıtların Karşılaştırılması

Comparing Large Language Models and Responses Provided by Nursing Students in Pediatric Drug Dosage Calculations

Eda GÜNEY ŞAHİN¹, Ayla İrem AYDIN²

ÖZ

Amaç: Bu çalışmanın amacı, pediatrik ilaç dozu hesaplamalarında farklı büyük dil modellerinin (ChatGPT-4, ChatGPT-4o, Gemini ve Gemini Advanced) performansını hemşirelik öğrencilerinin performansıyla karşılaştırmaktır.

Gereç ve Yöntem: Kesitsel ve karşılaştırmalı desenle yürütülen çalışma, 2024–2025 güz döneminde bir devlet üniversitesinin Hemşirelik Bölümü öğrencileriyle (n=162) gerçekleştirilmiştir. Literatüre dayalı 19 pediatrik ilaç dozu sorusu öğrencilere ve dört büyük dil modeline yöneltilmiştir. Öğrencilerin ve modellerin yanıtları karşılaştırılarak analiz edilmiştir.

Bulgular: Öğrencilerin doğruluk oranı %78,95 belirlenmiş, Gemini dil modeliyle karşılaştırıldığında anlamlı bir farklılık göstermemiştir (p=0,252). ChatGPT-4, ChatGPT-4o ve Gemini Advanced tüm soruları %100 doğru yanıtlamıştır. Tüm sorulara verilen cevaplar karşılaştırıldığında, öğrenciler ile ChatGPT-4, ChatGPT-4o ve Gemini Advanced arasında anlamlı fark tespit edilmiştir (p<0,001).

Sonuç: Bulgular; büyük dil modellerinin pediatrik ilaç dozu hesaplamalarında hemşirelik öğrencilerinden daha doğru sonuçlar verdiğini göstermektedir. ChatGPT-4, ChatGPT-4o ve Gemini Advanced tüm soruları doğru yanıtlamıştır. Gemini'nin öğrencilere yakın sonuçlar vermesi, insan denetiminin gerekliliğini vurgulamaktadır; gelecekte güncel modellerle güvenilirlik daha iyi değerlendirilebilir.

Anahtar Kelimeler: Pediatrik ilaç dozu hesaplama, Büyük dil modelleri, ChatGPT, Gemini, Hemşirelik öğrencileri

GİRİŞ

Doğal Dil İşleme (NLP), bilgisayarlar ve insanlar arasında bulunan etkileşimi taklit ederek işlev gösteren bir yapay zeka alanıdır. NLP'nin amacı insan dilini doğru ve doğal bir şekilde anlayıp üretebilen ve işleyebilen algoritmalar ve modeller

ABSTRACT

Title: Comparison of Large Language Models and Responses Given by Nursing Students in Paediatric Drug Dose Calculations

Aim: The aim of this study is to compare the performance of different large language models (ChatGPT-4, ChatGPT-4o, Gemini, and Gemini Advanced) in paediatric drug dose calculations with that of nursing students.

Method: The cross-sectional and comparative study was conducted with students (n=162) from the Nursing Department of a state university during the autumn term of 2024–2025. Nineteen paediatric medication dosage questions based on the literature were posed to the students and four major language models. The responses of the students and models were compared and analysed.

Results: The students' accuracy rate was determined to be 78.95%, showing no significant difference when compared to the Gemini language model (p=0.252). ChatGPT-4, ChatGPT-4o, and Gemini Advanced answered all questions correctly at a rate of 100%. When comparing the answers provided to all questions, a significant difference was detected between the students and ChatGPT-4, ChatGPT-4o, and Gemini Advanced (p<0.001).

Conclusion: The findings indicate that large language models produce more accurate results than nursing students in paediatric drug dosage calculations. ChatGPT-4, ChatGPT-4o, and Gemini Advanced answered all questions correctly. Gemini's results being close to those of students highlights the necessity of human oversight; reliability can be better assessed in the future with updated models.

Keywords: Pediatric drug dose calculation, Large language models, ChatGPT, Gemini, Nursing students

ortaya koymaktır (1,2). Araştırmacılar tarafından çok sayıda metin verisinden öğrenebilen yeni nesil istatistiksel yapay zeka modelleri oluşturulmuştur ve bu modeller, önceki kural tabanlı sistemlere göre önemli bir gelişmeyi ortaya koymaktadır. Bu modeller geniş metin veri kümeleri üzerinde eğitilerek okuduğunu anlama, soru cevaplama, kod üretme, metin özetleme, anlamsal benzerlik eşleştirme ve meydana getirmek gibi son derece gerçekçi insan benzeri yanıtlar üretebilmektedir (1,3). Bu programların öncülerinden biri Chat Generative Pre-Trained Transformer (ChatGPT)

1- Yüksek Lisans Öğrencisi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Pediatri Hemşireliği Bölümü, Bursa, Türkiye

E-mail: guneyshahineda@gmail.com ORCID ID: 0009-5846-7578

2- Dr. Öğr. Üyesi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Fakültesi Hemşirelik Bölümü, Bursa, Türkiye

E-mail: iremaydin@uludag.edu.tr ORCID ID: 0000-0002-9402-9861

Gönderim Tarihi:24.09.2025 - Kabul Tarihi: 03.08.2026

Kasım 2022 tarihinde, Generative AI uygulaması olan Gemini, Google tarafından Mart 2023'te piyasaya sürmüştür (4). Bu modeller piyasaya sürüldükten çok kısa bir süre içinde ücretsiz erişimin de etkisiyle birlikte yüksek oranda kullanıcı sayısına ulaşmıştır (5-7). Bu dil modelleriyle ilgili literatürde bilimsel çalışmalar yapılmaya başlanmıştır (4,8,9).

İlaç uygulaması, hemşirelerin sıklıkla yerine getirdiği en kritik görevlerden biridir ve hekim tarafından reçete edilen ilaçlar, hemşireler tarafından uygulanmaktadır (10). Çocuklara herhangi bir yolla ilaç uygulanması özel dikkat ve potansiyel hataları azaltmak veya bunların önüne geçmek için çok yönlü müdahaleler gerektiren karmaşık bir süreç olabilir (11). Çocuklarda dozlar erişkin bireylere göre çok daha düşüktür. Bu sebeple hemşireler ilaç uygulamasında daha dikkatli ve hassas olmalıdırlar (12). Gereğinden fazla uygulanan ilaç dozlarında çocuklar bunu metabolize edemeyebilir, gereğinden az uygulanan dozlarda ise tedavi gecikmeleri yaşanabilir (13). İlaçların çoğu erişkinlere yönelik hazırlanmış olup birçok ilacın pediatrik formu bulunmamaktadır. Pediatrik ilaç uygulamalarındaki problemlere bakıldığında pediatriye uygun kolay doz formlarının bulunmaması, miligram ve mikrogram gibi ölçüm birimlerinin birbirine çevrilmesi, çocukların yaş, kilo ve yüzey alanının yorumlanmasını gerektiren kişisel doz ayarlamaları yer almaktadır. İlaç uygulama hataları ile ilgili yapılan çalışmalar incelendiğinde ise çoğunluğun doz hatalarından oluştuğu görülmektedir (13-15). Pediatrik ilaç uygulamalarında ortaya çıkan hatalar ile ilgili yapılan bir çalışmada, hemşireler arasında en yaygın hata, %39.3'lük bir oran ile yanlış dozda ilaç uygulaması olarak tespit edilmiştir (16). Ersun ve arkadaşlarının (2013) yaptığı başka bir çalışmada ise pediatri hemşirelerinden %27.7'sinin daha önce ilaç hatası yaptığı, ilaç hatası yapanlardan %67.8'inin yanlış doz ilaç uyguladığı saptanmıştır (17).

Hemşireler, ilaç uygulama hatalarını önleyerek hasta güvenliğinin sağlanmasında ve olası zararların engellenmesinde temel bir işlev üstlenir (18). Hemşirelerin henüz öğrenciyken

ilaç uygulama hatalarını önlemeyi öğrenmeleri gerekmektedir (19). Bu konuda günümüz teknolojisi olan yapay zekadan faydalanılabilir. Pediatri hemşireliğinde yapay zeka kullanımı ile ilişkili yapılan çalışma alanı oldukça sınırlıdır (20). Büyük dil modelleri çeşitli tıbbi ortamlarda test edilmiş olsa da, bildiğimiz kadarıyla hiçbir çalışmada bu dil modellerinin pediatrik ilaç doz hesaplama performansı özel olarak değerlendirilmemiştir. Büyük dil modellerinin sağlık alanındaki kullanımı oldukça yaygınlaşmış, fırsatları ve sınırlılıkları araştırılarak sağlık çalışmalarına konu olmuştur. Bu çalışmadaki amacımız ChatGPT-4, ChatGPT-4o, Gemini ve Gemini Advanced uygulamalarına sorular sorarak hemşirelerin pediatrik ilaç dozu hesaplarırken yardım alabileceği farklı büyük dil modellerinin güvenilirliğini test etmek ve hemşirelik öğrencilerinin cevapları ile karşılaştırmaktır.

Gereç ve Yöntemler

Araştırmanın Türü ve Örneklemi

Bu çalışma Ekim-Aralık 2024 tarihleri arasında bir devlet üniversitesinin Hemşirelik Bölümü 2024-2025 eğitim öğretim yılı güz döneminde Çocuk Sağlığı ve Hastalıkları Hemşireliği dersini alan öğrencilerle kesitsel karşılaştırmalı bir çalışma olarak yapılmıştır. Araştırmanın evrenini Çocuk Sağlığı ve Hastalıkları Hemşireliği dersini alan 171 öğrenci oluştururken, örneklemi dahil edilme kriterlerine uyan 162 öğrenci oluşturmuştur.

Verilerin Toplanması

Bu aşamada, ilgili literatür taranarak oluşturulan 19 adet pediatrik ilaç dozu hesaplama sorusu, öğrencilere 40 dakikalık süre tanınarak yüz yüze anket formu ile uygulanmıştır. Aynı sorular ChatGPT-4, ChatGPT-4o, Gemini ve Gemini Advanced dil modellerine her defasında yeni bir sayfa açıp bilgiyi hafızaya kaydetmesine karşı önlem alınacak şekilde sorulmuştur. Yapay zeka uygulamalarına "Pediatri servisinde staja çıkan öğrenci hemşiresin. Birazdan sana pediatrik ilaç dozu hesaplamasıyla ilgili sorular soracağım. Bu sorulara cevap vermeni istiyorum." talimatı verilerek sorulara yanıt istenmiştir. Sorular pediatrik ilaç dozu hesaplamalarından oluşan

birim dönüşüm hesaplamaları, güvenli doz hesaplamaları, vücut ağırlığına göre doz hesaplamaları, infüzyon hızı hesaplamaları ve milimetre-oran hesaplamaları olmak üzere 5 alt kategoride incelenmiştir. Her bir soru 5 puan olmak üzere en yüksek puan 95 olarak belirlenmiştir. Boş bırakılan sorular yanlış olarak değerlendirilmiştir.

İstatistiksel Analiz

Çalışmanın verileri IBM SPSS Statistics Version 29.0.2.0 (20) programı ile analiz edilmiştir. Tanımlayıcı istatistikler için yüzde ve frekans dağılımları kullanılmıştır. Verilerin dağılımı normallik testleriyle değerlendirilmiş, parametrik test varsayımları sağlanmadığı için analizlerde One-Sample Wilcoxon Signed Rank Test gibi parametrik olmayan testler tercih edilmiştir. Anlamlılık düzeyi $p < 0,05$ olarak kabul edilmiştir.

Araştırmanın Etik Yönü

Bu çalışmanın yürütülmesi için bir devlet üniversitesinin Sağlık Bilimleri Araştırma ve Yayın Etik Kurulu'ndan onay (25.09.2024 tarihli ve 2024-07 sayılı) ve çalışmanın yapılacağı kurumdan izin alınmıştır.

Bulgular

Bu çalışmada pediatrik ilaç dozu hesaplama sorularına verilen cevapların tanımlayıcı verileri Tablo 1'de gösterildi. Öğrencilerin tüm sorulara verdikleri doğru yanıt oranı %78,95 bulundu. ChatGPT-4, ChatGPT-4o ve Gemini Advanced tüm sorulara %100 doğru yanıt verdi. Gemini modeli ise %78,95 oranında performans göstererek öğrencilerle aynı başarı ortalamasına sahipti. Yapılan istatistiksel analizler sonucunda, öğrencilerin tüm sorulara verdiği yanıtlar ile ChatGPT-4, ChatGPT-4o ve Gemini Advanced'in verdiği yanıtlar arasında istatistiksel olarak anlamlı farklılık ortaya çıkmıştır ($p < 0,001$). Gemini modeli ile öğrenci yanıtları kıyaslandığında istatistiksel anlamlı bir fark bulunmamıştır ($p = 0,252$).

Soruların alt kategorilere göre yapılan analizinde, birim dönüşüm, güvenli doz hesaplama, vücut ağırlığına ilişkin hesaplama soruları, infüzyon hızı hesaplamaları ve milimetre-oran hesaplamalarında, çalışmaya dahil edilen büyük dil modelleri (ChatGPT-4, ChatGPT-4o, Gemini, Gemini Advanced) ile öğrencilerin verdikleri yanıtlar arasında istatistiksel olarak anlamlı farklar bulunmuştur (Tablo 2; $p < 0,001$). Bu durum, söz konusu üç dil modelinin (ChatGPT-4, ChatGPT-4o ve Gemini Advanced), öğrencilerle karşılaştırıldığında sistematik olarak daha yüksek düzeyde doğruluk oranları verdiğini göstermektedir. Öğrenci ve Gemini modelinin kategorilere göre verdikleri yanıtların doğruluğu incelendiğinde; güvenli doz, vücut ağırlığı ve mililitre-oran hesaplamaları kategorilerinde öğrenciler lehine pozitif yönde; birim dönüşüm ve infüzyon hızı kategorilerinde ise negatif yönde istatistiksel olarak anlamlı fark saptanmıştır (Tablo 2; $p < 0,001$).

Dil modellerinin pediatrik ilaç doz kategorisi bazında başarı yüzdeleri şekil 1'de verilmiştir. ChatGPT-4, ChatGPT-4o ve Gemini Advanced modelleri tüm kategorilerdeki sorulara doğru yanıt verirken, Gemini modeli daha düşük bir performans sergilemiştir.

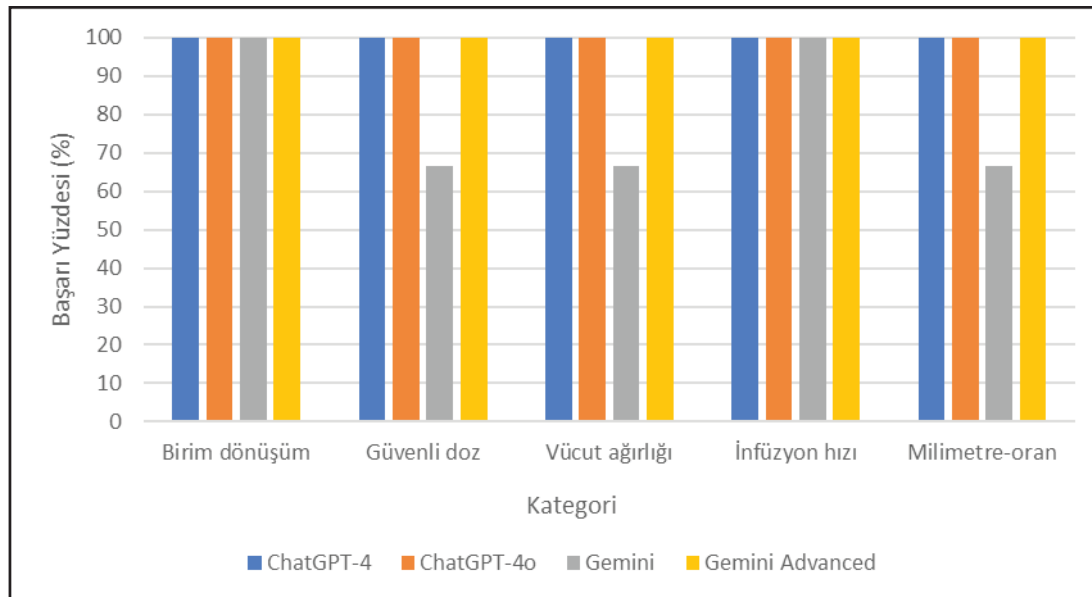
Tablo 1. Öğrenciler ile dil modellerinin genel başarı yüzdeleri ve istatistiksel karşılaştırması (p-değeri), Bursa, 2025

Katılımcı/Model	Ortalama Başarı Yüzdesi (%)	Öğrenci ile Karşılaştırma p-değeri
Öğrenci (n=162)	78,95	-
ChatGPT-4	100	<0,001
ChatGPT-4o	100	<0,001
Gemini	78,95	,252
Gemini Advanced	100	<0,001

Tablo 2. Kategorilere göre öğrenciler ve dil modellerinin istatistiksel karşılaştırması (p-değeri), Bursa, 2025

Kategori	Birim dönüşüm	Güvenli doz	Vücut ağırlığı	İnfüzyon hızı	Milimetre-oran
Öğrenci ve ChatGPT-4	Z test = -5,088; p < 0,001	Z test = -5,756 p < 0,001	Z test = -4,414 p < 0,001	Z test = -9,680 p < 0,001	Z test = -9,700 p < 0,001
Öğrenci ve ChatGPT-4o	Z test = -5,088 p < 0,001	Z test = -5,756 p < 0,001	Z test = -4,414 p < 0,001	Z test = -9,680 p < 0,001	Z test = -9,700 p < 0,001
Öğrenci ve Gemini	Z test = -5,088 p < 0,001	Z test = 8,257 p<0,001	Z test = 10,658 p < 0,001	Z test = -9,680 p < 0,001	Z test = 3,278 p < 0,001
Öğrenci ve Gemini Advanced	Z test = -5,088 p < 0,001	Z test = -5,756 p < 0,001	Z test = -4,414 p < 0,001	Z test = -9,680 p < 0,001	Z test = -9,700 p < 0,001

*p değeri < 0,001



Şekil 1. Dil Modellerinin İlaç Doz Kategorisi Bazında Başarı Yüzdeleri (%), Bursa, 2025

Tartışma

Büyük dil modelleri hızlı bir şekilde bilgiye ulaşabilmenin bir yolu olmakla birlikte, güvenli kullanım için yanıtlarının doğruluğunun değerlendirilmesi gerekmektedir. Yapay zeka teknolojilerinin gelişmesiyle birlikte, bu teknolojiler özellikle sağlık alanında oldukça yaygın kullanılmaya başlanmıştır (21–23). Bu çalışmada pediatrik ilaç dozu hesaplama sorularına verilen yanıtlar analiz edilmiştir. Öğrencilerin ve Gemini dil modelinin doğru yanıt oranı %78,95 bulunurken, ChatGPT-4, ChatGPT-4o ve Gemini Advanced modelleri tüm soruları %100 doğrulukla yanıtlamıştır. Yapılan istatistiksel analizlerde, ChatGPT-4, ChatGPT-4o ve Gemini Advanced ile öğrenciler arasında anlamlı farklar saptanırken ($p < 0,001$), Gemini modeli ile öğrenciler arasında anlamlı bir fark bulunmamıştır ($p = 0,252$). Bu çalışma bulgularına benzer şekilde, çocuk acillerde hemşirelik yaklaşımına ilişkin sınav sorularının değerlendirildiği bir araştırmada en yüksek başarıyı sırasıyla ChatGPT 4.0, öğrenciler ve ChatGPT 3.5 göstermiştir (Aydın ve Reis, 2025). Asker ve arkadaşları (2023) tarafından yapılan çalışmada ise öğrencilere, ChatGPT-3.5 ve ChatGPT-4'e biyoistatistik soruları yöneltilmiş; ChatGPT-3.5 %80, ChatGPT-4 %100 başarı oranına ulaşırken, öğrencilerin başarı oranı %67,9 olarak bulunmuştur (24). Bu sonuçlar çalışma bulgularımızla benzerdir. Ortodontik sorulara verilen yanıtların karşılaştırıldığı bir çalışmada, öğrencilerin doğruluk oranı ChatGPT-4o'ya kıyasla anlamlı derecede düşük bulunmuştur. Ancak Google Gemini 1.5 Pro ve ChatGPT-4 ile karşılaştırıldığında doğruluk oranı daha düşük olsa da bu fark istatistiksel olarak anlamlı bulunmamıştır (25). Talan ve Kalinkara (2023), anatomi sorularına verilen yanıtlarda ChatGPT'nin (%67,5) öğrencilerden (%52,8) daha iyi performans gösterdiğini ortaya koymuştur (26). Bu çalışmalar, yapay zeka destekli LLM'lerin birçok alanda yüksek performans gösterdiğini ortaya koymaktadır. Ancak performans düzeyleri; soru kapsamı, alan bilgisi, model versiyonu ve değerlendirme kriterlerine bağlı olarak değişkenlik gösterebilmektedir. Çalışma bulgularının aksine

Huh (2023) tarafından yapılan bir çalışmada, parazitoloji dersi sınavında öğrenciler (%90,8), ChatGPT'ye (%60,8) kıyasla daha yüksek başarı göstermiştir (27). Bu durumun, yapılan çalışmada kullanılan dil modelinin henüz gelişim aşamasında olması ve mevcut versiyonunun karmaşık disiplinler arası bilgi gerektiren sınavlarda yeterli performans gösterememesiyle ilişkili olabileceği düşünülmektedir. Buna karşılık, bu araştırmada en güncel dil modeli versiyonları kullanılmıştır.

Dil modellerinin pediatrik ilaç dozu hesaplamalarında klinik uygulamalarda kullanılabilirliğinin umut verici olmasıyla birlikte klinik alanlardaki kullanımı mutlaka sağlık profesyonelinin denetimi altında olmalıdır. Bu çalışmada hemşirelik öğrencilerinin dil modellerine göre daha düşük başarı düzeyine sahip olması, klinik deneyim yetersizliği ve aktif bir şekilde ilaç dozu hesaplamaya yeterince maruz kalmamış olmalarına bağlanabilir. Genel anlamda büyük dil modelleri, bilgiye dayalı sorularda yüksek doğruluk oranlarıyla yanıt verebilmektedir. Bu durum, onların klinik uygulama alanlarında sağlık profesyonellerine tamamlayıcı bir destek aracı olarak kullanılabileceğini göstermektedir. Ancak dil modellerinin farklı çalışmalarda performanslarının değişkenlik göstermesi, bu alanda daha fazla sayıda ve geniş kapsamlı araştırmalar gerektiğini ortaya koymaktadır.

Sonuç ve Öneriler

Bu çalışma, dört farklı büyük dil modelinin pediatrik ilaç dozu hesaplama performanslarının öğrenci performanslarıyla karşılaştırıldığı ilk çalışmadır. Hesaplama sorularının, klinikte karşılaşılabilecek gerçekçi senaryolar üzerinden hazırlanarak uygulama temelli bir yaklaşım benimsenmiştir. Bunlar çalışmanın güçlü yönlerini oluşturmaktadır. Bu çalışmada, pediatrik ilaç dozu hesaplamalarına yönelik olarak hemşirelik öğrencileri ile dört büyük dil modeli (ChatGPT-4, ChatGPT-4o, Gemini ve Gemini Advanced) karşılaştırılmış ve 19 soruya verdikleri yanıtların doğrulukları değerlendirilmiştir. ChatGPT-4, ChatGPT-4o ve Gemini Advanced modelleri %100 doğruluk oranıyla tüm soruları doğru yanıtlamıştır. Bu üç model ile hemşirelik

öğrencileri arasında istatistiksel olarak anlamlı fark bulunmuştur. Buna karşılık, Gemini modeli bazı sorularda hata yapmış ve öğrenci performansına daha yakın sonuçlar vermiştir. Bu durum, dil modellerinin pediatrik ilaç dozu hesaplama gibi kritik işlemlerde destek olarak kullanılabileceğini, ancak hata yapma olasılığı göz önünde bulundurularak bu kullanımın insan denetimi altında sürdürülmesi gerektiğini ortaya koymaktadır.

Bu çalışmanın bazı sınırlılıkları da mevcuttur. Sadece belli bir üniversitenin hemşirelik öğrencileri ile gerçekleştirilmesi genellenebilirliği sınırlamaktadır. Öğrencilerin demografik verilerinin olmaması, bireysel farklılıkların (yaş, klinik deneyim vb.) değerlendirilememesine sebep olmuştur. Değerlendirme yalnızca 19 soru üzerinden yapılmış olup, bu durum bazı bilgi alanlarının yeterince temsil edilememesine ve değerlendirme kapsamının sınırlı kalmasına neden olmuştur. Ek olarak dil modellerinin sürekli güncellenmesi nedeniyle, elde edilen sonuçların zamanla değişiklik gösterebileceği ve güncel modellerle yapılan çalışmalarda farklı sonuçlar elde edilebileceği göz önünde bulundurulmalıdır. Gelecekte yapılacak çalışmalarda, daha güncel dil modellerinin kullanılması ve soru içeriklerinin daha geniş kapsamlı olacak şekilde yapılandırılması önerilmektedir. Ayrıca, elde edilen bulguların genellenebilirliğini artırmak adına farklı üniversitelerde benzer yöntemlerle gerçekleştirilecek çok merkezli araştırmaların yürütülmesi faydalı olacaktır.

Çıkar Çatışması

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

KAYNAKÇA

1. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biol Sport*. 2023;40(2):615-22.
2. Patwardhan N, Marrone S, Sansone C. Transformers in the real world: a survey on NLP applications. *Information*. 2023;14(4):242.
3. Zhang M, Li J. A commentary of GPT-3 in MIT

- technology review 2021. *Fundam Res*. 2021;1(6):831-3.
4. Baytak A. The content analysis of the lesson plans created by ChatGPT and Google Gemini. *Res Soc Sci Technol*. 2024;9(1):329-50.
5. Eriç A, Özgür EG, Asker ÖF, Bekiroğlu N. ChatGPT ve sağlık bilimlerinde kullanımı. *Celal Bayar Üniversitesi Sağlık Bilim Enstitüsü Derg*. 2024;11(1):176-82.
6. Yağar SD. ChatGPT'nin sağlık alanındaki potansiyel kullanımına ilişkin çıkarımlar. *Bus Manag Stud Int J*. 2023;11(3):1226-40.
7. Yıldız MS, Alper A. Potential functions of artificial intelligence chatbot ChatGPT in health management: Scoping review. *Turk Bull Hyg Exp Biol*. 2023;80(4):523-30.
8. Aydın Aİ, Reis D. ChatGPT 3.5, ChatGPT 4.0 ve hemşirelik öğrencilerinin çocuk acillerde hemşirelik yaklaşımı dersi sınavındaki performans karşılaştırmaları. *Bandırma Onyedil Eylül Üniversitesi Sağlık Bilim Ve Araştırmaları Derg*. 2025;7(1):73-9.
9. Naz R, Akacı O, Erdoğan H, Açıkgöz A. Can large language models provide accurate and quality information to parents regarding chronic kidney diseases? *J Eval Clin Pract*. 2024;30(8):1556-64.
10. Lee SE, Quinn BL. Incorporating medication administration safety in undergraduate nursing education: A literature review. *Nurse Educ Today*. 2019;72:77-83.
11. Marufu TC, Bower R, Hendron E, Manning JC. Nursing interventions to reduce medication errors in paediatrics and neonates: Systematic review and meta-analysis. *J Pediatr Nurs*.
12. Liang HF, Wu KM, Weng CH, Hsieh HW. Nurses' views on the potential use of robots in the pediatric unit. *J Pediatr Nurs*. 2019;47:e58-64.
13. Özyazıcıoğlu N, Aydın Aİ, Sürenler S, Çınar HG, Yılmaz D, Arkan B, vd. Evaluation of students' knowledge about paediatric dosage calculations. *Nurse Educ Pract*. 2018;28:34-9.
14. Çavuşoğlu H. Çocuklarda ilaç uygulama hataları. *Turk Klin J Nurs Sci*. 2015;7(2):121-7.
15. Abolhassani N, Vollenweider P, Servet J, Marques-Vidal P. Trend and characteristics of medication errors in a Swiss academic hospital: an observational retrospective study. *Drugs Ther Perspect*. 2021;37:536-44.
16. Başer S, Manav G. Çocuk hemşirelerinin ilaç hatası yapma durumları ve eğilimlerinin incelenmesi. *Gümüşhane Üniversitesi Sağlık Bilim Derg*. 2018;7(3):41-9.
17. Ersun A, Başbakkal Z, Yardımcı F, Muslu G, Beytut D. Çocuk hemşirelerinin tıbbi hata yapma eğilimlerinin incelenmesi. *Ege Üniversitesi Hemşire Fakültesi Derg*.

2013;29(2):33-45.

18. Latimer S, Hewitt J, Stanbrough R, McAndrew R. Reducing medication errors: Teaching strategies that increase nursing students' awareness of medication errors and their prevention. *Nurse Educ Today*. 2017;52:7-9.
19. Hewitt J, Tower M, Latimer S. An education intervention to improve nursing students' understanding of medication safety. *Nurse Educ Pract*. 2015;15(1):17-21.
20. Ünal AS, Avcı A. Pediatri Hemşireliğinde Yapay Zeka. *Akdeniz Hemşire Derg*. 2024;3(1):36-43.
21. Lorenzi A, Pugliese G, Maniaci A, Lechien JR, Allevi F, Boscolo-Rizzo P, vd. Reliability of large language models for advanced head and neck malignancies management: a comparison between ChatGPT 4 and Gemini Advanced. *Eur Arch Otorhinolaryngol*. 2024;281(9):5001-6.
22. Tong L, Zhang C, Liu R, Yang J, Sun Z. Comparative performance analysis of large language models: ChatGPT-3.5, ChatGPT-4 and Google Gemini in glucocorticoid-induced osteoporosis. *J Orthop Surg*. 2024;19(1):574.
23. Gunesli I, Aksun S, Fathelbab J, Yildiz BO. Comparative evaluation of ChatGPT-4, ChatGPT-3.5 and Google

Gemini on PCOS assessment and management based on recommendations from the 2023 guideline. *Endocrine*. 2025;88(1):315-22.

24. Asker ÖF, Özgür EG, Eriç A, Bekiroğlu N. Comparing the performance of medical students, ChatGPT-3.5 and ChatGPT-4 in biostatistics exam: pros and cons as an education assistant. *Uluslar Önetim Bilişim Sist Ve Bilgi Bilim Derg*. 2023;7(2):85-94.
25. Metin U, Goymen M. Information from digital and human sources: A comparison of chatbot and clinician responses to orthodontic questions. *Am J Orthod Dentofacial Orthop*. 2025.
26. Talan T, Kalınkara Y. The Role of Artificial Intelligence in Higher Education: ChatGPT Assessment for Anatomy Course. *Int J Manag Inf Syst Comput Sci*. 2023;7(1):33-40.
27. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof*. 2023;20(1).