2025

7(1), 1-39

Üretken Yapay Zeka Performans Kıyaslaması: ChatGPT ve DeepSeek'in Kullanıcı Etkileşimleri Üzerindeki Etkisi

Koray DEMİROK* [1] Fatih EREN [2] Yusuf Samet AKKAYA [3] Dilek ÇAYIRLI [4]

Anahtar Sözcükler

Üretken Yapay Zekâ
ChatGPT
DeepSeek
Kullanıcı Etkileşimi
Performans Kıyaslaması

Makale Hakkında

Gönderim Tarihi
31 Mayıs 2025
Kabul Tarihi
27 Haziran 2025
Yayın Tarihi
30 Haziran 2025

Makale Türü
Araştırma Makalesi

Öz

Bu çalışma, üretken yapay zekâ modelleri olan ChatGPT ve DeepSeek'in kullanıcı etkileşimleri üzerindeki performansını ve etkisini karşılaştırmalı bir analizle değerlendirmeyi amaçlamaktadır. Araştırmada, Kaggle.com'dan temin edilen ve Temmuz 2023 ile Şubat 2025 arasını kapsayan, 10.000'den fazla kullanıcı etkileşim verisi içeren sentetik bir veri seti kullanılmıştır. Makine öğrenmesi algoritmaları (regresyon, karar ağaçları, kümeleme) ve istatistiksel testler (t-test, ANOVA) aracılığıyla yanıt doğruluğu, oturum süresi, kullanıcı memnuniyeti gibi değişkenler analiz edilmiştir. Bulgular, DeepSeek'in özellikle teknik görevlerde (hata ayıklama, kod optimizasyonu) daha yüksek kullanıcı derecelendirmeleri aldığını, ChatGPT'nin ise içerik üretimi ve eğitim gibi alanlarda öne çıktığını göstermiştir. DeepSeek kullanıcıları genel olarak daha yüksek deneyim puanları ve derecelendirmeler bildirirken, daha düşük terk etme oranlarına sahip olduğu gözlemlenmiştir. Çalışma, her iki platformun güçlü yönlerini, sınırlılıklarını ve kullanıcı etkileşim dinamiklerini ortaya koyarak, platformların kullanım bağlamına göre performans farklılıkları sergilediğini ve veri gizliliği ile etik konuların önemli olduğunu vurgulamaktadır.

Generative AI Performance Benchmarking: The Impact of ChatGPT and DeepSeek on User Interactions

Keywords

Generative Artificial Intelligence
ChatGPT
DeepSeek
User Interaction
Performance Comparison

Article Info

Received
May 31, 2025
Accepted
June 27, 2025
Published
June 30, 2025

Article Type
Research Paper

Abstract

This study aims to evaluate and compare the performance and impact of generative AI models ChatGPT and DeepSeek on user interactions through a comparative analysis. The research utilizes a synthetic dataset obtained from Kaggle.com, encompassing over 10,000 user interaction records collected between July 2023 and February 2025. Variables such as response accuracy, session duration, and user satisfaction were analyzed using machine learning algorithms (regression, decision trees, clustering) and statistical tests (t-test, ANOVA). The findings indicate that DeepSeek received higher user ratings particularly in technical tasks (e.g., debugging, code optimization), while ChatGPT stood out in areas such as content generation and education. DeepSeek users generally reported higher experience scores and ratings, along with lower abandonment rates. The study highlights the strengths, limitations, and user interaction dynamics of both platforms, emphasizing that performance varies depending on the context of use and underlining the importance of data privacy and ethical considerations.

Atıf: Demirok, K., Eren, F., Akkaya Y. S., Çayırlı, D., (2025). Üretken yapay zeka performans kıyaslaması: ChatGPT ve DeepSeek'in kullanıcı etkileşimleri üzerindeki etkisi *Bilgi ve İletişim Teknolojileri Dergisi*, 7(1), 1-39.

Cite: Demirok, K., Eren, F., Akkaya Y. S., Çayırlı, D., (2025). Generative AI performance benchmarking: The impact of ChatGPT and DeepSeek on user interactions. *Journal of Information and Communication Technologies*, 7(1), 1-39.

***Sorumlu Yazar/Corresponding Author** demirkoray74@gmail.com

[1] M.Sc. Student, Bartin University, Science Faculty, Bartin/Turkey, demirkoray74@gmail.com, https://orcid.org/0009-0003-2796-845X

[2] M.Sc. Student, Bartin University, Science Faculty, Bartin/Turkey, fatiheren.15@outlook.com, https://orcid.org/0009-0009-2259-0420

[3] M.Sc. Student, Bartin University, Science Faculty, Bartin/Turkey, yusufsametakkaya@gmail.com, https://orcid.org/0009-0006-3927-9100

[4] M.Sc. Student, Bartin University, Science Faculty, Bartin/Turkey, dicayirli@gmail.com, https://orcid.org/0000-0003-0621-2074

Extended Abstract

Introduction

Artificial Intelligence (AI) has progressively evolved from rule-based systems to advanced generative models capable of mimicking human-like reasoning and creativity. Generative Artificial Intelligence (GenAI), as a subdomain of AI, focuses on producing novel content such as text, images, and code by learning patterns from large datasets (Al-Busaidi et al., 2024). The current study examines the comparative performance of two prominent GenAI platforms—OpenAI's ChatGPT and DeepSeek—through the lens of user interaction metrics, leveraging synthetic datasets derived from realistic user behavior.

As GenAI systems become increasingly integrated into education, software development, and digital content creation, understanding their performance and interaction dynamics is critical (Rivas & Liang, 2023). ChatGPT has garnered widespread use due to its fluency in natural language understanding and generation, while DeepSeek, albeit less popular, is recognized for its exceptional accuracy in programming and logic-intensive tasks (Guo et al., 2024; Manik, 2025). Despite their common foundations in transformer architecture, the models differ in architectural specialization, data access policies, and reinforcement learning strategies (Du et al., 2022; Ouyang et al., 2022).

Method

This research adopts a comparative quantitative methodology, utilizing a synthetic dataset titled "DeepSeek vs ChatGPT: AI Platform Comparison" retrieved from Kaggle.com. The dataset simulates over 10,000 user interactions from July 2023 to February 2025 and includes diverse variables such as session duration, user satisfaction, response accuracy, platform retention rate, and device type. Data preprocessing involved encoding categorical variables (e.g., Query_Type, Device_Type), handling missing values, and normalization using StandardScaler. Outlier detection employed IQR-based trimming.

Machine learning techniques such as linear regression, random forest, and gradient boosting were implemented to predict user satisfaction and session behaviors. Statistical analyses, including independent samples t-tests and one-way ANOVA, evaluated significant differences between the two platforms in terms of user ratings and experience scores.

Findings

User Satisfaction and Experience

Descriptive and inferential analyses indicated notable differences in the quality of user interactions between the two platforms. ChatGPT sessions were generally associated with moderate satisfaction scores, commonly ranging between 3 and 4 stars. These moderate evaluations were frequently observed in use cases such as content creation and educational assistance. In contrast, DeepSeek consistently achieved higher user ratings, predominantly between 4 and 5 stars, especially in sessions focused on debugging, algorithm development, and code optimization.

The average user experience score for DeepSeek was approximately 2.03, compared to 1.23 for ChatGPT. A strong positive correlation ($r \approx 0.75$) was observed between user ratings and experience scores, indicating that higher interaction satisfaction typically coincided with a more positive overall experience. This correlation was evident across both platforms, with DeepSeek exhibiting a more pronounced relationship between the two metrics.

Device-Type Variations

Interaction quality varied depending on the type of device used. Sessions conducted on desktop and laptop computers yielded the highest levels of both satisfaction and experience scores for both platforms. In contrast, interactions via tablets and smart speakers displayed greater variability and lower average experience values. These results suggest that device type plays a considerable role in shaping user perceptions and experiences during interactions with generative AI platforms.

Inferential Statistics

Statistical tests confirmed that the observed differences between ChatGPT and DeepSeek were not due to chance. Independent samples t-tests revealed statistically significant differences ($p < 0.001$) in both user satisfaction and experience metrics. Analysis of variance (ANOVA) further supported these findings, with F-statistics exceeding 4000 for satisfaction and 20,000 for experience scores. These results affirm that the type of AI platform used has a significant effect on user outcomes.

Machine Learning Modeling

Regression Models

Among the various regression approaches tested, the Gradient Boosting Regression model achieved the best performance, with a Root Mean Squared Error (RMSE) of approximately 0.3825 and an R^2 value of 0.2954. Despite this relatively better performance, all models yielded R^2 values below 0.30, indicating that a large portion of the variance in user satisfaction could not be explained by the available features. This suggests the influence of other unmeasured or subjective factors that affect user perceptions.

Classification Models

In attempts to classify user satisfaction into categories such as 'Excellent', 'Fair', and 'Good', the Logistic Regression model demonstrated the highest overall accuracy, around 57.5%. The model performed particularly well in predicting the most frequently occurring class but had lower accuracy in identifying the less common categories. Support Vector Machines, by comparison, showed weaker performance, especially for underrepresented satisfaction classes, highlighting the potential need for future work to address class imbalance in the dataset.

Churn and Loyalty Modeling

A Random Forest classifier was employed to predict user churn based on interaction data, achieving perfect accuracy on the test set. This model successfully identified users who were likely to discontinue usage of the

platform. Additionally, regression models were used to predict session duration and user return frequency, with RMSE values of approximately 14.57 seconds and 3.16 respectively. These results indicate a reasonable level of predictive performance, though improvements are still possible.

Response Accuracy and Correction Need

Analysis showed that DeepSeek generally provided more accurate responses and required fewer corrections compared to ChatGPT. A classification model was used to predict whether a given response would require correction. While the model achieved high accuracy in identifying correct responses, its ability to detect errors was limited. The macro-average F1 score was approximately 0.51, suggesting that performance was moderate and could be enhanced through further optimization and data balancing strategies.

Discussion and Conclusion

This study affirms that platform-specific architectural and training differences manifest distinctly in user interactions. ChatGPT's broad adaptability (Bahrini et al., 2023; Hariri, 2023) makes it effective in educational and creative scenarios, while DeepSeek's MoE-based architecture (Du et al., 2021) provides efficiency and precision in logic-based tasks, especially programming. DeepSeek’s superior performance in debugging, its use of large-scale token windows (128K), and high efficiency in energy consumption (IBM, 2025b) reflect an optimization strategy tailored to technical use-cases. However, its closed training datasets and limited transparency in RLHF protocols (Sapkota et al., 2025; CTOL Digital, 2025) raise ethical concerns around reproducibility and bias. In contrast, ChatGPT, while offering broader linguistic fluency, is challenged by hallucination problems (Alkaissi & McFarlane, 2023) and ethical constraints related to misinformation, bias, and data privacy (Rozado, 2023; Zirpoli, 2023). Both models face scrutiny regarding user data handling under GDPR and KVKK regulations, as reflected in the temporary ban of ChatGPT in Italy (Reuters, 2024).

This study provides a comprehensive comparative analysis of ChatGPT and DeepSeek based on synthetic user interaction data. While both platforms excel in distinct domains, DeepSeek demonstrates a performance edge in technical accuracy and user satisfaction. These findings suggest that user context and task type should guide the deployment of GenAI systems. Future research should integrate real-world datasets, address classification imbalance, and further explore ethical implications such as transparency, data security, and hallucination mitigation.

The implications of this research extend to developers, AI ethicists, and organizational decision-makers seeking to optimize GenAI adoption. As generative models become increasingly embedded in daily workflows, ongoing assessment of their sociotechnical impact remains essential.

1.Giriş

"Yapay zekâ (YZ), insan davranışındaki zekâ ile ilişkilendirdiğimiz özellikleri sergileyen sistemler geliştirme teori ve pratiği ile ilgili Bilim ve Mühendislik alanıdır" (Tecuci, 2012).

Yapay zekâ teknolojileri, insan zekasını taklit etme yetenekleriyle karakterize edilir. Eren (2024), bu tür sistemlerin öğrenme, mantıksal çıkarım yapma ve karar verme becerilerine sahip olduğunu belirtmektedir. Yapay zekâ ise, algılama, doğal dil işleme, problem çözme, planlama, öğrenme, uyum sağlama ve çevreyle etkileşim kurma gibi insan zekasına özgü davranışları sergileyen sistemlerin geliştirilmesine odaklanan bir Bilim ve Mühendislik disiplinidir (Tecuci, 2012). Yapay zekâ'nın temel bilimsel amacı, insanlarda, hayvanlarda ve yapay zekalarda akıllı davranışı mümkün kılan prensipleri anlamaktır. Bu bilimsel amaç, akıllı ajanlar geliştirmek, bilgiyi biçimlendirmek ve muhakemeyi mekanikleştirmek, bilgisayarlarla çalışmayı insanlarla çalışmak kadar kolaylaştırmak ve insan ve otomatik muhakemenin tamamlayıcılığından yararlanan insan-makine sistemleri geliştirmek gibi çeşitli mühendislik hedeflerini doğrudan destekler (Tecuci, 2012).

Üretken Yapay Zekâ (GenAI), mevcut veri kümelerinden örüntüler çıkararak metin, görsel, müzik, yazılım kodu ve video gibi özgün içerikler üretebilen özel bir yapay zekâ biçimini tanımlar (Al-Busaidi vd., 2024). (Bell ve Bell, 2023) ile (Dwivedi vd., 2023), geleneksel yapay zekâ yaklaşımlarının çoğunlukla veri analizi ve sınıflandırma gibi belirli görevlerle sınırlı kaldığını belirtmektedir. Buna karşın üretken yapay zekâ, insan tarafından oluşturulmuş içeriklere oldukça yakın ve yaratıcı nitelikte çıktılar üretebilme yeteneğiyle öne çıkmaktadır. Yapay zekâ, çok katmanlı ve öğrenme tabanlı bilgisayar programlarının geliştirilmesi ve kullanılmasıyla insanların yapabileceği her şeyi yapabilen bir teknolojidir. Yapay zekâ aynı zamanda, insan davranışını ve zihinsel yeteneklerini taklit eden algoritmalar kullanarak, öğrenmek ve karar vermek için insanların yapamayacağı şeyleri yapabilen bir bilgisayar sistemi olarak da tanımlanabilir (Şentürk, 2023).

Generatif yapay zekâ modellerinin hızla gelişmesi, insanların teknolojiyle etkileşim biçimlerinde köklü bir değişimi beraberinde getirmiştir (Rivas ve Liang 2023). Bu dönüşümün öncüsü olarak ChatGPT ve DeepSeek gibi modellerin kilit bir rol oynadığını vurgulamaktadır. Doğal dil işleme ve makine öğrenmesi algoritmalarından yararlanan bu yapay zekâ sistemleri, çeşitli sektörleri dönüştürmekte ve kullanıcı beklentilerini yeniden şekillendirmektedir (Rivas ve Liang, 2023). OpenAI tarafından geliştirilen ChatGPT, doğal dili anlama konusundaki yetkinliği, bağlama duyarlılığı ve insan benzeri yanıtlar üretebilme kapasitesiyle, çok çeşitli kullanım alanlarına hitap eden esnek ve etkili bir yapay zekâ aracı olarak dikkat çekmektedir (Bahrini vd., 2023; Hariri, 2023). ChatGPT; müşteri hizmetleri, sağlık ve eğitim gibi birçok alanda uygulama bulmuştur (Hariri, 2023; Ray, 2023). DeepSeek ise belki daha az bilinen bir model olmakla birlikte, generatif yapay zekâ alanında önemli bir gelişmeyi temsil etmekte olup, kendine özgü yetenekleri ve performans özellikleriyle karşılaştırmalı bir analiz yapılmasını gerektirmektedir (Wang vd., 2024). Bu teknolojiler gelişimini sürdürürken, güçlü yönlerinin, sınırlılıklarının ve kullanıcı etkileşimleri üzerindeki etkilerinin anlaşılması araştırmacılar, geliştiriciler ve son kullanıcılar açısından büyük önem taşımaktadır. Bu modellerin performansını incelemek, yalnızca anlamlı ve bağlama uygun metinler üretme yeteneklerini değil, aynı zamanda kullanıcılarla anlamlı ve verimli diyaloglar kurabilme kapasitelerini de dikkate alan çok yönlü bir yaklaşımı gerektirir (Nazir & Wang, 2023).

1.2.Üretken Yapay Zekâ ile Kullanıcı Etkileşimi Dinamikleri

Bireylerin ChatGPT gibi üretken yapay zekâ araçlarıyla etkileşime girmeyi neden seçtiklerini anlamak, performanslarını ve etkilerini değerlendirmek için çok önemlidir (Skjuve vd., 2024). Kullanıcı etkileşimi, algılanan değer, kullanım kolaylığı ve etkileşimden elde edilen keyif dahil olmak üzere çeşitli faktörlerden etkilenir (Al-Abdullatif ve Alsubaie, 2024). Yapay zekanın eğitim ortamlarına entegrasyonu hem fırsatlar hem de zorluklar sunar ve pedagojik olarak güçlü ve etik açıdan sağlam yapay zekâ entegreli öğrenme ortamlarını teşvik etmek için kullanıcı deneyimlerinin eleştirel bir şekilde incelenmesini gerektirir (Mogavi vd., 2023; Qawqzeh, 2024). Yapay zekâ ile kullanıcı etkileşimleri, özellikle yapay zekâ tarafından üretilen içeriklerin sıklıkla görüldüğü sosyal medya bağlamlarında giderek daha yaygın hale geliyor (Chein vd., 2024). Yapay zekanın eğitimi devrim niteliğinde değiştirme potansiyeli önemlidir; kişiselleştirilmiş öğrenme deneyimleri sunar ve idari görevleri otomatikleştirir (Adıgüzel vd., 2023). Yapay zekâ sistemleri daha karmaşık hale geldikçe, sanal asistanlar olarak hizmet verebilir, öğrencilere anında geri bildirim ve destek sağlayabilir ve eğitimcilerin daha karmaşık öğretim etkinliklerine odaklanmasını sağlayabilir (Sharples, 2023). Üniversitelerde yapay zekâ sohbet robotlarının konuşlandırılması, öğrenci katılımını artırmanın ve öğrenme başarılarını geliştirmenin olası bir yolu olarak önemli ilgi görmektedir (Fırat, 2023). Öğrencilerin yapay zekâ ile etkileşim kurma biçimi, özellikle işbirlikçi öğrenme senaryolarında, kritik öneme sahiptir.

1.3.ChatGPT

ChatGPT'nin Tanımı: ChatGPT, OpenAI tarafından geliştirilen güçlü bir GPT modeline dayanan ve özellikle optimize edilmiş konuşma yeteneklerine sahip bir sohbet modelidir (Zhao vd, 2023). OpenAI tarafından geliştirilen üretken bir yapay zekâ dil modelidir. Transformer mimarisini kullanarak tutarlı ve bağlamsal olarak uygun, insan benzeri yanıtlar üreten bir OpenAI üretken yapay zekâ dil modelidir. Gelişmiş bir doğal dil işleme modelidir (Markov vd, 2023).

ChatGPT ve GPT modellerinin temel prensibi, dünya bilgisini dil modellemesi yoluyla yalnızca-kod çözücülü (decoder-only) Transformer modeline sıkıştırmaktır. Bu sayede dünya bilgisinin semantiğini geri kazanabilir (veya ezberleyebilir) ve genel amaçlı bir görev çözücü olarak işlev görebilir. Başarısının iki temel noktası şunlardır: (I) Bir sonraki kelimeyi doğru bir şekilde tahmin edebilen yalnızca-kod çözücülü Transformer dil modellerinin eğitilmesi ve (II) dil modellerinin boyutunun ölçeklendirilmesi. ChatGPT, InstructGPT'ye benzer bir şekilde eğitilmiştir. Ancak ChatGPT'nin eğitiminde, insan tarafından oluşturulan konuşmalar (kullanıcı ve AI rollerini oynayan) InstructGPT veri setiyle diyalog formatında birleştirilmiştir (Zhao vd, 2023). GPT modelleri, geniş miktarda metin verisi üzerinde derin öğrenme tekniği olan transformer'ları kullanarak eğitilmiştir. ChatGPT'nin farklı versiyonları bulunmaktadır. Kasım 2022'de genel erişime açılan ChatGPT-3.5, geniş kullanıcı kitlesiyle buluşarak dil modelleme alanında önemli bir adım olmuştur. Bunu takiben, Mart 2023'te sunulan GPT-4 modeli; artırılmış doğruluk, daha güvenli yanıtlar ve kullanıcı niyetine daha duyarlı tepkiler verme kabiliyetiyle öne çıkmıştır. Özellikle akademik yazım, proje üretimi ve içerik geliştirme gibi alanlarda tercih edilen bu modellerin eğitim verilerindeki zamansal sınırlama ise ChatGPT-3.5 için Eylül 2021 itibarıyla belirlenmiştir. Bu tarih, modelin bilgi tabanının güncellik sınırını oluşturmaktadır, bir araştırma sürecinde bu sınır göz önünde bulundurulmalıdır. Bu, ChatGPT'nin bu tarihten sonraki olayları, keşifleri veya görüşleri entegre edemeyeceği anlamına gelir. ChatGPT4o'nun en son veriye erişim tarihi konusunda çelişkili bilgiler bulunmaktadır, bazı kaynaklar Ekim 2023 derken, ChatGPT4o'nun kendisi Eylül 2021'i belirtmektedir (Spennemann, 2025).

1.3.1.ChatGPT'nin Temel Özellikleri:

İnsanlarla mükemmel düzeyde iletişim kurma kapasitesine sahip olup, geniş bir bilgi birikimiyle donatılmıştır. Matematiksel problemler üzerinde akıl yürütebilir, çok turlu diyaloglarda bağlamı doğru bir şekilde takip edebilir ve insan değerleriyle uyum sağlayarak güvenli kullanım sunar. Ayrıca eklenti mekanizmasını desteklemesi, mevcut araçlar ve uygulamalarla entegrasyonunu mümkün kılarak kapasitesini daha da genişletir. Doğal dil konuşmaları üretme konusunda güçlü bir temele sahip olan sistem, doğal sesli yanıtlar oluşturabilir. Dil çevirisi, görüntü tanıma ve tahmine dayalı modelleme gibi çeşitli yapay zekâ uygulamalarında etkin biçimde kullanılabilir. Bununla birlikte, metin kaynaklarından veri çıkarma ve özetleme, insan benzeri bağlamsal yanıtlar sunma, metin sınıflandırma ve duygu analizi gibi doğal dil görevlerini başarıyla yerine getirme yeteneğine sahiptir. E-posta ya da programlama kodu gibi temel bağlamsal metinlerin taslağını hazırlayabilmesi de önemli bir avantajdır. Tüm bu özelliklerinin ötesinde, GPT-4 ile birlikte daha yüksek olgusal doğruluk ve kullanıcı niyetine daha isabetli yanıtlar verme gibi gelişmiş yetenekler de sunmaktadır (OpenAI, 2023).

Yapay zekâ teknolojileri sağlık alanında biyolojik verilerden bilgi çıkarma, tıbbi tavsiye danışmanlığı, ruh sağlığı analizi ve raporların sadeleştirilmesi gibi görevleri yerine getirebilmekte; hatta USMLE sınavlarında uzman düzeyinde başarı gösteren Med-PaLM modelleri gibi özel sistemler geliştirilmiştir (Mesko, 2023). Eğitimde ise öğretim materyali üretme, sunum hazırlama, araştırma fikirleri ve metodolojileri önerme, istatistiksel analiz ve veri yorumlama, yazma becerilerini geliştirme ve akademik yazım desteği sağlama gibi çok yönlü katkılar sunmaktadır (Zheldibayeva, 2024). İş dünyasında operasyon ve yönetim süreçlerinden tedarik zinciri yönetimine, iş analitiğinden pazarlama ve e-ticaret uygulamalarına kadar geniş bir yelpazede kullanılmaktadır. Bilimsel araştırmalarda ise araştırma fikri geliştirme, yöntem belirleme, literatür taraması yapma ve hipotez oluşturma gibi çalışmalarda etkin rol oynamaktadır (Acemoglu vd., 2023). Hukuk alanında dava taslağı hazırlama, yasal araştırmaları hızlandırma, sözleşme analizi ve belge inceleme süreçlerinde kullanılabilirken; finans alanında da çeşitli uygulamalara entegre edilebilmektedir. Sanat ve kültür alanında müzik, sanat ve tasarım, yaratıcı yazarlık, oyun geliştirme ve sanal gerçeklik gibi yaratıcı süreçlere katkı sağlamaktadır. Programlama konusunda ise kod üretimi, hata ayıklama, akademik projelerde destek ve mülakatlara hazırlık gibi işlevleriyle yaygın şekilde kullanılmaktadır. Müşteri hizmetlerinde geçmiş konuşmaları da dikkate alarak müşteri etkileşimlerini otomatikleştirme, soruları yanıtlama ve destek sağlama görevlerini üstlenirken; veri bilimi alanında kişisel bir veri bilimcisi gibi işlev görebilmektedir. Ayrıca akademik yazımda makale üretimine katkı sağlayabilmekte ve hakemlik süreçlerinde makale değerlendirme çalışmalarında da kullanılabilmektedir (Liu vd., 2024).

1.3.2.ChatGPT ile İlişkili Etik Endişeler:

ChatGPT'nin kullanımında bazı önemli sınırlılıklar ve riskler bulunmaktadır. Bunların başında, zaman zaman gerçek dışı veya mantıksız çıktılar üretmesiyle ortaya çıkan "hallucination" sorunu gelmektedir; özellikle akademik referanslar gibi bilgi temelli içeriklerde, uydurma veriler sunabilme eğilimi dikkat çekicidir (Alkaissi & McFarlane, 2023). Ayrıca, eğitim verilerindeki önyargıları yansıtma riski taşıyan sistem, din, kültür, cinsiyet, ırk ve sosyoekonomik durum gibi çeşitli konularda istemeden taraflı ifadeler üretebilir; bazı analizler, ChatGPT'nin siyasi yönelim testlerinde özgürlükçü, ilerici ve sol eğilimli görüşlere yatkınlık sergilediğini göstermektedir (Rozado, 2023). Fikri mülkiyet ve telif hakları da bir başka önemli kaygı alanıdır; zira üretken yapay zekâ modelleri, telif hakkına tabi içeriklerden öğrenmekte ve bu da potansiyel olarak hukuki sorumluluk doğurabilecek

durumlar yaratmaktadır (Zirpoli, 2023). Kötüye kullanım potansiyeli de göz ardı edilemez; bu teknolojiyle kötü amaçlı yazılım üretimi, toksik dil kullanımı, yanlış bilgi yayma ve akademik etik ihlalleri gibi olumsuz senaryolar mümkün hale gelebilmektedir (Weidinger vd., 2021). Sistemlerin arkasındaki algoritmik yapıların karmaşıklığı nedeniyle şeffaflık sınırlıdır; bu da tahminlerin ya da önerilerin nasıl üretildiğini anlamayı güçleştirmektedir. Dahası, yapay zekâ tarafından oluşturulan içerikler konusunda sorumluluğun kime ait olduğu hâlâ net değildir; bu nedenle ChatGPT'nin bilimsel makalelerde yazar olarak kabul edilmemesi gerektiği belirtilmektedir. Son olarak, bu tür sistemlere aşırı güven duymak, kullanıcıların eleştirel düşünme ve doğrulama becerilerini zayıflatma riskini beraberinde getirmektedir (Stokel-Walker, 2023).

ChatGPT ve benzeri yapay zekâ tabanlı sohbet sistemleri, veri gizliliği ve güvenliği açısından çeşitli riskler barındırmaktadır. Kullanıcıların hassas veya kişisel bilgilerini ticari sunuculara yüklemesi, özellikle sağlık verileri gibi yüksek derecede özel içeriklerde ciddi gizlilik ihlali potansiyeli taşımaktadır (Borgesius, 2018). Bu sistemlerin işleyebilmesi için büyük miktarda veriye ihtiyaç duyulması ise, onları siber saldırılara karşı daha savunmasız hale getirmekte ve veri güvenliği risklerini artırmaktadır (ICO, 2023). Ayrıca, ChatGPT gibi modellerin kişisel verileri işlemesi, özellikle eğitimlerinde doğruluğu tartışmalı ya da eksik veri setleri kullanılmışsa, hatalı veya yanıltıcı sonuçlar üretme olasılığını beraberinde getirmektedir (Weidinger vd., 2021). Avrupa Birliği Genel Veri Koruma Tüzüğü - General Data Protection Regulation (GDPR) ve Türkiye'deki Kişisel Verilerin Korunması Kanunu (KVKK) gibi düzenlemeler bağlamında bu tür sistemlerin veri işleme uygulamalarının ne ölçüde uyumlu olduğu tartışmalı olup, İtalya'nın ChatGPT'yi gizlilik gerekçesiyle geçici olarak yasaklaması bu alandaki kaygıların altını çizmektedir (Reuters, 2024). Ayrıca, kullanıcıların sohbet geçmişi ve girdilerinin nasıl saklandığı, kimlerle ve ne amaçla paylaşıldığı konusundaki şeffaflık eksikliği önemli bir endişe kaynağıdır; nitekim Google'ın Gemini modeli için yayımladığı gizlilik bildiriminde, kullanıcıların kişisel bilgilerini yapay zekâ ile paylaşmamaları yönündeki uyarı, bu endişeleri açıkça yansıtmaktadır (Search Engine Journal, 2024).

1.4.DeepSeek

DeepSeek, mimari olarak OpenAI'nin GPT serisi ya da Meta'nın LLaMA modellerine benzer şekilde transformatör tabanlı büyük dil modelleri sınıfına ait olsa da, özellikle kod üretimi ve mantıksal çıkarım gerektiren görevlerde optimize edilmiş yapısıyla öne çıkmaktadır (DeepSeek-AI, 2024). GPT-4 gibi modeller, geniş kapsamlı doğal dil anlama ve üretimi süreçlerinde genelci bir paradigma benimseyerek çok çeşitli görevlerde etkin performans sergilemeyi amaçlamaktadır. Buna karşın, DeepSeek daha dar kapsamlı ve belirli görev odaklı senaryolarda, doğruluk ve bağlamsal bütünlük açısından daha yüksek başarı elde etmeye yönelik, özelleştirilmiş ve hedef odaklı bir mühendislik yaklaşımını benimsemektedir. Kod üretiminde, özellikle Python, Java ve C++ gibi dillerde yüksek başarı gösteren DeepSeek, CodeLlama ya da DeepMind'ın AlphaCode modeliyle aynı ligde değerlendirilmekte, ancak daha istikrarlı bağlam takibi ve hata ayıklama hassasiyetiyle öne çıkmaktadır (Guo vd., 2024). Modelin açık ağırlıklı olması, araştırma ve geliştirici topluluğu için erişim imkânı sunsa da, tam anlamıyla açık kaynak olmaması nedeniyle yeniden eğitim, veri incelemesi veya modifikasyon gibi işlemler sınırlı düzeyde gerçekleştirilebilmektedir. Performans açısından bakıldığında, DeepSeek'in tümdengelimli akıl yürütme, çok adımlı mantıksal çözümleme ve programatik problem çözme gibi görevlerde GPT-3.5'i geçtiği; ancak GPT-4 ya da Claude 3 Opus gibi modellerle karşılaştırıldığında bağlam genişliği ve genel dil üretim kalitesinde hâlen sınırlı kaldığı görülmektedir. Bununla birlikte, kaynak verimliliği, inference (çıktı üretme) hızı ve geliştirici dostu

arayüzü sayesinde dar amaçlı yüksek doğruluk gerektiren görevlerde tercih edilebilir bir çözüm sunmaktadır (Manik, 2025).

DeepSeek ekosistemi, farklı görev ve kullanım senaryolarına uyarlanmış çeşitli modeller içermektedir. Örneğin, DeepSeek-R1 modeli, özellikle matematiksel akıl yürütme ve programlama gibi bilişsel yoğunluk gerektiren görevler doğrultusunda optimize edilmiş olmakla birlikte, aynı zamanda genel amaçlı bir dil modeli olarak da işlev görmektedir (Guo vd., 2024). Buna karşılık, DeepSeek-V3 modeli, çoklu disiplinleri kapsayan geniş ölçekli bir veri kümesi üzerinde eğitilmiş olup, genel amaçlı kullanım senaryolarında yüksek performans sergileyen kapsamlı bir yapay zeka çözümü olarak öne çıkmaktadır. Kodlama alanına odaklanan DeepSeek Coder V2, 338 dili destekleyen ve gelişmiş kodlama yeteneklerine sahip bir kod dil modeli olup, açık kaynaklı bir Karışık Uzmanlar (Mixture-of-Experts, MoE) mimarisine sahiptir. Bunun yanında, DeepSeek VL, gerçek dünya görme ve dil görevlerinde başarılı olabilmek üzere tasarlanmış açık kaynaklı bir Görsel-Dil (Vision-Language, VL) modelidir (DeepSeek-AI, 2024).

Ayrıca, DeepSeek LLM modelleri de dikkat çekmektedir. DeepSeek LLM, 67 milyar parametreye sahip olup tescilli ağırlıklarla eğitilmiş ve 2023'e kadar olan Kitaplar+Wiki verileri üzerinde geliştirilmiştir. Daha büyük ölçekteki DeepSeek LLM V2 modeli ise 236 milyar parametreye ve tescilli bir Karışık Uzmanlar (MoE) mimarisine sahiptir (Guo vd., 2024). Son olarak, DeepSeek-V3 modeli 671 milyar parametre kapasitesine ulaşmakta ve FP8 eğitimi gibi yüksek verimli öğrenme tekniklerini kullanan tescilli bir Karışık Uzmanlar mimarisiyle tasarlanmıştır. Bu çeşitlilik, DeepSeek ekosisteminin hem genel amaçlı hem de uzmanlaşmış görevlerde güçlü performans sunmasına olanak tanımaktadır (DeepSeek-AI, 2024).

1.4.1.DeepSeek'in Temel Özellikleri

MoE (Mixture-of-Experts) mimarisi, büyük dil modellerinin hesaplama verimliliğini artırmak amacıyla geliştirilen ve yalnızca belirli parametre alt kümelerini etkinleştirerek çalışan dinamik bir hesaplama yöntemi olarak tanımlanmaktadır. Bu mimari, her bir katmanda çok sayıda "uzman" (expert) bileşen barındırmakta olup, model her bir girdi için yalnızca sınırlı sayıda uzmanın çıktısını hesaba katarak işlemi gerçekleştirmektedir. Bu sayede, hem hesaplama maliyetleri azaltılmakta hem de modelin ölçeklenebilirliği önemli ölçüde artırılmaktadır (Du vd., 2021). Bu yaklaşımın başlıca avantajlarından biri kaynak verimliliğidir: yüz milyarlarca parametreli modellerin eğitimi ve çalıştırılması oldukça maliyetli iken, MoE yalnızca gerekli uzmanları aktif ederek hesaplama yükünü önemli ölçüde azaltmaktadır. Ayrıca, ölçeklenebilirlik açısından parametre sayısı artırılarak daha yüksek kapasiteli modeller tasarlanabilirken, çıkarım (inference) sırasında maliyet sabit tutulabilmektedir (Artetxe vd., 2021).

MoE mimarisinin bir diğer avantajı da uyarlanabilirlik özelliğidir; çünkü her uzman, farklı görevler veya bağlamlar için eğitilebilmekte ve bu sayede model çoklu görevlerde daha esnek bir şekilde çalışabilmektedir. Örneğin, DeepSeek-R1 modeli, MoE mimarisi temel alınarak geliştirilmiş olup, toplamda 671 milyar parametre içermesine rağmen, çıkarım aşamasında yalnızca 37 milyar parametreyi etkinleştirmektedir. Bu yaklaşım, modelin yüksek parametre kapasitesine sahip olmasına rağmen hesaplama verimliliğinin korunmasını sağlamaktadır. Bu durum hem yüksek kapasiteyi hem de maliyet avantajını birlikte sunmakta olup, muhtemelen 16 veya 32 uzmandan oluşan bir top-k yönlendirme mekanizması (örneğin k=2) kullanılarak gerçekleştirilmektedir (DeepSeek R1, t.y.).

Pekiştirmeli öğrenme (Reinforcement Learning), modelin çıktılarını optimize etmek amacıyla bir "ödül sinyali" kullanılarak gerçekleştirilen bir öğrenme yöntemidir (Du vd., 2021). Büyük dil modellerinde (LLM'lerde), özellikle insan geri bildirimiyle pekiştirmeli öğrenme (Reinforcement Learning from Human Feedback, RLHF) formu öne çıkmaktadır. Bu yaklaşım, muhakeme ve karar verme becerilerinin geliştirilmesi, daha insani, bağlamsal ve faydalı yanıtlar üretilmesi ve zararlı ya da ilgisiz içeriklerin azaltılması gibi önemli kullanım alanlarıyla dikkat çekmektedir (Ouyang vd., 2022). DeepSeek-R1 modeli, yalnızca gözetimli öğrenme yöntemleriyle değil, aynı zamanda pekiştirmeli öğrenme süreçleriyle de geliştirilmiştir. Böylece, model yalnızca verileri taklit etmekle kalmayıp, amaca yönelik mantıksal çıkarımlarda bulunmayı da öğrenebilmektedir (Gupta, 2025). Örneğin, uzun adımlı çıkarım süreçlerinde ortaya çıkan hataları analiz ederek kendi yanıtlarını düzeltebilme kapasitesine sahiptir. Bu sürecin teknik olarak uygulanmasında sıklıkla gözetimli ince ayar (supervised fine-tuning), ödül modeli eğitimi (reward model training) ve Proximal Policy Optimization (PPO) gibi algoritmalarla gerçekleştirilen RLHF aşamaları yer almaktadır (Neptune.ai, 2025).

DeepSeek-V3 modeli, büyük olasılıkla GPT mimarisi temel alınarak geliştirilmiş olup, rotary embedding, katman normalizasyonu ve dikkat mekanizması optimizasyonları gibi güncel teknikleri bünyesinde barındırmaktadır (DeepSeek-V3, t.y.). Bu temel yapının üzerine inşa edilen DeepSeek-R1 ise, MoE (Mixture-of-Experts) mimarisi ile entegre edilmiş ve pekiştirmeli öğrenme (RL) yöntemleriyle desteklenmiş, performans ve verimlilik açısından iyileştirilmiş özel bir varyant olarak dikkat çekmektedir (Gupta, 2025). Böylece, hem model kapasitesi artırılmış hem de öğrenme süreçlerinde adaptasyon yeteneği güçlendirilmiştir. DeepSeek-R1 modeli, kod anlama, tümdengelimli muhakeme ve yüksek doğruluk gerektiren görevlerde geliştirilmiş olup, çok büyük bir parametre havuzuna sahip olmasına rağmen çıkarım sırasında yalnızca küçük bir kısmını kullanarak verimliliği artırmaktadır (Du vd., 2021). Ayrıca, DeepSeek-Coder, kod üretimi, analiz ve hata ayıklama gibi görevler için özelleştirilmiş bir alt model olarak konumlanmakta ve OpenAI Codex, CodeLlama veya AlphaCode gibi modellerle rekabet edebilecek düzeyde performans sunmayı hedeflemektedir (DeepSeek-Coder, t.y.).

DeepSeek-R1'in eğitimi, sentetik ve gerçek dünya verilerinden derlenen toplamda 9.8 trilyonluk geniş kapsamlı ve karma bir veri seti kullanılarak gerçekleştirilmiştir (DeepSeek-R1, t.y.). Söz konusu veri seti tescilli ve gizli bilgi içermesi sebebiyle, yalnızca yetkilendirilmiş kişiler tarafından erişilebilmekte ve sıkı güvenlik protokolleriyle korunmaktadır. Bu yaklaşım, modelin hem çeşitlilik hem de kalite açısından zengin bir eğitim materyaliyle desteklenmesini sağlamaktadır. Eğitim sürecini daha verimli hâle getirmek amacıyla, DeepSeek Gelişmiş Yeniden Parametrelendirme Optimizasyonu (GRPO) yöntemini kullanarak Pekiştirmeli Öğrenme (RL) için gerekli olan denetimli ince ayar (SFT) gibi ön adımları ortadan kaldırmış ve böylece maliyetleri önemli ölçüde düşürmüştür (Gupta, 2025). Şirketin başarısının temel faktörlerinden biri, titiz veri etiketleme uygulamalarıdır; liderlik kadrosu bu etiketleme süreçlerine doğrudan katılarak kaliteyi sürekli olarak denetlemektedir. Öte yandan, DeepSeek-R1 modeli, MIT Lisansı kapsamında açık ağırlık (open-weights) politikasıyla yayınlanmış olup, bu sayede ticari ve akademik araştırma amaçlı kullanımlar için nöral ağ parametrelerine sınırsız erişim imkânı sağlamaktadır (DeepSeek-R1, t.y.). Bu yaklaşım, geniş kullanıcı kitlesinin model üzerinde özgürce çalışabilmesine ve yenilikçi uygulamalar geliştirebilmesine olanak tanımaktadır. Ayrıca, DeepSeek'in bazı modelleri (özellikle V2 ve V3 sürümleri), FP8 eğitimi gibi yüksek verimli eğitim tekniklerini kullanarak daha hızlı ve verimli bir öğrenme süreci sağlamaktadır (DeepSeek-V3, t.y.).

DeepSeek-R1, zorlu problem çözme görevlerinde dikkat çekici muhakeme yetenekleri sergileyerek güçlü bir performans göstermektedir. Özellikle DeepSeek-Coder-V2, kod üretimi ve kod analizi alanlarında yüksek doğruluk oranları sergilemekte olup, bu başarısını GPT-4 Turbo ile karşılaştırılabilir düzeyde gerçekleştirmektedir (BytePlus, 2025). Model, karmaşık programlama görevlerinde etkin performans göstererek yazılım geliştirme süreçlerinde önemli bir araç olarak öne çıkmaktadır. 338 programlama dilini desteklemesi ve 128K'ya kadar uzayan bağlam uzunluğu, modelin esneklik ve derinlik açısından güçlü bir kapasite sunduğunu göstermektedir (Dataloop, t.y.). Matematiksel muhakeme alanında, DeepSeek-R1'in MATH veri seti üzerinde gerçekleştirdiği 30 zorlu matematik problemi, ChatGPT ve Gemini'yi geride bırakarak üstün bir performans ortaya koymuştur (Shirani, 2025). Ayrıca, bilimsel hesaplama görevlerinde de yüksek başarı elde etmiştir, özellikle nümerik integrasyon gibi alanlarda başarılı sonuçlar elde etmiştir. Bağlamsal anlama ve mantıksal çıkarım yetenekleri açısından ileri düzeyde olan DeepSeek, bazı modellerinde 128K token'a kadar uzayabilen uzun bağlam pencereleri sayesinde karmaşık ve kapsamlı metinleri etkin şekilde işleyebilmektedir (Dataloop, t.y.). Ayrıca, eğitim sürecinde sağladığı maliyet etkinliği de önemli bir rekabet avantajı sunmaktadır; DeepSeek-R1 modeli, yalnızca 5.5 milyon dolarlık bir bütçe ile geliştirilmiş olup, performans açısından birçok muadilini geride bırakmayı başarmıştır (IBM, 2025a). Karışık Uzmanlar (MoE) mimarisi sayesinde, yoğun eşdeğerlerine kıyasla enerji tüketimini %58 oranında azaltarak enerji verimliliği konusunda da önemli bir adım atmıştır (IBM, 2025b). DeepSeek V3'ün teknik özellikleri arasında, 7/8 aktivasyon parametresi ve toplamda 67 milyar parametre ile oldukça güçlü bir model yapısına sahiptir. Son olarak, DeepSeek modelleri, genellikle yapılandırılmış ve tutarlı referanslar sunabilme kapasitesiyle öne çıkmakta; bu özellik, modelin bilgi doğruluğunu artırmakta ve elde edilen çıktılara yönelik güvenilirliği önemli ölçüde yükseltmektedir (DeepSeek-V3, t.y.).

1.4.2.DeepSeek ile İlişkili Etik ve Veri Koruma Endişeleri

DeepSeek, ağırlıklarını ve bazı mimari detaylarını açıkça paylaşmasına rağmen (açık ağırlık), eğitim veri seti kompozisyonu ve Reinforcement Learning from Human Feedback (RLHF) gibi kritik eğitim detaylarını tescilli tutmaktadır (Sapkota vd., 2025). Bu durum, modelin potansiyel önyargılarını ve etik kaynak kullanımını incelemeyi zorlaştırmakta, aynı zamanda modelin tam anlamıyla yeniden üretilebilirliğini sınırlamaktadır. Açık kaynak belirsizliği ise, DeepSeek-R1'in MIT lisansı altında sunulmasına rağmen, eğitim verileri ve metodolojilerinin tam olarak paylaşılmaması nedeniyle, aslında tam anlamıyla "açık kaynak" tanımına uymadığını göstermektedir. Bu durum, "açık ağırlık" ve "açık kaynak" kavramları arasındaki karışıklığı ortaya koymaktadır (CTOL Digital, 2025). Önyargılar açısından, eğitim verilerinin tescilli olması, DeepSeek modellerinde olası önyargı kaynaklarını belirlemeyi ve bu önyargıları azaltmayı zorlaştırmaktadır (Witness AI, 2025). Veri gizliliği konusunda ise, DeepSeek'in veri işleme uygulamaları ve kullanıcı verilerinin nasıl saklandığına dair yeterli bilgi bulunmamaktadır; bu durum, kullanıcıların hassas verilerini bu tür platformlara yüklerken gizlilik endişeleri yaratmaktadır (Security Magazine, 2025). Bunun yanı sıra, diğer dil modellerinde olduğu gibi, DeepSeek'te de zaman zaman yanlış veya uydurma bilgi üretme (hallucination) riski mevcuttur; model, bazı durumlarda gerçeklikten sapmış veya mantıksal tutarsızlık içeren çıktılar üretebilmekte ve referans gösterme süreçlerinde hatalar yapabilmektedir (Vectara, 2025). Kötüye kullanım potansiyeli ise, DeepSeek'in kod üretme yeteneği sayesinde kötü amaçlı yazılım geliştirme gibi etik dışı kullanım endişelerini gündeme getirmektedir (SecurityWeek, 2025). Özetlemek gerekirse, DeepSeek; gelişmiş mantıksal muhakeme, matematiksel işlem ve kodlama becerileriyle dikkat çeken, potansiyeli yüksek bir dil modeli olarak öne çıkmaktadır. Açık ağırlık stratejisi

sayesinde geniş kullanıcı kitlesine erişim ve kapsamlı ince ayar imkânları sağlarken, eğitim süreçlerinin tam anlamıyla şeffaf olmaması bazı etik kaygılar ve yeniden üretilebilirlik sorunlarını gündeme getirmektedir. Kullanıcıların DeepSeek'i kullanırken bu potansiyel risklerin farkında olmaları önemlidir.

Bu çalışmanın temel amacı, üretken yapay zekâ modelleri olan ChatGPT ve DeepSeek'in kullanıcı etkileşimleri üzerindeki performansını ve etkisini karşılaştırmalı bir analizle değerlendirmektir. Kaggle.com'dan elde edilen ve 10.000'den fazla sentetik kullanıcı etkileşim verisini içeren bir veri seti kullanılarak, bu iki platformun yanıt doğruluğu, oturum süresi, kullanıcı memnuniyeti, farklı görevlerdeki (örneğin içerik üretimi, teknik görevler) performansları ve kullanıcı bağlılığı gibi çeşitli metrikler açısından farklılıkları incelenmiştir. Makine öğrenmesi algoritmaları ve istatistiksel testler aracılığıyla, platformların güçlü yönleri, sınırlılıkları ve kullanım bağlamına göre performans farklılıkları ortaya konulmaya çalışılmış, aynı zamanda veri gizliliği ve etik konuların önemi vurgulanmıştır.

Bu araştırma sonucunda aşağıdaki sorulara cevap verilmesi hedeflenmektedir:

1. Kullanıcı memnuniyeti ve deneyim puanları açısından ChatGPT ve DeepSeek platformları arasında istatistiksel olarak anlamlı bir fark var mıdır?
2. Cihaz türü (mobil, masaüstü, tablet, akıllı hoparlör) platformlar kullanılırken nasıl farklılaşmaktadır?
3. Makine öğrenmesi ile geliştirilen regresyon ve sınıflandırma modelleri kullanıcı memnuniyetini ve bağlılık davranışlarını ne ölçüde doğru tahmin edebilmektedir?
4. Üretken yapay zekâ platformlarının kullanıcı sadakati ve terk oranları üzerindeki etkileri nelerdir ve bu etkiler hangi kullanıcı segmentlerinde daha belirgindir?
5. Yanıt doğruluğu ile kullanıcı tarafından bildirilen deneyim ve memnuniyet skorları arasında nedensel bir ilişki kurulabilir mi?
6. Üretken yapay zekâ platformlarının kullanıcı etkileşimi performansları zamansal olarak nasıl değişim göstermektedir? Mevsimsellik veya eğilim (trend) etkisi var mıdır?
7. Farklı coğrafi bölgelerdeki kullanıcıların ChatGPT ve DeepSeek'e yönelik memnuniyet düzeyleri arasında anlamlı farklar bulunmakta mıdır?
8. Üretken yapay zekâ platformlarında (ChatGPT ve DeepSeek) kullanıcıların tercih ettiği diller ile platform seçimi arasında anlamlı bir ilişki var mıdır ve bu ilişki kullanıcı yoğunluğunun dağılımı açısından ne tür farklılıklar ortaya koymaktadır?

2.Yöntem

Bu çalışma, karşılaştırmalı bir analiz yöntemiyle, ChatGPT ve DeepSeek'in kullanıcı etkileşimleri üzerindeki performansını değerlendirmeyi amaçlayan nicel bir araştırmadır. Araştırmada, Kaggle.com'dan temin edilen bir veri seti kullanılacak ve bu veriler üzerinde makine öğrenmesi algoritmaları, özellikle regresyon, karar ağaçları ve kümeleme teknikleriyle analizler yapılacaktır. Katılımcılar, farklı coğrafi bölgelerden, çeşitli cihaz türleri (mobil, masaüstü, tablet, akıllı hoparlör) aracılığıyla ChatGPT ve DeepSeek platformlarıyla etkileşime gireceklerdir. Kullanıcı etkileşimleri, yanıt doğruluğu, oturum süresi, kullanıcı memnuniyeti (1-5 puan aralığında), sorgu türü ve cihaz türü gibi değişkenler üzerinden ölçülüp kaydedilecektir. Verilerin görselleştirilmesi, Google Looker Studio aracılığıyla yapılacak, böylece her iki yapay zeka platformunun performans karşılaştırılması görsel olarak

sunulacaktır. Elde edilen veriler, deskriptif istatistikler ile analiz edilecek ve platformlar arasındaki performans farkları, t-test veya ANOVA gibi karşılaştırmalı istatistiksel testler kullanılarak değerlendirilecektir. Çalışma, her iki platformun kullanıcı etkileşimlerine olan etkilerini anlamak için kapsamlı bir karşılaştırmalı değerlendirme sunacak ve araştırma, nicel analiz yöntemleriyle veri toplama, analiz etme ve görselleştirme aşamalarını içeren bir deneysel araştırma türüne girmektedir.

2.1.Veri Toplama Araçları

Bu çalışmada kullanılan veri seti, çevrimiçi veri platformu Kaggle.com üzerinden erişilen "DeepSeek vs ChatGPT: AI Platform Comparison" başlıklı kaynaktan temin edilmiştir.

Söz konusu veri seti, önde gelen yapay zeka modelleri olan ChatGPT (GPT-4-turbo) ve DeepSeek (DeepSeek-Chat 1.5) arasındaki performansı ve kullanıcı davranışlarını karşılaştırmak amacıyla Python programlama dili (özellikle "`faker`" kütüphanesi kullanılarak) ve SQL aracılığıyla yapay olarak oluşturulmuştur. Gerçekçi bir yapay zeka etkileşimi temsili sunmayı hedefleyen veri seti, Temmuz 2023 ile Şubat 2025 arasındaki yaklaşık 1.5 yıllık bir dönemi kapsamakta olup, 10.000'den fazla satır içermektedir.

Veri seti, analiz kapsamında kritik öneme sahip çeşitli kullanıcı etkileşim metriklerini, platform performans göstergelerini ve yapay zeka yanıt karakteristiklerini barındırmaktadır. Veri setinin temel özellikleri arasında şunlar yer almaktadır:

- **Zaman Serisi Uygunluğu:** Eğilim (trend) ve mevsimsellik analizleri yapılmasına olanak tanıyan ayrıntılı tarih ve saat sütunları içermektedir.
- **Karşılaştırmalı YZ Analizi:** Kullanıcı etkileşimi, platforma elde tutma oranları ve yanıt kalitesi gibi kıyaslama metriklerini karşılaştırma imkanı sunmaktadır.
- **Kullanıcı Davranışı İçgörüleri:** Oturum süreleri, kullanıcı girdilerinin metinsel karmaşıklığı ve kullanıcılar tarafından verilen derecelendirmeler gibi kullanıcı davranışlarına dair bilgiler sağlamaktadır.
- **Teknik Performans Metrikleri:** Yapay zeka yanıtlarının doğruluğu ve işlem/yanıt hızı gibi teknik performans göstergelerini içermektedir.
- **Veri Ön İşleme İmkanı:** Veri temizleme ve ön işleme adımları için pratik sunmak amacıyla bilerek eklenmiş eksik (null) değerler de mevcuttur.

Bu veri seti, yapay zeka modellerinin kıyaslanması, platform performansının değerlendirilmesi, zaman serisi analizleri ve kullanıcı davranışı araştırmaları gibi konular için uygun bir yapıya sahiptir. Veri seti, araştırma, proje ve analiz amaçlı kullanımlar için MIT Lisansı ile lisanslanmıştır. Çalışmamızın kapsamına uygunluğu ve geniş veri içeriği nedeniyle bu veri seti tercih edilmiştir.

2.2.Veri Analizi

Bu çalışmada, "Üretken Yapay Zeka Performans Kıyaslaması: ChatGPT ve DeepSeek'in Kullanıcı Etkileşimleri Üzerindeki Etkisi" başlığı altında, iki farklı yapay zeka platformunun (ChatGPT ve DeepSeek) kullanıcı etkileşim verileri analiz edilmiştir. Araştırmanın bu aşamasındaki temel amaç, platformlar arası performans farklılıklarını ve kullanıcı etkileşim örüntülerini ortaya koymak için veriyi hazırlamak ve temel metrikler üzerinden tanımlayıcı analizler gerçekleştirmektir. Analizde kullanılan veri seti, ChatGPT ve DeepSeek platformlarından elde edilen kullanıcı etkileşimlerini içermektedir. Veri seti; Date, Month_Num, Weekday, AI_Platform, AI_Model_Version, Active_Users, New_Users, Churned_Users, Daily_Churn_Rate, Retention_Rate, User_ID, Query_Type, Input_Text, Input_Text_Length, Response_Tokens, Topic_Category, User_Rating, User_Experience_Score, Session_Duration_sec, Device_Type, Language, Response_Accuracy, Response_Speed_sec, Response_Time_Category, Correction_Needed, User_Return_Frequency, Customer_Support_Interactions ve Region gibi çeşitli değişkenleri kapsamaktadır. Bu değişkenler, kullanıcı demografisi, etkileşim yoğunluğu, platform performansı ve kullanıcı memnuniyeti gibi farklı boyutları temsil etmektedir. Analizlerin güvenilirliğini ve tutarlılığını sağlamak amacıyla ham veri seti üzerinde kapsamlı bir ön işleme süreci uygulanmıştır. Bu süreçte veri setindeki eksik değerler tespit edilmiş ve değişkenin tipine ve eksiklik oranına bağlı olarak uygun stratejiler kullanılarak yönetilmiştir. AI_Platform (örneğin, ChatGPT=0, DeepSeek=1 şeklinde), Weekday, Query_Type, Topic_Category, Device_Type, Language, Response_Time_Category ve Region gibi nominal veya ordinal yapıda olan kategorik değişkenler, analizlerde kullanılabilir sayısal formata dönüştürülmüştür. Bu dönüşüm için genellikle "Label Encoding" veya "One-Hot Encoding" teknikleri tercih edilmiştir. Özellikle, iki platformun doğrudan karşılaştırılabilmesi için AI_Platform sütunu sayısal olarak kodlanmıştır. Date sütunu, zaman içindeki eğilimlerin analiz edilebilmesi için datetime objelerine dönüştürülmüş; gerekirse analizlerde kullanılmak üzere Yıl, Ay, Haftanın Günü gibi ek zaman özellikleri türetilmiştir. Session_Duration_sec gibi süre bildiren metrikler, doğrudan sayısal değerler olarak kullanılmıştır. Session_Duration_sec, Input_Text_Length, Response_Tokens gibi sürekli sayısal değişkenlerdeki aykırı değerler, istatistiksel yöntemler (örneğin, IQR – Çeyrekler Açıklığı yöntemi) ve görsel araçlar (örneğin, kutu grafikleri) kullanılarak tespit edilmiştir. Tespit edilen aykırı değerlerin analiz sonuçları üzerindeki potansiyel etkisini azaltmak amacıyla, veri setinin doğasına uygun olarak kırpma (winsorizing) veya logaritmik dönüşüm gibi teknikler uygulanmıştır. Farklı ölçeklerdeki sayısal değişkenlerin analizler sırasında eşit ağırlıkta değerlendirilmesi ve bazı görselleştirme tekniklerinin performansını artırmak amacıyla, MinMaxScaler veya StandardScaler gibi ölçeklendirme yöntemleri kullanılarak özellikler belirli bir aralığa (örneğin, 0-1) veya standart normal dağılıma (ortalama=0, standart sapma=1) dönüştürülmüştür. Veri ön işleme aşamasının tamamlanmasının ardından, ChatGPT ve DeepSeek platformlarının performansını ve kullanıcı etkileşimlerini karşılaştırmak amacıyla tanımlayıcı istatistiksel analizler yapılması planlanmıştır. Bu kapsamda, her iki platform için temel performans göstergeleri (örneğin, ortalama User_Rating, ortalama Session_Duration_sec, Daily_Churn_Rate, Response_Speed_sec vb.) hesaplanmış ve karşılaştırılmıştır. Görselleştirmelerde, zaman serisi grafikleri, çubuk grafikler, pasta grafikler, dağılım grafikleri ve ısı haritaları gibi çeşitli grafik türlerinden faydalanılarak iki platform arasındaki temel farklılıklar, benzerlikler ve zaman içindeki eğilimler vurgulanacaktır.

3. Bulgular

3.1. Kullanıcı Memnuniyeti ve Deneyimi Karşılaştırmasına İlişkin Bulgular

ChatGPT ve DeepSeek platformları arasındaki kullanıcı memnuniyeti ve deneyimi karşılaştırması analizi, kullanıcı derecelendirmeleri ve deneyim puanlarına ilişkin değerli içgörüler sunmuştur. Bulguların ayrıntılı bir dökümü aşağıda sunulmaktadır.

ChatGPT oturumlarının analizinde, en yaygın derecelendirmelerin 3, 4 ve 5 olduğu gözlemlenmiştir. Özellikle 5 puanlık yüksek derecelendirmelerin, "İyi Uygulamalar", "İçerik Üretimi", "Profesyonel Yazarlık" ve "Eğitim" gibi başlıklarla ilişkilendirildiği tespit edilmiştir. Buna karşın, 3 puanlık düşük derecelendirmeler ise genellikle "İçerik Üretimi", "Eğitim" ve "Profesyonel Yazarlık" gibi konularla bağlantılı olmakla birlikte, daha kısa oturum sürelerinde yoğunlaşmaktadır. DeepSeek oturumlarında ise derecelendirmelerin ağırlıklı olarak 4 ile 5 arasında yoğunlaştığı görülmektedir. "Hata Ayıklama", "Kod Optimizasyonu", "Uygulama Yardımı" ve "Algoritma Açıklaması" gibi konularda tutarlı bir şekilde yüksek derecelendirmeler (4-5 puan) elde edilmiştir. Bununla birlikte, deneyim puanlarının 2.0 ve üzeri olduğu durumlarda derecelendirmelerde daha fazla değişkenlik gözlemlenmektedir.

ChatGPT oturumlarına ilişkin deneyim puanlarının genellikle 0.5 ile 2.2 arasında değiştiği görülmektedir. Bu bağlamda, daha yüksek deneyim puanlarının (2.0 ve üzeri) genellikle daha yüksek kullanıcı derecelendirmeleri (4-5 puan) ile pozitif bir ilişki sergilediği dikkat çekmektedir. Öte yandan, yaklaşık 0.5 ile 1.0 arasındaki daha düşük deneyim puanlarının ise sıklıkla 3 puan gibi daha düşük kullanıcı derecelendirmeleriyle örtüştüğü tespit edilmiştir.DeepSeek oturumlarında ise deneyim puanlarının genellikle daha yüksek seviyelerde (2.0 ve üzeri) olduğu gözlemlenmektedir. Bu yüksek deneyim puanları, kullanıcı derecelendirmeleriyle güçlü bir pozitif korelasyon göstermekte, özellikle 2.0'ın üzerindeki puanlarda bu ilişki daha da belirginleşmektedir.

Kullanıcı derecelendirmeleri ile kullanıcı deneyimi puanları arasında dikkate değer bir korelasyon tespit edilmiştir. Özellikle daha yüksek deneyim puanlarının, genellikle daha yüksek kullanıcı derecelendirmeleri ile pozitif bir ilişki içerisinde olduğu gözlemlenmiştir. Bu iki değişken arasındaki ilişkinin gücünü göstermek amacıyla yapılan analizde, Pearson korelasyon katsayısı yaklaşık olarak r≈0.75r \approx 0.75r≈0.75 olarak hesaplanmıştır. Bu sonuç, deneyim puanları ile kullanıcı derecelendirmeleri arasında güçlü bir pozitif korelasyon olduğunu ortaya koymaktadır.

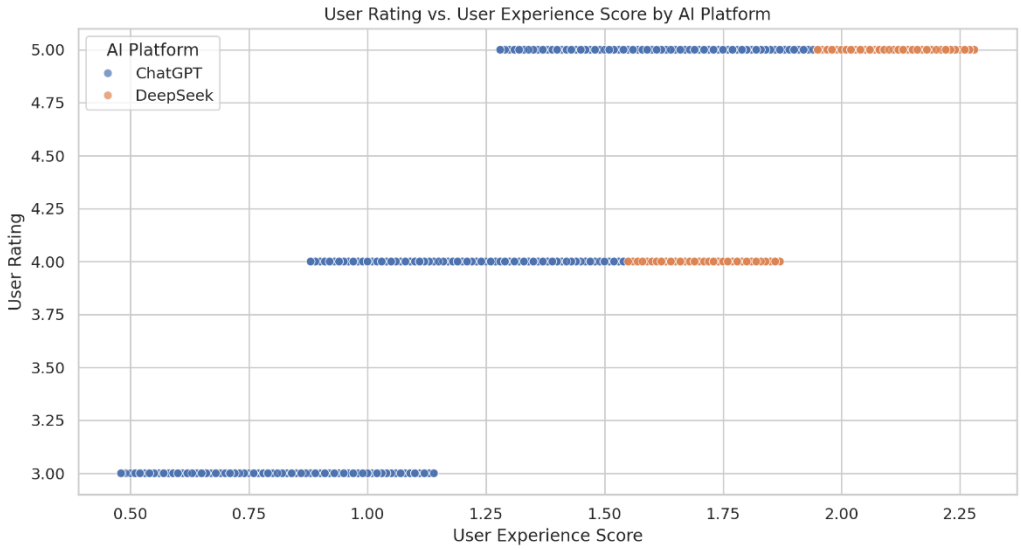
Profesyonel yazarlık oturumlarında, kullanıcıların ortalama deneyim puanlarının nispeten daha düşük olduğu (yaklaşık 1.0-1.5 aralığında) belirlenmiştir. Bu oturumlarda kullanıcı derecelendirmelerinin çoğunlukla 3 ile 4 arasında yoğunlaştığı gözlemlenmiştir. İçerik üretimi oturumlarında ise deneyim puanlarının orta ile yüksek düzeyde (yaklaşık 1.5-2.0 aralığında) olduğu ve kullanıcı derecelendirmelerinin genellikle 4 veya 5 olma eğiliminde bulunduğu dikkat çekmektedir. Hata ayıklama ve uygulama yardımı oturumlarında deneyim puanlarının sıklıkla 2.0'ı aştığı ve kullanıcı derecelendirmelerinin ağırlıklı olarak 4 ile 5 arasında olduğu görülmektedir. Bu durum, bu alanlarda kullanıcı deneyimlerinin pozitif bir algı yarattığını ortaya koymaktadır.

Mobil cihazlarda yapılan oturumlarda, deneyim puanlarında biraz daha fazla değişkenlik gözlemlenmektedir. Buna rağmen, kullanıcı derecelendirmeleri genellikle yüksek düzeylerde (3-5 puan) seyretmektedir. Dizüstü ve masaüstü bilgisayar oturumlarında ise ortalama deneyim puanlarının daha yüksek olduğu (yaklaşık 2.0 ve üzeri)

ve kullanıcı derecelendirmelerinin tutarlı bir şekilde yüksek düzeylerde (4-5 puan) gerçekleştiği tespit edilmiştir. Tablet ve akıllı hoparlör oturumlarında ise ortalama deneyim puanlarının nispeten daha düşük olduğu dikkat çekmektedir. Buna karşın, kullanıcı derecelendirmeleri yine 3-5 arasında yoğunlaşmakla birlikte, daha fazla dalgalanma ve değişkenlik göstermektedir.

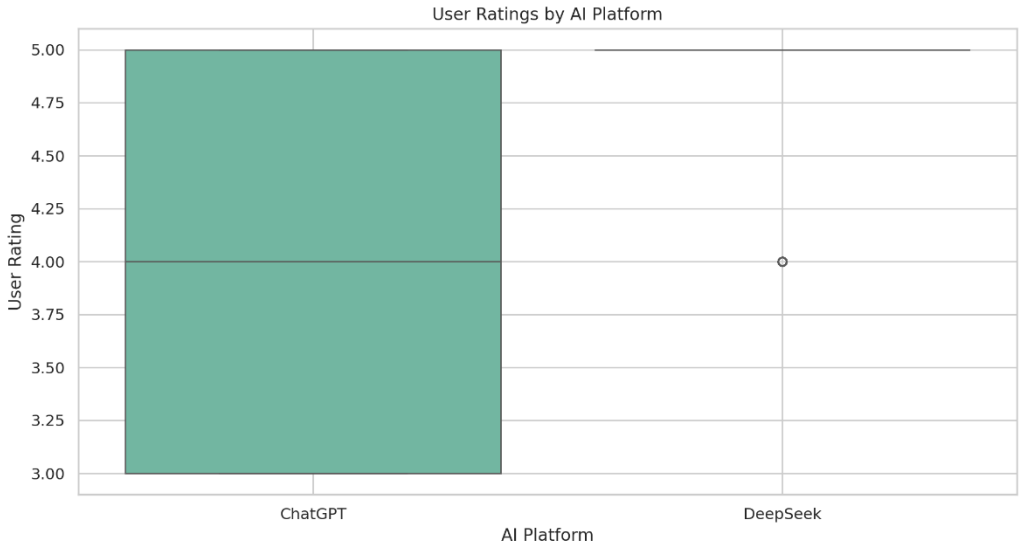
Genel olarak, daha yüksek kullanıcı deneyimi puanları, daha yüksek kullanıcı derecelendirmeleri ile korelasyon göstermektedir. DeepSeek kullanıcıları, ChatGPT kullanıcılarına kıyasla daha yüksek deneyim puanları ve derecelendirmeler bildirme eğilimindedir. "Hata Ayıklama", "Uygulama Yardımı" ve "Kod Optimizasyonu" gibi teknik konular, deneyim puanları ve kullanıcı derecelendirmeleri arasında güçlü pozitif ilişkiler sergilemektedir. Cihaz türü, deneyim puanlarını kısmen etkilemekte olup, dizüstü/masaüstü bilgisayarlar daha iyi kullanıcı deneyimleriyle ilişkilendirilmektedir.

Şekil 1'de sunulan saçılım grafiği, kullanıcı oyları ile kullanıcı deneyimi skorları arasındaki ilişkiyi platform bazında göstermektedir. Grafikte yer alan her bir nokta, analiz kapsamındaki bir oturumu temsil etmekte olup, noktalar kullanılan yapay zeka platformuna (ChatGPT veya DeepSeek) göre farklı renklendirilmiştir.



Şekil 1. Kullanıcı Oyları ve Kullanıcı Deneyimi Skoru İlişkisi

Şekil 2 ise, her bir yapay zeka platformu için kullanıcı oylarının dağılımını özetleyen bir kutu grafiğidir. Bu grafik, platformlara göre kullanıcı oylarının medyanını, çeyrekler açıklığını ve genel değişkenliğini vurgulamaktadır.



Şekil 2. Kullanıcı Oyları Dağılımı

Elde edilen bulgular incelendiğinde:

Saçılım grafiği (Şekil 1), kullanıcı deneyimi skorları ile kullanıcı oyları arasında pozitif bir korelasyon olduğunu, özellikle DeepSeek platformu için bu ilişkinin daha belirgin olduğunu göstermektedir. DeepSeek'in hem kullanıcı deneyimi skorlarının hem de kullanıcı oylarının genel olarak daha yüksek değerlerde yoğunlaştığı gözlenmiştir. Kutu grafiği (Şekil 2) bulgularını destekler nitelikte, DeepSeek platformunun ChatGPT'ye kıyasla genel olarak daha yüksek kullanıcı oyları aldığı ve bu oylardaki değişkenliğin (yayılımın) daha az olduğu tespit edilmiştir.

3.2. ChatGPT ve DeepSeek Platformlarının Kullanıcı Oyu ve Deneyimi Skorları Açısından İstatistiksel Karşılaştırılması

Bu bölümde, farklı yapay zeka platformlarının (ChatGPT ve DeepSeek) Kullanıcı Derecelendirmeleri ve Kullanıcı Deneyimi Puanları ortalamaları arasındaki farkların istatistiksel anlamlılığını değerlendirmek amacıyla Bağımsız Örneklemler T-Testi ve Tek Yönlü Varyans Analizi (ANOVA) uygulanmıştır.

Her iki değişken için platformlar arası ortalamaların karşılaştırılması amacıyla yapılan bağımsız örneklemler T-testi sonuçları Tablo 1'de sunulmuştur.

Tablo 1. T-testi Sonuçlarına Göre Kullanıcı Derecelendirmesi ve Deneyimi

Ölçüm Değişkeni	***t*(sd)**	**p-değeri**
Kullanıcı Derecelendirmesi	–65.37 (9998)	< .001
Kullanıcı Deneyimi Puanı	–142.12 (9998)	< .001

Not. Negatif t-istatistik değerleri, ikinci grubun ortalamasının daha yüksek olduğunu göstermektedir. $p < .001$ değeri istatistiksel olarak yüksek derecede anlamlı sonuçlara işaret eder.

T-testi sonuçları, hem Kullanıcı Derecelendirmeleri hem de Kullanıcı Deneyimi Puanları için istatistiksel olarak anlamlı düzeyde düşük p-değerleri (< 0,001) vermiştir. Bu durum, ChatGPT ve DeepSeek platformlarının ilgili

ortalama değerleri arasındaki gözlemlenen farkların istatistiksel olarak anlamlı olduğunu göstermektedir. Ortalama Kullanıcı Derecelendirmesi ChatGPT için yaklaşık 4,00 iken, DeepSeek için yaklaşık 4,80 olarak hesaplanmıştır. Benzer şekilde, ortalama Kullanıcı Deneyimi Puanı ChatGPT için yaklaşık 1,23 iken, DeepSeek için yaklaşık 2,03'tür. Bu bulgular, kullanıcıların DeepSeek platformunu ChatGPT'ye kıyasla genel olarak daha yüksek derecelendirdiğini ve DeepSeek'in daha olumlu bir kullanıcı deneyimi sunduğunu işaret etmektedir.

Platformun (ChatGPT vs. DeepSeek) Kullanıcı Derecelendirmeleri ve Kullanıcı Deneyimi Puanları üzerindeki etkisini incelemek için yapılan ANOVA sonuçları Tablo 2'de gösterilmiştir.

Tablo 2. ANOVA Sonuçlarına Göre Kullanıcı Derecelendirmesi ve Deneyimi

Ölçüm Değişkeni	F-İstatistik	p-değeri
Kullanıcı Derecelendirmesi	4273.78	< .001
Kullanıcı Deneyimi Puanı	20199.47	< .001

Not. F-İstatistik değerleri ve p-değerleri, gruplar arasında anlamlı farklar olduğunu göstermektedir. $p < .001$ değeri, istatistiksel olarak yüksek derecede anlamlı farklara işaret eder.

ANOVA sonuçları, hem Kullanıcı Derecelendirmeleri ($F(1, 9998) = 4273,78$, $p < 0,001$) hem de Kullanıcı Deneyimi Puanları ($F(1, 9998) = 20199,47$, $p < 0,001$) açısından platformlar arasında istatistiksel olarak anlamlı farklılıklar olduğunu doğrulamaktadır. Yüksek F-istatistik değerleri, yapay zeka platformunun söz konusu değişkenler üzerinde güçlü bir etkiye sahip olduğunu göstermektedir. Elde edilen p-değerlerinin belirlenen anlamlılık düzeyi olan 0,05'ten düşük olması nedeniyle, sıfır hipotezi reddedilmekte ve ChatGPT ile DeepSeek arasında kullanıcı memnuniyeti düzeylerinde önemli farklılıklar olduğu sonucuna varılmaktadır. Gözlemlenen farklılıkların rastgele şanstan kaynaklanma olasılığı çok düşüktür.

3.3. Regresyon Modeli Sonuçları

Bu bölümde, Kullanıcı Derecelendirmelerindeki değişkenliği açıklamak amacıyla uygulanan farklı regresyon modellerinin performans sonuçları sunulmaktadır. Doğrusal Regresyon, Rastgele Orman ve Gradyan Artırma modelleri test edilmiş ve modellerin başarıları Ortalama Karesel Hata (OKH) ve R^2 Skoru metrikleri üzerinden değerlendirilmiştir. Modellere ait performans sonuçları Tablo 3'te gösterilmiştir.

Tablo 3. Modellere Ait Performans Sonuçları

Model	Ortalama Karesel Hata (OKH)	R^2 Skoru
Doğrusal Regresyon	0.4089	0.2467
Rastgele Orman	0.4090	0.2466
Gradyan Artırma	0.3825	0.2954

Not. Ortalama Karesel Hata (OKH), modelin tahminlerinin gerçek değerlere olan ortalama karesel uzaklığını temsil eder. R^2 Skoru, modelin veri üzerindeki açıklayıcılığını ölçer.

Elde edilen regresyon sonuçları incelendiğinde, test edilen üç model arasında en iyi performansı Gradyan Artırma modelinin sergilediği görülmektedir. Gradyan Artırma modeli, en düşük Ortalama Karesel Hataya (0.3825) ve en

yüksek R^2 skoruna (0.2954) sahip olup, bu durum modelin Kullanıcı Derecelendirmelerindeki varyansın diğer modellere kıyasla daha büyük bir oranını açıkladığını düşündürmektedir.

Bununla birlikte, elde edilen R^2 skorlarının (hepsi 0.3'ün altında olması) tüm modeller için nispeten düşük olduğu dikkat çekmektedir. Bu durum, modellerin Kullanıcı Derecelendirmelerindeki toplam varyansın büyük bir kısmını açıklayamadığını göstermektedir. Düşük R^2 değerleri, mevcut özellik setinde yakalanamayan, kullanıcı derecelendirmelerini etkileyen başka önemli faktörlerin olabileceğine işaret etmektedir.

3.4. Sınıflandırma Modelleri Sonuçları

Bu çalışmada, sınıflandırma görevi için Lojistik Regresyon, Rastgele Orman Sınıflandırıcısı (Random Forest Classifier) ve Destek Vektör Makinesi (Support Vector Machine - SVM) olmak üzere üç farklı sınıflandırma modeli uygulanmıştır. Modellerin performansları, Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall) ve F1 Skoru metrikleri kullanılarak değerlendirilmiştir. Her bir model için genel doğruluk değeri ve her bir sınıf ('Excellent', 'Fair', 'Good') bazında elde edilen detaylı performans metrikleri aşağıda sunulmuştur.

Yapılan sınıflandırma analizleri sonucunda, lojistik regresyon modeli %57.51 doğruluk oranı ile en yüksek performansı göstermiştir. Rastgele orman (random forest) algoritması %57.25 doğruluk oranı ile lojistik regresyona oldukça yakın bir performans sergilemiştir. Buna karşılık, destek vektör makineleri (SVM) %52.88 doğruluk oranı ile diğer iki modele kıyasla daha düşük bir başarı elde etmiştir. Bu sonuçlar, kullanılan veri seti ve özellikler doğrultusunda lojistik regresyon modelinin sınıflandırma görevinde diğer modellere göre daha etkili olabileceğini göstermektedir.

Tablo 4. Sınıflandırma Modellerinin Performans Metrikleri

Sınıf	Model	Kesinlik	Duyarlılık	F1 Skoru	Destek
Mükemmel	Lojistik Regresyon	0.6497	0.8418	0.7334	1018
	Rastgele Orman	0.6934	0.7908	0.7389	1018
	SVM	0.5288	1.0000	0.6918	1018
Orta	Lojistik Regresyon	0.3386	0.1448	0.2028	297
	Rastgele Orman	0.2672	0.2088	0.2344	297
	SVM	0.0000	0.0000	0.0000	297
İyi	Lojistik Regresyon	0.4322	0.3393	0.3802	610
	Rastgele Orman	0.4417	0.3852	0.4116	610
	SVM	0.0000	0.0000	0.0000	610
Makro Ort.	Lojistik Regresyon	0.4735	0.4420	0.4388	1925
	Rastgele Orman	0.4674	0.4616	0.4616	1925
	SVM	0.1763	0.3333	0.2306	1925
Ağırlık. Ort.	Lojistik Regresyon	0.5328	0.5751	0.5396	1925

Rastgele Orman	0.5479	0.5725	0.5573	1925
SVM	0.2797	0.5288	0.3659	1925

Not. Kesinlik (Precision), Duyarlılık (Recall), ve F1 Skoru, sınıflandırma modellerinin doğruluk ölçütlerini temsil eder. Destek, her sınıftaki örnek sayısını göstermektedir.

Elde edilen sınıflandırma sonuçları incelendiğinde Lojistik Regresyon Modelinin genel olarak makul bir performans sergilediği görülmüştür. Bu model özellikle 'Excellent' (Mükemmel) kategorisini tahmin etmede daha başarılı iken, 'Fair' (Orta) ve 'Good' (İyi) kategorilerinde performansı daha düşük kalmıştır.

Rastgele Orman Modeli, Lojistik Regresyon modeline benzer bir performans göstermiştir. Genel doğruluk değeri hafifçe daha düşük olmasına rağmen, Kesinlik ve Duyarlılık gibi diğer performans metrikleri karşılaştırılabilir düzeydedir. SVM modelinin ise diğer modellere kıyasla genel doğruluğunun daha düşük olduğu ve özellikle 'Fair' ve 'Good' kategorilerini etkili bir şekilde tahmin edemediği tespit edilmiştir. Bu sonuçlar, kullanılan modellerin özellikle yaygın olan 'Excellent' sınıfını tahmin etmede bir miktar başarı gösterdiğini, ancak daha az temsil edilen sınıflarda (Fair, Good) sınıflandırma performansının belirgin şekilde düştüğünü ortaya koymaktadır.

3.5.Kullanıcı Etkileşimi ve Sadakati Karşılaştırması

Bu bölümde, kullanıcı etkileşimi ve bağlılığına ilişkin farklı yönleri modellemek amacıyla geliştirilen makine öğrenmesi modellerinin sonuçları sunulmaktadır. Analizler hem sınıflandırma hem de regresyon görevlerini içermektedir.

Günlük Terk Oranı (Daily_Churn_Rate) değişkeni, belirlenen bir eşik değeri (0.1) kullanılarak ikili bir sınıflandırma problemine dönüştürülmüştür. Bu eşik değere göre, günlük terk oranı 0.1'i aşan kullanıcılar 'terk etmiş' (1), aşmayanlar ise 'terk etmeyen' (0) olarak sınıflandırılmıştır. Bu ikili hedef değişkeni tahmin etmek üzere bir Rastgele Orman Sınıflandırıcısı (Random Forest Classifier) eğitilmiş ve modelin değerlendirme sonuçları aşağıda sunulmuştur. Modelin performansına ilişkin detaylı sınıflandırma metrikleri — kesinlik (precision), duyarlılık (recall), F1 skoru ve destek (support) — Tablo 5'te sunulmuştur. Bu metrikler, modelin sınıflandırma başarısını daha kapsamlı bir şekilde değerlendirmek amacıyla raporlanmıştır. Rastgele Orman Sınıflandırıcısının genel doğruluk değeri ise %100 olarak elde edilmiştir.

Tablo 5. Model Performansına Göre Sınıf 0 ve Ortalama Değerler

Sınıf / Ortalama	Kesinlik	Duyarlılık	F1 Skoru	Destek
0	1.000	1.000	1.000	1925
Makro Ortalama	1.000	1.000	1.000	1925
Ağırlıklı Ortalama	1.000	1.000	1.000	1925

Not. Tüm metriklerin 1.000 olması, sınıflandırma modelinin hata yapmadan tüm örnekleri doğru tahmin ettiğini göstermektedir.

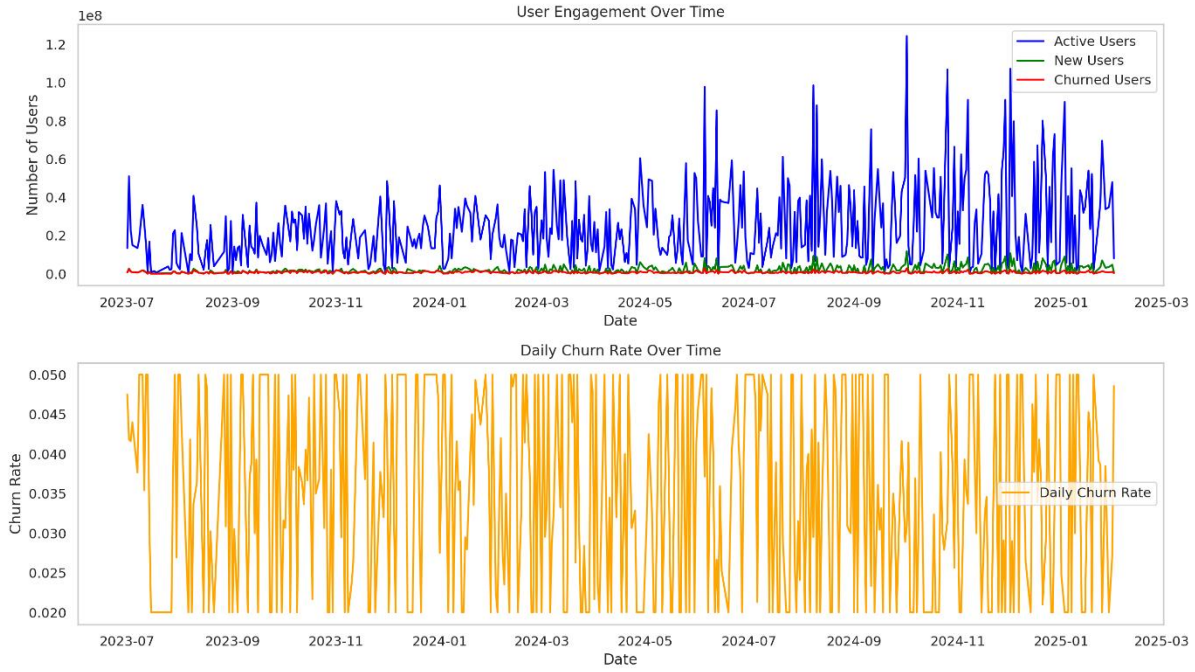
Elde edilen sonuçlara göre, Rastgele Orman sınıflandırma modeli test setinde %100 doğruluk elde etmiştir, bu da modelin tahminlerinin mükemmel olduğunu göstermektedir. Sunulan verilere göre, '0' sınıfı (terk etmeyen) için

Kesinlik, Duyarlılık ve F1 Skoru değerlerinin tamamı 1.0'dır. Bu durum, modelin test setindeki 'terk etmeyen' kullanıcıları hatasız bir şekilde belirleyebildiğini işaret etmektedir.

Sürekli değişkenler olan Oturum Süresi (Session_Duration_sec) ve Kullanıcı Dönüş Sıklığı (User_Return_Frequency) gibi kullanıcı davranışlarını tahmin etmek amacıyla bir Rastgele Orman Regresyoncusu (Random Forest Regressor) eğitilmiştir. Regresyon modellerinin performansını değerlendirmek için Ortalama Karesel Hatanın Karekökü (Root Mean Squared Error - RMSE) metriği kullanılmıştır.

Regresyon modellerine ait KOKH (RMSE) sonuçları aşağıda sunulmuştur:

- **Oturum Süresi KOKH (RMSE):** Yaklaşık 14.57 saniye
- **Kullanıcı Dönüş Sıklığı KOKH (RMSE):** Yaklaşık 3.16



Şekil 3. Oturum Süresi ve Kullanıcı Dönüş Sıklığı

Elde edilen KOKH (RMSE) değerleri, oturum süresi tahmini için yaklaşık 14.57 saniye ve kullanıcı dönüş sıklığı tahmini için yaklaşık 3.16 olarak hesaplanmıştır. KOKH değeri, tahmin edilen değerlerin gerçek değerlerden ortalama sapmasını gösterir; dolayısıyla daha düşük KOKH değerleri daha iyi performansı işaret eder. Bu sonuçlar, Rastgele Orman Regresyon modellerinin oturum süresi ve kullanıcı dönüş sıklığını tahmin etmede makul derecede etkili olduğunu düşündürmektedir.

3.6.Yanıt Kalitesi ve Performansı Karşılaştırması

Yanıt doğruluğunu tahmin etmek amacıyla uygulanan regresyon modelinin performansını değerlendirmek için Kök Ortalama Karesel Hata (RMSE) metriği kullanılmıştır. Yanıt doğruluğu tahmini için elde edilen KOKH (RMSE) değeri yaklaşık **0.064**'tür. Bu düşük ortalama sapma değeri, modelin yanıt doğruluğu skorlarını tahmin etmede iyi bir performans sergilediğini düşündürmektedir.

Yanıtın düzeltmeye ihtiyacı olup olmadığını (0 = Hayır, 1 = Evet) sınıflandırmak amacıyla uygulanan modelin performansını gösteren sınıflandırma raporu aşağıdadır:

Tablo 6. Sınıflandırma Performans Metrikleri (Sınıf 0 ve 1)

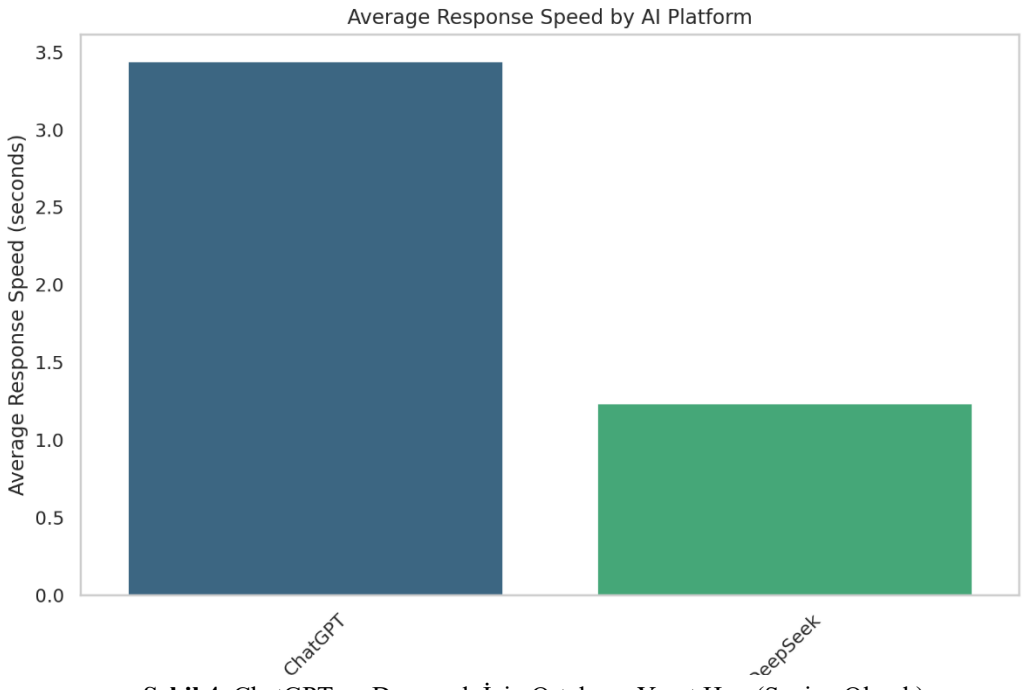
Sınıf / Ortalama	Kesinlik	Duyarlılık	F1 Skoru	Destek
0	0.863	0.913	0.888	1655
1	0.177	0.115	0.139	270
Makro Ortalama	0.520	0.514	0.513	1925
Ağırlıklı Ortalama	0.767	0.801	0.783	1925

Not. Makro ortalamalar sınıflar arasında eşit ağırlıkla hesaplanırken, ağırlıklı ortalamalar sınıf destek (örnek sayısı) oranında ağırlıklandırılarak elde edilir.

Sınıflandırma modeli, düzeltme gerektirmeyen yanıtlar (Sınıf 0) için yaklaşık **%86.34** Kesinlik ve yaklaşık **%91.30** Duyarlılık sergilemiştir. Düzeltme gerektiren yanıtlar (Sınıf 1) için ise Kesinlik yaklaşık **%17.71**, Duyarlılık ise yaklaşık **%11.48** olarak bulunmuştur. Bu metrikler, modelin düzeltme gerektirmeyen durumları başarılı bir şekilde belirleyebildiğini, ancak düzeltme gerektiren durumları belirlemede zorlandığını göstermektedir. Modelin genel doğruluğu ise yaklaşık **%80.10**'dur.

Yanıt kalitesi ve performansına ilişkin analizler, farklı yapay zeka platformlarının etkinliği hakkında değerli içgörüler sağlamıştır. Yanıt doğruluğu ve düzeltme gereksinimi değerlendirmelerinin ardından, çeşitli platformlar arasındaki ortalama yanıt hızı görselleştirilmiştir.

Aşağıdaki çubuk grafik, her bir yapay zekâ platformu için ortalama yanıt hızını (saniye olarak) göstermektedir:

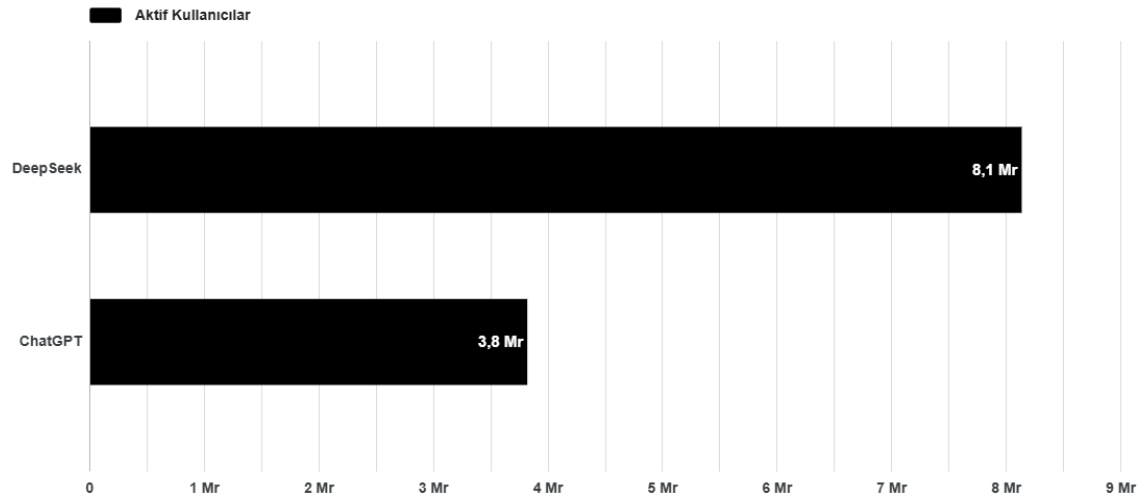


Şekil 4. ChatGPT ve Deepseek İçin Ortalama Yanıt Hızı (Saniye Olarak)

Bu grafik (Şekil 4), her bir yapay zekâ platformunun kullanıcı sorgularına ne kadar hızlı yanıt verdiğini karşılaştırmamıza olanak tanımaktadır. Bu bilgi, kullanıcı memnuniyeti için kritik öneme sahip olan yanıt süresi açısından hangi platformların daha iyi performans gösterdiğini belirlemeye yardımcı olabilir.

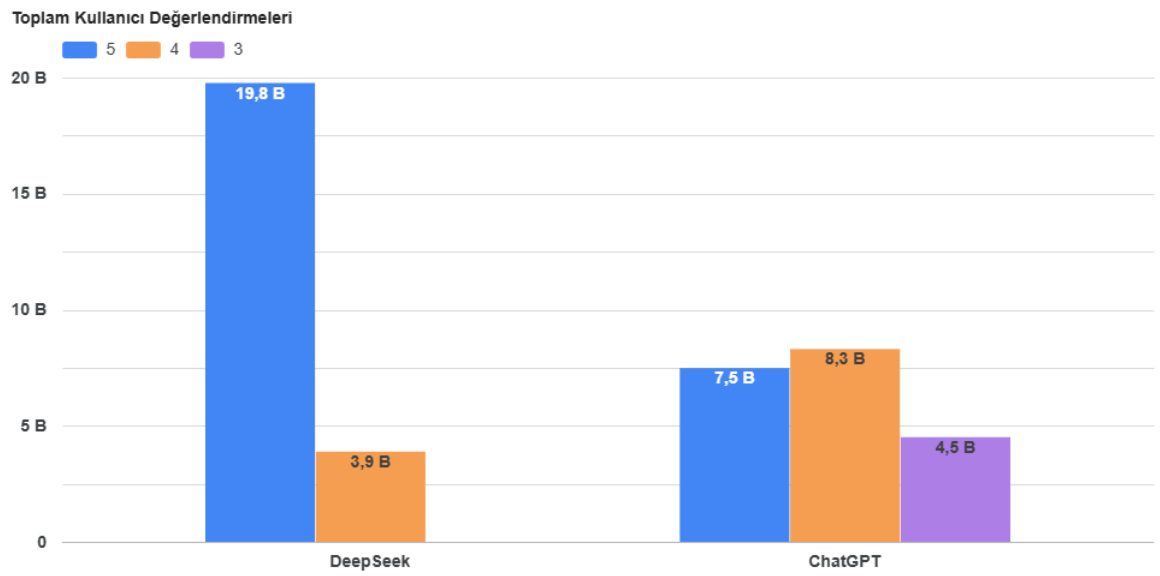
3.7.Aktif Kullanıcı Sayısı ve Toplam Kullanıcı Değerlendirmeleri

Şekil 5, DeepSeek ve ChatGPT platformlarının aktif kullanıcı sayılarını karşılaştırmaktadır. Elde edilen verilere göre, DeepSeek platformu 8.1 milyon aktif kullanıcıya sahipken, ChatGPT'nin aktif kullanıcı sayısı 3.8 milyon olarak belirlenmiştir. Bu sonuç, DeepSeek'in ChatGPT'ye kıyasla daha geniş bir aktif kullanıcı kitlesine ulaştığını göstermektedir.



Şekil 5. Aktif Kullanıcı Sayısı

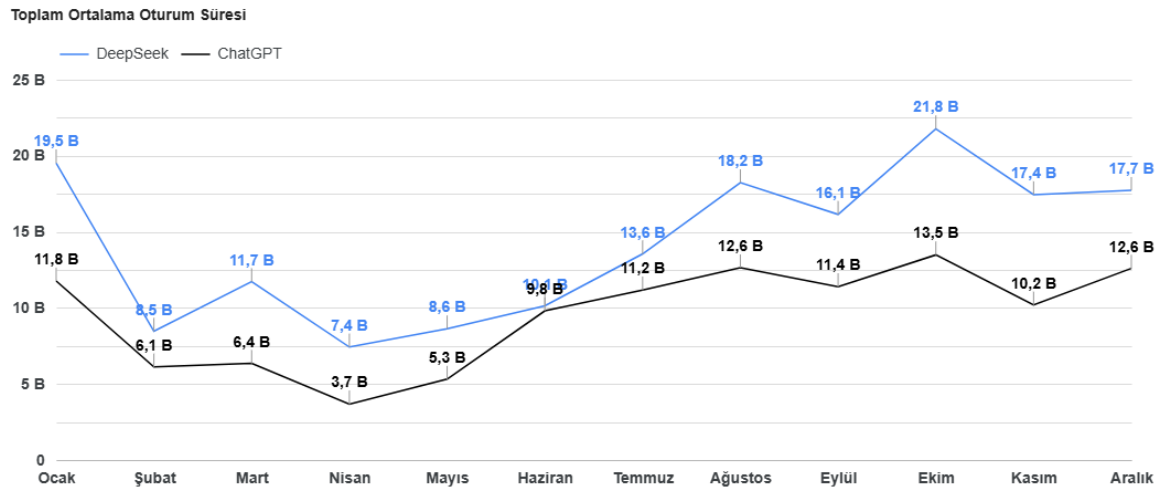
Şekil 6, DeepSeek ve ChatGPT platformlarının toplam kullanıcı değerlendirmelerini karşılaştırmaktadır. Elde edilen verilere göre, DeepSeek platformu 5 üzerinden ortalama 19.8 milyar, 4 üzerinden 3.9 milyar değerlendirme almıştır. ChatGPT ise 5 üzerinden 7.5 milyar, 4 üzerinden 8.3 milyar ve 3 üzerinden 4.5 milyar değerlendirme almıştır. Bu sonuçlar, her iki platformun da yüksek kullanıcı etkileşimi ve geri bildirimi aldığını göstermektedir. DeepSeek'in 5 yıldız üzerinden aldığı yüksek değerlendirme sayısı dikkat çekicidir.



Şekil 6. Toplam Kullanıcı Değerlendirmeleri Karşılaştırması

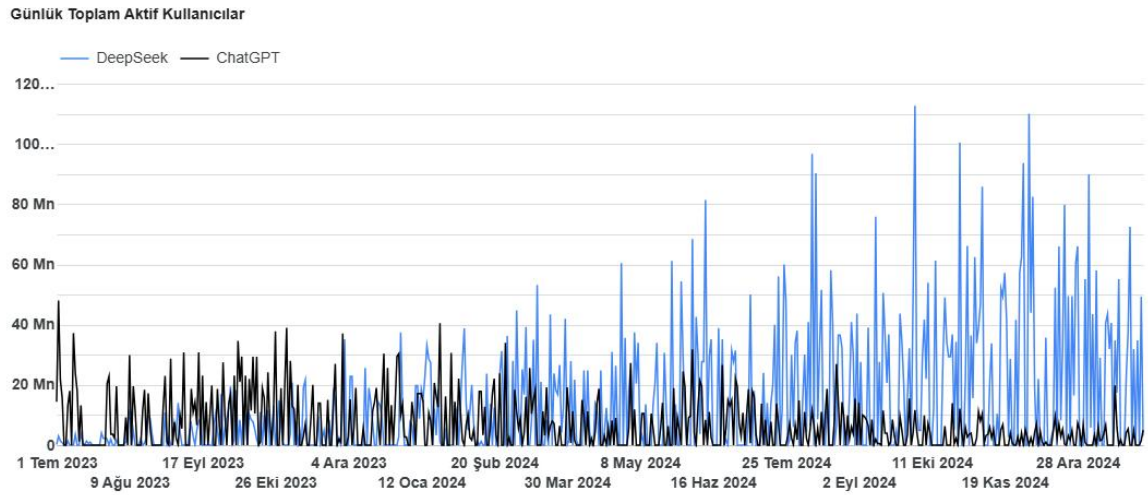
3.8.Toplam Ortalama Oturum Süresi ve Günlük Toplam Aktif Kullanıcılar

Şekil 7, DeepSeek ve ChatGPT platformlarının aylık bazda toplam ortalama oturum sürelerini milyar cinsinden göstermektedir. Yıl boyunca her iki platformda da oturum sürelerinde dalgalanmalar gözlemlenmektedir.



Şekil 7. Toplam Ortalama Oturum Süresi Karşılaştırması

DeepSeek'in oturum süreleri genellikle ChatGPT'den daha yüksek seyretmekle birlikte, bazı aylarda (örneğin Şubat, Nisan) ChatGPT'nin oturum sürelerinin DeepSeek'i geçtiği görülmektedir. Özellikle Ekim ayında DeepSeek'in oturum süresinde belirgin bir artış yaşanırken, Kasım ayında düşüş gözlemlenmiştir. ChatGPT'nin oturum süreleri ise daha istikrarlı bir seyir izlemekle birlikte, Aralık ayında belirgin bir artış göstermiştir. Bu bulgular, kullanıcıların platformlarla etkileşim sürelerinin zaman içinde değişiklik gösterebildiğini ve platform özelliklerinin bu değişikliklerde etkili olabileceğini düşündürmektedir.

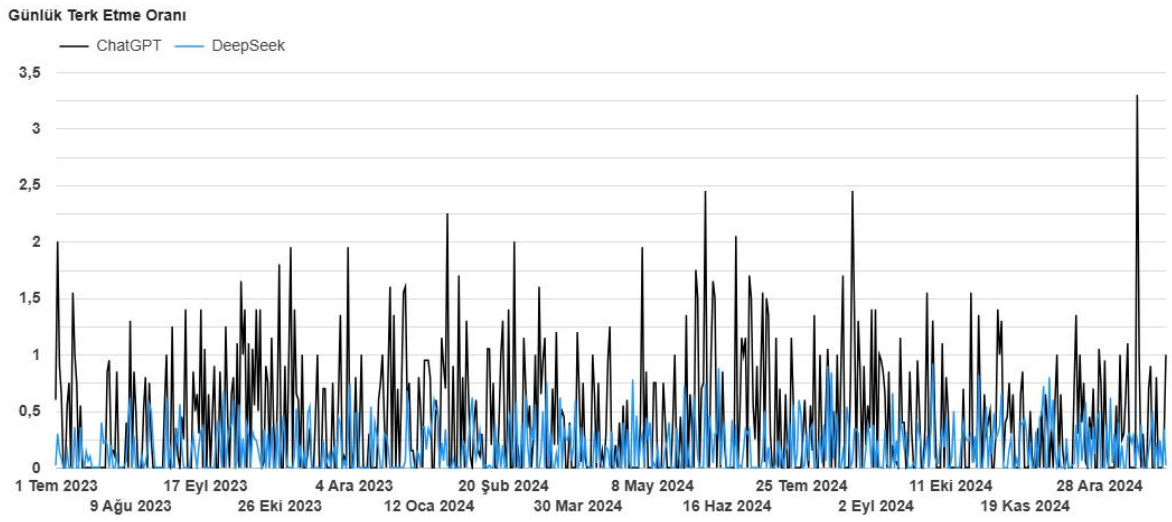


Şekil 8. Günlük Toplam Aktif Kullanıcı Sayısı Karşılaştırması

Şekil 8, DeepSeek ve ChatGPT platformlarının günlük toplam aktif kullanıcı sayılarını milyon cinsinden göstermektedir. Grafik incelendiğinde, DeepSeek'in günlük aktif kullanıcı sayısında ChatGPT'ye kıyasla genel olarak daha yüksek ve daha dalgalı bir seyir izlendiği görülmektedir. Özellikle 2024 yılının ortalarından itibaren DeepSeek'in aktif kullanıcı sayısında belirgin pikler yaşanmıştır. ChatGPT'nin günlük aktif kullanıcı sayısı ise daha stabil bir görünüm sergilemekte ve DeepSeek'in zirve yaptığı dönemlerde bile daha düşük seviyelerde kalmaktadır. Bu durum, DeepSeek'in zaman zaman önemli ölçüde daha fazla kullanıcı etkileşimi yarattığını, ancak bu etkileşimin ChatGPT'ye göre daha değişken olabileceğini düşündürmektedir.

3.9.Günlük Terk Etme Oranı ve Ortalama Oturum Süresi

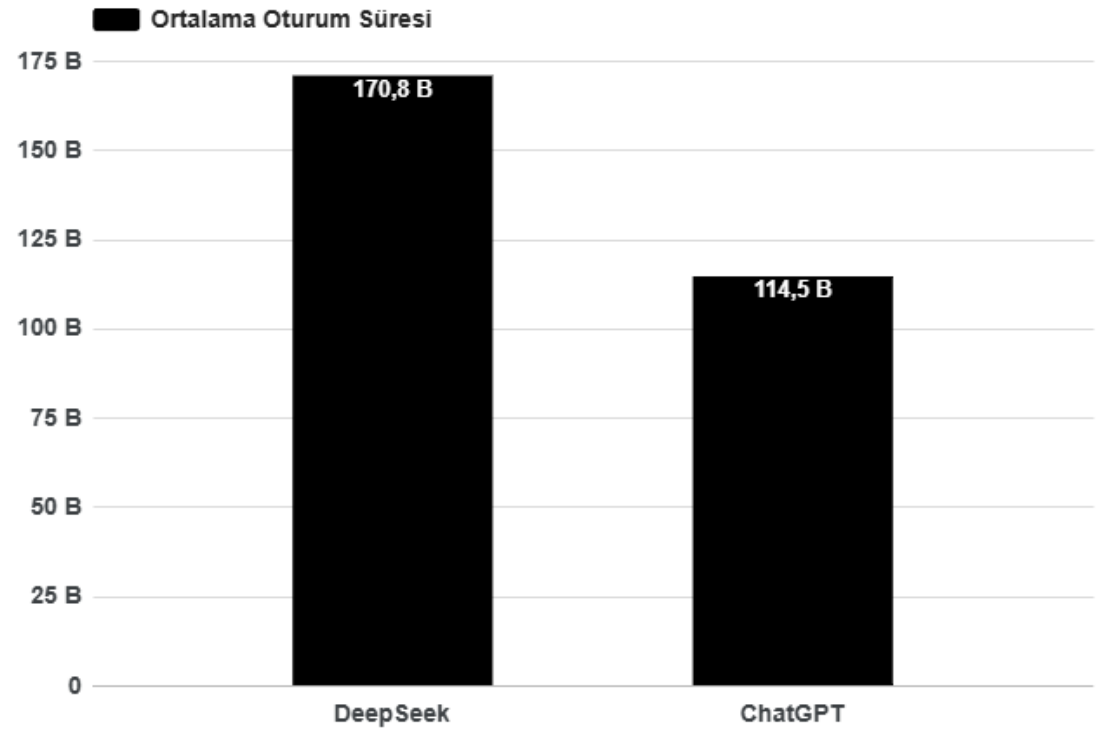
Şekil 9, DeepSeek ve ChatGPT platformlarının günlük terk etme oranlarını göstermektedir. Grafik incelendiğinde, her iki platformda da zaman zaman belirgin yükselişler olsa da, genel olarak DeepSeek'in terk etme oranının ChatGPT'ye göre daha düşük seyrettiği görülmektedir.



Şekil 9. Günlük Terk Etme Oranı Karşılaştırması

ChatGPT'nin terk etme oranında özellikle 2023'ün sonlarına doğru ve 2024'ün ilk aylarında daha yüksek değerler gözlemlenirken, DeepSeek'te bu dönemde daha stabil ve düşük bir oran söz konusudur. 2024 yılı boyunca her iki platformda da terk etme oranlarında inişli çıkışlı bir seyir izlenmekle birlikte, DeepSeek'in genel ortalamasının ChatGPT'den daha iyi olduğu söylenebilir. Bu bulgu, kullanıcıların DeepSeek ile etkileşimlerinden ChatGPT'ye göre daha fazla memnun kaldıkları veya platformun kullanıcı beklentilerini daha iyi karşıladığı şeklinde yorumlanabilir.

Şekil 10, DeepSeek ve ChatGPT platformlarının ortalama oturum sürelerini milyar cinsinden karşılaştırmaktadır.

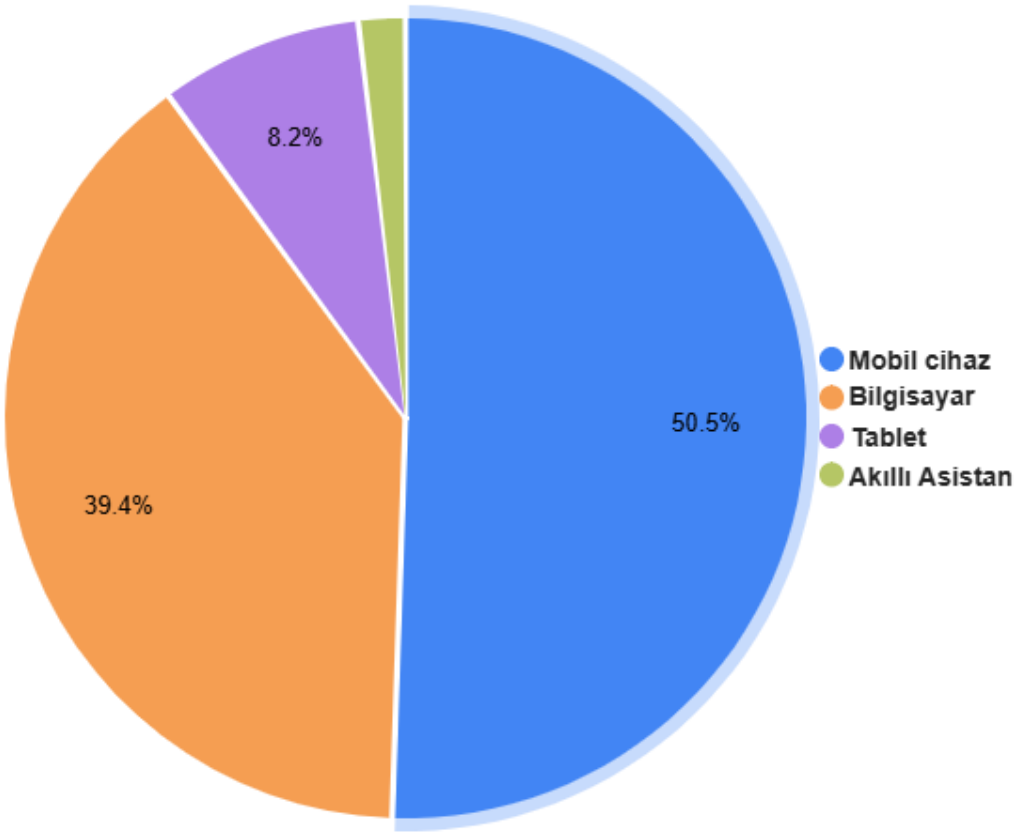


Şekil 10. Ortalama Oturum Süresi Karşılaştırması

Elde edilen verilere göre, DeepSeek platformunun ortalama oturum süresi 170.8 milyar olarak belirlenirken, ChatGPT'nin ortalama oturum süresi 114.5 milyardır. Bu sonuç, kullanıcıların DeepSeek ile etkileşimlerinin ChatGPT'ye kıyasla önemli ölçüde daha uzun sürdüğünü göstermektedir. Bu durum, DeepSeek'in kullanıcıları daha fazla etkileşimde tutabilen veya daha karmaşık sorgulara yanıt verebilen özelliklere sahip olabileceği şeklinde yorumlanabilir.

3.10. Kullanıcıların Cihaz Tercihleri ve Dil Bazında Kullanıcı Dağılımı

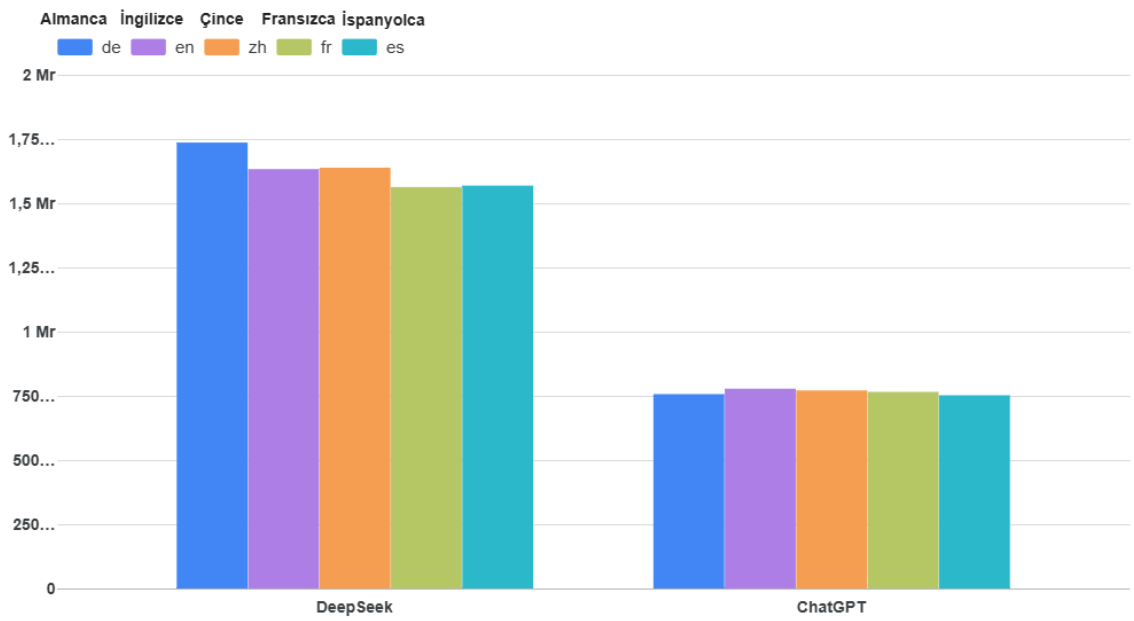
Şekil 11, kullanıcıların platforma erişim için kullandıkları cihazların dağılımını göstermektedir.



Şekil 11. Kullanıcıların Cihaz Tercihleri

Elde edilen verilere göre, kullanıcıların büyük çoğunluğu (%50.5) platforma mobil cihazlar aracılığıyla erişmektedir. Bilgisayar (%39.4) ise en sık kullanılan ikinci erişim yöntemi olarak belirlenmiştir. Tablet (%8.2) ve akıllı asistanlar (%1.9) üzerinden erişim oranları ise daha düşüktür. Bu bulgu, platforma erişimde mobil cihazların önemli bir rol oynadığını ve mobil uyumluluğun kullanıcı deneyimi açısından kritik olduğunu göstermektedir.

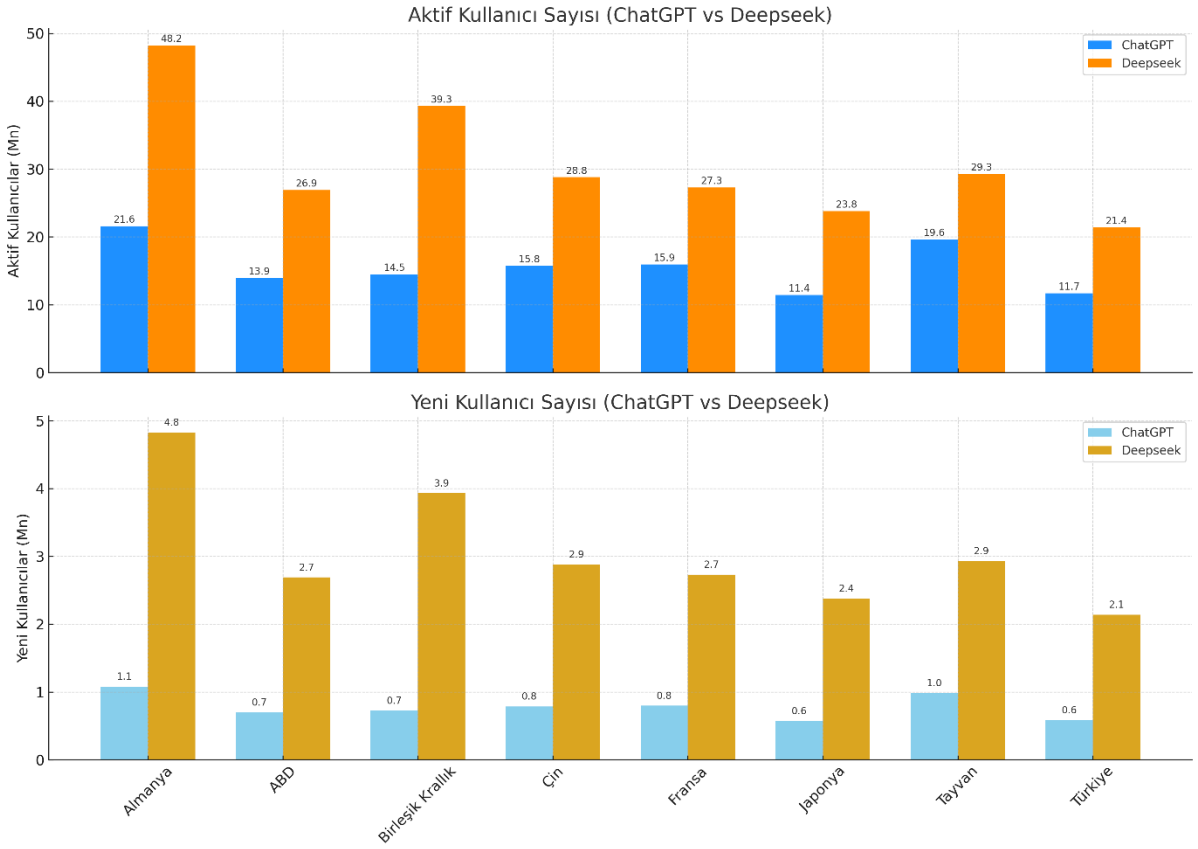
Şekil 12, DeepSeek ve ChatGPT platformlarının farklı dillerdeki kullanıcı dağılımlarını milyon cinsinden göstermektedir.



Şekil 12. Dil Bazında Kullanıcı Dağılımı

DeepSeek platformunda en fazla kullanıcının İngilizce (en) konuştuğu (yaklaşık 1.75 milyon) görülmektedir. Bunu sırasıyla Almanca (de), Çince (zh) ve İspanyolca (es) takip etmektedir. Fransızca (fr) konuşan kullanıcı sayısı ise diğer dillere göre daha düşüktür. ChatGPT platformunda ise İngilizce (en) ve Çince (zh) konuşan kullanıcı sayıları birbirine yakın (yaklaşık 0.75 milyon) ve diğer dillere göre daha yüksektir. Almanca (de), Fransızca (fr) ve İspanyolca (es) konuşan kullanıcı sayıları ise benzer seviyelerde ve İngilizce ile Çinceye göre daha düşüktür. Bu bulgular, DeepSeek'in İngilizce konuşan kullanıcılar arasında daha popüler olduğunu gösterirken, ChatGPT'nin İngilizce ve Çince konuşan kullanıcılar arasında daha dengeli bir dağılıma sahip olduğunu düşündürmektedir. Her iki platformda da diğer Avrupa dillerini konuşan kullanıcı sayısının İngilizce ve Çinceye göre daha az olması dikkat çekicidir.

3.11.Ayrıntılı İncelenen 8 Ülkenin Aktif Kullanıcı Sayısı ve Yeni Kullanıcı Sayısı Karşılaştırılması (Platform Bazında)



Şekil 13. Ayrıntılı İncelenen 8 Ülkenin Aktif Kullanıcı Sayısı ve Yeni Kullanıcı Sayısı Karşılaştırılması (Platform Bazında)

Şekil 13'ün üst panelinde yer alan bar grafik, incelenen sekiz ülkedeki (Almanya, ABD, Birleşik Krallık, Çin, Fransa, Japonya, Tayvan, Türkiye) ChatGPT ve DeepSeek'in milyon cinsinden aktif kullanıcı sayılarını karşılaştırmaktadır. Genel olarak, DeepSeek'in çoğu ülkede ChatGPT'ye kıyasla önemli ölçüde daha yüksek aktif kullanıcı sayısına sahip olduğu görülmektedir. Özellikle Almanya'da DeepSeek'in 48.7 milyon aktif kullanıcısı varken, ChatGPT'nin aktif kullanıcı sayısı 21.6 milyonda kalmıştır. Benzer şekilde, Çin (DeepSeek: 38.8 Mn, ChatGPT: 15.8 Mn), Fransa (DeepSeek: 27.5 Mn, ChatGPT: 15.9 Mn) ve Tayvan (DeepSeek: 29.3 Mn, ChatGPT: 21.6 Mn) gibi ülkelerde DeepSeek belirgin bir üstünlük sağlamıştır. ABD'de ise DeepSeek (26.9 Mn) aktif kullanıcı sayısında ChatGPT'yi (13.9 Mn) geride bırakmasına rağmen, aradaki fark diğer bazı ülkelere göre daha düşüktür. Türkiye'de ise DeepSeek'in (21.4 Mn) ChatGPT'ye (12.7 Mn) kıyasla daha fazla aktif kullanıcıya sahip olduğu tespit edilmiştir. Bu veriler, DeepSeek'in küresel çapta, özellikle belirli pazarlarda, mevcut kullanıcı tabanı açısından ChatGPT'ye göre daha geniş bir erişime sahip olduğunu göstermektedir.

Şekil 13'ün alt panelindeki bar grafik ise aynı ülkelerdeki yeni kullanıcı sayılarını milyon cinsinden kıyaslamaktadır. Bu paneldeki bulgular, aktif kullanıcı sayısındaki eğilime paralel olarak, DeepSeek'in yeni kullanıcı kazanımında da ChatGPT'ye kıyasla dikkate değer bir başarı elde ettiğini ortaya koymaktadır. Almanya'da DeepSeek 4.8 milyon yeni kullanıcı kazanırken, ChatGPT'nin yeni kullanıcı sayısı yalnızca 1.3 milyondur. ABD'de DeepSeek 2.7 milyon yeni kullanıcıya ulaşırken, ChatGPT 0.7 milyonda kalmıştır. Birleşik Krallık (DeepSeek: 3.9 Mn, ChatGPT: 0.7 Mn), Çin (DeepSeek: 2.9 Mn, ChatGPT: 0.9 Mn), Fransa (DeepSeek:

2.7 Mn, ChatGPT: 0.8 Mn), Japonya (DeepSeek: 2.4 Mn, ChatGPT: 0.6 Mn), Tayvan (DeepSeek: 2.9 Mn, ChatGPT: 1.0 Mn) ve Türkiye (DeepSeek: 2.1 Mn, ChatGPT: 0.8 Mn) verileri de DeepSeek'in yeni kullanıcı kazanma oranlarının ChatGPT'ye göre belirgin şekilde yüksek olduğunu teyit etmektedir. Bu durum, DeepSeek'in pazarlama stratejilerinin, ürün özelliklerinin veya yerel pazar uyumluluğunun yeni kullanıcıları çekme konusunda daha etkili olduğunu düşündürmektedir.

Her iki grafiğin birleşik analizi, DeepSeek'in incelenen pazarlarda hem mevcut kullanıcı tabanını genişletme hem de yeni kullanıcıları çekme konusunda ChatGPT'ye karşı genel bir üstünlük sağladığını göstermektedir. Bu bulgu, üretken yapay zekâ pazarındaki rekabetin dinamiklerini ve DeepSeek gibi platformların belirli bölgelerde önemli bir kullanıcı tabanı oluşturma potansiyelini ortaya koymaktadır.

4.Tartışma ve Sonuç

Bu çalışmada, ChatGPT ve DeepSeek üretken yapay zekâ modellerinin kullanıcı etkileşimleri üzerindeki etkileri çok boyutlu olarak analiz edilmiştir. Elde edilen bulgular, her iki platformun da güçlü yönleri ve geliştirilmesi gereken alanları olduğunu ortaya koymakta, literatürdeki benzer çalışmalarla da paralellikler ve farklılıklar göstermektedir.

4.1.Bulguların Yorumlanması ve Alanyazınla İlişkisi

Bulgulara göre, DeepSeek'in özellikle teknik görevlerde (örneğin hata ayıklama, kod optimizasyonu gibi) kullanıcılar tarafından daha yüksek derecelendirmelerle değerlendirilmesi, modelin uzmanlaşmış işlevsellikleriyle uyumludur. Bu durum, DeepSeek'in mimari olarak transformatör tabanlı kodlama görevleri için optimize edilmesinin (Wang vd., 2024) ve özellikle kod üretimi ile mantıksal çıkarım gerektiren görevlerde öne çıkmasının (DeepSeek-AI, 2024) doğal bir sonucu olarak değerlendirilebilir. Nitekim, Guo vd. (2024), DeepSeek'in Python, Java ve C++ gibi dillerde yüksek başarı gösterdiğini ve CodeLlama ya da DeepMind'ın AlphaCode modeliyle aynı ligde değerlendirildiğini, hatta daha istikrarlı bağlam takibi ve hata ayıklama hassasiyetiyle öne çıktığını belirtmektedir. Bu uzmanlaşma, kullanıcıların belirli ve karmaşık teknik problemlere çözüm ararken DeepSeek'i daha tatmin edici bulmasını açıklayabilir ve Manik (2025) tarafından da işaret edildiği gibi, dar amaçlı yüksek doğruluk gerektiren görevlerde tercih edilmesine yol açabilir. DeepSeek'in bu teknik üstünlüğü, BytePlus (2025) tarafından da DeepSeek-Coder-V2 modelinin kod üretimi ve analizi alanlarındaki GPT-4 Turbo ile karşılaştırılabilir yüksek doğruluk oranlarıyla desteklenmektedir.

ChatGPT ise özellikle içerik üretimi, profesyonel yazarlık ve eğitim senaryolarında daha geniş bir kullanım alanına sahiptir. Bu, Hariri (2023) ve Ray (2023) tarafından da ifade edilen, ChatGPT'nin çok yönlü doğasına dair literatür bulgularını desteklemektedir. ChatGPT'nin doğal dili anlama konusundaki yetkinliği ve insan benzeri yanıtlar üretebilme kapasitesi (Bahrini vd., 2023), onu eğitim materyali üretme, sunum hazırlama ve yazma becerilerini geliştirme gibi alanlarda (Zheldibayeva, 2024) popüler kılmaktadır. Ancak, bu geniş kullanım alanı, bazen teknik derinlik gerektiren konularda DeepSeek kadar odaklanmış bir performans sunamamasına neden olabilmektedir.

ChatGPT kullanıcılarının daha kısa oturum süresine rağmen daha yüksek memnuniyet bildirdiği bazı oturumlar, yapay zekâ ile olan etkileşimin sadece teknik doğrulukla değil, aynı zamanda iletişim tarzı, dil doğallığı ve bağlamsal uyumla da şekillendiğini ortaya koymaktadır. Bu durum, Nazir ve Wang (2023) tarafından da vurgulandığı üzere, üretken yapay zekâ sistemlerinin "anlamlı diyalog" kurma kapasitesinin kullanıcı deneyiminde belirleyici olduğunu göstermektedir. Skjuve vd. (2024) de kullanıcıların ChatGPT gibi araçları kullanma motivasyonlarının çeşitliliğine işaret ederek, algılanan değer ve kullanım kolaylığının önemini vurgular.

4.2.Kullanıcı Etkileşimi ve Cihaz Türleri Üzerindeki Etkiler

Kullanıcıların platformlara erişim için kullandıkları cihaz türleri de deneyim puanları üzerinde anlamlı etkiler yaratmıştır. Masaüstü/dizüstü bilgisayarlarda gerçekleşen etkileşimlerde kullanıcı deneyimi puanlarının daha yüksek olması, bu cihazların hem işlem gücü hem de ekran boyutu açısından daha uygun bir kullanıcı ortamı sunmasından kaynaklanabilir. Bu bulgu, Qawqzeh (2024) ve Sharples (2023) tarafından belirtilen "pedagojik etkileşim ortamının fiziksel yapısının öğrenme verimliliğine etkisi" kavramıyla örtüşmektedir; daha geniş ekranlar ve klavye kullanımı, özellikle karmaşık görevlerde ve içerik üretiminde daha verimli bir etkileşim sağlayabilir. Mobil cihazlar üzerinden yapılan etkileşimlerde ise deneyim puanlarının daha değişkenlik göstermesi, arayüz farklılıkları, ekran boyutu kısıtlamaları ve etkileşim süresi ile açıklanabilir. Bu durum, Adıgüzel vd. (2023) tarafından belirtilen yapay zekanın eğitimde kişiselleştirilmiş deneyimler sunma potansiyelinin, kullanılan cihazın ergonomisi ve işlevselliği ile yakından ilişkili olduğunu düşündürmektedir.

4.3.Regresyon ve Sınıflandırma Modelleri Işığında Derinlemesine Değerlendirme

Çalışmada uygulanan regresyon analizleri, kullanıcı derecelendirmelerinin düşük R^2 değerleri ile modellenebildiğini ortaya koymuştur. Bu durum, kullanıcı memnuniyetini etkileyen faktörlerin yalnızca oturum süresi, cihaz türü veya yanıt doğruluğu gibi gözlemlenebilir değişkenlerle sınırlı olmadığını, aksine bireysel kullanıcı özellikleri, duygusal bağlam, önceki etkileşim geçmişi gibi daha derin psikometrik faktörlerin de önemli olduğunu düşündürmektedir. Al-Abdullatif ve Alsubaie (2024), kullanıcı deneyiminin çok boyutlu olduğunu ve kişisel beklenti düzeylerinin, yapay zekâ okuryazarlığının ve algılanan değerin memnuniyet üzerinde belirleyici rol oynadığını vurgulamaktadır. Bu bulgu, yapay zekâ etkileşimlerinde insan faktörünün karmaşıklığını ve standart metriklerle ölçülmesinin zorluğunu bir kez daha teyit etmektedir.

Sınıflandırma modelleri özelinde, "Excellent" olarak etiketlenen kullanıcı geri bildirimlerinin görece başarıyla tahmin edilebilmesine karşın, "Fair" ve "Good" sınıflarının ayırt edilememesi, kullanıcı değerlendirme ölçeklerinin sübjektifliğini ve potansiyel dengesiz veri dağılımını ortaya koymaktadır. Kullanıcıların "orta" ve "iyi" gibi derecelendirmeleri farklı yorumlayabileceği ve bu durumun model performansını etkileyebileceği düşünülmektedir. Bu bağlamda, kullanıcıların değerlendirme yaparken hangi kriterleri temel aldıkları ve bu kriterlerin platformdan platforma nasıl değiştiği daha derinlemesine incelenmelidir.

4.4.Kullanıcı Sadakati, Terk Etme Oranı ve Platform Performansı

Çalışmanın dikkat çeken sonuçlarından biri, DeepSeek'in genel olarak daha düşük terk etme oranlarına sahip olmasıdır. Bu durum, kullanıcıların DeepSeek ile kurdukları etkileşimin sürdürülebilirliğini ve platformun belirli kullanıcı ihtiyaçlarını karşılama başarısını göstermektedir. Mogavi vd. (2023), yapay zekâ destekli dijital araçların

sürekli kullanımının, sadece işlevsel başarı değil, aynı zamanda psikolojik tatmin ile de ilişkili olduğunu belirtmektedir. Bu bağlamda, DeepSeek'in daha teknik, hedef odaklı ve kodlama görevlerine dayalı yapısı, profesyonel kullanıcılar için daha cazip ve tatmin edici bir seçenek olarak değerlendirilebilir ve bu da daha düşük terk oranlarını açıklayabilir. Ayrıca, DeepSeek'in aktif kullanıcı sayısının ve ortalama oturum süresinin ChatGPT'ye kıyasla daha yüksek olması, bu platformun kullanıcıları daha uzun süre etkileşimde tutabildiğini veya daha karmaşık ve zaman alıcı görevler için tercih edildiğini göstermektedir.

Öte yandan ChatGPT'nin özellikle yıl sonlarına doğru kullanıcı bağlılığında artış göstermesi, eğitim-öğretim dönemlerinin etkisiyle ve genel amaçlı kullanımının yaygınlığıyla açıklanabilir. Hariri (2023), ChatGPT'nin akademik takvimle uyumlu kullanım örüntülerine sahip olduğunu ifade etmiş ve bu modeli öğrenci destek aracı olarak tanımlamıştır. Fırat (2023) da üniversitelerde yapay zekâ sohbet robotlarının öğrenci katılımını artırma potansiyeline dikkat çekmektedir. Bu durum, ChatGPT'nin geniş bir kullanıcı kitlesine hitap etme ve çeşitli ihtiyaçlara yanıt verme esnekliğinden kaynaklanıyor olabilir.

4.5.Etik ve Veri Koruma Açısından Değerlendirme

Her iki modelin de kullanıcı verilerini işlerken önemli etik sorumluluklar taşıdığı görülmektedir. ChatGPT için belirtilen en önemli sorunlardan biri, "hallucination" adı verilen yanlış veya uydurma bilgi üretme eğilimidir; bu durum Alkaissi & McFarlane (2023) tarafından da vurgulanmıştır. Bu sorun özellikle akademik ya da tıbbi içeriklerde ciddi sonuçlar doğurabilir (Bahrini vd., 2023; Spennemann, 2025). Ayrıca, eğitim verilerindeki önyargıları yansıtma riski (Rozado, 2023) ve fikri mülkiyet hakları (Zirpoli, 2023; Al-Busaidi vd., 2024) ChatGPT ile ilişkili önemli etik kaygılardır. DeepSeek'in eğitim verilerinin tescilli olması (Sapkota vd., 2025), modelin potansiyel önyargılarını (Witness AI, 2025) ve eğitim seti kalitesini değerlendirmeyi güçleştirmektedir. "Açık ağırlık" politikasının (DeepSeek-R1, t.y.) tam anlamıyla "açık kaynak" anlamına gelmediği (CTOL Digital, 2025) ve bu durumun yeniden üretilebilirlik ve şeffaflık açısından sınırlamalar getirdiği belirtilmelidir. DeepSeek'in de, diğer dil modelleri gibi, zaman zaman yanlış bilgi üretme (hallucination) riski (Vectara, 2025) ve kod üretme yeteneği nedeniyle kötü amaçlı yazılım geliştirme gibi kötüye kullanım potansiyeli (SecurityWeek, 2025) bulunmaktadır. Kullanıcı verilerinin güvenliğine ilişkin olarak, DeepSeek ve ChatGPT platformlarının şeffaflık düzeylerinin farklılık gösterdiği görülmektedir. Özellikle Avrupa Birliği Genel Veri Koruma Tüzüğü (GDPR) ve Türkiye'deki Kişisel Verilerin Korunması Kanunu (KVKK) gibi düzenlemelere uyum düzeylerinin platformlar bazında açıklığa kavuşturulması gerekmektedir, nitekim İtalya'nın OpenAI'ye gizlilik kuralları ihlali nedeniyle ceza vermesi (Reuters, 2024) bu konunun hassasiyetini göstermektedir. Kullanıcıların hassas bilgilerini bu platformlara yüklemesi, özellikle sağlık verileri gibi özel içeriklerde ciddi gizlilik ihlali potansiyeli taşımaktadır (Borgesius, 2018). Google'ın Gemini modeli için yayımladığı gizlilik uyarısı (Search Engine Journal, 2024) ve ICO (2023) tarafından belirtilen yapay zekâ ve veri koruma rehberliği, bu alandaki genel endişeleri yansıtmaktadır. Kullanıcı sohbet geçmişinin nasıl saklandığı, hangi amaçlarla analiz edildiği ve üçüncü taraflarla paylaşılıp paylaşılmadığı gibi konular, etik sorumluluklar açısından kritik önemdedir.

4.6.Araştırmanın Sınırlılıkları

Bu çalışmanın bazı metodolojik sınırlılıkları bulunmaktadır. İlk olarak, kullanılan veri seti sentetik olarak oluşturulmuş olup, kullanıcı davranışlarını tamamen gerçekçi biçimde yansıtmayabilir. Bu nedenle, çalışmanın dış

geçerliği (external validity) sınırlı olabilir. İkinci olarak, zaman içinde yapay zekâ modellerinde gerçekleşen güncellemeler (örneğin, GPT-4'ten GPT-4o'ya geçiş gibi), model performansını etkilemekte ve elde edilen sonuçların geçerlilik süresini sınırlamaktadır. Çalışmada kullanılan modellerin versiyonları (GPT-4-turbo ve DeepSeek-Chat 1.5) en son çıkan modeller olmayabilir ve bu da güncel karşılaştırmalar açısından bir kısıtlılık oluşturabilir. Ayrıca, çalışmada kullanıcıların demografik özelliklerine, kültürel geçmişlerine veya yapay zekâ okuryazarlık düzeylerine ilişkin detaylı veri bulunmaması, analizlerin daha özelleştirilmiş ve derinlemesine çıkarımlar sunmasını zorlaştırmıştır. Bu tür faktörlerin kullanıcı deneyimi ve memnuniyeti üzerinde önemli etkileri olabileceği (Al-Abdullatif ve Alsubaie, 2024) bilinmektedir. Son olarak, "hallucination" veya önyargı gibi etik sorunların tespiti ve analizi için kullanılan metriklerin bu çalışmada sınırlı olması, bu konulardaki değerlendirmelerin derinliğini kısıtlamıştır.

4.7.Sonuçlar ve Öneriler

Bu çalışmadan elde edilen bulgular, üretken yapay zekâ platformları olan ChatGPT ve DeepSeek'in performanslarının ve kullanıcı etkileşimlerinin; kullanım bağlamı, görevin türü ve erişim sağlanan cihazın özelliklerine bağlı olarak anlamlı farklılıklar gösterdiğini kapsamlı bir şekilde ortaya koymaktadır. Her iki platform da kendi içinde belirli alanlarda üstünlükler sergilemekle birlikte, etik ve veri gizliliği gibi konular tüm yapay zekâ sistemlerinin geliştirilmesinde temel öncelikler arasında yer almaktadır.

Genel bulgular doğrultusunda, ChatGPT'nin içerik üretimi, akademik yazım desteği, eğitim uygulamaları ve sosyal etkileşim gibi alanlarda geniş kullanım alanına sahip olduğu ve yüksek kullanıcı memnuniyeti sağladığı görülmüştür. Platformun güçlü doğal dil işleme yetenekleri ve esnek yapısı, farklı kullanıcı profilleri için cazip bir seçenek haline gelmesine katkı sağlamaktadır. Öte yandan, DeepSeek'in özellikle teknik ve uzmanlık gerektiren görevlerde —örneğin kodlama, hata ayıklama, algoritma açıklamaları— optimize edilmiş mimarisi sayesinde daha yüksek doğruluk oranları sunduğu ve bu bağlamda kullanıcı sadakatini artırdığı tespit edilmiştir. DeepSeek, aktif kullanıcı sayısı ve ortalama oturum süresi gibi metriklerde de dikkat çekici üstünlükler göstermiştir.

Kullanıcı deneyimini etkileyen önemli değişkenlerden biri olan cihaz türü de çalışmada öne çıkan bir başka bulgudur. Masaüstü ve dizüstü bilgisayar kullanıcılarının, daha gelişmiş arayüz etkileşimi ve işlem kapasitesi nedeniyle daha pozitif deneyimler yaşadığı rapor edilmiştir. Buna karşılık, mobil cihazlarda kullanıcı deneyimi puanlarının daha yüksek oranda değişkenlik gösterdiği gözlemlenmiştir.

Çalışmada yürütülen regresyon analizleri, kullanıcı memnuniyetini etkileyen çok sayıda değişken bulunduğunu ve mevcut bağımsız değişkenlerin bu varyansın yalnızca belirli bir kısmını açıklayabildiğini ortaya koymuştur. Bu durum, kullanıcı memnuniyetinin öznel doğasını ve kişisel beklentiler gibi ölçülmesi zor faktörlerin etkisini vurgulamaktadır.

Etik hususlar ve veri gizliliği hem ChatGPT hem de DeepSeek için kritik endişe alanları olarak öne çıkmaktadır. Model şeffaflığı, önyargıların azaltılması, "hallucination" riski ve zararlı içerik üretimi gibi konular, yapay zekâ teknolojilerinin güvenilirliğini ve toplumsal kabulünü doğrudan etkilemektedir. Bu bağlamda platformların etik kurallara uyumu ve yasal düzenlemelere (örneğin GDPR, KVKK) uygunluğu, kullanıcı güveni açısından büyük önem taşımaktadır.

Bu bulgular doğrultusunda geliştirilen öneriler ise üç temel paydaş grubu etrafında şekillenmiştir. İlk olarak, platform geliştiricileri için; eğitim veri setlerinin içeriği, önyargılar ve bu önyargıların azaltılmasına yönelik stratejiler konusunda daha şeffaf olunması gerekmektedir. Kullanıcı verilerinin işlenmesi, saklanması ve korunmasına ilişkin politikalar açık biçimde sunulmalı ve veri koruma yasalarına tam uyum sağlanmalıdır. Ayrıca, farklı cihazlar için optimize edilmiş arayüzler geliştirilerek kullanıcı geri bildirimleri sistematik şekilde değerlendirilmelidir. Sürekli Ar-Ge çalışmalarıyla etik risklerin (örneğin hallucination, kötüye kullanım) önlenmesine yönelik mekanizmalar güçlendirilmelidir. DeepSeek örneğinde görüldüğü üzere, genel amaçlı ve uzmanlaşmış modeller arasında bir denge kurulması da önemli bir strateji olacaktır.

İkinci olarak, eğitim kurumları ve profesyonel kullanıcılar için; yapay zekâ okuryazarlığını artırmak amacıyla kullanıcıların farklı araçların kapasite ve sınırlılıklarını anlayabilecekleri rehberler hazırlanmalı, eleştirel düşünme ve doğrulama becerileri teşvik edilmelidir. Kullanım bağlamına uygun araç seçiminde farkındalık oluşturulmalı; örneğin yaratıcı içerik üretimi için ChatGPT, teknik görevler için ise DeepSeek tercih edilmelidir. Ayrıca, akademik ve profesyonel alanlarda yapay zekânın etik kullanımına ilişkin ilkeler açık biçimde tanımlanmalı ve bu ilkelere uyum teşvik edilmelidir.

Son olarak, araştırmacılar için; demografik faktörlerin etkisini göz önünde bulunduran daha kapsamlı ve karşılaştırmalı çalışmalar yapılması önerilmektedir. Kullanıcıların üretken yapay zekâ ile olan uzun vadeli etkileşimlerinin; öğrenme, problem çözme ve yaratıcılık üzerindeki etkileri derinlemesine araştırılmalıdır. Önyargıların tespiti, hallucination probleminin çözümü ve yapay zekâ güvenliği konularında disiplinlerarası çalışmalar teşvik edilmeli; sentetik veri setlerinin ötesine geçilerek gerçek dünya senaryolarına dayalı risk analizleri yapılmalıdır.

Teşekkür ve Bilgilendirme / Acknowledgements

Makale, kısmen veya tamamen yayınlanmaması şartıyla, bir proje veya tez olarak devam eden bir bildiri olarak sunuluyorsa, bu bölümde belirtilmelidir. Makale bir araştırma kurumu veya bir fon tarafından destekleniyorsa, vakfın adı, proje numarası ve tamamlanma tarihi burada belirtilmelidir. İstenirse, makale bağlamında bir kişi veya vakıf için takdirler burada belirtilmelidir. / If the article is submitted as a proceeding, on the condition of not being partially or fully published, a project or dissertation, it should be stated in this section. If the article is supported by a research institution or a fund, the name of the foundation, project number and completion date should be stated here. If desired, appreciations to a person or a foundation within the context of the article should be stated here.

Yayın Etiği Bildirimi / Research Ethics

Yazar araştırmanın etik dışı bir sorunu olmadığını, araştırma ve yayın etiği konusunu gözlemlediğini beyan etmektedir. / The author declares that the research has no unethical problem and observes the research and publication ethics.

Araştırmacıların Katkı Oranı / Contribution Rate of Researchers

Çalışmanın her aşamasına tüm yazarlar eşit derecede katkı sunmuştur. / All authors contributed equally to every stage of the study.

Çıkar Çatışması / Conflict of Interest

Çalışmada herhangi bir çıkar çatışması bulunmamaktadır. / The study has no conflict of interest.

Fon Bilgileri / Funding

Bu çalışmada herhangi bir fon kullanılmamıştır. / There is no funding for this study.

Etik Kurul Onayı / The Ethical Commitee Approval

Etik kurul kararı: Bu araştırmada, tüm araştırmacılara açık, uluslararası veri tabanında yer alan veriler kullanıldığından etik kurul kararı gerektirmemektedir. / The Ethical Committee Approval: This research does not require an ethics committee decision, since data in an international database open to all researchers are used.

Kaynakça / References

Acemoglu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker ai. https://shapingwork.mit.edu/wp-content/uploads/2023/09/Pro-Worker-AI-Policy-Memo.pdf

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3).

Al-Abdullatif, A. M., & Alsubaie, M. A. (2024). ChatGPT in learning: Assessing students' use intentions through the lens of perceived value and the influence of AI literacy. *Behavioral Sciences*, *14*(9), 845.

Al-Busaidi, A. S., Raman, R., Hughes, L., Albashrawi, M. A., Malik, T., Dwivedi, Y. K., ... & Walton, P. (2024). Redefining boundaries in innovation and knowledge domains: Investigating the impact of generative artificial intelligence on copyright and intellectual property rights. *Journal of Innovation & Knowledge*, *9*(4), 100630.

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, *15*(2).

Artetxe, M., Bhosale, S., Goyal, N., Mihaylov, T., Ott, M., Shleifer, S., & Stoyanov, V. (2021). Efficient large scale language modeling with mixtures of experts. https://arxiv.org/pdf/2112.10684

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, opportunities, and threats. *In 2023 Systems and Information Engineering Design Symposium (SIEDS)*, 274-279.

Bell, R., & Bell, H. (2023). Entrepreneurship education in the era of generative artificial intelligence. *Entrepreneurship Education*, *6*(3), 229–244. https://doi.org/10.1007/s41959-023-00099-x

Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. *Council of Europe, Directorate General of Democracy*.

BytePlus. (2025). DeepSeek-Coder-V2: Revolutionizing AI-powered code generation. https://www.byteplus.com/en/topic/375561?title=deepseek-coder-v2-revolutionizing-ai-powered-code-generation

Chein, J., Martinez, S., & Barone, A. (2024). Can human intelligence safeguard against artificial intelligence? Exploring individual differences in the discernment of human from AI texts. *Research Square*, rs-3.

CTOL Digital. (2025). Is DeepSeek Truly Open Source or Just Following Industry Norms? https://www.ctol.digital/news/is-deepseek-truly-open-source-or-following-industry-norms/

Dataloop. (t.y.). DeepSeek Coder V2 Instruct · Models. https://dataloop.ai/library/model/deepseek-ai_deepseek-coder-v2-instruct/

DeepSeek R1: DeepSeek-V3. (t.y.). https://deepseeksr1.com/v3/

DeepSeek-AI. (2024). DeepSeek-V3 Technical Report. https://arxiv.org/pdf/2412.19437

DeepSeek-Coder. (t.y.). DeepSeek Coder. https://github.com/deepseek-ai/DeepSeek-Coder

DeepSeek-R1. (t.y.). DeepSeek-R1 Training Overview. https://github.com/deepseek-ai/DeepSeek-R1

DeepSeek-V3. (t.y.). DeepSeek-V3. https://github.com/deepseek-ai/DeepSeek-V3

Du, N., Huang, Y., Dai, A. M., Tong, S., Lepikhin, D., Xu, Y., ... & Cui, C. (2022, June). Glam: Efficient scaling of language models with mixture-of-experts. https://arxiv.org/pdf/2112.06905

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., & Wright, R. (2023). Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*.

Eren, F., Dülek, L. N., Uraz, Ö. A., Kuşcu, B., & Sakallı, M. (2024). Yapay zeka endeksi kapsamında ülke bazlı yapay zeka politika stratejilerinin etkinliği: Performans ve verimlilik değerlendirmesi incelenmesi. *Bilgi ve İletişim Teknolojileri Dergisi*, *6*(2), 113-148.

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, *6*(1), 57-63.

Guo, D., Zhu, Q., Yang, D., Xie, Z., Dong, K., Zhang, W., ... & Liang, W. (2024). DeepSeek-Coder: When the large language model meets programming -- The rise of code intelligence. https://arxiv.org/pdf/2401.14196

Gupta, M. (2025). RLHF (OpenAI) vs Simple RL (DeepSeek). Medium. https://medium.com/data-science-in-your-pocket/rlhf-openai-vs-simple-rl-deepseek-93054cb509a4

Hariri, W. (2023). Unlocking the potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing. https://arxiv.org/abs/2304.02017

IBM. (2025a). DeepSeek: sorting through the hype. https://www.ibm.com/think/topics/deepseek

IBM. (2025b). DeepSeek's reasoning AI shows power of small models, efficiently. https://www.ibm.com/think/news/deepseek-r1-ai

Information Commissioner's Office (ICO). (2023). Guidance on AI and data protection. ICO.

Liu, H., Zhou, Y., Li, M., Yuan, C., & Tan, C. (2024). Literature meets data: A synergistic approach to hypothesis generation. https://arxiv.org/pdf/2410.17309

Manik, M. M. H. (2025). ChatGPT vs. DeepSeek: A comparative study on AI-Based code generation. https://arxiv.org/pdf/2502.18467

Markov, T., Zhang, C., Agarwal, S., Eloundou, T., Lee, T., Adler, S., Jiang, A., & Weng, L. (2023). New and improved content moderation tooling. https://openai.com/blog/new-and-improved-content-moderation-tooling/

Mesko, B. (2017). The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development*, *2*(5), 239-241.

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., ... & Hui, P. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, *2*(1).

Nazir, A., & Wang, Z. (2023). A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. *Meta-radiology*.

Neptune.ai (2025). Reinforcement learning from human feedback for LLMs. https://neptune.ai/blog/reinforcement-learning-from-human-feedback-for-llms

OpenAI. (2023). GPT-4 Technical Report. https://arxiv.org/pdf/2303.08774

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. https://arxiv.org/pdf/2203.02155

Qawqzeh, Y. (2024). Exploring the influence of student interaction with ChatGPT on critical thinking, problem solving, and creativity. *International Journal of Information and Education Technology*, *14*(4), 596-601.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154.

Reuters. (2024). Italy fines OpenAI 15 million euros over privacy rules breach. Reuters. https://www.reuters.com/technology/italy-fines-openai-15-million-euros-over-privacy-rules-breach-2024-12-20/

Rivas, P., & Zhao, L. (2023). Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot technology. *AI*, *4*(2), 375-384.

Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences*, *12*(3), 148.

Sapkota, R., Raza, S., & Karkee, M. (2025). Comprehensive analysis of transparency and accessibility of ChatGPT, DeepSeek and other Sota large language models. https://arxiv.org/pdf/2502.18505

Search Engine Journal. (2024). Google gemini warning: Don't share confidential information. https://www.searchenginejournal.com/google-gemini-privacy-warning/507818/

Security Magazine. (2025). Dangers of DeepSeek's privacy policy: Data risks in the age of AI. https://www.securitymagazine.com/articles/101374-dangers-of-deepseeks-privacy-policy-data-risks-in-the-age-of-ai

SecurityWeek. (2025). DeepSeek's salware-generation capabilities put to test. https://www.securityweek.com/deepseeks-malware-generation-capabilities-put-to-test/

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159-167.

Shirani, M. (2025). Comparing the performance of ChatGPT 4o, DeepSeek R1, and Gemini 2 Pro in answering fixed prosthodontics questions over time. *The Journal of Prosthetic Dentistry*.

Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday*.

Spennemann, D. H. (2025). The origins and veracity of references 'Cited'by generative artificial intelligence applications: Implications for the quality of responses. *Publications*, *13*(1), 12.

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove.

Şentürk, Ö. (2023). İç denetim faaliyetlerinde yapay zekadan beklentiler: chatgpt uygulaması örneği. *TIDE AcademIA Research*, *4*(2), 51-82.

Tecuci, G. (2012). Artificial intelligence. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 168–180. https://doi.org/10.1002/wics.200

Vectara, (2025). DeepSeek-R1 hallucinates more than DeepSeek-V3. https://www.vectara.com/blog/deepseek-r1-hallucinates-more-than-deepseek-v3

Wang, K., Ruan, Q., Zhang, X., Fu, C., & Duan, B. (2024). Pre-service teachers' GenAI anxiety, technology self-efficacy, and TPACK: Their structural relations with behavioral intention to design GenAI-assisted teaching. *Behavioral sciences*, *14*(5), 373.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. https://arxiv.org/abs/2112.04359

Witness AI. (2025). Open R1 vs DeepSeek: security implications of open vs. closed data. https://witness.ai/blog-open-r1-vs-deepseek-security-implications-of-open-vs-closed-data/

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. https://arxiv.org/abs/2303.18223

Zheldibayeva, R. (2024). The impact of AI and peer feedback on research writing skills: a study using the CGScholar platform among Kazakhstani scholars. https://arxiv.org/pdf/2503.05820

Zirpoli, C. T. (2023). Generative artificial intelligence and copyright law.