



Examining Humanization Methods in AI Applications and Evaluating Them from an Ethical Perspective

Meriç TÜRKYILMAZ ^{1*} , Yakup KUTLU ¹ 

¹ İskenderun Teknik Üniversitesi, Bilgisayar Mühendisliği, İskenderun, Hatay, Türkiye.
meric.turkyilmaz.mdbf@iste.edu.tr, yakup.kutlu@iste.edu.tr

ABSTRACT

This study discusses how detectable AI-generated text is when modified using humanization methods and explores the ethical implications of this practice. First, the working principles of AI detectors and the criteria used are explained, followed by an examination of the techniques employed during humanization. Experiments conducted using Wikipedia and Kaggle data showed that AI-generated texts could be distinguished with 99.7% accuracy. The results indicate that such methods do not offer a permanent solution and that the real issue is the need to observe ethical principles.

ARTICLE INFO

Received 17.10.2025
Accepted 15.11.2025
Available online 20.11.2025
Publication Date 20.11.2025

Keywords:

Artificial intelligence,
LLM, ML, ethics

Distributed Under CC-BY 4.0



*Corresponding Author: Meriç TÜRKYILMAZ, E-mail: meric.turkyilmaz.mdbf@iste.edu.tr

Yapay Zeka Uygulamalarında İnsanlaştırma Yöntemlerinin İncelenmesi ve Etiksel Açıdan Değerlendirilmesi

ÖZET

Bu çalışmada yapay zeka tarafından üretilen metinlerin insanlaştırma yöntemi ile değiştirildiğinde ne kadar tespit edilebilir ve bunun etik sonuçları tartışılmıştır. Öncelikle yapay zeka dedektörlerinin çalışma mantığı ve kullanılan ölçütler açıklanmış ve ardında insanlaştırma yaparken kullanılan teknikler incelenmiştir. Wikipedia ve Kaggle verileri kullanılarak yapılan deneylerde, %99,7 oranında bir doğrulukla yapay zeka tarafından oluşturulan metinlerin ayırt edilebildiğini göstermiştir. Sonuçlara göre bu tür yöntemlerin kalıcı bir çözüm sunmadığını ve asıl sorunun etik ilkelerin gözetilmesi gerektiğini ortaya koymaktadır.

MAKALE BİLGİSİ

Keywords:

Yapay zeka, LLM, ML, etik

Distributed Under CC-BY 4.0



GİRİŞ

Doğal dil işleme modellerinin gelişimiyle bir çak kurum kendi eğitilmiş modellerini test edilmesi açısından web tabanlı servise açmışlardır. Araştırmacılar ChatGPT, Gemini, Copilot, Perplexity and dahası modellerin performans karşılaştırmasını veya uygulamalara karşı gösterdikleri cevapları analiz etmiş ve karşılaştırmışlardır (Abouelyazid, 2024; Atayolu ve Kutlu, 2024; Öncü ve Kutlu, 2024). Gelişmiş zeka modellerinin insan hayatına entegre olmasıyla beraber makaleler ve araştırmalar gibi kişinin özgünlüğünü ve yaratıcılığını koyarak ortaya çıkarması gereken durumda yapay zekadan destek alarak yada ona yazdırarak akademik dürüstlük ve etikliği yok saymak ve bunun anlaşılması için yapay zeka modeline yazdırılan metni insanlaştırma yoluna gidiliyor. Bir çok site yapay zeka modellerinin oluşturduğu metinleri insanlaştırma (humanized) yaparak yapay zeka detektörü tarafından yazıldığı anlaşılamayacak hale getirdiğini savunuyor. Bu durum insanlaştırma araçları ve yapay zeka dedektörleri arasında yarış ortamı doğurmuştur. Yapay zeka modellerinin insanlaştırma kullanarak textlerin değiştirilmesi gerçekten işe yarıyor mu? Bu yapılan insanlaştırma işlemi akademik etiklik tanımı içerisinde mi? Bu çalışmada yapay zeka dedektörlerini temel işleyişini, insanlaştırma yöntemlerinin ve etik açıdan doğurduğu problemleri ele almaktadır. Ayrıca deneysel veriler ile insanlaştırma yapılmasının etikliği sorgulanmış ve akademik dürüstlük bağlamında konunun ne şekilde değerlendirilmesi gerektiği tartışılmıştır. Bu şekilde yapay zekanın teknik ve etik boyutu incelenmesi amaçlanmıştır.

YÖNTEM

Öncelikle ai detector dediğimiz yapının nasıl çalıştığından bahselelim. Quillbot gibi ai dedektörleri makine öğrenimi ve doğal dil işleme ile çalışır ve dört temel kavramı baz alarak kontrol sağlar bunlar sınıflandırma, gömme, şaşkınlık, patlama olarak sınıflandırılmaktadır. Sınıflandırma da girdi verilerini önceden tanımlı sınıflara ayıran bir makine öğrenim modelidir. Yapay zeka metin dedektörleri için iki sınıftan oluşmaktadır. “Yapay zeka tarafından oluşturulan” ve “İnsan tarafından oluşturulan”. Sınıflandırıcı, farklı kelime kullanım sıklığı, cümlelerin uzunluğu ve karmaşıklığı gibi özellikleri analiz eder. Kümeler arası farklılıklar belirlenir ve sınıflar arası bir sınır oluşturulur. Sınıflandırıcıya metin verildiğinde yukarıda ki özellikleri değerlendirir ve aynı kalıpları arar. Bulunan kalıplara göre metnin yapay zeka tarafından mı yoksa insan tarafından mı yazıldığına karar verir. Gömme(Embedding) ise insan dilini bilgisayarlar anlayamaz. Sayısal ifadeye dönüştürerek anlamalarını sağlarız. Bu sayısal gösterime vektör adı verilir. Kelimenin anlamını, yapısını ve bağlamını makinelerin anlayacağı şekle getirir. Kelimeleri çok boyutlu bir alanda eşleyerek yapay zekanın kelimeler arasındaki ilişkiyi ve kalıpların tespitini sağlar. Şaşkınlık, metnin şaşırtma veya mantıksız gelme olasılığının ne kadar olduğunu ve bu metnin mantık sınırları içinde devam etmesi veya doğallığını koruyamaması durumunu inceler. Patlama, cümlelerin zenginliğini ve uzunluğunu analiz eder. Metinde cümlelerin yapısı uzun ve çeşitliliğe sahipse yüksek patlama, cümle kısa ve fazla çeşitlilik göstermiyorsa düşük patlama özelliğine sahiptir (Santoro, 2024).

Yapay zeka tarafından oluşturulan metni yapay zeka dedektörlerine yakalanmayacak şekilde değiştirmek için bazı algoritmaları kullanarak insanlaştırma yapan siteler ve modeller mevcuttur. Bunlar kural tabanlı eşanlamlı değiştirme, Yapay zeka yazım yapısını bozmak, stili basitleştirme, anlamı koruma. Kural tabanlı eşleştirme de cümle genel olarak değişmez belirli kelime ve ifadeleri değiştirir. Örneğin “O sahilde güneşleniyor” cümlesini “O kumsalda güneşleniyor” olarak değiştirir. Yapay zeka yazım yapısını bozmak için bu yöntemde cümleleri ayırıp tekrar birleştirebilir. Konuşma kalıpları ekleyebilir veya bir insan yazarın kararlarını taklit etmek için bilgileri yeniden sıralayabilir. Stili basitleştirme de yazılan metni daha düşük bir okuma seviyesine indirgeyebilir. Resmi bir yazıyı ilk okul öğrencisinin yazabileceği bir yazıya dönüştürür. Bu şekilde insanlaştırma işlemi gerçekleştirir. Anlamı koruma da değiştirilen metnin anlamını kaybetmemesi gerekir bunun için ek koşullar sağlar. İsim yerine isim ve fiil yerine fiil kullanmak, eski ve yeni

cümlelerin aynı anlam ifade ettiğini doğrulamak için cümle kodlayıcıları gibi teknikler ile cümlelerin anlamının korunması sağlanır (Buchert, 2025).

Doğal Dil İşleme denen NLP (Natural Language Processing), algoritmaların metni anlamasını sağlar. Çünkü çoğu algoritma matematikselidir ve etkili bir şekilde çalışması için metnin sayı olarak temsil edilmesine ihtiyaç duyar. Bunun için de oyuna TF IDF girer. TF-IDF (Term Frequency - Inverse Document Frequency); terim frekansı ve ters belge frekansı anlamına gelir. Metni anlamlı sayılar olarak göstermenin bir yoludur, vektör gösterimi olarak da bilinir. Basit matematiksel bir model olan liner regresyon sınıflandırma modeli elde edilen TFIDF vektörlerini sınıflandırmak için kullanılmıştır ("TF IDF Nedir", 2021).

BULGULAR

Bu çalışmada Wikipedia-API kullanılarak bir corpus oluşturulmuştur. Bunun yanında kaggle hazır açık erişimli insan ve yapay zekanın oluşturduğu 2 grup olan bir veri seti kullanılmıştır. Bu veri setleri ile TFIDF dönüşümleri yapılarak makine öğrenmesi modelleri sınıflandırılmıştır. Daha sonra yapay zeka modelleri ile insanlaştırma gerçekleştirme yapılmış ve elde edilen yeni metinler yine TFIDF dönüşümü yapılarak sınıflandırma sonuçları incelenmiştir (Bhagya, 2025.; Goldsborough, 2025).

Piyasada birçok yapay zeka metin insanlaştırma aracı var. Bu araçlar kullanılarak yapay zekaya üretilen metinlerin yapay zeka dedektörlerinden geçmesini sağlayacağını vaat ediyorlar. Bu yaptığımız örnekde yeni bir corpus oluşturduk. 4700 adet wikipedia'dan alınmış insan tarafından oluşturulan data, kullandığımız 1500 adet de yapay zeka tarafından oluşturulmuş ve 1500 adet de insan tarafından oluşturulmuş data kullandık.

Yapay zeka tarafından oluşturulan dataların bir kısmı ile ChatGPT ile farklı promptlarla ve farklı sitelerden insanlaştırma yapıldı: Humanize AI (Humanize AI, 2025), QuillBot (QuillBot, 2025), Grammarly (Grammarly, 2025), AI Humanize (AI Humanize, 2025), WriteHuman (WriteHuman, 2025), ChatGPT (ChatGPT, 2025).

Nitekin bizim kullandığımız model ile oluşturulan bütün datalara baktığımız zaman sınıflandırma performansı %99.7 oranında başarımlı elde etti. Yani insanlaştırma yöntemi hiçbir işe yaramadığını görebiliyoruz (Bhagya, 2025).

Ancak bu yaptığımız yöntemleri yada oluşturduğumuz dataları farklı finder yada Ai kontrol sitelerinden bazılarında geçebildiğini, belli oranda yapay zeka tarafından yazıldığını yada yakaladığını görebiliyoruz.

Buda şunu gösterir. Model her ne kadar insanlaştırma için geliştiriliyor olsada bunları yakalamak içinde farklı modeller geliştirilebilir. Daha gelişmiş modeller geliştirilebilir ve bunları yakalayabilir. Ve böylece insanlar bununla ilgili onlarca site ve insanlaştırma modeli geliştirilmiş olsada onun üst versiyonu modeller veya siteler tarafından yakalamamız mümkün bunu görüyoruz. Nitekim sonuçta bunu yapsakda yapmasakda aslında burada önemli olan etik kurallardır.

SONUÇ VE TARTIŞMA

Bir çok farklı yapay zeka metin insanlaştırma modeli ve sitesi ile insanlaştırma işlemi gerçekleştirdiğimizde gördük ki kullanılan yapay zeka metin dedektörüne göre sonuç değişiyor daha iyi insanlaştırma araçları, dedektörleri ve onları yakalayacak olan yapay zeka dedektörleride insanlaştırma modellerini çözüyor. Sonuç olarak bu araçları kullanarak %100 kesin bir çözüm üretemesekde %99.7 gibi iyi bir oranla insanlaştırma işlemi yapılsa dahi yapay zeka tarafından üretildiği anlaşıyor. Yapılacak en iyi şey çalışmamızda görüldüğü gibi başarı oranı yüksek ve güncel modeller ile kontrol etmek.

İş hayatının farklı alanlarında yapay zeka ile iç içe çalışıyoruz. Akademik çevreyi ele aldığımızda, zamandan kazanıp akademik baskıyı hafifletmek için birçok kişi yapay zekayı kullanıyor ancak nereye kadar etik sınırlar içindeyiz? YÖK tarafından sunulan Yükseköğretim Kurumları Bilimsel Araştırma Ve Yayınfaaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber şu şekilde ifade ediliyor. Üretken yapay zeka çıktılarından edinilen bilgilerin kaynaklarının gösterilmesi, öncelikle dürüstlük dediğer bilim insanlarının emeğine saygı gereği zorunludur. Bilimsel araştırma ve yayın yazım aşamalarında ÜYZ kullanılan bölümlerin yöntem kısmında açıklanması gereklidir (Yükseköğretim Kurulu, 2024).

Chatgpt, gemini gibi internette bulunan veriler ile eğitilen yapay zeka modelleri etiksel açıdan birden çok sorunu beraberinde getiriyor. Yapay zeka kullanarak makale yazan gazeteci, yazar veya akademisyen olsun yazılan makaleler kişinin özgünlüğünü, fikirlerini ve yaratıcılığını hissettirmiyor. Çünkü yapay zeka birebir kopyalamasada başkaları tarafından yazılmış içerikleri tekrar oluşturuyor. Buda aynı fikirlerin biçim değiştirerek tekrar kullanılmasına sebep oluyor.

İş hayatının bütününe yayılmış yapay zeka her alanda kullanılarak kişilerin kişisel asistanı haline geldi. Hayatın bir parçası haline gelen yapay zeka modellerini dışarda bırakmak pek mümkün gözüküyor ancak akademik çalışmalarda önlem olarak Yapay Zekâ Ve Akademik Yazımda Etik: Chatgpt Kullanımının Fırsatları Ve Tehlikeleri Üzerine Birdeğerlendirme makalesinin politika ve rehberlik önerileri kısmında şu şekilde ele alıyor. “Yükseköğretim kurumlarının alabileceği bir başka önemli önlem, ölçme-değerlendirmesistemlerinin yapay zekâ ile üretilmiş içerikleri bertaraf edecek biçimde güncellenmesidir. Sadece yazılıödevlere değil; sözlü sunumlara, proje temelli çalışmalara, araştırma günlüğüne dayalı süreç odaklıdeğerlendirmelere ağırlık verilmesi önerilmektedir (Eke, 2023). Özellikle tez savunmalarında öğrencinin kendi

metnini savunma ve açıklama becerilerinin ölçülmesi, etik dışı üretimlerin tespitini kolaylaştırabilir.” (Meydan & Akkaş, 2025; Okaibedi, 2023).

KAYNAKÇA

- AI Humanize. (2025, Eylül 20). *AI Humanize*. <https://aihumanize.io/>
- Atayolu, Y., Kutlu, Y. (2024). Similarity and Classification Analysis in Turkish Question Generation of AI Tools According to Bloom’s Taxonomy. *Tethys Environmental Science*, 1(2), 87-98,
- Abouelyazid, N. (2024). ChatGPT, Gemini, Copilot, Perplexity and More: A Comparison of Modern AI Models. *Journal of Artificial Intelligence with Applications*, 5(1), 8-12,
- Bhagya, A. (2025, Eylül 11). *Human vs AI 99% Acc [Veri seti]*. Kaggle. <https://www.kaggle.com/code/allelbhagya/human-vs-ai-99-acc>
- Buchert, J. (2025, Mart 10). *Do AI Humanizers Work? How Do They Work?*. *Intellectualead*. <https://intellectualead.com/do-ai-humanizers-work/>
- ChatGPT. (2025, Eylül 20). *ChatGPT*. <https://chatgpt.com/>
- Goldsborough, M. (2025, January 20). *wikipedia-api 0.7.1*. Python Package Index (PyPI). <https://pypi.org/project/Wikipedia-API/>
- Grammarly. (2025, Eylül 20). *AI Humanizer*. <https://www.grammarly.com/ai-humanizer>
- Humanize AI. (2025, Eylül 20). *Humanize AI*. <https://www.humanizeai.pro/>
- Meydan, C. H. & Akkaş, H. (2025 Mayıs 3). Yapay Zekâ Ve Akademik Yazımda Etik: Chatgpt Kullanımının Fırsatları Ve Tehlikeleri Üzerine Birdeğerlendirme. *Akademik Açı*, 5(1), 2-10. <https://dergipark.org.tr/tr/download/issue-file/86448>
- Okaibedi, D. (2023, Şubat 22). ChatGPT and the rise of generative AI: Threat to academic integrity?. *ScienceDirect*. <https://www.sciencedirect.com/science/article/pii/S2666659623000033>
- Öncü, M., Kutlu, Y. (2024). Performance Comparison of Known AI Translation Tools for Turkish Language. *Tethys Environmental Science*, 1(3), 117-126
- QuillBot. (2025, Eylül 20). *AI Humanizer*. <https://quillbot.com/ai-humanizer>
- Santoro, K. (2024, Kasım 27). *How Do AI Detectors Work? | Techniques & Accuracy*. QuillBot Blog. <https://quillbot.com/blog/ai-writing-tools/how-do-ai-detectors-work/>
- TF IDF Nedir. (2021, Kasım 26). Turhost. <https://www.turhost.com/blog/tf-idf-nedir/>
- WriteHuman. (2025, Eylül 20). *WriteHuman*. <https://writehuman.ai/>
- Yükseköğretim Kurulu. (2024, Mayıs). *Yükseköğretim Kurumları Bilimsel Araştırma Ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber*. <https://cdn.anadolu.edu.tr/files/anadolu-cms/GVz7XYek/announcement/yuksekogretim-kurumlari-bilimsel-arastirma-ve-yayin-faaliyetlerinde-uretken-yapay-zeka-kullanimina-dair-etik-rehber-93e43ab80ebf3202.pdf>