ALGORİTMİK ÖN YARGI VE ALGORİTMİK AYRIMCILIK: KAVRAMLAR, GÜNCEL YARGI KARARLARI VE ÖN YARGININ AZALTILMASINA YÖNELİK OLASI BİR YOL HARİTASI OLARAK AB YAPAY ZEKÂ TÜZÜĞÜ

ALGORITHMIC BIAS AND ALGORITHMIC DISCRIMINATION: CONCEPTS, RECENT CASE LAW, AND THE EU AI ACT AS A POSSIBLE ROADMAP TO BIAS MITIGATION

Araştırma Makalesi

Fatmanur CEBECİ ÇORUM*

ÖZ

Bu çalışma, yapay zekâ sistemlerinde ortaya çıkan algoritmik ön yargı ve ayrımcılık sorununu, bu risklerin hukuki düzenlemeler yoluyla azaltılıp azaltılamayacağı sorusu çerçevesinde incelemektedir. Çalışma, biçimsel olarak tarafsız görünen yapay zekâ sistemlerinin, uygulamada mevcut toplumsal eşitsizlikleri yeniden üretebildiği veya ayrımcı sonuçlar doğurabildiği tespitinden hareket etmektedir. Bu tür sonuçlar, ön yargılı veya tarihsel olarak şekillenmiş veri setlerinden, eşitsiz toplumsal koşulları yansıtan referans sonuçlardan ya da sistemlerin tasarım ve modelleme tercihlerinden kaynaklanabilmektedir.

Çalışmada algoritmik ön yargının başlıca kaynaklarını kısaca ortaya koyulmakta ve konut, istihdam, kredi, ceza adaleti ve kamu sektöründe kullanılan yapay zekâ uygulamalarına ilişkin ayrımcı örnekler üzerinden konunun pratik boyutu ele alınmaktadır. Ayrıca Amerika Birleşik Devletleri'nde verilen güncel yargı kararları ışığında, algoritmik ayrımcılığın giderek somut bir hukuki sorun haline geldiği vurgulanmaktadır. Bu çerçevede, ön yargının azaltılmasına yönelik teknik araçlar değerlendirilmekte; ancak bu araçların çoğu zaman ön yargı ortaya çıktıktan sonra devreye girdiği ve normatif varsayımlara dayandığı belirtilmektedir.

Çalışma, Avrupa Birliği Yapay Zekâ Tüzüğü'nü algoritmik ön yargı ve ayrımcılıkla mücadelede önemli ve yenilikçi bir hukuki girişim olarak ele almakta, Türkiye'de de bu şekilde bir istisna getirilmesinin olumlu etkileri olabileceğini belirtmekte; ancak bununla birlikte, böyle bir mevzuatın kabul edilmesinin algoritmik

DOI: 10.32957/hacettepehdf.1858048

Makalenin Geliş Tarihi: 06.01.2026 **Makalenin Kabul Tarihi:** 18.02.2026

* Dr. Öğretim Görevlisi, Ankara Yıldırım Beyazıt Üniversitesi Hukuk Fakültesi, Bilişim Hukuku Anabilim Dalı, https://ror.org/05ryemn72.

E-posta: fatmanurcebecicorum@aybu.edu.tr

ORCID: 0000-0002-2629-0320

Bu makale Hacettepe Üniversitesi Hukuk Fakültesi Dergisi Araştırma ve Yayın Etiği kurallarına uygun olarak hazırlanmıştır.

adaletin fiilen sağlanacağı anlamına gelmediğini vurgulamaktadır. Sonuç olarak çalışmada, düzenleyici müdahalelerin ön yargıyı teorik olarak azaltabileceğini, ancak tamamen ortadan kaldırmasının beklenmemesi gerektiğini savunmaktadır.

Anahtar Kelimeler: Algoritmik Ön Yargı, Algoritmik Ayrımcılık, AB Yapay Zekâ Tüzüğü, Otomatik Karar Verme, Ön Yargı Azaltma

ABSTRACT

This study examines algorithmic bias and discrimination in artificial intelligence systems, focusing on whether these risks can be reduced through legal regulation. It proceeds from the observation that AI systems which appear formally neutral may, in practice, reproduce existing social inequalities or generate discriminatory outcomes. Such outcomes may stem from biased or historically shaped datasets, reference outputs that reflect unequal social conditions, or design and modelling choices embedded in AI systems.

The study briefly sets out the main sources of algorithmic bias and addresses the practical dimension of the problem through examples of discriminatory AI use in housing, employment, credit, criminal justice, and the public sector. Drawing on recent judicial decisions in the United States, it emphasises that algorithmic discrimination is increasingly treated as a concrete legal issue. In this context, the study also considers technical tools aimed at bias mitigation, while noting that such tools often operate only after bias has emerged and rely on normative assumptions.

The study addresses the European Union Artificial Intelligence Act as a significant and innovative legal initiative in combating algorithmic bias and discrimination, and notes that the introduction of a similar exception in Türkiye could yield positive effects. However, it emphasises that the adoption of such legislation does not necessarily mean that algorithmic justice will be effectively achieved in practice. In conclusion, the study argues that while regulatory interventions may theoretically reduce bias, they should not be expected to eliminate it entirely.

Keywords: Algorithmic Bias, Algorithmic Discrimination, EU Artificial Intelligence Act, Automated Decision-Making, Bias Mitigation

EXTENDED SUMMARY

This article examines the growing prominence of algorithmic bias and discrimination in contemporary uses of artificial intelligence, focusing on whether these risks can be effectively mitigated through a legal regulation. The study proceeds from the observation that Artificial Intelligence (AI) systems which appear neutral or objective in formal terms may, in practice, reproduce existing social inequalities or generate discriminatory outcomes. Such effects may arise from biased datasets, historically conditioned social patterns, or modelling and design choices that shape how decisions are produced. For this reason, the

increasing reliance on AI in decision-making processes is assessed not merely as a technological development, but as a phenomenon that raises broader concerns relating to equality, justice, and fundamental rights.

Against this background, the article addresses several interrelated questions. It considers whether algorithmic bias can be meaningfully mitigated in practice; whether developers and deployers are capable of preventing discriminatory outcomes through interventions at different stages of system development and deployment; and whether legal frameworks can provide effective safeguards to detect, correct, and prevent algorithmic discrimination. In this context, the article pursues four main objectives. First, it outlines the basic elements of algorithmic bias and algorithmic discrimination through selected examples. Second, it situates these issues within recent judicial developments, illustrating that algorithmic discrimination has increasingly become a concrete legal concern rather than a purely theoretical one. Third, it reviews selected technical initiatives aimed at identifying and mitigating bias in AI systems. Fourth, it evaluates the regulatory approach adopted by the European Union (EU) through the EU Artificial Intelligence Act (EU AI Act) and considers whether this framework may inform future legislative developments in Türkiye and other jurisdictions.

The article identifies three broad sources from which such biases may arise. First, bias may originate in data, including skewed or historically patterned datasets, underrepresentation of certain groups, or the use of variables that appear neutral but nonetheless encode social inequalities. Second, bias may stem from the 'ground truth' a model is trained to predict, where the reference outcomes themselves reflect structurally unjust social conditions. In such cases, proxy variables correlated with protected characteristics may indirectly reproduce discrimination even where sensitive attributes are formally excluded. Third, bias may be introduced through design and modelling choices, including how a problem is framed, which features are prioritised, and which performance metrics are adopted, all of which may systematically disadvantage particular groups.

To demonstrate that these issues are not merely abstract, the article refers to a range of well-documented examples of discriminatory AI outputs. These include demeaning or abusive content generated by conversational systems, patterns of inadequate or stereotypical representation in generative models, and discriminatory effects observed in systems used for recruitment, credit assessment, criminal justice, and image classification. Taken together, these examples suggest that algorithmic bias can be systemic rather than incidental and may reproduce existing inequalities when embedded in automated decision-making.

The legal dimension of algorithmic discrimination is explored through a brief examination of selected recent judicial decisions. In the United States, cases such as Mary Louis v SafeRent and Mobley v Workday, Inc demonstrate how AI-assisted systems used in housing and employment contexts have been challenged for producing disparate impacts, as well as how courts have begun to engage with questions of responsibility and accountability in algorithmic decision-making.

Building on this case law, the article assesses existing strategies for mitigating algorithmic bias. On the technical side, it reviews fairness-oriented toolkits and methods, including AIF360, AEQUITAS, and Fairlearn, as well as resampling techniques and model explainability tools such as SHAP and LIME. While these tools provide valuable mechanisms for identifying and reducing discriminatory effects, the article notes that they often operate after bias has already emerged and do not necessarily guarantee fairness from the outset. Moreover, technical interventions inevitably rely on normative assumptions regarding what constitutes 'fair' data or 'neutral' outcomes, limiting the possibility of purely technical solutions.

On the legal side, the article focuses on the EU AI Act as a novel attempt to address algorithmic bias through binding regulation. While the General Data Protection Regulation (GDPR) generally restricts the processing of special category data, the EU AI Act introduces a narrowly framed exception allowing providers of high-risk AI systems to process such data where strictly necessary for bias detection and mitigation, subject to appropriate safeguards. Although this exception is limited in scope and applies only to certain actors and systems, it represents an important regulatory experiment and one of the first legislative attempts to reconcile data protection and non-discrimination in the context of AI and the introduction of a similar exception in Türkiye could yield positive effects.

However, the article emphasises that the adoption of AI-specific legislation does not, by itself, ensure the realisation of algorithmic justice or fairness. While technical tools and legal mechanisms may contribute, achieving algorithmic justice in practice remains far more complex. Determining which data should be treated as discriminatory, which outcomes may be regarded as 'correct' or 'bias-free', and which evaluative standards should prevail is particularly difficult in a global digital ecosystem shaped by divergent socio-cultural, political, and moral perspectives. Empirical observations showing that different AI models provide divergent answers to politically sensitive questions make this complexity especially visible, suggesting that systems may reflect, to some extent, the values of the contexts in which they are developed and trained. The article concludes that the EU AI Act constitutes a valuable and forward-looking regulatory initiative and may serve as a reference point for future regulation, including in Türkiye; nevertheless, while such regulation may reduce algorithmic bias to a degree, it should not be expected to eliminate discriminatory outcomes entirely.

GİRİŞ

Yapay zekânın genel kabul görmüş bir tanımı olmasa da,[1] Britannica yapay zekâyı bir bilgisayarın veya bilgisayar kontrollü bir sistemin, akıl yürütme gibi insanın bilişsel

[1] Ahmet Esad Berktaş ve Saide Begüm Feyzioğlu, "Regulating AI Against Discrimination: From Data Protection Legislation to AI-Specific Measures," *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde, ed. Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu (Singapore: Springer,

süreçleriyle yaygın olarak ilişkilendirilen görevleri yerine getirme kapasitesi olarak tanımlamaktadır.[2] Doktrinde de yapay zekâ kavramı bir makinenin insan zihinsel süreçlerini modelleyerek öğrenme, akıl yürütme ve problem çözme gibi insana özgü bilişsel faaliyetleri taklit etmesi[3] olarak tanımlanmıştır. Daha genel olarak, yapay zekâ doğal sistemlerin gerçekleştirebildiği bilişsel faaliyetlerin, yapay sistemler aracılığıyla nasıl ve hangi yollarla daha yüksek performans düzeylerinde yerine getirilebileceğini araştıran bir bilim dalı olarak tanımlanabilir.[4]

Yapay zekânın kökenleri 1956 yılında gerçekleştirilen Dartmouth Konferansı'na kadar uzanmakla birlikte,[5] kavram son yıllarda gerek popülerliği gerekse etrafında şekillenen tartışmaların çeşitliliği bakımından kayda değer bir görünürlük kazanmıştır. Günümüzde yapay zekâ, en basit işlevlerden en karmaşık operasyonlara kadar uzanan geniş bir yelpazede kullanılan ve önemsiz ya da son derece önemli olsun, çok sayıda kararı etkileyen, gündelik hayatın ayrılmaz bir parçası haline gelmiştir. Özellikle yapay zekâ sistemleri tarafından verilen kararlar, bu sistemlerin ürettiği bilgiler ve söz konusu kararların dayandığı veriler; şeffaflık, hesap verebilirlik ve tarafsızlık bakımından sıklıkla sorgulanmaktadır. Bu durum, yapay zekâ sistemlerinin yalnızca teknik doğruluğunun değil, aynı zamanda etik ve hukuki meşruiyetinin de değerlendirilmesini gerekli kılmakta; bu çerçevede algoritmik ön yargı, algoritmik ayrımcılık ve nihayetinde algoritmik eşitlik kavramlarını güncel tartışmaların merkezine taşımaktadır.

Son dönem yargı kararlarının da ortaya koyduğu üzere, yapay zekâ uygulamalarının kullanımı özellikle ayrımcılık bağlamında dikkat çekici bir hal almıştır. Bu sistemlerde kullanılan verilerin içerdiği açık veya örtük ön yargılar, karar alma

---

2024), 57; Sedat Ayaz, "Yapay Zekânın Hukuk Alanındaki Uygulamaları ve Hukuk Alanında Kullanılmasının Olası Sonuçları", *Sakarya Üniversitesi Hukuk Fakültesi Dergisi* 13/1 (2025) 82.

[2] Encyclopaedia Britannica, 'Artificial Intelligence', erişim 12 Kasım 2025, https://www.britannica.com/technology/artificial-intelligence.

[3] Hüseyin Can Aksoy, "Kişisel Verilerin Korunması Yönüyle Algoritmik Karar Verme", *Kişisel Verileri Koruma Dergisi* 4/2 (2022) 71.

[4] Cem Say, *50 Soruda Yapay Zekâ,* 25. Baskı, (İstanbul: Bilim ve Gelecek Kitaplığı, 2025) 83.

[5] Nathalie A. Smuha, "An Introduction to the Law, Ethics and Policy of Artificial Intelligence", *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* içinde, ed. Nathalie A. Smuha (Cambridge: Cambridge University Press, 2025) 1.

süreçlerinin sonuçlarını doğrudan etkileyebilmektedir. Buna ek olarak, verilerde açıkça yer almayan toplumsal eşitsizlikler de algoritmik modelleme süreçleri aracılığıyla dolaylı biçimde yeniden üretilebilmekte ve mevcut adaletsizlikler pekiştirilebilmektedir. Bu nedenle, biçimsel olarak tarafsız görünen sistemler, uygulamada belirli grupları dezavantajlı konuma düşüren veya mevcut toplumsal ön yargıları sürdüren sonuçlar doğurabilmektedir. Dolayısıyla yapay zekâ sistemlerinin karar alma süreçlerine dahil edilmesi, salt bir teknolojik yenilik olarak değil; adalet, eşitlik ve insan hakları perspektiflerinden dikkatle incelenmesi gereken bir gelişme olarak değerlendirilmelidir.

Bu bağlamda, algoritmik ön yargının gerçekten azaltılıp azaltılamayacağı, bu sistemlerin tasarımcıları ve geliştiricilerinin veri seçimi, işlenmesi, eğitimi veya kullanım sonrası geri bildirimler aşamalarında ayrımcılığı bütünüyle ortadan kaldırıp kaldıramayacağı ve hatta hukuki düzenlemelerin bu tür ön yargılara karşı ne ölçüde etkili güvenceler sağlayabileceği yönünde temel sorular ortaya çıkmaktadır.

Makalenin dört temel amacı vardır. İlk amaç, algoritmik ön yargı ve algoritmik ayrımcılık kavramlarının örnekleriyle birlikte açıklanması ve okuyucunun bu kavramlar ve konunun özü hakkında bilgi sahibi olmasının sağlanmasıdır. İkinci amaç, konu kapsamındaki Amerika Birleşik Devletleri'nde (ABD) verilmiş güncel yargı kararlarının kısaca belirtilerek konunun hukuki boyutunun da önem kazandığı ve düzenlemelerin gerektiği hususunun belirtilmesidir. Üçüncü amaç, algoritmik ön yargının tasarım aşamasında azaltılmasına ilişkin dünyadaki çeşitli teknik girişimlerin ortaya konmasıdır. Dördüncü ve son amaç ise Avrupa Birliği'ndeki (AB) mevcut mevzuatın işlerliği ve Türkiye başta olmak üzere dünyadaki yeni yasama faaliyetlerinde kullanılabilecek bir örnek olup olmadığı ve algoritmik ön yargının bu düzenlemeler sayesinde azaltılabilme hatta ayrımcılığın algoritmik düzeyde yok olabilme, algoritmik eşitliğin sağlanabilme ihtimalidir.

Bu amaçlar doğrultusunda makalenin ilk bölümünde algoritmik ön yargı ve algoritmik ayrımcılık kavramları kısaca ele alınmakta; yapay zekâ sistemlerinin neden ayrımcı sonuçlar üretebildiği özetle ve örneklerle açıklanmaktadır. İkinci bölümde, algoritmik ön yargı ve algoritmik ayrımcılık bağlamında uluslararası düzeydeki güncel yargı kararları incelenmektedir. Üçüncü bölümde ise algoritmik ön yargının hem teknik

hem de hukuki yaklaşımlar yoluyla azaltılmasına yönelik çabalar ele alınmakta; yapay zekâ geliştiricilerinin ve mühendislerin mevcut teknolojik ve etik araçları kullanarak daha adil sistemler tasarlama imkânları ile ön yargıyı uygulamada tespit etme ve azaltma kapasiteleri değerlendirilmektedir. Bu çerçevede ayrıca üçüncü bölümde, özellikle AB Yapay Zekâ Tüzüğü'ne odaklanılarak, algoritmik ayrımcılığın tespiti, giderilmesi ve önlenmesine imkân tanıyan normatif bir çerçevenin oluşturulmasında hukuki düzenlemelerin rolü ve söz konusu düzenlemenin etkinliği analiz edilmektedir. Son bölümde genel bir değerlendirme yapılmakta ve AB Yapay Zekâ Tüzüğü'nün algoritmik ayrımcılığın tespiti ve azaltılmasına ilişkin hükümlerinin, başta Türkiye olmak üzere diğer ülkeler bakımından yol gösterici olup olamayacağına dair sonuç niteliğinde tespitlere yer verilmektedir.

I. ALGORİTMİK ÖN YARGI VE ALGORİTMİK AYRIMCILIK

A. Kavramsal Çerçeve ve Nedenler

Algoritmik ön yargı ve algoritmik ayrımcılık kavramları zaman zaman birbirinin yerine kullanılsa da, gerçekte birbirinden farklı ancak birbiriyle ilişkili olgulara işaret etmektedir.[6] Algoritmik ön yargı (algorithmic bias), yapay zekâ temelli sistemlerde verilerin toplanması, seçilmesi, işlenmesi veya modellenmesi sırasında ortaya çıkan sistematik hatalar sonucunda, algoritmanın çıktılarının bazı kişileri veya grupları diğerlerine kıyasla gerekçesiz biçimde avantajlı ya da dezavantajlı bir konuma yerleştirmesi durumu olarak tanımlanabilir.[7] Algoritmik ayrımcılık (algorithmic discrimination) ise, etnik köken, cinsiyet, yaş veya engellilik gibi özelliklere dayanarak, karar alma süreçlerinde algoritmaların belirli kişi veya gruplara karşı adaletsiz veya önyargılı muamelede bulunmasını ifade eder.[8] Başka bir ifadeyle, algoritmik ayrımcılık;

---

[6] Trang Anh Mac, "Bias and Discrimination in ML-based Systems of Administrative Decision-Making and Support", *Computer Law & Security Review* 55 (2024) 1; Yeliz Bozkurt Gümrükçüoğlu ve Gülnihal Ahter Yakacak, "Yapay Zekanın İşe Alım Süreçlerinde Kullanımı ve Algoritmik Ayrımcılık", *Ankara Üniversitesi Hukuk Fakültesi Dergisi* 72,4 (2024) 1719, https://doi.org/10.33629/auhfd.1403311.

[7] Zoya Yousef, "How Can the Law Address the Effects of Algorithmic Bias in the Healthcare Context?", *LSE Law Review* 9,2 (2024), 203, erişim 1 Ocak 2026, https://doi.org/10.61315/lselr.651.

[8] Mac, "Bias and Discrimination", 2.

ön yargılı verilerle eğitilen veya tasarımında ön yargı barındıran algoritmaların, nesnel adalet ilkesini ihlal eden sonuçlar üretmesi hâlinde ortaya çıkar.[9]

Algoritmik ön yargı hem verinin niteliğinden hem de algoritmanın tasarım sürecinden kaynaklanabilir. Bu ön yargı, veriye ilişkin isabetsizliklerden doğabileceği gibi, sistem tasarımı sırasında yapılan değer yüklü tercihlerden de kaynaklanabilir. Literatürde[10] genellikle üç temel ön yargı kaynağı tespit edilmektedir: veriye dayalı ön yargı (data-driven bias), eşitsiz temel gerçeklik ön yargısı (unequal ground truth bias) ve tasarım temelli ön yargı (design-based bias).

1. Veriye Dayalı Ön Yargı

Kavram çoğu zaman 'algoritmik ön yargı' olarak adlandırılsa da, uygulamada bu tür ön yargının en yaygın kaynağı algoritmanın kendisinden ziyade verinin bizzat kendisidir.[11] Veri, algoritmik ön yargının ortaya çıkmasında temel unsuru oluşturur. Veriye dayalı ön yargı beş ana kategori altında incelenebilir: (i) ön yargılı eğitim verisinin kullanımı (biased training data), (ii) örnekleme ön yargısı (sampling bias), (iii) temsili olmayan veriden kaynaklanan ön yargı (bias arising from non-representative data), (iv) tarihsel ön yargı (historical bias) ve (v) örtük ön yargı içeren verinin kullanımı (implicit bias in data).

i. Ön yargılı eğitim verisi: Yapay zekâ sistemlerinin eğitilmesinde ön yargılı veri kümelerinin kullanılması, algoritmik ön yargının en yaygın kaynaklarından biridir.[12] Algoritmalar, insan ön yargılarını yansıtan veya geçmişteki ayrımcı örüntüleri yeniden üreten verilerden öğrenebilir. Örneğin, bir yapay zekâ modelinin eğitim verisinde belirli ırksal gruplara yönelik aşağılayıcı ya da dışlayıcı ifadeler yer alıyorsa, model bu kalıpları

---

9 Bozkurt ve Yakacak, "Yapay Zekanın İşe Alım Süreçlerinde Kullanımı", 1720.

10 Literatürde bahsedilen farklı ön yargı türleri yazar tarafından üçe ayrılarak sınıflandırılmıştır. Özellikle bkz: Philipp Hacker, "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law", *Common Market Law Review* 55,4 (2018) 1148; Mac, "Bias and Discrimination", 5; Yousef, "How Can the Law Address", 208.

11 Trisha Mahoney, Kush R. Varshney ve Michael Hind, *AI Fairness: How to Measure and Reduce Unwanted Bias in Machine Learning* (IBM Corporation, 2020) VII.

12 Hacker, "Teaching Fairness to Artificial Intelligence", 1151.

içselleştirerek çıktılarında benzer nitelikte sonuçlar üretme eğilimi gösterebilir. Nitekim Birleşik Krallık'ta 1980'lerde okul başvurularının değerlendirilmesi amacıyla geliştirilen bir programın, kadınlar ve göçmen kökenli bireyler aleyhine önyargılı verilerle eğitildiği ve bu nedenle ayrımcı sonuçlar ürettiği sonradan ortaya konulmuştur.[13]

ii. Örnekleme ön yargısı: Bu tür ön yargı, veri kümesinde bazı grupların eksik temsil edilmesi veya hatalı temsil edilmesi hâlinde ortaya çıkar.[14] Ağırlıklı olarak beyaz bireylerin görüntüleriyle eğitilen yüz tanıma sistemlerinin, beyaz olmayan bireylerin yüzlerini daha yüksek oranda yanlış sınıflandırabilmesi buna örnek olarak verilebilir.[15]

iii. Temsili olmayan veriden kaynaklanan ön yargı: Eğitim veri kümesi, algoritmanın fiili veya hedeflenen kullanıcı kitlesini yeterince yansıtmadığında da ön yargı ortaya çıkabilir.[16] Örneğin, esasen standart Amerikan İngilizcesi aksanına sahip erkek konuşmacı kayıtlarıyla eğitilen bir ses tanıma sistemi, farklı aksanlara sahip kullanıcılar veya veri kümesinde temsili daha zayıf olan kadın konuşmacılar bakımından daha düşük doğrulukla çalışabilecektir.

iv. Tarihsel ön yargı: Bu tür ön yargı, önceden var olan toplumsal ve yapısal eşitsizliklerin veri üretim sürecine yerleşmesiyle ortaya çıkar. Örnekleme ve özellik seçimi büyük bir özenle yapılsa dahi varlığını sürdürebilir.[17] Nitekim bir çalışma, 'CEO' terimi için yapılan görsel arama sonuçlarında kadınların ciddi ölçüde eksik temsil edilmesinin, ilgili yılda Fortune 500 şirket yöneticilerinin yalnızca yüzde 4,8'inin kadın olmasının bir yansıması olduğunu ortaya koymuştur.[18]

---

[13] Selin Çetin Kumkumoğlu ve Ahmet Kemal Kumkumoğlu, "Rethinking Non-discrimination Law in the Age of Artificial Intelligence," *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde, ed. Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu (Singapore: Springer, 2024) 37.

[14] Hacker, "Teaching Fairness to Artificial Intelligence", 1152.

[15] Lethiwe Nzama-Sithole, "Managing Artificial Intelligence Algorithmic Discrimination: The Internal Audit Function Role," *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde, ed. Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu (Singapore: Springer, 2024) 210.

[16] Mac, "Bias and Discrimination", 5.

[17] Mac, "Bias and Discrimination", 6.

[18] Mac, "Bias and Discrimination", 6.

v. Örtük ön yargı: Örtük ön yargı, bireylerin farkında olmaksızın taşıdığı ve düşünce ile davranışlarını şekillendiren kalıp yargılar veya tutumlar anlamına gelir. Bu tür ön yargılar, hedef değişkenlere ayrımcı unsurlar eklenmesi veya dışlayıcı ölçütlerin tanımlanması suretiyle kasıtlı biçimde sisteme dâhil edilebileceği gibi, teknik hatalar yoluyla istemeden de sistemlere sızabilir.[19] Örneğin, özgeçmiş ve mülakat notlarından kültürel uyuma önem veren işe alım sistemleri, değerlendiricilerin farkında olmaksızın 'kendine benzeyeni daha uygun görme' eğilimini model çıktısına taşıyarak, aynı yetkinlikteki adaylar arasında belirli profilleri, örneğin belirli okul mezunlarını veya belirli ülkede yaşayanları, sistematik biçimde daha avantajlı konuma getirebilir.

2. Eşitsiz 'Temel Gerçeklik'

Bir veri kümesi görünüşte tarafsız olsa dahi, algoritmanın 'temel gerçeklik' olarak öğrendiği referansın bizzat kendisi adaletsiz olabilir.[20] Toplumsal eşitsizlikler, 'doğru sonuçlar' olarak kodlanabilir. Bu bağlamda, vekil değişken üzerinden ayrımcılık meselesi (proxy discrimination) (istatistiksel ayrımcılık olarak da adlandırılmaktadır) özel bir önem kazanır.[21] Algoritmalar, korunan niteliklerle, örneğin ırk veya cinsiyet, istatistiksel olarak ilişkili değişkenleri vekil olarak kullanabilir. Yinelemeli kodlama yoluyla (redundant encoding)[22], bu hassas nitelikler modele açıkça dâhil edilmese bile, onlarla güçlü biçimde ilişkili değişkenler üzerinden ayrımcı sonuçlar ortaya çıkabilir. Bir işe alım modeli, eğitim geçmişi gibi değişkenleri kullanarak, örneğin yalnızca erkek öğrencilerin kabul edildiği bir okuldan mezuniyet bilgisini adayın erkek olduğuna işaret eden bir gösterge olarak değerlendirebilir. Benzer şekilde posta kodu gibi değişkenler üzerinden, örneğin ikamet edilen bölgedeki yoğunlaşma örüntülerine dayanarak etnik kökene ilişkin çıkarımlar yapılması, cinsiyet veya etnik kökenin karar alma sürecine dolaylı biçimde dâhil edilmesine yol açabilir.

---

19 Hacker, "Teaching Fairness to Artificial Intelligence", 1150.

20 Hacker, "Teaching Fairness to Artificial Intelligence", 1150.

21 Hacker, "Teaching Fairness to Artificial Intelligence", 1149.

22 Hacker, "Teaching Fairness to Artificial Intelligence", 1149.

3. Tasarım Temelli Ön Yargı

Bazı durumlarda, yapay zekâ sistemlerinde kullanılan veriler doğası gereği ön yargılı olmasa dahi, sistem tasarıma ilişkin hatalar veya makinenin yanlış öğrenme süreçleri nedeniyle ön yargı sergileyebilir. Tasarım temelli ön yargılar genel olarak iki ana kategoriye ayrılabilir: makine öğrenmesi kaynaklı ön yargı (machine laerning bias) ve kodlama kaynaklı ön yargı (coding bias).

i. Makine öğrenmesi kaynaklı ön yargı: Makine öğrenmesi kaynaklı ön yargı, bir algoritmanın hedeflenen değişkeni tahmin etmek yerine vekil değişkenlere dayanması ve sonuçta amaçlanan hedeften farklı bir olguyu öngörmesi hâlinde ortaya çıkar.[23] Sağlık alanında kullanılan algoritmaların, geçmiş hastalıklara ilişkin risk profillerine dayanarak hastaneye kabul veya yoğun bakım önceliğini belirlemesi; oysa yeni ortaya çıkan bir hastalıkta, örneğin Covid-19'da, farklı kişi gruplarının gerçekte daha yüksek risk altında bulunmasına rağmen bu durumun hesaba katılmaması, buna örnek olarak gösterilebilir.[24] Yine aynı şekilde, Amerika Birleşik Devletleri'nde kalp hastalığı riskini değerlendirmek amacıyla kullanılan bir algoritma, hastaları sağlık harcamaları temelinde değerlendirmiştir. Beyaz hastalar daha çok sağlık harcaması yapmıştır. Bu durum, beyaz olmayan hastaların daha düşük sağlık riski taşıdığı yönünde hatalı bir sonuca varılmasına yol açmış ve söz konusu hastaların tıbbi tedavi önceliğinin haksız biçimde düşürülmesine neden olmuştur.[25]

ii. Kodlama kaynaklı ön yargı: Kodlama kaynaklı ön yargı, geliştiricinin problemi nasıl tanımladığı, hangi verileri seçtiği ve hangi başarı ölçütlerini benimsediği gibi

---

[23] Yousef, "How Can the Law Address", 208.

[24] Merve Ayşegül Kulular İbrahim, "Legal Challenges of Artificial Intelligence in Healthcare," *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde, ed. Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu (Singapore: Springer, 2024) 150.

[25] Çalışmanın tamamı için bkz: Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations" *Science* 366, 6464 (2019): 447–453; Council of Europe, Steering Committee for Human Rights in the Fields of Biomedicine and Health (CDBIO), 'Report on the Application of Artificial Intelligence in Healthcare and Its Impact on the "Patient–Doctor" Relationship' (September 2024) 26; Yousef, "How Can the Law Address", 208.

tasarıma ilişkin tercihlerden kaynaklanır.[26] Örneğin, kalp hastalığını tespit etmeye yönelik bir algoritmanın, erkeklerde daha yaygın görülen bir belirti olan göğüs ağrısını temel tanı göstergesi olarak esas alması, belirtileri çoğu zaman farklı biçimlerde ortaya çıkan kadın hastalar bakımından ayrımcı sonuçlar üretmesine neden olur.[27]

B. Uygulamada Algoritmik Ayrımcılık Örnekleri

Bu noktada, algoritmik ön yargının bir sonucu olarak yapay zekâ sistemleri tarafından üretilen ayrımcı çıktıları incelemek de yerinde olacaktır. Öğretide Philip Hacker tarafından dile getirilen görüşe göre, bu tür ayrımcı sonuçlar temel olarak iki ana biçimde ortaya çıkmaktadır: (i) aşağılayıcı veya küçük düşürücü içerik (demeaning and abusive content) ve (ii) yetersiz temsil (inadequate representation).[28]

Yapay zekâ tarafından üretilen çıktılar, belirli gruplara yönelik aşağılayıcı veya küçük düşürücü ifadelere yer verebilmektedir.[29] Nitekim Güney Kore'de geliştirilen ve Facebook üzerinden kullanıma sunulan bir sohbet botunun, lezbiyenleri 'ürkütücü' olarak nitelendirdiği ve onlara yönelik nefret içeren ifadeler kullandığı tespit edilmiştir.[30]

Yetersiz temsil sorunu ise, her somut durumda doğrudan ayrımcılık teşkil etmese dahi, belirli grupların algoritmik çıktılarda aşırı ya da yetersiz biçimde temsil edilmesine yol açabilmektedir. Bazı grupların algoritmik çıktılarda sistematik biçimde eksik temsil edilmesi hâlinde, istatistiksel ayrımcı etkiler ortaya çıkabilmektedir.[31] Örneğin, bir yapay zekâ tabanlı görsel üretim modeline 'üretken bir insan çizmesi' talimatı verildiğinde,

---

[26] Yousef, "How Can the Law Address", 208.

[27] Yousef, "How Can the Law Address", 208.

[28] Philipp Hacker, Frederik Zuiderveen Borgesius, Brent Mittelstadt ve Sandra Wachter, 'Generative Discrimination: What Happens When Generative AI Exhibits Bias, and What Can Be Done About It', *The Oxford Handbook of the Foundations and Regulation of Generative AI* içinde ed. Philipp Hacker vd., çevrim içi edisyon (Oxford: Oxford University Press, 22 Nisan 2025), erişim 11 Kasım 2025, https://doi.org/10.1093/oxfordhb/9780198940272.013.0016 2.

[29] Hacker, "Generative Discrimination", 2.

[30] Justin McCurry, 'South Korean AI Chatbot Pulled from Facebook after Hate Speech towards Minorities', The Guardian, 13 Ocak 2021, erişim 11 Kasım 2025, https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook.

[31] Hacker, "Generative Discrimination", 4.

sıklıkla beyaz, erkek ve üst düzey yönetici profillerini yansıtan görseller üretildiği gözlemlenmiştir.[32]

Algoritmik ayrımcılığın tezahürlerine ilişkin çok sayıda güncel örnek mevcuttur. Amerika Birleşik Devletleri'nde mahkemeler tarafından yeniden suç işleme olasılığını tahmin etmek amacıyla kullanılan COMPAS modeli, beyaz olmayan sanıklar aleyhine sistematik bir ön yargı sergilediği gerekçesiyle eleştirilmiştir.[33] Benzer şekilde, Google Photos'un kişileri ve nesneleri tanımlayarak görselleri otomatik olarak etiketleyen sistemi, 2015 yılında bazı kişilere ait fotoğrafları 'goril' olarak etiketlemesi üzerine ciddi bir tartışmaya yol açmıştır.[34] Bu durumun, ağırlıklı olarak beyaz yüzlerden oluşan eğitim verilerinden ve veri kümelerini oluşturan kişilerin büyük ölçüde beyaz bireyler olmasından kaynaklanmış olabileceği; dolayısıyla ırksal ön yargının yeniden üretildiği ileri sürülmüştür. [35]

Bir diğer dikkat çekici örnek ise Google Translate'e ilişkindir. Sistem, cinsiyet içermeyen Türkçe 'o' zamirini İngilizceye çevirirken, 'he is a doctor' ve 'she is a nurse' şeklinde cinsiyetçi çeviriler üretmiş; Türkçede cinsiyet belirtilmemesine rağmen doktorların erkek, hemşirelerin ise kadın olduğu yönündeki kalıp yargılara dayalı varsayımlarda bulunmuştur.[36] Bu hata daha sonra düzeltilmiş olmakla birlikte, benzer bir sorunun yakın zamanda Danca çevirilerde de ortaya çıktığı gözlemlenmiştir.[37] Buna göre, 'partnerim doktordur' ifadesi erkek özneyle, 'partnerim hemşiredir' ifadesi ise kadın

---

[32] Hacker, "Generative Discrimination", 4.

[33] Martini, Mario, "Algorithmen als Herausforderung fürdie Rechtsordnung", *Juristen Zeitung,* 72/21 (2017) 1018; Hacker, "Teaching Fairness to Artificial Intelligence", 3: Mehmet Çoğalan ve Zeynep Ebrar Kaya, "Chatbotların Yargı Bağımsızlığı ve Adil Yargılanmak Hakkı Boyutuyla Değerlendirilmesi", *İnÜHFD*, 16, 2 (2025) 634; Nzama-Sithole "Managing Artificial Intelligence Algorithmic Discrimination", 210.

[34] Bozkurt ve Yakacak, "Yapay Zekanın İşe Alım Süreçlerinde Kullanımı", 1720.

[35] Nicolas Kayser-Bril, 'Google Apologizes after its Vision AI Produced Racist Results', AlgorithmWatch, 7 Nisan 2020, erişim 11 Kasım 2025, https://algorithmwatch.org/en/google-vision-racism/.

[36] Aksoy, "Kişisel Verilerin Korunması", 73.

[37] Sune Selsbæk-Reitz, LinkedIn gönderisi, erişim 11 Kasım 2025, https://www.linkedin.com/posts/suneselsbaekreitz_in-danish-the-word-kæreste-is-gender-neutral-activity-7383384358679699456-OA8P/

özneyle çevrilmiştir. Ayrıca literatürde, benzer cinsiyetçi varsayımların Macarca diline yapılan çevirilerde de ortaya çıktığı ifade edilmektedir.[38]

Amazon da bu bağlamda çarpıcı bir örnek sunmaktadır. Şirket, 2015 yılında özgeçmişleri otomatik olarak değerlendirmek üzere geliştirdiği yapay zekâ temelli işe alım sisteminin, kadın adayları sistematik biçimde dezavantajlı konuma düşürdüğünü tespit etmiştir.[39] Model, teknik pozisyonlara başvuran adayların büyük çoğunluğunun erkek olduğu bir döneme ait tarihsel işe alım verileriyle eğitilmiş; bu durum, erkek adaylara yönelik bir tercihin sistem tarafından içselleştirilmesine yol açmıştır. Sonuç olarak algoritma, tarihsel cinsiyet eşitsizliklerini yeniden üretmiş ve kadınlar aleyhine yapısal bir ön yargı geliştirmiştir.[40]

Eğitim verisinin belirleyici rolü bir başka örnekle daha açık biçimde ortaya konulabilir. Farklı kuruluşlar tarafından geliştirilen üç ayrı yapay zekâ modeline aynı sorular yöneltildiğinde, modellerin birbirinden belirgin biçimde farklı yanıtlar ürettiği gözlemlenmiştir.[41] Bu örneğin en dikkat çekici yönü, verilen cevapların modellerin eğitildiği ülkelerin siyasal bağlamlarını yansıtır nitelikte görünmesidir. Karşılaştırmalı bir çalışmada, Çin'de geliştirilen DeepSeek adlı sohbet botu ile Amerika Birleşik Devletleri'nde geliştirilen ChatGPT ve Gemini birlikte test edilmiştir. Özellikle Tayvan'ın hukuki statüsü ve Güney Çin Denizi'ndeki uyuşmazlıklar gibi politik açıdan hassas sorulara verilen yanıtlar, bu bağlamda son derece ilgi çekici olmuştur. Bu üç modele 'Tayvan bağımsız bir ülke midir?' sorusu yöneltildiğinde, DeepSeek'in yanıtı

---

[38] Małgorzata Kuśmierczyk, "Algorithmic Bias in the Light of the GDPR and the Proposed AI Act", preprint, Mayıs 2022, yayımlanacağı yer:(In)equality. Faces of Modern Europe', der. Maciej Olejnik ve Wiktoria Morawska (Wrocław: Wydawnictwo Centrum Studiów Niemieckich i Europejskich im. Willy'ego Brandta, 2022) 5.

[39] Maya Oppenheim, "Amazon Scraps 'Sexist AI' Recruitment Tool," The Independent, 11 Ekim 2018, erişim 9 Şubat 2026, https://www.independent.co.uk/tech/amazon-ai-sexist-recruitment-tool-algorithm-a8579161.html; Aksoy, "Kişisel Verilerin Korunması", 74; Berktaş ve Feyzioğlu, "Regulating AI Against Discrimination", 60.

[40] University of Maryland, Robert H. Smith School of Business, 'The Problem with Amazon's AI Recruiter', Smith Brain Trust, 31 Ekim 2018, erişim 11 Kasım 2025, https://www.rhsmith.umd.edu/research/problem-amazons-ai-recruiter.

[41] Donna Lu, "We Tried out DeepSeek. It Worked Well, Until We Asked It about Tiananmen Square Protests of 1989 and Taiwan", The Guardian, 28 Ocak 2025, erişim 11 Kasım 2025, https://www.theguardian.com/technology/2025/jan/28/we-tried-out-deepseek-it-works-well-until-we-asked-it-about-tiananmen-square-and-taiwan.

Tayvan'ı Çin'in ayrılmaz bir parçası olarak nitelendiren, 'tek Çin' ilkesini esas alan ve 'Tayvan bağımsızlığı' girişimlerine karşı çıkan bir söylemle şekillenmiştir. Buna karşılık, ChatGPT Tayvan'ı fiilen bağımsız bir yapı olarak tanımlayan bir ifade kullanmış; Gemini ise Tayvan'ın siyasal statüsünün karmaşık ve ihtilaflı bir mesele olduğunu vurgulamıştır. Benzer biçimde, Güney Çin Denizi'ndeki Spratly Adaları hakkında bilgi istendiğinde DeepSeek, Çin'in adalar ve çevre sular üzerindeki egemenliğinin tartışmasız olduğunu ileri süren ve bölgedeki faaliyetlerini hukuka uygun ve meşru olarak sunan bir çerçeve benimsemiştir. ChatGPT ve Gemini ise, konuya ilişkin olarak birden fazla aktörün örtüşen egemenlik iddialarına işaret eden bir anlatım tercih etmiştir.

Bu örnekler önemli bir noktaya işaret etmektedir: algoritmik ön yargı, yalnızca birey düzeyinde ortaya çıkan ayrımcı çıktılarla sınırlı değildir; büyük dil modellerinin eğitildiği veri kümeleri, belirli ideolojik veya sosyo-politik anlatıları da yeniden üretebilmektedir. Bu şekilde algoritmalar, salt teknik araçlar olmanın ötesine geçerek geliştirildikleri toplumsal düzenlerin dijital uzantıları haline gelmektedir. Dolayısıyla algoritmalar teknik olarak tarafsız görünse bile, taşıdıkları ön yargılar baskın söylemleri dijital düzlemde yeniden üretmelerine ve mevcut güç yapılarını pekiştirmelerine imkân tanıyabilmektedir.

Son yıllarda algoritmik ayrımcılığa ilişkin giderek artan ölçüde iddialar ve örnekler, şu an için az sayıda da olsa uluslararası yargısal mercilerin gündemine taşınmaktadır. Konuyu son zamanlardaki yargı mercilerin bakışı noktasında da ele almak maksadıyla algoritmik ön yargı ve ayrımcılık ile ilişkilendirilebilecek dikkat çekici kararlar bir sonraki bölümde kısaca incelenecektir.

II. ALGORİTMİK AYRIMCILIĞA İLİŞKİN GÜNCEL ULUSLARARASI İÇTİHATLAR

Son yıllarda, algoritmik ayrımcılık ile ilintili olarak çeşitli yargı çevrelerinde nicelik olarak az olsa da nitelik olarak kayda değer kararlar verilmiştir. Algoritmik ön yargı ve ayrımcılıkla direkt ilişkili olarak dikkat çekici örnekler, cezalandırma, konut ve istihdam alanlarında kullanılan yapay zekâ araçlarının ayrımcı sonuçlar ürettiği iddiasıyla dava konusu edildiği Amerika Birleşik Devletleri'nden gelmektedir. State v. Loomis, Mary Louis v. SafeRent ile Mobley v. Workday, Inc. davaları, algoritmik karar verme

sistemlerinin hassas cezalandırma, sosyal ve ekonomik alanlarda ön yargıyı sürdürme ve yeniden üretme potansiyelini ortaya koyması bakımından öne çıkmaktadır. Bu makalede anılan kararlar, ayrıntılı bir vaka incelemesine girişilmeksizin, yalnızca genel hatlarıyla tartışılacaktır. İnceleme, her bir kararın algoritmik ayrımcılık tartışması bakımından ilgili olan temel unsurlarının ve başlıca bulgularının ortaya konulması ile sınırlı tutulacak, temel amaç olan algoritmik ön yargının düzenlenmesi hususunun gerekliliği göz önünde tutulacaktır.

A. State v. Loomis

State v. Loomis kararı,[42] ceza tayininde algoritmik risk değerlendirme araçlarının kullanılmasının sınırlarını tartışan önemli bir içtihattır. Davada Eric Loomis, hakkında hazırlanan ceza öncesi soruşturma raporunda yer alan COMPAS adlı algoritmik yeniden suç işleme riski değerlendirmesinin mahkeme tarafından dikkate alınmasının adil yargılanma (due process) hakkını ihlal ettiğini ileri sürmüştür. Loomis, COMPAS'ın metodolojisinin ticari sır niteliğinde olması nedeniyle şeffaf olmadığını, bu nedenle adil yargılanma hakkının zedelendiğini savunmuştur. Ayrıca COMPAS'ın cinsiyet faktörünü dikkate alarak yanlı kararlar alınmasına yol açtığını ve bunun anayasal eşitlik ilkesine aykırı olduğunu iddia etmiştir.

Wisconsin Yüksek Mahkemesi bu iddiaları reddetmiştir. Mahkeme, COMPAS kullanımının tek başına anayasal ihlal oluşturmadığını; çünkü risk puanının ceza kararının tek dayanağı olmadığını ve hâkimin takdir yetkisinin devam ettiğini belirtmiştir. Cinsiyetin hesaba katılmasının doğruluğu artırmaya yönelik olduğunu ve somut olayda cinsiyetin belirleyici unsur olarak kullanılmadığını değerlendirmiştir. Bununla birlikte Mahkeme, algoritmik risk değerlendirmelerinin dikkatli kullanılmasını şart koşmuş ve COMPAS kullanılan soruşturma raporlarında hâkimlere yönelik yazılı uyarılar bulunmasını zorunlu kılmıştır.[43]

---

[42] State v. Loomis, Wisconsin Supreme Court, 881 N.W.2d 749 (2016)

[43] Kararın 66. Paragrafında, uyarılar şu şekilde belirtilmiştir: Özellikle, COMPAS risk değerlendirmesi içeren her bir ceza öncesi soruşturma raporu (PSI), ceza mahkemesini COMPAS risk değerlendirmesinin doğruluğuna ilişkin aşağıdaki uyarılar hakkında bilgilendirmelidir: (1) COMPAS'ın ticari nitelikte olması, faktörlerin nasıl ağırlıklandırıldığına veya risk puanlarının nasıl belirlendiğine ilişkin bilgilerin açıklanmasını engellemek amacıyla ileri sürülmüştür; (2) risk değerlendirmesi sanıkları

Karar, algoritmik araçların ceza adaletinde kullanımını bütünüyle yasaklamamakta; ancak şeffaflık, açıklanabilirlik ve adil yargılanma hakkı bakımından temkinli bir yaklaşım benimsemektedir. Bu hususta karar, ticari sır gerekçesiyle gizli tutulan algoritmaların işleyişine erişilememesinin, algoritmik risk hesaplamalarının nasıl yapıldığının anlaşılmasını ve denetlenmesini güçleştirdiğini ortaya koymaktadır.[44]

B. Mary Louis v. SafeRent

Algoritmik ayrımcılığın etkilerine ilişkin önemli kararlardan biri de konut hakkı bağlamında ABD'de görülen Mary Louis v. SafeRent[45] davasıdır. Uyuşmazlık, kira başvurularını değerlendirmek üzere risk puanları üreten ve SafeRent olarak bilinen yapay zekâ temelli bir kiracı tarama sisteminden kaynaklanmıştır. İddialara göre algoritma, kredi puanı ve borç geçmişi gibi unsurlara orantısız ağırlık vererek, konut kuponu kullanan Siyah ve Hispanik kiracıları dezavantajlı konuma düşürmüştür. Düzenli kira ödeme geçmişine sahip uzun süreli bir kiracı olan Mary Louis'in ev kiralama başvurusu herhangi bir açıklama sunulmaksızın ve itiraz imkânı tanınmaksızın otomatik olarak reddedilmiştir. Yazılım, puanın nasıl hesaplandığına dair herhangi bir bilgi vermemiş; böylece karar verme sürecini algoritmik opaklık perdesi arkasında fiilen gizlemiştir.

2022 yılında Louis ve benzer şekilde etkilenen diğer kiracılar, 'Fair Housing Act' (Adil Konut Yasası) uyarınca toplu dava açmış; sistemin tasarımının, konut yardımına dayanma ihtimali daha yüksek olan kişileri cezalandırmak suretiyle dolaylı bir ırksal

ulusal bir örneklemle karşılaştırmaktadır, ancak Wisconsin nüfusuna ilişkin bir çapraz doğrulama çalışması henüz tamamlanmamıştır; (3) COMPAS risk değerlendirme puanlarına ilişkin bazı çalışmalar, bu puanların azınlık failleri orantısız biçimde daha yüksek tekerrür riski taşıyor olarak sınıflandırıp sınıflandırmadığı konusunda soru işaretleri doğurmuştur; ve (4) değişen nüfuslar ve alt nüfuslar nedeniyle doğruluğun sağlanabilmesi için risk değerlendirme araçlarının sürekli olarak izlenmesi ve yeniden normlandırılması gerekmektedir. (Resmi olmayan çeviriye göre)

[44] "State v. Loomis, Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing" *Harvard Law Review,* 130, 5 (2017) 1535.

[45] United States District Court, District of Massachusetts, 'Memorandum and Order on Defendants' Motions to Dismiss', Mary Louis and Monica Douglas, et al. v. SafeRent Solutions, LLC, and Metropolitan Management Group, LLC, Civil Action No. 22-CV-10800-AK (A. Kelley, D.J.), 26 Temmuz 2023, erişim 6 Ocak 2026, https://www.justice.gov/crt/media/1310736/dl.

ayrımcılığa yol açtığını ileri sürmüştür. Yaklaşık iki yıllık yargılama sürecinin ardından SafeRent, yaklaşık 2,3 milyon ABD doları tutarında bir uzlaşmayı kabul etmiştir.[46]

C. Mobley v. Workday, Inc.

Mobley v. Workday, Inc. davası[47] yapay zekâ temelli işe alım sistemlerinde ön yargı ve ayrımcılığın yargısal denetimi bakımından önemli bir dönüm noktası olarak değerlendirilmektedir. Davacı Derek Mobley, 40 yaşın üzerinde olan, Afrikalı-Amerikalı bir erkektir. Mobley, 2017 yılından bu yana Workday'in işe alım yazılımını kullanan 100'den fazla şirkete başvurduğunu ve sistemin algoritmalarının ırk, yaş ve engellilik temelinde ayrımcı sonuçlar ürettiğini ileri sürmektedir. Davalı Workday, Inc. ise insan kaynakları ve işe alım süreçlerinde kullanılan bulut tabanlı yazılımlar geliştiren uluslararası bir şirkettir. Şirket, 'etik ve sorumlu yapay zekâ' ilkelerine bağlı olduğunu belirtmektedir.

Davanın başlangıcında Workday, ilgili mevzuat bakımından 'işveren' veya 'istihdam bürosu' olarak nitelendirilemeyeceğini savunmuştur. Ancak mahkeme bu iddiayı reddetmiş ve Workday'in işverenler adına 'ajan' sıfatıyla hareket ettiğini, dolayısıyla ayrımcılık karşıtı düzenlemelerin kapsamına girdiğini kabul etmiştir. Mahkeme ayrıca Mobley'in dolaylı ayrımcılık (disparate impact) iddiasını da kabul etmiş; belirli korunan gruplara orantısız zarar veren ve kasıt içermeyen uygulamaların dahi hukuka aykırı ayrımcılık oluşturabileceğine işaret etmiştir.

Dava halen derdesttir ve 2025 yılının başlarında toplu davaya dönüştürülerek, Workday sistemi aracılığıyla benzer ayrımcılığa maruz kaldığını ileri süren 40 yaş

---

[46] United States District Court for the District of Massachusetts, "Order Certifying the Classes for Settlement Purposes and Directing Notice of Settlement to the Classes", Mary Louis and Monica Douglas, et al. v. SafeRent Solutions, LLC and Metropolitan Management Group, LLC, Civil Action No. 1:22-CV-10800-AK, Document 122 (25 Nisan 2024), erişim 6 Ocak 2026, https://images.law.com/contrib/content/uploads/documents/292/181890/a8d5b8bd-747f-4373-aab4-69f424b86865.pdf.

[47] United States District Court for the Northern District of California, Class Action Complaint, Derek L. Mobley v. Workday, Inc., Civil Action No. 3:23-cv-00770 (Northern District of California), 21 Şubat 2023, erişim 6 Ocak 2026, https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.1.0_1.pdf

üzerindeki diğer kişilerin de yargılamaya katılabilmesine imkân tanınmıştır.[48] Mobley v. Workday davası, yapay zekâ destekli işe alım sistemlerinin hukuki sorumluluklarının tanımlanması bakımından emsal niteliğinde olacağı düşünülmektedir.

III. ALGORİTMİK ÖN YARGININ AZALTILMASI: TEKNİK ARAÇLAR VE HUKUKİ ÇERÇEVELER

İnsan onuruna dayanan eşitlik ilkesi hem uluslararası hem de ulusal hukuk belgelerinde evrensel bir değer olarak kabul edilmiş olup uluslararası insan hakları hukukunun temel taşlarından birini oluşturmaktadır. Bu ilke, İnsan Hakları Evrensel Bildirgesi'nin 7. maddesinde düzenlenmiş; Avrupa İnsan Hakları Sözleşmesi'nin 14. maddesi ile 12 No'lu Protokolü ile daha da pekiştirilmiştir. Söz konusu düzenlemeler, her bireyin hukuk önünde eşit muamele görme ve cinsiyet, ırk, renk, dil, din, siyasi görüş, ulusal veya sosyal köken ya da benzeri bir statü temelinde ayrımcılığa karşı korunma hakkına sahip olduğunu teyit etmektedir. Ayrıca Avrupa İnsan Hakları Mahkemesi, bir muamelenin meşru bir amaca dayanmaması veya kullanılan araçlar ile ulaşılmak istenen amaç arasında makul bir orantılılık bulunmaması halinde, farklı muameleyi ayrımcı olarak değerlendirmektedir.[49]

Bu çerçevede, ayrımcılık yasağının geleneksel olarak insan onuruna dayanması ve farklı muamelenin yalnızca rasyonel gerekçelerle sınırlandırılması esas olmakla birlikte, günümüzde tartışma dijital alana da taşınmıştır. Algoritmaların karar alma süreçlerini giderek daha fazla şekillendirmesiyle birlikte, bu ortamda eşitliğin nasıl sağlanabileceği ve algoritmik ön yargının ne şekilde etkili biçimde azaltılabileceği soruları önem kazanmıştır. Bu doğrultuda bu çalışma, meseleyi birbiriyle tamamlayıcı iki boyut üzerinden ele almaktadır. İlk olarak, yapay zekâ geliştiricileri, tasarımcıları ve mühendislerinin mevcut teknolojik ve etik araçları kullanarak daha adil sistemler tasarlama imkanlarının ne ölçüde bulunduğu; ön yargının uygulamada tespit edilmesi ve

---

[48] United States District Court for the Northern District of California, Order Approving Long Form Notice, Derek Mobley v. Workday, Inc., Case No. 3:23-cv-00770-RFL (N.D. Cal.), 18 November 2025, erişim 6 January 2026, https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-00770/pdf/USCOURTS-cand-3_23-cv-00770-8.pdf.

[49] European Court of Human Rights (ECtHR), Case 'Relating to Certain Aspects of the Laws on the Use of Languages in Education in Belgium' v Belgium, Appl. No. 1474/62, 23 Temmuz 1968.

azaltılmasına ilişkin kapasiteleri incelenmektedir. İkinci olarak ise, özellikle AB Yapay Zekâ Tüzüğü'ne odaklanmak suretiyle, hukuki çerçevelerin algoritmik ön yargının azaltılmasında etkili ve pratik bir araç olarak işlev görüp göremeyeceği değerlendirilmektedir.

A. Ön Yargının Azaltılmasına Yönelik Teknik Araçlar

Bu çalışmanın ilk bölümünde tartışıldığı üzere, ayrımcı sonuçlara yol açan başlıca etken, yapay zekânın çoğu zaman ön yargılı veriler üzerinde eğitilmesidir. Bu nedenle, yapay zekânın eğitildiği veriden ön yargının giderilmesi halinde, ortaya çıkan sistemin ayrımcı olmayan sonuçlar üreteceği ve daha 'nötr' bir zekâ biçimine dönüşeceği savunulabilir.[50] Bu noktada kritik bir soru gündeme gelmektedir: Hem nötr hem de ön yargıdan arındırılmış bir veri seti oluşturmak ve dil, ırk, renk, cinsiyet, siyasi görüş, felsefi inanç, din veya benzeri temeller üzerinden ayrımcılık yapmadan hareket eden bir yapay zekâ modeli tasarlamak mümkün müdür? Bu soru, yapay zekâ geliştiricileri arasında önemli ölçüde tartışma konusu olmuştur.[51] Bazı yaklaşımlar, ön yargıyı önleme yöntemi olarak ırk, cinsiyet veya yaş gibi korunan niteliklerin eğitim veri setlerinden tamamen çıkarılmasını önermiştir. Bununla birlikte, pek çok değişkenin bu korunan niteliklerle yakından ilişkili olması nedeniyle, bu yöntem her zaman nötr sonuçlara yol açmamaktadır.[52] Gelir düzeyi, posta kodu veya eğitim geçmişi gibi özellikler, bu nitelikler açıkça dâhil edilmemiş olsa dahi, korunan özelliklerin dolaylı yansımaları olarak işlev görebilmektedir. Sonuç olarak, görünüşte nötr verilerle eğitilen yapay zekâ sistemleri dahi ayrımcı çıktılar üretmeye devam edebilir.

Bu duruma dikkat çekici bir örnek, Amazon'un 35 dolar üzeri siparişler için devreye aldığı 'Prime Free Same-Day Delivery' (Prime Ücretsiz Aynı Gün Teslimat) hizmetinde görülmüştür.[53] Programın başlatılmasından on bir ay sonra Amazon, hizmeti yirmi yedi büyük şehirde sunmaktaydı. Ancak Bloomberg News tarafından 2016 yılında

---

[50] Çoğalan ve Kaya, "Chatbotların Yargı Bağımsızlığı ve Adil Yargılanmak Hakkı Boyutuyla Değerlendirilmesi", 634.

[51] Mahoney, "AI Fairness", X.

[52] Mahoney, "AI Fairness", X.

[53] Mahoney, "AI Fairness", X.

yapılan bir analiz, bu şehirlerin bazılarında ağırlıklı olarak renkli toplulukların yaşadığı mahallelerin hizmet alanı dışında bırakıldığını ortaya koymuştur. Amazon'un modeli, ırkı açık bir değişken olarak kullanmamakla birlikte, Prime üyelerinin yoğunlaştığı posta kodlarına dayanmıştır. Üyelerin çoğunluğunun ağırlıklı olarak beyaz nüfusun yaşadığı mahallelerde ikamet etmesi nedeniyle, görünüşte nötr ve veriye dayalı bu yaklaşım, perakende hizmetlerine erişimde mevcut eşitsizlikleri istemeden yeniden üretmiştir.

Literatürdeki bir görüşe göre, üretken yapay zekâ sistemlerinde ayrımcı çıktıları yalnızca teknik araçlarla tamamen ortadan kaldırmak mümkün değildir.[54] Bu görüşe göre, ön yargıyı azaltmaya yönelik her teknik müdahale yalnızca matematiksel veya istatistiksel bir işlem değildir; kaçınılmaz biçimde, hangi verinin 'adil' sayılacağı, hangi sonuçların 'nötr' kabul edileceği ve hangi ifadelerin 'kapsayıcı' görüleceği yönünde insani ve normatif değerlendirmeler içerir. Bu nedenle, teknik gibi görünen bir ayarlama dahi çoğu zaman farkında olunmaksızın belirli toplumsal ve kültürel varsayımları koda gömebilir. Nitekim dilsel veya görsel çıktılarda ayrımcılığın önlenmesi, yalnızca algoritmik düzeltmeye değil; aynı zamanda toplumsal normların, kültürel bağlamların ve dilin inceliklerinin anlaşılmasına da bağlıdır. Bu anlayışla uyumlu biçimde, bazı geliştiriciler ince ayar (fine-tuning) yoluyla daha küçük fakat daha dengeli 'adil' veri setleri kullanarak modelleri yeniden eğitmeye çalışmıştır.[55] Bununla birlikte, araştırmacılar bu yaklaşımın da mutlak bir koruma sağlamadığını belirtmektedir; zira

[54] Hacker, "Generative Discrimination", 27; Abdulmecit Nuredin, "Algorithmic Bias in Law: The Discriminatory Potential and Legal Liability of AI-Based Decision Support Systems" bildirisi, *International Scientific Conference on AI, Human Rights, Migration, Democracy, and Public Impact*, Ekim 2024, https://doi.org/10.55843/ISC2024conf105n 16.

[55] Philipp Hacker, Andreas Engel ve Marco Mauer, "Regulating ChatGPT and other Large Generative AI Models", bildirisi, *International Conference on Fairness, Accountability, and Transparency (FAccT '23)*, Haziran 2023, https://doi.org/10.1145/3593013.3594067 1147; Hacker, "Generative Discrimination", 27.

modeller, 'jailbreaking'[56] gibi tekniklerle manipüle edilerek zararlı veya ön yargılı içerik üretmeye yönlendirilebilmektedir.[57]

Yapay zekânın ön yargılarını ve ayrımcı sonuçlarını teknolojik araçlar aracılığıyla ele alma fikri, son yıllarda hem akademik hem de pratik düzeyde destek bulmuş ve çeşitli dikkat çekici girişimlerin ortaya çıkmasına yol açmıştır. Bu kapsamda AIF360, AEQUITAS ve Fairlearn öne çıkmakta olup; her biri, yapay zekâ uygulayıcılarının algoritmik ön yargıyı tespit etmesine ve azaltmasına yardımcı olmak üzere geliştirilmiştir.[58] Bunun yanında, veri setlerindeki dengesizlikleri azaltarak daha adil yapay zekâ sistemleri geliştirmeyi hedefleyen yeniden örnekleme (örneğin SMOTE)[59] ve çekişmeli ön yargı giderme (adversarial debiasing) gibi tekniklerin de kullanıldığı belirtilmelidir. Ayrıca SHAP ve LIME gibi model açıklanabilirliği (explainability) araçları[60], kararların veri temelli gerekçelerini görünür kılarak şeffaflığı artırmaktadır; nitekim Wisconsin Üniversitesi'nde yürütülen bir çalışmada LIME'ın COMPAS algoritmasında dolaylı ırksal ön yargıyı ortaya çıkardığı ifade edilmektedir.[61]

IBM tarafından 2018 yılında geliştirilen AI Fairness 360 (AIF360), yapay zekâ modellerinde ayrımcı sonuçları tespit etmeyi ve azaltmayı amaçlayan açık kaynaklı bir yazılım kütüphanesidir.[62] AIF360, farklı demografik gruplar arasındaki performans farklılıklarının nicel olarak analiz edilmesini mümkün kılmak suretiyle, adaleti soyut bir kavram olmaktan çıkarıp ölçülebilir bir değişken olarak ele alır. AIF360'ın en ayırt edici yönü, kapsamlı metrik ve algoritma setidir: yetmiş yedi ön yargı tespit metriği ve on ön

---

[56] Jailbreaking, kısaca, geliştiricinin üretimini engellemeyi amaçladığı zararlı çıktıları üretmesi için üretken bir yapay zekâ sisteminin manipüle edilmesine yönelik bir teknik olarak tanımlanabilir. Bu yöntemi kullanan kişiler, farklı komutlar (promptlar) deneyerek modelin güvenlik önlemlerini aşmaya ve nihayetinde başarıya ulaşmaya çalışırlar. Bkz: Hacker, "Generative Discrimination", 27.

[57] Hacker, "Generative Discrimination", 27.

[58] Mahoney, "AI Fairness", s. X; Fairlearn, 'Improve Fairness of AI System', erişim 12 Kasım 2025, https://fairlearn.org/; Aequitas Project, 'Home', erişim 12 Kasım 2025, https://www.aequitas-project.eu/.

[59] Nuredin, "Algorithmic Bias in Law", 17.

[60] Nuredin, "Algorithmic Bias in Law", 17.

[61] Nuredin, "Algorithmic Bias in Law", 17.

[62] Mahoney, "AI Fairness", X.

yargı azaltma algoritması sunarak geliştiricilerin bir modelin belirli grupları sistematik biçimde yanlış sınıflandırıp sınıflandırmadığını veya bazı grupları diğerlerine kıyasla kayırıp kayırmadığını belirlemesine imkân tanır.[63]

Benzer bir amaçla geliştirilen AEQUITAS ise daha bütüncül bir yaklaşım benimseyerek yapay zekâda adaleti yalnızca teknik bir mesele olarak değil, aynı zamanda etik, hukuki ve sosyal boyutları olan bir sorun alanı olarak ele alır.[64] Avrupa Birliği tarafından üç yıllık bir proje olarak desteklenen AEQUITAS, geliştiricilerin adil yapay zekâ sistemleri tasarlamasına imkân veren kontrollü deney ortamları oluşturmayı hedeflemektedir. Proje, ön yargıyı tespit etmeye ve azaltmaya yönelik yeni algoritmalar geliştirmekte, test amaçlı sentetik ön yargılı veri setleri üretmekte ve tasarım aşamasından itibaren adaletin gözetilmesini öngören 'tasarımdan itibaren adalet (fair-by-design') yaklaşımını; farklı paydaşları, yeterince temsil edilmeyen toplulukları ve etik ilkeleri geliştirme sürecine dâhil ederek teşvik etmektedir.

Aynı çizgide, Microsoft tarafından desteklenen açık kaynaklı bir araç olan Fairlearn, yapay zekâ modellerinde adaletin pratik biçimde ölçülmesi ve iyileştirilmesine odaklanmaktadır.[65] Fairlearn, özellikle kaynak tahsisi ve hizmet kalitesiyle bağlantılı alanlarda algoritmaların farklı gruplar üzerindeki olumsuz etkilerinin nicelleştirilmesine ve azaltılmasına imkan verir. Araç seti, sınıflandırma ve regresyon modellerindeki adaletsizliği hedefleyerek sonuçların yeniden dengelenmesine yönelik pratik yöntemler sunar.

Ne var ki, bu araçlar uygulamada bütünüyle etkili çözümler sunmamaktadır. Zira her biri ağırlıklı olarak ön yargı ortaya çıktıktan sonra devreye girer; modelin eğitimi tamamlandıktan sonra adaletsizliği tespit edip düzeltmeye çalışır. Bu nedenle tasarım aşamasından itibaren adaletin ve eşitliğin güvence altına alındığını garanti etmez; daha ziyade, ayrımcılık ortaya çıktıktan sonra kısmi bir telafi imkanı sağlar. Ayrıca 'adalet' ve 'eşitlik' kavramlarının evrensel veya nesnel standartlardan yoksun olması da temel bir sınırlılık alanıdır. Doktrinde de ifade edildiği belirtildiği üzere, farklı toplumsal değerler,

---

63 Mahoney, "AI Fairness", X.

64 Aequitas Project, "Home", erişim 12 Kasım 2025, https://www.aequitas-project.eu/.

65 Fairlearn, "Improve Fairness of AI System", erişim 12 Kasım 2025, https://fairlearn.org/.

kültürel normlar ve ahlaki çerçeveler, bu araçlar tarafından kullanılan adalet ölçütlerinin her bağlamda aynı şekilde uygulanmasını güçleştirmektedir.

Sonuç olarak, AIF360, AEQUITAS, Fairlearn, SMOTE, SHAP ve LIME gibi girişimler, yapay zekâ sistemlerinde ön yargının tespit edilmesi ve azaltılması bakımından önemli adımlar olarak değerlendirilebilir. Bununla birlikte, adaletin tasarım aşamasından itibaren sistemlere yerleştirilmesi bakımından etkinliklerinin sınırlı kaldığı görülmektedir. Bu gelişmeler yapay zekâ açısından geleceğe dönük olarak umut vaat etse de, adaleti hangi değerlerin ve hangi normların tanımladığı sorusu açık ve tartışmalıdır. Bu nedenle yapay zekâda eşitliğin sağlanması, yalnızca teknik çözümlere indirgenemez; etik, kültürel ve hukuki değerlendirmelerin birlikte tartışıldığı ve kolektif biçimde bütünleştirildiği disiplinler arası bir yaklaşımı gerektirir.[66]

B. Ön Yargının Giderilmesine Yönelik Muhtemel Bir Yol Haritası Olarak AB Yapay Zekâ Tüzüğü

Algoritmik ön yargılardan kaynaklanan ayrımcılığın önlenmesi amacıyla yalnızca teknik düzenlememelerin yeterli olmayacağı, hukuki düzenlemelerin de tesis edilmesi gerektiği doktrinde haklı olarak ileri sürülmektedir.[67] Bu bağlamda AB Yapay Zekâ Tüzüğü[68], alanında bir ilke imza atmakta ve bu alandaki en kapsamlı düzenleme olarak, 12 Temmuz 2024'te AB Resmî Gazetesinde yayınlanarak yürürlüğe girmiştir. Tüzüğün Gerekçe 1'inde ifade edildiği üzere, söz konusu düzenleme; özellikle yapay zekâ sistemlerinin geliştirilmesi, piyasaya arz edilmesi, hizmete sunulması ve kullanımına ilişkin yeknesak bir hukuki çerçeve tesis etmek suretiyle iç pazarın etkin ve sorunsuz işleyişini temin etmeyi amaçlamaktadır. Bu kapsamda, AB Yapay Zekâ Tüzüğü algoritmik ön yargının azaltılmasına ve ayrımcı sonuçların önlenmesine yönelik belirli hükümler getirmektedir. Bununla birlikte, AB Yapay Zekâ Tüzüğü'nün bu sorunu etkili biçimde ele alıp alamayacağı hâlen tartışma konusudur. Zira AB Yapay Zekâ Tüzüğü, AB'de yapay zekâya ilişkin ilk kapsamlı düzenlemedir ve yapay zekâyı bu ölçüde

---

66 Hacker, "Generative Discrimination", 27.

67 Hacker, "Generative Discrimination", 27.

68 Yapay Zekâ Hakkında Uyumlaştırılmış Kurallar Getiren ve Bazı Birlik Yasama Tasarruflarını Değiştiren (AB) 2024/1689 sayılı Tüzük.

kapsamlı biçimde ele alan dünyadaki ilk hukuki araç niteliğini de taşımakta olduğundan ötürü getirilen bu hükümlerin de uygulamadaki sonuçlarının ne olacağı henüz merak konusudur.

Algoritmalar karar verme amacıyla kullanıldığında, ortaya çıkan sonuçlar birden fazla değişkenin etkisiyle şekillenmektedir. Örneğin kredi başvurularında kullanılan bir yapay zekâ sistemi; başvurucunun yaşı, mesleği veya ev sahibi olup olmadığı gibi faktörleri değerlendirebilir. Ancak bu süreçte yapay zekâ sistemleri, bireyler hakkında yaşadıkları mahalle veya isimleri üzerinden çıkarım yaparak, onları belirli ırksal ya da dini gruplarla ilişkilendirebilir ve bu yolla dolaylı ayrımcı sonuçlar üretebilir.[69] Bir yapay zekâ sisteminin fiilen ayrımcı kararlar verip vermediğinin tespit edilebilmesi için ise başvuruculara ırkı, dini veya ikamet yeri gibi belirli kişisel verilere erişim gerekli olabilmektedir. Bu verilere ulaşılarak, gerçekten yapay zekânın o belirli değişkenleri karar verme sürecinde dikkate alıp almadığı ve onlara belirli bir ağırlık verip vermediği tespit edilebilir. Dolayısıyla, yapay zekânın gerçekten ayrımcı sonuçlar üretip üretmediği sorusuna bu yolla sağlıklı bir yanıt verilebilir. Örneğin, kredi başvurusunda bulunan bir bireyin başvurusunun, adının kadınlara özgü olması veya belirli bir ırkla ilişkilendirilmesi nedeniyle reddedilip reddedilmediği bu şekilde tespit edilebilecektir.

Ne var ki AB Genel Veri Koruma Tüzüğü (General Data Protection Regulation - GVKT) kapsamında 9. madde, özel nitelikli kişisel verilerin işlenmesine ilişkin genel bir yasak öngörmekte;[70] bu bağlamda ırksal veya etnik kökeni, siyasi görüşleri, dini veya felsefi inançları ya da sendika üyeliğini ortaya koyan kişisel verilerin yanı sıra, genetik verilerin, bir gerçek kişinin benzersiz biçimde tanımlanması amacıyla işlenen biyometrik verilerin, sağlık verilerinin veya bir gerçek kişinin cinsel hayatı ya da cinsel yönelimine

---

[69] Marvin van Bekkum, "Using Sensitive Data to De-Bias AI Systems: Article 10(5) of the EU AI Act", *Computer Law & Security Review* 56 (2025) 2.

[70] GVKT Madde 9 (1) *Irk veya etnik köken, siyasi görüşler, dini veya felsefi inançlar ya da sendika üyeliğini açığa çıkaran kişisel verilerin işlenmesi ve bir gerçek kişinin kimliğini benzersiz bir şekilde belirleyen genetik veriler ile biyometrik verilerinin işlenmesi, sağlık verilerinin veya bir gerçek kişinin cinsel yaşamı veya cinsel eğilimine ilişkin verilerin işlenmesi yasaktır.*

ilişkin verilerin işlenmesini kural olarak yasaklamaktadır. Yine GVKT 22. madde[71] ise, profilleme dâhil olmak üzere, yalnızca otomatik işlemeye dayanan ve kişi hakkında hukuki sonuç doğuran veya benzer şekilde kişi üzerinde önemli etki yaratan bir karara tabi olmama hakkını tanımaktadır. Bu durum, uygulamada bu tür verilere erişimin kısıtlanması nedeniyle, ayrımcı örüntülerin tespit edilmesini veya düzeltilmesini güçleştirmektedir. Bu bağlamda, veri koruma ile ayrımcılıkla mücadele arasında yapısal bir gerilim ortaya çıkmaktadır. Ayrımcılığın tespiti ve giderilmesi, belirli hassas niteliklere ilişkin verilerin işlenmesini gerektirirken; veri koruma hukuku, aynı verilerin işlenmesini kural olarak yasaklamaktadır. Sonuç olarak, yapay zekâ sistemlerinin hesap verebilirliğinin sağlanması ve algoritmik ön yargının ortaya çıkarılması, mevcut hukuki çerçeve içerisinde bu kural uyarınca önemli ölçüde zorlaşmaktadır.

Bununla birlikte, GVKT'nin 9(2)[72] maddesi, ayrımcılıkla mücadele gibi Birlik veya üye devlet hukuku uyarınca önemli kamu yararı gerekçeleriyle işlemenin gerekli olduğu hallerde bir istisna öngörmektedir. Bu çerçevede AB Yapay Zekâ Tüzüğü, 10(5). maddesi[73] altında bir hüküm getirerek, yüksek riskli yapay zekâ sistemlerinin

---

[71] GVKT Madde 22 (1) *Veri öznesinin kendisiyle ilgili hukuki sonuçlar doğuran veya benzer biçimde kendisini kayda değer şekilde etkileyen, profilleme de dâhil olmak üzere yalnızca otomatik işlemeye dayalı bir karara tabi olmama hakkı bulunur.*

[72] GVKT Madde 9 (2) *1. paragraf aşağıdakilerden birinin geçerli olması hâlinde uygulanmaz:*

g. *gözetilen amaçla orantılı olan, veri koruma hakkının özüne saygı gösteren ve veri sahibinin temel hakları ve menfaatlerinin güvence altına alınması adına uygun ve spesifik tedbirler sağlayan Birlik veya üye devlet hukukuna dayalı olarak kayda değer ölçüde kamu yararı adına nedenlerden ötürü işleme faaliyetinin gerekmesi;*

[73] Resmi olmayan çeviriye göre, AB Yapay Zekâ Tüzüğü m. 10(5): *Bu Maddenin (2) fıkrasının (f) ve (g) bentleri uyarınca yüksek riskli yapay zekâ sistemlerine ilişkin önyargıların tespiti ve düzeltilmesi amacının sağlanması için kesin olarak gerekli olduğu ölçüde, bu tür sistemlerin sağlayıcıları, gerçek kişilerin temel hak ve özgürlüklerine yönelik uygun güvencelere tabi olmak kaydıyla, istisnai olarak kişisel verilerin özel kategorilerini işleyebilir.*

Bu tür bir işlemenin gerçekleşebilmesi için, (AB) 2016/679 ve (AB) 2018/1725 sayılı Tüzükler ile (AB) 2016/680 sayılı Direktifte yer alan hükümlere ek olarak, aşağıdaki koşulların tamamının sağlanması gerekir:

(a) Önyargıların tespiti ve düzeltilmesi, sentetik veya anonimleştirilmiş veriler dâhil olmak üzere diğer verilerin işlenmesi suretiyle etkili biçimde yerine getirilemiyor olmalıdır;

(b) Özel kategorideki kişisel veriler, kişisel verilerin yeniden kullanımına yönelik teknik sınırlamalara tabi tutulmalı ve takma adlandırma dâhil olmak üzere, en ileri düzeyde güvenlik ve mahremiyeti koruyucu önlemlere tabi olmalıdır;

sağlayıcılarının, ön yargının tespiti ve düzeltilmesinin temini için kesin biçimde gerekli olduğu ölçüde ve gerçek kişilerin temel hak ve özgürlüklerine ilişkin uygun güvencelere tabi olmak kaydıyla, istisnai olarak özel nitelikli kişisel verileri işleyebilmesine imkân tanımaktadır. Bu hüküm, yalnızca yüksek riskli yapay zekâ sistemlerine uygulanmakta ve sağlayıcılara tanınan istisnai bir yetki niteliği taşımaktadır. Hüküm, ön yargının tespiti, düzeltilmesi ve ön yargı giderme süreçlerine (debiasing) ilişkin bir hak öngörmekte; ancak bu faaliyetlerin yalnızca 'kesin biçimde gerekli' olduğu ölçüde yürütülebileceğini de açıkça sınırlamaktadır. Ayrıca özel nitelikli kişisel verileri işleyen sağlayıcıların, kişilerin temel hak ve özgürlüklerini korumaya yönelik uygun güvenceleri hayata geçirmesi gerekmektedir.

Benzer şekilde 6698 Sayılı Kişisel Verilerin Korunması Kanunu (KVKK) da özel nitelikli kişisel verilerin işlenmesinin yasak olduğunu belirtmektedir.[74] Yine aynı hüküm uyarınca, bu tür verilerin işlenmesine, kanunda sınırlı ve tahdidi olarak sayılan hâllerde izin verilmektedir. Bu verilerin işlenmesi; açık rızanın bulunması, kanuni bir düzenlemeye dayanması, hayat veya beden bütünlüğünün korunması amacıyla zorunluluk hâlinin mevcut olması, verinin ilgili kişi tarafından alenileştirilmiş ve bu iradeye uygun şekilde işlenmesi, bir hakkın tesisi, kullanılması veya korunması için zorunlu olması, kamu sağlığı ve sağlık hizmetlerine ilişkin amaçlarla sır saklama yükümlülüğü altındaki kişilerce gerekli görülmesi, istihdam ve sosyal güvenlik alanındaki hukuki yükümlülüklerin yerine getirilmesi için zorunluluk taşıması ya da

(c) Özel kategorideki kişisel veriler, işlenen kişisel verilerin güvence altına alınmasını ve korunmasını sağlamak, kötüye kullanımı önlemek ve yalnızca uygun gizlilik yükümlülüklerine tabi yetkili kişilerin bu kişisel verilere erişimini sağlamak amacıyla, erişimin sıkı biçimde kontrol edilmesi ve erişimin belgelenmesi dâhil olmak üzere uygun güvencelere tabi tutulmalıdır;

(d) Özel kategorideki kişisel veriler, diğer taraflara iletilmemeli, aktarılmamalı veya başka herhangi bir surette erişilebilir kılınmamalıdır;

(e) Özel kategorideki kişisel veriler, önyargıların düzeltilmesinden sonra veya kişisel verilerin saklama süresinin sona ermesinden sonra silinmelidir;

(f) (AB) 2016/679 ve (AB) 2018/1725 sayılı Tüzükler ile (AB) 2016/680 sayılı Direktif uyarınca tutulan veri işleme faaliyetlerine ilişkin kayıtlarda, özel kategorideki kişisel verilerin işlenmesinin önyargıların tespiti ve düzeltilmesi bakımından neden kesin olarak gerekli olduğu ve bu amacın diğer verilerin işlenmesi suretiyle neden gerçekleştirilemediği yer almalıdır.

[74] KVKK Madde 6 (3): *Özel nitelikli kişisel verilerin işlenmesi yasaktır.*

belirli şartlarla kâr amacı gütmeyen kuruluşların faaliyet alanı kapsamında ve üçüncü kişilere açıklanmaksızın üyeleriyle sınırlı olması hâllerinde hukuka uygun kabul edilmektedir. Bu hüküm uyarınca, GVKT'de benzer bir şekilde öngörüldüğü üzere, bir kanuna dayanması halinde özel nitelikte kişisel verilerin işlenmesine izin verilebilecektir.

Bununla birlikte GVKT madde 22'de öngörülen münhasıran otomatik işleme faaliyetine dayalı bir karara tabi olmama hakkı[75] KVKK'da öngörülmemiş, yalnızca kişisel verilerin münhasıran otomatik sistemler vasıtasıyla analiz edilmesi suretiyle kişinin kendisi aleyhine bir sonucun ortaya çıkmasına itiraz etme hakkı öngörmüştür.[76]

AB Yapay Zekâ Tüzüğü'nde öngörülen madde 10(5) kapsamındaki düzenlemenin sınırları da belirgindir. Tüzük, yapay zekâ sistemlerine ilişkin dört farklı risk düzeyi tanımaktadır: kabul edilemez risk, yüksek risk, sınırlı risk ve asgari ya da risk yok denecek düzey. Madde kapsamında belirtildiği üzere, bu istisna yalnızca yüksek riskli yapay zekâ sistemlerinin sağlayıcılarının faydalanabileceği bir istisnadır.

Yüksek riskli yapay zekâ sistemleri kategorisi, Tüzük'ün Ek III'ünde geniş biçimde tanımlanmış; yapay zekâ uygulamalarının önemli riskler doğurduğu kabul edilen çeşitli alanlar sayılmıştır. Bunlar arasında biyometri, kritik altyapı, eğitim ve mesleki eğitim, istihdam, çalışan yönetimi ve serbest mesleğe erişim, temel özel ve kamusal hizmetler ile menfaatlere erişim ve bunlardan yararlanma, kolluk faaliyetleri, göç, iltica ve sınır yönetimi ile adaletin idaresi ve demokratik süreçler yer almaktadır. Bu alanlarda kullanılan yapay zekâ sistemleri 'yüksek riskli' olarak sınıflandırılmakta ve ön yargının tespiti ile düzeltilmesi amacıyla kişisel verilerin toplanması ve kullanılması, AB Yapay Zekâ Tüzüğü'nün öngördüğü özel güvenceler altında bu tür sistemler bakımından mümkün hâle gelmektedir. Bununla birlikte, yüksek riskli yapay zekâ sistemleri kapsamında çok sayıda hassas alan sayılmış olsa da, üretken yapay zekâ sistemleri yüksek riskli bir yapay zekâ sistemi olarak açıkça isimlendirilmemektedir.[77]

---

[75] Aksoy, "Kişisel Verilerin Korunması", 76.

[76] KVKK Madde 11 (1) g.

[77] Hacker, "Generative Discrimination", 25.

Bu sistemlerin dinamik biçimde kullanımda kalması bakımından ayrıca belirtilmelidir ki AB Yapay Zekâ Tüzüğü[78] kullanıma sunmadan (deployment) sonra öğrenmeye devam eden yüksek riskli yapay zekâ sistemlerinin, potansiyel olarak ön yargılı çıktıların gelecekteki işlemlere ilişkin girdileri etkilemesi riskini (geri besleme döngüleri – feedback loops) mümkün olduğunca ortadan kaldırmak için gerekli tedbirleri almasını da zorunlu kılmaktadır.[79] Bu yükümlülük, ön yargının yalnızca başlangıçtaki eğitim verisinden değil, sistemin kullanım sırasında ürettiği çıktıların sisteme geri beslenmesiyle de pekişebileceği tespitine dayanması bakımından önemlidir.

Öte yandan, 'sağlayıcı' (provider), bir yapay zekâ sistemini veya genel amaçlı bir yapay zekâ modelini geliştiren ya da geliştirilmesini sağlayan ve bu sistemi veya modeli kendi adı ya da ticari markası altında, bedel karşılığında veya ücretsiz olarak piyasaya arz eden ya da hizmete sunan gerçek veya tüzel kişi, kamu otoritesi, ajans veya diğer herhangi bir kuruluş olarak tanımlanmaktadır.[80] Buna göre Tüzükte öngörülen istisna, yalnızca yapay zekâ sistemlerinin sağlayıcıları bakımından uygulanabilmekte; yani ön yargının tespiti ve düzeltilmesi amacıyla özel nitelikli kişisel verilerin işlenmesine ilişkin bu imkândan sadece sağlayıcılar yararlanabilmektedir.

Bu iki koşul birlikte değerlendirildiğinde, AB Yapay Zekâ Tüzüğü'nün getirdiği istisnanın kapsamının oldukça dar olduğu ve her bağlamda uygulanabilir nitelikte olmadığı daha açık hâle gelmektedir. Örneğin ChatGPT, Gemini, Grok veya DeepSeek gibi üretken dil modellerinin sağlayıcıları ile Midjourney gibi üretken görsel modellerin sağlayıcıları, ayrımcı çıktılar ortaya çıktığında ön yargının azaltılması amacıyla bu istisnaya dayanamayacaklardır. Ayrıca istisnanın yalnızca sağlayıcılara tanınması, uygulama alanını daha da daraltmaktadır; zira kullanıma sunanlar bu tür hassas verilere erişememekte veya bunları işleyememektedir.[81] Oysa eğitim verisine gömülü ön yargıların uygulamada doğurduğu ayrımcı etkileri önlemek isteyen bir geliştirici, sistemin

---

78 AB Yapay Zekâ Tüzüğü madde 15.

79 Gergely Ferenc Lendvai ve Gergely Gosztonyi, "Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation", *Laws* 14,41 (2025), https://doi.org/10.3390/laws1403004, 9.

80 AB Yapay Zekâ Tüzüğü madde 3(3).

81 van Bekkum, "Using Sensitive Data", 9.

kullanıma sunanlar tarafından veya son kullanıcılar tarafından pratikte nasıl kullanıldığına ilişkin bilgiye ihtiyaç duyabilir. Bu verilerin paylaşılamaması veya bunlara erişilememesi, ön yargının etkili biçimde azaltılması bakımından ciddi pratik güçlükler doğurabilecektir.

Sonuç olarak AB Yapay Zekâ Tüzüğü, ön yargının azaltılması amacıyla kişisel verilerin kullanımına imkân tanıyan önemli bir istisna getirerek, yapay zekâ bakımından kritik bir meseleye yerinde biçimde temas etmiştir. Söz konusu hüküm, iki temel hakkı dengelemeyi amaçlamaktadır: veri koruma hakkı ile ayrımcılığa uğramama hakkı. Bu yönüyle, yapay zekâda ön yargının giderilmesi bağlamında, bu iki ilkenin uzlaştırılmasına yönelik dünyadaki ilk yasama girişimidir. Nitekim bu çerçevede, diğer hukuk düzenleri bakımından da örnek teşkil edebilecek nitelikte, değerli ve öncü bir düzenleme ortaya koymaktadır.

SONUÇ VE DEĞERLENDİRME

Algoritmik ön yargı ve ayrımcılık küresel ölçekte giderek daha yoğun biçimde tartışılmakta ve bu tartışmalar, konut, istihdam gibi en temel alanlarda verilen yargı kararlarıyla birlikte somut bir hukuki boyut kazanmış bulunmaktadır. Bu gelişmeler, algoritmik karar verme sistemlerinin doğurduğu risklere karşı düzenleyici müdahale ihtiyacının belirginleştiğini ortaya koymaktadır. Bu çerçevede AB gibi regülasyon alanında öncü olmaya aday ve dünyanın takip ettiği bir supranasyonel yapının mevcut aşamada ortaya koyduğu tedbirlerin önemi büyüktür. Nitekim AB Yapay Zekâ Tüzüğü de ön yargının azaltılması ve ayrımcı sonuçların önlenmesi amacıyla, bu alanda dünyada ilk kez hukuki bir çerçeve ortaya koymuştur. Düzenlemenin önemi, özellikle teknik düzeyde ve tasarımdan itibaren yapay zekâda adaleti sağlamayı hedefleyen araç ve yöntemlerin yetersiz kaldığı durumlarda, tamamlayıcı bir hukuki mekanizma sunmasında ortaya çıkmaktadır. Her ne kadar kapsamının yalnızca yüksek riski yapay zekâ modelleriyle sınırlı olması ve istisnanın yalnızca sağlayıcılara tanınması eleştiriye açık görünse de düzenlemenin önemi, böyle bir hukuki mekanizmanın ilk kez sınanması ve sistemin bu müdahaleye nasıl tepki vereceğinin gözlemlenebilmesi noktasında ortaya çıkmaktadır. Bu

deneyim, ileride yapılabilecek mevzuat geliştirmelerine yön verebilecek önemli bir veri ve değerlendirme zemini sağlayacaktır.

Bu kapsamda Türkiye'nin de benzer bir mevzuat kabul etmesinin gerekli olup olmadığı sorusu gündeme gelebilecektir. Belirtildiği üzere KVKK'da öngörülen düzenleme uyarınca özel nitelikli verilerin işlenmesi kural olarak yasak olmakla birlikte, bir kanuna dayanması halinde istisnai olarak izin verilebilecektir. Dolayısıyla, AB Yapay Zekâ Tüzüğü ile öngörülen istisnai hüküm KVKK'ya da teorik olarak eklenebilecek veya ayrık bir Yapay Zekâ Kanunu ile bu hüküm getirilebilecektir. Nitekim 7 Eylül 2025 tarihli ve 33010 (mükerrer) sayılı Resmî Gazete'de yayımlanan Orta Vadeli Program (2026–2028) uyarınca, Türkiye'deki mevzuatın AB Yapay Zekâ Tüzüğü ile uyumlaştırılması kapsamında yürütülecek çalışmaların 2026'nın üçüncü çeyreğinde tamamlanması öngörülmektedir. Bu takvim dikkate alındığında, yakın dönemde kabul edilmesi muhtemel Türk yapay zekâ mevzuatının algoritmik ön yargı ve ayrımcı sonuçlar problematiğini de gündemine alması ve bu alanda düzenleyici bir çerçeve benimsemesi şaşırtıcı olmayacaktır.

Bununla birlikte, böyle bir mevzuatın kabul edilmesi, 'algoritmik adalet'in fiilen sağlanacağı anlamına gelmemektedir. Bu noktada özellikle belirtmek gerekir ki 'algoritmik adalet'in sağlanması, teorik düzeyde teknik araçların yardımı ve hukuki mekanizmalar, başta AB Yapay Zekâ Tüzüğü olmak üzere farklı hukuk düzenlerinde ortaya çıkabilecek yasama faaliyetleri, yoluyla kısmen mümkün görünse de, pratikte çok daha karmaşık bir uğraştır. Zira bu çalışmanın diğer bölümlerinde de vurgulandığı üzere, hangi verilerin ayrımcı sayılacağına ve hangi verilerin 'doğru' veya 'ön yargıdan arındırılmış' olarak kabul edileceğine karar vermek, özellikle coğrafi sınırların büyük ölçüde anlamını yitirdiği dijital bir ekosistem söz konusu olduğunda, son derece güçtür. Buna ek olarak, bu değerlendirmeyi yapacak mekanizmaların kimler tarafından ve hangi usullerle belirleneceği; kolektif bir adalet anlayışının nasıl tesis edileceği; hangi ülkenin, dinin, dilin, kültürün veya yaşam biçiminin 'ön yargısız veri' ölçütünü belirleyici kabul edileceği soruları da cevaplandırılması zor meselelerdir. Bu makalenin ilk bölümünde aktarılan ve Tayvan hakkında yöneltilen sorulara farklı yapay zekâ modellerinden farklı yanıtlar alındığını gösteren çalışma, konunun karmaşıklığını ve sübjektifliğini daha da

görünür kılmaktadır. Dolayısıyla algoritmalar her ne kadar tarafsızlık iddiasıyla tasarlansa dahi, eğitildikleri veya geliştirildikleri toplumun sosyo-kültürel, ekonomik, dini ya da ahlaki değer yargılarını belirli ölçüde yansıtmaları kuvvetle muhtemeldir.

Sonuç olarak yazar, AB Yapay Zekâ Tüzüğü'nün algoritmik ön yargıları azaltmaya ve ayrımcı sonuçları önlemeye yönelik attığı adımların bu alanda kıymetli bir girişim olduğunu değerlendirmektedir. En azından, gelecekte daha kapsamlı bir mevzuatın geliştirilmesi bakımından, üretken yapay zekâ sistemlerini ve bu çalışmada tartışılan diğer istisna alanlarını da kapsayabilecek bir düzenleme zemini ve hazırlığı olarak görülmesi mümkündür. Bununla birlikte, sosyo-kültürel, dini, politik ve benzeri alanlarda yapay zekâ sistemlerinin uyabileceği ortak ve evrensel bir adalet anlayışının, en azından mevcut teknolojik gelişmişlik düzeyi itibarıyla, kolaylıkla tesis edilemeyeceğini belirtmek gerekir. Bu nedenle, söz konusu düzenlemenin yapay zekâların ön yargılarını ve ayrımcı sonuçlarını belirli ölçüde azaltabilmesi mümkün olsa da, bunları tamamen ortadan kaldırmasının beklenmesi güçtür.

KAYNAKÇA

Aequitas Project. "Home". https://www.aequitas-project.eu/. (erişim 12 Kasım 2025).

Aksoy, Hüseyin Can. "Kişisel Verilerin Korunması Yönüyle Algoritmik Karar Verme". *Kişisel Verileri Koruma Dergisi* 4,2 (2022): 69–87.

Ayaz, Sedat. "Yapay Zekânın Hukuk Alanındaki Uygulamaları ve Hukuk Alanında Kullanılmasının Olası Sonuçları." *Sakarya Üniversitesi Hukuk Fakültesi Dergisi* 13,1 (2025): 80-108.

Berktaş, Ahmet Esad, ve Saide Begüm Feyzioğlu. "Regulating AI Against Discrimination: From Data Protection Legislation to AI-Specific Measures." *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde. Editörler Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu. Singapore: Springer, 2024.

Bozkurt Gümrükçüoğlu, Yeliz ve Gülnihal Ahter Yakacak. "Yapay Zekanın İşe Alım Süreçlerinde Kullanımı ve Algoritmik Ayrımcılık". *Ankara Üniversitesi Hukuk Fakültesi Dergisi* 72,4 (Şubat 2024): 1701–1757. https://doi.org/10.33629/auhfd.1403311.

Çoğalan, Mehmet, ve Zeynep Ebrar Kaya. "Chatbotların Yargı Bağımsızlığı ve Adil Yargılanmak Hakkı Boyutuyla Değerlendirilmesi." İnÜHFD 16,2 (2025): 629–640.

Encyclopaedia Britannica. "Artificial Intelligence". https://www.britannica.com/technology/artificial-intelligence. (erişim 12 Kasım 2025).

European Court of Human Rights (ECtHR). Case "Relating to Certain Aspects of the Laws on the Use of Languages in Education in Belgium" v Belgium. Appl. No. 1474/62. 23 Temmuz 1968.

Fairlearn. "Improve Fairness of AI System". https://fairlearn.org/. (erişim 12 Kasım 2025).

Hacker, Philipp. "Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law". *Common Market Law Review* 55,4 (2018): 1143–1185.

Hacker, Philipp, Frederik Zuiderveen Borgesius, Brent Mittelstadt ve Sandra Wachter. "Generative Discrimination: What Happens When Generative AI Exhibits Bias, and What Can Be Done About It". *The Oxford Handbook of the Foundations and Regulation of Generative AI* içinde. ed., Philipp Hacker vd. Oxford: Oxford University Press, 2025. https://doi.org/10.1093/oxfordhb/9780198940272.013.0016. (erişim 11 Kasım 2025).

Hacker, Philipp, Andreas Engel ve Marco Mauer. "Regulating ChatGPT and other Large Generative AI Models". Bildiri, International Conference on Fairness, Accountability, and Transparency (FAccT '23), Haziran 2023. https://doi.org/10.1145/3593013.3594067.

Kayser-Bril, Nicolas. "Google Apologizes after its Vision AI Produced Racist Results". *AlgorithmWatch*, 7 Nisan 2020. https://algorithmwatch.org/en/google-vision-racism/. (erişim 11 Kasım 2025).

Kulular İbrahim, Merve Ayşegül. "Legal Challenges of Artificial Intelligence in Healthcare." *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde. Editörler Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu. Singapore: Springer, 2024.

Kumkumoğlu, Selin Çetin, ve Ahmet Kemal Kumkumoğlu. "Rethinking Non-discrimination Law in the Age of Artificial Intelligence." *Algorithmic Discrimination and Ethical Perspective*

of Artificial Intelligence içinde. Editörler Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu. Singapore: Springer, 2024.

Kuśmierczyk, Małgorzata. "Algorithmic Bias in the Light of the GDPR and the Proposed AI Act". Preprint, Mayıs 2022. Yayımlanacağı yer: *'(In)equality. Faces of Modern Europe'*. Der., Maciej Olejnik ve Wiktoria Morawska. Wrocław: Wydawnictwo Centrum Studiów Niemieckich i Europejskich im. Willy'ego Brandta, 2022.

Lendvai, Gergely Ferenc ve Gergely Gosztonyi. "Algorithmic Bias as a Core Legal Dilemma in the Age of Artificial Intelligence: Conceptual Basis and the Current State of Regulation". *Laws* 14, 41 (2025): https://doi.org/10.3390/.

Lu, Donna. "We Tried out DeepSeek. It Worked Well, Until We Asked It about Tiananmen Square Protests of 1989 and Taiwan". *The Guardian*, 28 Ocak 2025. https://www.theguardian.com/technology/2025/jan/28/we-tried-out-deepseek-it-works-well-until-we-asked-it-about-tiananmen-square-and-taiwan. (erişim 11 Kasım 2025).

Mac, Trang Anh. "Bias and Discrimination in ML-based Systems of Administrative Decision-Making and Support". *Computer Law & Security Review* 55 (2024): 1–17.

Mahoney, Trisha, Kush R. Varshney ve Michael Hind. *AI Fairness: How to Measure and Reduce Unwanted Bias in Machine Learning*. IBM Corporation, 2020.

Martini, Mario. "Algorithmen als Herausforderung für die Rechtsordnung." Juristen Zeitung 72,21 (2017): 1017–1025.

McCurry, Justin. "South Korean AI Chatbot Pulled from Facebook after Hate Speech towards Minorities". *The Guardian*, 13 Ocak 2021. https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook. (erişim 11 Kasım 2025).

Nuredin, Abdulmecit. "Algorithmic Bias in Law: The Discriminatory Potential and Legal Liability of AI-Based Decision Support Systems". Bildiri, International Scientific Conference on AI, Human Rights, Migration, Democracy, and Public Impact, Ekim 2024. https://doi.org/10.55843/ISC2024conf105n.

Nzama-Sithole, Lethiwe. "Managing Artificial Intelligence Algorithmic Discrimination: The Internal Audit Function Role." *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence* içinde. Editörler Muharrem Kılıç ve Sezer Bozkuş Kahyaoğlu. Singapore: Springer, 2024.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, ve Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Science 366/6464 (2019): 447–453.

Say, Cem. *50 Soruda Yapay Zekâ*. İstanbul: Bilim ve Gelecek Kitaplığı, 2025.

Selsbæk-Reitz, Sune. LinkedIn gönderisi. Erişim 11 Kasım 2025. https://www.linkedin.com/posts/suneselsbaekreitz_in-danish-the-word-kæreste-is-gender-neutral-activity-7383384358679699456-OA8P/.

Smuha, Nathalie A. 'An Introduction to the Law, Ethics and Policy of Artificial Intelligence'. *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* içinde ed., Nathalie A. Smuha. Cambridge: Cambridge University Press, 2025: 1–14. https://doi.org/10.1017/9781009367783.001.

"State v. Loomis, Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing." Harvard Law Review 130,5 (2017): 1530–1537.

United States District Court, District of Massachusetts. 'Memorandum and Order on Defendants' Motions to Dismiss'. *Mary Louis and Monica Douglas, et al. v. SafeRent Solutions, LLC, and Metropolitan Management Group, LLC*. Civil Action No. 22-CV-10800-AK. 26 Temmuz 2023. https://www.justice.gov/crt/media/1310736/dl. (erişim 6 Ocak 2026).

United States District Court for the District of Massachusetts. 'Order Certifying the Classes for Settlement Purposes and Directing Notice of Settlement to the Classes'. *Mary Louis and Monica Douglas, et al. v. SafeRent Solutions, LLC and Metropolitan Management Group, LLC*. Civil Action No. 1:22-CV-10800-AK. Document 122. 25 Nisan 2024. https://images.law.com/contrib/content/uploads/documents/292/181890/a8d5b8bd-747f-4373-aab4-69f424b86865.pdf. (erişim 6 Ocak 2026).

United States District Court for the Northern District of California. *Class Action Complaint. Derek L. Mobley v. Workday, Inc.* Civil Action No. 3:23-cv-00770. 21 Şubat 2023.

https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.1.0_1.pdf. (erişim 6 Ocak 2026).

United States District Court for the Northern District of California. *Order Approving Long Form Notice*. *Derek Mobley v. Workday, Inc.* Case No. 3:23-cv-00770-RFL. 18 Kasım 2025. https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-00770/pdf/USCOURTS-cand-3_23-cv-00770-8.pdf. (erişim 6 Ocak 2026).

University of Maryland, Robert H. Smith School of Business. "The Problem with Amazon's AI Recruiter". *Smith Brain Trust*, 31 Ekim 2018. https://www.rhsmith.umd.edu/research/problem-amazons-ai-recruiter. (erişim 11 Kasım 2025).

van Bekkum, Marvin. "Using Sensitive Data to De-Bias AI Systems: Article 10(5) of the EU AI Act". *Computer Law & Security Review* 56 (2025): 1–11.

Yousef, Zoya. "How Can the Law Address the Effects of Algorithmic Bias in the Healthcare Context?". *LSE Law Review* 9/2 (2024): 197–253. https://doi.org/10.61315/lselr.651. (erişim 1 Ocak 2026).