Ankara Üniversitesi
**SBF Dergisi,**
Cilt 81, No. 2, 2026, s. 373-388
Türkçe

**Araştırma Makalesi / Research Article**
**DOI**

# DİJİTAL DÖNÜŞÜMDE ADALETİN KAVRAMSAL ÇERÇEVESİ: ALGORİTMİK ÖNYARGI VE YAPISAL YÖNETİŞİM ARAYIŞI *

**Merve Asiloğulları Ayan**
Milli Savunma Üniversitesi
ORCID: 0000-0001-7522-0911
ROR ID: ror.org/010t24d82

● ● ●

## Öz

Dijital dönüşümün beraberinde getirdiği algoritmik karar alma sistemleri, toplumsal adalet ve eşitlik idealleri karşısında yeni ve karmaşık tehditler oluşturmaktadır. Bu çalışma, algoritmik önyargı sorununu merkeze alarak, mevcut teknoloji yönetişimi paradigmasının yapısal sınırlılıklarını eleştirel bir bakışla analiz etmekte ve yerine Yapısal Adalet Odaklı Teknoloji Yönetişimi adını verdiği yeni bir kavramsal çerçeve önermektedir. Çalışmada öncelikle, algoritmik önyargının salt teknik bir mesele olmadığı, toplumsal eşitsizliklerin bir yansıması olduğu argümanı işlenmekte ve dijital adalet kavramı dağıtımsal, işlemsel ve yapısal boyutlarıyla incelenmektedir. Ardından, hakim yönetişim yaklaşımının teknokratik, bireyci ve etnosentrik sınırlılıkları ortaya konmakta, bu sınırlılıkları aşmak üzere dört temel bileşenden oluşan (önleyici adalet ve katılımcı tasarım, sistemsel hesap verebilirlik, bağlama duyarlılık ve yeni demokratik aktörler) alternatif bir model detaylandırılmaktadır. Kamu sektöründeki otomatik fayda tahsisi sistemleri üzerinden yapılan vaka analizi, önerilen çerçevenin pratikte nasıl işleyebileceğini somutlaştırmaktadır. Çalışma, teknoloji yönetişiminde paradigmatik bir dönüşüm çağrısıyla son bulmakta ve gelecek araştırmalar için bir gündem önermektedir.

**Anahtar Sözcükler:** Algoritmik önyargı, Dijital adalet, Teknoloji yönetişimi, Yapısal adalet, Katılımcı tasarım

*The Conceptual Framework of Justice in Digital Transformation: Algorithmic Bias and the Quest for Structural Governance*

## Abstract

Digital transformation has ushered in algorithmic decision-making systems that generate new and complex threats to the ideals of social justice and equality. This study places the issue of algorithmic bias at its core, offering a critical analysis of the structural limitations of the prevailing technology governance paradigm. It introduces an alternative conceptual framework, termed *Structural Justice-Oriented Technology Governance*. The paper argues that algorithmic bias is not merely a technical problem but rather a reflection of underlying social inequalities, and examines the notion of digital justice across its distributive, procedural, and structural dimensions. It then identifies the technocratic, individualistic, and ethnocentric constraints of dominant governance approaches and elaborates a four-pillar alternative model, composed of preventive justice and participatory design, systemic accountability, contextual sensitivity, and the inclusion of new democratic actors. A case study on automated welfare allocation systems in the public sector illustrates how this framework can be operationalized in practice. The study concludes with a call for a paradigmatic shift in technology governance and proposes an agenda for future research.

**Keywords:** Algorithmic bias, Digital justice, Technology governance, Structural justice, Participatory design

---

* Makale geliş tarihi / Received date: 28.10.2025
Makale kabul tarihi / Accepted date: 06.04.2026
Erken görünüm tarihi / Published online: 11.05.2026

**Yazar email:** merveasilogullari@gmail.com

CC BY NC ND

OPEN ACCESS

# Dijital Dönüşümde Adaletin Kavramsal Çerçevesi: Algoritmik Önyargı ve Yapısal Yönetişim Arayışı

## Giriş

Yirminci yüzyılı şekillendiren endüstriyel rasyonalite, yerini yirmi birinci yüzyılda algoritmik rasyonaliteye bırakırken, toplumsal yaşamın yeniden düzenlenmesinde yapay zekâ ve otonom karar alma sistemleri kritik bir role sahip olmuştur. Bu dönüşüm, benzeri görülmemiş verimlilik vaatleri ve yeni kamusal fayda alanları yaratsa da, aynı zamanda dijital adalet olarak kavramsallaştırılabilecek yeni ve karmaşık bir sorunsalı da beraberinde getirmiştir (Hoffmann, 2019). Özellikle, insan failliğinin giderek arka plana çekildiği bu sistemlerin, tarihsel ve toplumsal eşitsizlikleri yeniden üreterek derinleştirdiğine dair ampirik bulgular, bu alana dair eleştirel bir sorgulama zorunluluğunu doğurmaktadır (Eubanks, 2018; Noble, 2018).

Bu çalışmanın çıkış noktasını oluşturan temel problematik, dijital dönüşüm sürecinde teknoloji yönetişimine hâkim olan mevcut paradigmanın, algoritmik sistemlerin neden olduğu yapısal adaletsizlikleri kavramakta ve çözmekte yetersiz kalmasıdır. Şeffaflık, hesap verebilirlik ve etik ilkeler odaklı mevcut yaklaşım, sorunu genellikle teknik ve bireysel düzleme hapsetmekte; toplumsal güç ilişkileri, kurumsal sorumluluk ve sistematik tahakküm boyutlarını yeterince ele almamaktadır (Zarsky, 2016). Bu yetersizlik, adaleti istatistiksel bir denge problemi olarak kodlayan teknokratik bir anlayışa işaret etmekte ve sonuçta etik ilkeler ile somut düzenlemeler arasında derin bir uçurum oluşmasına neden olmaktadır (Jobin, Ienca, & Vayena, 2019).

Buradan hareketle, bu makale iki temel amaca hizmet etmektedir: İlk olarak, algoritmik önyargı ve dijital adalet kavramlarını, eleştirel teori perspektifinden de beslenerek anatomik bir titizlikle analiz etmek ve bu kavramların salt teknik değil, özünde toplumsal ve siyasi olduğunu ortaya koymaktır. İkinci olarak ise, mevcut yönetişim paradigmasının sınırlılıklarını bu kavramsal çerçeve üzerinden eleştirerek, yapısal adaleti merkezine alan yeni ve bütüncül bir teknoloji yönetişimi modeli önermektir. Bu çerçeve, önleyici adalet,

katılımcı tasarım, sistemsel hesap verebilirlik ve bağlama duyarlı yeni demokratik mekanizmalar üzerine inşa edilmektedir.

Çalışma, bu amaçlar doğrultusunda, öncelikle algoritmik önyargının teknik bir arızadan ziyade nasıl toplumsal bir yansıma olduğunu ve dijital adaletin dağıtımsal, işlemsel ve yapısal boyutlarını mercek altına alacaktır. Bu kavramsal anatomi, makalenin teorik temelini oluşturacak ve devamında gelecek eleştirel analiz ile model önerisi için sağlam bir zemin hazırlayacaktır.

## 1. Algoritmik Önyargı ve Dijital Adalet

Dijital dönüşümün toplumsal hayatı şekillendiren yeni bir güç olarak ortaya çıkışı, geleneksel adalet ve eşitlik kavramlarını derinden sarsmaktadır. Bu bölümde, algoritmik önyargı olgusu merkezinde, dijital adalet arayışının kavramsal temellerini, disiplinlerarası bir perspektifle incelenecektir.

Algoritmik karar alma mekanizmalarının, ırk, cinsiyet, yaş ve sosyo-ekonomik statü temelinde sistematik ve ayırt edici sonuçlar üretme eğilimi, literatürde algoritmik önyargı olarak tanımlanmaktadır (Noble, 2018). Bu önyargının doğasını anlamak için onun sadece bir kod hatası olmadığını, daha ziyade üç temel kaynaktan beslenen yapısal bir sorun olduğunu kavramak esastır.

İlk olarak, veri temelli önyargı, tarihsel ve toplumsal eşitsizliklerin veri kümelerine yansımasından kaynaklanır. Bu durum, makine öğrenimi modellerinin geçmişteki adaletsizlikleri öğrenerek geleceğe taşımasına neden olur. Örneğin, işe alım için kullanılan bir algoritma, tarihsel olarak erkek egemen bir sektörün verileriyle eğitildiğinde, kadın adaylara karşı sistematik bir önyargı geliştirebilir (Bolukbasi vd., 2016).

İkinci kaynak, tasarım ve geliştirme önyargısıdır. Algoritmaları tasarlayan mühendis ve veri bilimcilerinin, kaçınılmaz olarak kendi dünya görüşleri, kültürel değer yargıları ve bilinçdışı önyargılarını ürüne aktarmalarıdır. O'Neil (2016), bu tür modelleri, ölçülmesi zor başarı kriterleri üzerinden gücü meşrulaştıran ve zarar veren İnsanlığa Zarar Veren Matematik Silahları (Weapons of Math Destruction) olarak adlandırmıştır.

Üçüncü ve en az sorgulanan kaynak ise bağlamsal önyargıdır. Bir bağlam için geliştirilen ve başarılı addedilen bir algoritmanın, farklı bir toplumsal veya kültürel bağlamda uygulanmasıyla ortaya çıkar. Bu, teknolojinin evrensel ve tarafsız olduğu varsayımının naifliğini gözler önüne serer (Crawford, 2021).

Algoritmik önyargı sorununa bir yanıt olarak gelişen dijital adalet kavramı, geleneksel adalet teorilerinin dijital alana uyarlanmasının ötesine

geçerek yeni bir paradigma sunmaktadır. Bu paradigmanın anatomisini, birbiriyle kesişen üç temel boyutta analiz etmek mümkündür.

Dağıtımsal Adalet (Distributive Justice): Bu boyut, dijital sistemlerin yarattığı fayda ve zararların toplumda nasıl dağıldığı sorusuna odaklanır. Geleneksel anlamda kaynak dağılımına ilişkin olan bu kavram, dijital bağlamda fırsat, erişim ve sonuç adaletine kaymaktadır. Örneğin, bir yapay zekâ sisteminin kararlarından kimlerin olumlu/olumsuz etkilendiği, kimlerin dijital kamu hizmetlerine adil şekilde erişebildiği bu boyutun inceleme alanına girer (Hoffmann, 2019).

İşlemsel Adalet (Procedural Justice): Bu boyut, algoritmik karar alma süreçlerinin nasıl işlediği ile ilgilidir. İşlemsel adalet, şeffaflık, hesap verebilirlik ve itiraz mekanizmalarının varlığını merkeze koyar. Bir bireyin, kendisi hakkında verilen otomatik bir kararın nasıl alındığını anlayamaması, sorgulayamaması ve itiraz edememesi, işlemsel adaletin ihlali anlamına gelir. Bu, yurttaşı sadece bir veri nesnesine indirgeyen bir süreçtir (Zarsky, 2016).

Yapısal Adalet (Structural Justice): En derin ve en az görünür olan bu boyut, algoritmik sistemlerin toplumsal güç ilişkilerini nasıl inşa ettiği, meşrulaştırdığı veya dönüştürdüğü sorusunu sorar. Bu, adaleti bireyler arası bir mesele olmaktan çıkarıp, sistemik ve kurumsal bir analiz düzeyine taşır. Eubanks (2018), yoksul ve hassas grupları hedef alan otomatik değerlendirme sistemlerinin, yeni bir dijital cezaevleri (digital poorhouse) inşa ettiğini iddia ederek, yapısal adaletsizliğin çarpıcı bir resmini çizer.

Algoritmik önyargıyı salt bir teknik arıza veya veri kalitesi sorunu olarak okumak, meselenin özünü ıskalamak anlamına gelir. Sorunun köklerine inmek için, teknolojinin toplumsal ilişkiler içinde nasıl konumlandığına dair eleştirel bir perspektif benimsemek elzemdir. Bu noktada, Eleştirel Teori'nin sağladığı lensler, algoritmaların mevcut iktidar yapılarını ve toplumsal eşitsizlikleri nasıl yeniden ürettiğini ve meşrulaştırdığını açığa çıkarmakta kilit rol oynar.

Foucaultcu İktidar Analizi ve Algoritmik Gözetim: Michel Foucault'nun iktidar kavramsallaştırması, algoritmik sistemleri anlamak için son derece verimli bir çerçeve sunar. Foucault'ya (1977) göre modern iktidar, egemenin doğrudan zoruyla değil, bilginin üretimi, sınıflandırılması ve normalleştirilmesi yoluyla, bireyleri dizginlemez ama yönlendirir. Panoptikon metaforuyla betimlediği bu yönlendirici iktidar, sürekli bir gözetim ve normatif kategoriler aracılığıyla işler. Günümüzün algoritmik sistemleri, dijital birer süper-panoptikon işlevi görerek bu dinamikleri katbekat güçlendirmektedir (Cheney-Lippold, 2017). Örneğin, kredi skorlama sistemleri veya önleyici polislik algoritmaları, bireyleri güvenilir/riskli, masum/potansiyel suçlu gibi dijital kategorilere ayırarak onların toplumsal hareketliliklerini ve yaşam olanaklarını

yapılandırır. Bu yumuşak iktidar biçimi, bireyleri kendi davranışlarını bu normlara uyacak şekilde öz-düzenlemeye (self-regulation) iter. Bu süreçte algoritmalar, sadece önyargıyı yansıtan araçlar değil, aynı zamanda tahakkümün kendisini yeniden üreten aktif aktörler haline gelir.

Habermas ve İletişimsel Eylem Çerçevesinde Tekno-Rasyonalitenin Eleştirisi: Jürgen Habermas'ın iletişimsel eylem ve sistem dünyası kavramları ise, algoritmik karar alma süreçlerinin demokratik meşruiyetini sorgulamak için kullanılabilir. Habermas (1987), modern toplumlarda, öznel deneyimlerin ve normatif beklentilerin yer aldığı yaşam dünyasının (lifeworld), para ve iktidar gibi soyut mekanizmalarla işleyen sistem tarafından giderek istila edildiğini (colonization) savunur. Algoritmik sistemler, bu istilanın en sofistike biçimlerinden biridir. İnsan ilişkilerine dair karmaşık, bağlama duyarlı ve normatif meseleleri (örneğin, adalet, toplumsal fayda), salt teknik ve veriye dayalı bir rasyonaliteye (tekno-rasyonalite) indirger. Bu indirgeme, Habermas'ın ideal konuşma durumu için gerekli gördüğü şeffaflık, önyargısızlık ve herkesin katılımı ilkelerini tamamen dışlar (Mühlhoff, 2020). Sonuç olarak, toplumun ortak değerleri üzerine yapılması gereken kamusal iletişimsel tartışma, uzmanlar ve algoritmalar tarafından verilen teknik kararlarla ikame edilir. Örneğin, bir sosyal yardım algoritmasının kimin muhtaç olduğuna dair verdiği otomatik karar, toplumun bu kavram üzerine yürütmesi gereken etik ve siyasi müzakerenin yerini alarak, yaşam dünyasının sistem tarafından istila edilmesine tipik bir örnek oluşturur.

Bu eleştirel lensler, algoritmik önyargının, basit bir düzeltilebilir hata olmanın ötesinde, toplumsal tahakküm mekanizmalarının yeni ve etkili bir enstrümanı olduğunu gözler önüne serer. Bu perspektif, bir sonraki bölümde ele alınacak olan mevcut teknoloji yönetişimi paradigmasının neden yüzeysel kaldığını anlamamız için de sağlam bir zemin hazırlamaktadır.

Bu bölümde, algoritmik önyargı olgusunu üç temel kaynak üzerinden (veri, tasarım, bağlam) anatomik bir titizlikle inceledik ve dijital adalet arayışını dağıtımsal, işlemsel ve yapısal boyutlarıyla kavramsallaştırılmıştır. Ancak bu analiz, salt bir kavram haritalandırması olmanın ötesine geçmektedir. Eleştirel teori lenslerinden yaptığımız okuma, bu kavramların canlı birer toplumsal ve siyasi gerçeklik olduğunu ortaya koymuştur.

Foucaultcu iktidar analizi, algoritmaların bireyleri dizginlemeden yönlendiren, onları dijital kategoriler içinde özneler haline getiren ve davranışlarını öz-düzenlemeye iten modern bir tahakküm aracı olarak işlev gördüğünü gösterirken; Habermas'ın iletişimsel eylem kuramı, algoritmik karar alma süreçlerinin, toplumsal meselelere dair kamusal iletişimsel tartışmanın yerini nasıl tekno-rasyonalist bir sisteme bıraktığını ve yaşam dünyasının nasıl istila edildiğini açık etmiştir.

Dolayısıyla, algoritmik önyargıyı anlamak, onun sadece teknik veya istatistiksel köklerini değil, aynı zamanda toplumsal güç ilişkilerini yeniden üreten ve meşrulaştıran yapısal ve siyasi köklerini de görmeyi gerektirir. Buna karşılık dijital adalet arayışı ise, dağıtımsal, işlemsel ve yapısal boyutları bir arada ele alan, bu boyutları eleştirel teori ile beslenen derin bir kavrayışla yorumlayan bütüncül bir çerçeve sunmaktadır.

Bir sonraki bölümde, bu kavramsal ve teorik incelenmesinin ışığında, mevcut teknoloji yönetişimi paradigmasının, bu karmaşık sorunlar karşısında neden yetersiz ve sığ kaldığını eleştirel bir bakışla analiz edilecektir. Temel argümanımız, mevcut yaklaşımın, yukarıda tartışılan bu geniş ve eleştirel tabloyu kavramakta başarısız olarak, sorunu yönetilmesi daha kolay ancak etkisi daha sınırlı teknik ve bireyci bir meseleye indirgediği yönünde olacaktır.

## 2. Mevcut Teknoloji Yönetişimi Paradigmasının Eleştirel Analizi: Yapısal Sınırlılıklar

Algoritmik sistemlerin doğurduğu eşitsizliklere yönelik hâkim yönetişim yaklaşımları, büyük ölçüde etik ilkeler ve teknik düzenleme ekseninde şekillenmektedir. Son on yıl içerisinde devletler, uluslararası kuruluşlar ve teknoloji şirketleri tarafından yayımlanan yüzlerce Yapay Zekâ Etik İlkeleri belgesi, şeffaflık, adalet, sorumluluk, hesap verebilirlik ve zarar vermeme gibi normlarda büyük ölçüde yakınsama göstermektedir (Jobin, Ienca & Vayena, 2019, s. 389). Ancak bu belgeler çoğunlukla soyut ilkeler düzeyinde kalmakta ve pratikte uygulanabilir mekanizmalar ortaya koymamaktadır. Bu durum, literatürde frame problem olarak tanımlanan kavramsal bir açmazı gündeme getirmektedir: Örneğin adalet ilkesinin bir algoritmada ne şekilde karşılık bulacağı konusunda ortak bir operasyonel mutabakat yoktur. Bir işe alım algoritmasının adil olması, demografik pariteyi (seçim sonuçlarının nüfus dağılımını yansıtması) mi, yoksa eşit fırsatı (aynı niteliklere sahip bireylerin grup aidiyetinden bağımsız eşit şansa sahip olması) mı gerektirir? Bu ölçütler kimi durumlarda birbirleriyle çelişebilir (Kleinberg, Mullainathan & Raghavan, 2016, ss. 2–4). Böylelikle, hâkim teknokratik yaklaşım adaleti istatistiksel bir denge problemi olarak kodlayarak, onun toplumsal, tarihsel ve siyasi boyutlarını görünmez kılmaktadır.

Etik söylemin soyutluğunu eleştiren birçok araştırmacı, bu belgelerin büyük bölümünün pratikte bağlayıcı bir güce sahip olmadığını, çoğu zaman şirketlerin ve hükümetlerin kamuoyunu ikna etmeye dönük etik PR belgeleri olarak işlev gördüğünü belirtmektedir (Metzinger, 2019, ss. 7–8). Floridi ve Cowls'un (2019) ileri sürdüğü gibi, etik ilkeler önemli bir başlangıç olmakla

birlikte, bunlar kurumsal düzenlemeler ve somut yaptırımlar ile desteklenmedikçe yalnızca iyi niyet beyanı olarak kalmaktadır.

Mevcut düzenleyici girişimlere bakıldığında, Avrupa Birliği'nin Genel Veri Koruma Tüzüğü (GDPR) gibi düzenlemeler özellikle bireysel haklara odaklanmaktadır. Açıklama hakkı maddesi, bireylerin kendilerini etkileyen otomatik kararların gerekçelerine erişebilmesini ve itiraz edebilmesini amaçlamaktadır (Goodman & Flaxman, 2017, s. 97; Wachter, Mittelstadt & Russell, 2017, ss. 843–844). Bu bireysel hak temelli yaklaşım önemli bir kazanım olmakla birlikte, algoritmik sistemlerin daha geniş ve yapısal etkilerini kavramakta yetersiz kalmaktadır. Çünkü pek çok algoritma yalnızca bireyleri değil, mahalleleri puanlayan risk skorları, toplumsal kaynakların dağılımını şekillendiren kamu algoritmaları ya da işe alımda grupları sistematik biçimde dışlayan modeller gibi nüfus düzeyinde etkiler üretmektedir. Bu noktada, birey merkezli hak mekanizmaları kolektif zararı görünmez kılmaktadır.

Virginia Eubanks'ın (2018, ss. 6–8) Automating Inequality adlı çalışması bu açığı çarpıcı biçimde ortaya koyar. Eubanks, ABD'de sosyal yardımların otomatik değerlendirme sistemlerinin, bireysel vakaları değerlendirmekten çok, yoksulları toplumsal gözetim ve disiplin altına alan bir dijital yoksulluk yönetimi aracı olarak işlev gördüğünü göstermektedir. Bu tür örnekler, adaletin bireysel haklar düzeyinde kavranmasının, algoritmik sistemlerin kolektif düzeydeki yapısal etkilerini görünmezleştirdiğini kanıtlamaktadır.

Hakim yönetişim yaklaşımı yalnızca bireyci değil, aynı zamanda etnosentriktir. Teknoloji politikalarının büyük bölümü Batılı, liberal ve piyasa odaklı bir perspektiften yürütülmekte; teknolojinin evrensel ve nötr olduğu varsayılmaktadır. Ancak algoritmik sistemler, üretildikleri bağlamların kültürel değerlerini ve güç ilişkilerini kodlamaktadır. Mohamed, Png ve Isaac'in (2020) ileri sürdüğü dekolonyal yapay zekâ perspektifi, mevcut paradigmaların sömürgecilik sonrası güç asimetrilerini dijital teknolojiler yoluyla yeniden ürettiğini savunmaktadır. Benzer biçimde, Abeba Birhane (2020, s. 392) algoritmik sistemlerin Afrika bağlamında epistemik ve ekonomik sömürgeleştirme biçimleri olarak işlediğini göstermektedir.

Somut bir örnek olarak, Batı'da geliştirilmiş ve şüpheliyi tanımlayan bir yüz tanıma algoritmasının, farklı demografik yapılara sahip bir ülkede doğrudan belirli etnik grupları hedef alabilmesi mümkündür. Bu tür bağlamsal önyargılar, teknoloji transferi yoluyla küresel eşitsizlikleri derinleştirmekte, algoritmaların yalnızca yerel değil küresel adalet meseleleri yaratabileceğini göstermektedir. Ne var ki hâkim paradigma, kültürel çeşitliliği ve jeopolitik farklılıkları hesaba katan dekolonyal teknoloji yönetişimi perspektifinden yoksundur.

Sonuç olarak, mevcut teknoloji yönetişimi paradigmasının eleştirel analizi, bu yaklaşımın algoritmik gücün yapısal ve sistematik doğasını kavramakta yetersiz kaldığını ortaya koymaktadır. Teknokratik yaklaşım adaleti matematiksel formüllere indirgemekte, bireyci yaklaşım kolektif zararı göz ardı etmekte, etnosentrik yaklaşım ise kültürel ve tarihsel çeşitlilikleri yok saymaktadır. Bu nedenle algoritmik adaletsizliği yalnızca teknik bir problem olarak değil, dağıtımsal, işlemsel ve yapısal boyutları olan siyasal bir mesele olarak kavrayan, eleştirel ve bağlama duyarlı yeni bir paradigmanın inşası zorunludur.

## 3. Yeni Bir Paradigma Önerisi: Yapısal Adalet Odaklı Teknoloji Yönetişimi Çerçevesi

Mevcut teknoloji yönetişimi modelleri, çoğu zaman teknokratik, bireyci ve Batı merkezli yaklaşımlara dayanmakta; bu durum, algoritmik sistemlerin yol açtığı yapısal eşitsizliklerle başa çıkmada ciddi sınırlılıklar doğurmaktadır. Adalet, yalnızca bireysel hakların korunması ya da şeffaflık ilkesi üzerinden kavramsallaştırıldığında, sorunun toplumsal güç ilişkileri ve kurumsal tasarılarla iç içe geçmiş doğası görünmez hale gelmektedir (Mittelstadt vd., 2016, s. 3; Selbst vd., 2019, s. 60). Bu nedenle, algoritmik adaletsizliklere yönelik mevcut yaklaşımlar semptomlara odaklanırken, kök nedenlerin yani tarihsel eşitsizliklerin ve kurumsal tahakküm biçimlerinin göz ardı edildiği görülmektedir. Bu çalışmada önerilen "Yapısal Adalet Odaklı Teknoloji Yönetişimi" çerçevesi, adaletin yalnızca teknik bir düzenleme sorunu değil, aynı zamanda toplumsal ve siyasal bir süreç olduğunu kabul ederek, dört temel bileşen üzerinden paradigmatik bir dönüşüm çağrısı yapmaktadır.

İlk olarak, telafi edici mekanizmalara dayanan mevcut yönetişim anlayışlarının ötesine geçerek önleyici adaletin merkezileştirilmesi gerekmektedir. Algoritmik sistemler, zarar ortaya çıktıktan sonra düzeltilmesi gereken araçlar olarak değil, tasarım aşamasından itibaren toplumsal ihtiyaçlarla uyumlu kılınması gereken mekanizmalar olarak düşünülmelidir. Bu bağlamda, etkilenmesi muhtemel toplulukların ve özellikle marjinal grupların karar süreçlerine katılımı zorunludur (Sloane, Moss & Awomolo, 2020, ss. 3–5). Katılımcı tasarımın yalnızca danışma mekanizması olmaktan çıkarılıp, doğrudan karar verme süreçlerine entegre edilmesi, sistemlerin daha adil sonuçlar üretmesini sağlayabilir. Noble'ın (2018, ss. 23–24) Algorithms of Oppression adlı çalışmasında ortaya koyduğu gibi, Google aramalarında siyah kadınlara ilişkin önyargılı ve aşağılayıcı içeriklerin sistematik biçimde öne çıkarılması, algoritmaların nasıl toplumsal eşitsizlikleri yeniden üretebildiğini

göstermektedir. Bu türden örnekler, katılımcı ve önleyici yaklaşımların önemini açık biçimde ortaya koymaktadır.

İkinci olarak, şeffaflık sıklıkla yeterli bir çözüm olarak sunulsa da, tek başına anlamlı bir hesap verebilirlik sağlamaktan uzaktır. Algoritmaların kaynak kodlarının ifşa edilmesi, teknik detayların açığa çıkarılması, çoğu zaman kamuya gerçek anlamda hesap verme kapasitesi yaratmaz. Çünkü algoritmik adaletsizlik yalnızca teknik kodlama hatalarından değil, bu sistemlerin gömülü olduğu kurumsal tasarı ve karar zincirlerinden kaynaklanmaktadır. Pasquale'nin (2015, ss. 8–9) The Black Box Society kitabında belirttiği gibi, algoritmalar finans, sağlık ve kamusal hizmetlerde adeta kara kutular olarak işlev görmekte ve sorumluluğu belirsizleştirmektedir. Bu nedenle önerilen çerçeve, şeffaflık odaklı bireysel çözümler yerine sistemsel hesap verebilirliği öne çıkarır. Kurumların kullandıkları algoritmik araçların toplumsal etkilerini düzenli olarak ölçmeleri, raporlamaları ve eşitlik odaklı etki değerlendirmeleri sunmaları gerekmektedir (Raji vd., 2020, ss. 35–37). Ayrıca, model belgeleri (Model Cards) ve veri seti belgeleri (Datasheets for Datasets) gibi araçların hesap verebilirlik rejimlerinin parçası haline getirilmesi, yalnızca teknik değil kurumsal düzeyde sorumluluğu görünür kılmaktadır (Mitchell vd., 2019, s. 222; Gebru vd., 2021, ss. 87–89).

Üçüncü olarak, yönetişim evrenselci yaklaşımlardan uzaklaşarak bağlama duyarlı ve dekolonyal bir perspektif kazanmalıdır. Algoritmalar nötr ya da evrensel araçlar değildir; değerler, veri pratikleri ve performans ölçütleri tarihsel eşitsizlikler ve kültürel normlar tarafından şekillenir. Mohamed, Png ve Isaac'in (2020) Decolonial AI makalesinde belirttikleri üzere, Batı merkezli etik ilkelerin evrenselleştirilmesi epistemik tahakkümü yeniden üretmekte, farklı bağlamlarda farklı değer önceliklerini görmezden gelmektedir (Mohamed, Png & Isaac, 2020, ss. 662–663). Benzer şekilde Birhane (2020, s. 392), algoritmik sömürgeleştirme kavramıyla, Afrika'da veri teknolojilerinin Batılı güç ilişkilerini yeniden ürettiğini göstermektedir. Bu nedenle, algoritmik sistemlerin adalet anlayışının, her bağlamın kendi kültürel değerleri ve toplumsal öncelikleri doğrultusunda şekillendirilmesi gerekir. Green ve Viljoen (2020) da algoritmik realizm kavramıyla, algoritmaların yalnızca teknik nesneler değil, toplumsal bağlamların ürünü olarak değerlendirilmesi gerektiğini vurgulamaktadır (Green & Viljoen, 2020, ss. 25–26).

Son olarak, yapısal dönüşüm için yeni demokratik aktörlerin sürece dahil edilmesi gerekmektedir. Sandvig, vd., (2014), algoritma denetimlerinin ayrımcılığı tespit etmede kritik rol oynadığını göstermektedir (Sandvig, vd., 2014, ss. 5–6). Bu çerçevede, teknik uzmanlık ve bağımsızlıkla donatılmış bağımsız algoritmik denetçiler, kamuya hizmet eden algoritmalar üzerinde düzenli eşitlik denetimleri (equity audits) gerçekleştirmeli; bu denetimler yalnızca teknik uyum raporları değil, toplumsal etkileri ölçen kurumsal belgeler

olmalıdır. Bunun yanı sıra, yüksek riskli algoritmik sistemlerin (ör. yüz tanıma, sosyal yardım tahsisi) kullanımına ilişkin kararlara vatandaş jürilerinin dahil edilmesi, teknokratik karar alma süreçlerine demokratik meşruiyet kazandıracaktır. Bu tür uygulamaların, Raji vd.'nin (2022) belirttiği gibi, yalnızca denetim süreçlerini iyileştirmekle kalmayıp, ticari aktörlerin davranışlarını da dönüştürebildiği görülmektedir (Raji, vd., 2022, ss. 3–4).

Sonuç olarak, Yapısal Adalet Odaklı Teknoloji Yönetişimi çerçevesi, adaleti teknik bir uyum sorunu değil, dinamik bir siyasal ve toplumsal süreç olarak yeniden tanımlamaktadır. Önleyici ve katılımcı tasarım, kurumsal düzeyde hesap verebilirlik, bağlama duyarlı ve dekolonyal perspektif ile yeni demokratik aktörlerin dâhil edilmesi, teknolojinin yalnızca daha verimli değil, aynı zamanda daha özgürleştirici ve eşitlikçi bir toplum inşa etmesine hizmet edecektir. Bu çerçeve, etik ilkeler ile somut düzenlemeler arasındaki uçurumu kapatma, algoritmaları yalnızca adil değil, aynı zamanda demokratik ve toplumsal sorumlulukla uyumlu hale getirme iddiasını taşımaktadır.

## 4. Vaka Analizi: Kamu Sektöründe Otomatik Fayda Tahsisi Sistemleri ve Yapısal Adalet

Kamu hizmetlerinin (sosyal yardım, işsizlik ödeneği, konut desteği vb.) dağıtımında giderek artan bir şekilde kullanılan otomatik karar alma sistemleri, devletin vatandaşlarıyla olan ilişkisini temelden dönüştürmektedir. Bu sistemler, verimlilik ve nesnellik vaadiyle sunulsa da, pratikte Eubanks (2018)'in dijital cezaevleri (digital poorhouse) olarak adlandırdığı, yoksulluğu suçlaştıran ve yapısal eşitsizlikleri pekiştiren bir mekanizmaya dönüşebilmektedir. Bu vaka, önerilen yönetişim çerçevesinin neden gerekli olduğunu tüm çıplaklığıyla ortaya koymaktadır.

Mevcut otomatik fayda sistemleri, çoğunlukla yukarıdan aşağıya, teknokratik bir mantıkla tasarlanır. Sistemden etkilenecek olan yoksul bireylerin, sosyal hizmet uzmanlarının ve topluluk temsilcilerinin tasarım sürecine anlamlı katılımı nadiren sağlanır. Bu durum, katılım açığına (participation gap) yol açar ve sistem, gerçek ihtiyaçlar ve bağlamsal incelikler yerine, kolayca ölçülebilir ancak yanıltıcı olabilecek vekil göstergelere (proxy variables) dayanır (Sloane vd., 2020). Bir konut yardımı önceliklendirme algoritması, posta kodunu bir vekil değişken olarak kullanabilir. Tarihsel olarak düşük gelirli ve belirli bir etnik grubun yoğunlukta olduğu bir posta kodunda yaşamak, kişinin yüksek riskli olarak sınıflandırılmasına ve yardım başvurusunun otomatik olarak reddedilmesine veya daha uzun bir inceleme sürecine tabi tutulmasına neden olabilir. Bu, tarihsel konut ayrımcılığını (redlining) dijital çağda yeniden

üretmekten başka bir işe yaramaz (Benjamin, 2019). Önleyici adalet ve katılımcı tasarım ilkeleri uygulansaydı, bu tür vekil değişkenlerin yıkıcı etkisi daha baştan fark edilebilir ve tasarım buna göre revize edilebilirdi.

Bu sistemler genellikle ticari sır gerekçesiyle şeffaf değildir. Ancak sorun, sadece algoritmanın kendisinin kara kutu olması değildir. Daha derindeki sorun, kurumsal sorumluluğun bir teknoloji tedarikçisine devredilmesiyle ortaya çıkan sorumluluk boşluğudur (accountability gap). Kurum, kararı algoritma verdi diyerek kendi sorumluluğundan sıyrılabilmektedir (Zarsky, 2016). Örneğin bir işsizlik ödeneği sisteminde, karmaşık bir veri kümesine dayanarak başvuruları potansiyel sahtekarlık şüphesiyle işaretleyen bir algoritma kullanıldığını varsayalım. Vatandaş, kendisine ulaşan red kararının nedenini anlayamaz. Geleneksel açıklama hakkı, ona sadece modelinizin %85 kesinlikle sahtekarlık olduğunu tahmin ettiği gibi anlamsız bir çıktı verebilir. Oysa sistemsel hesap verebilirlik, kurumun şu sorulara yanıt vermesini gerektirir: Algoritmayı kim, hangi kriterlerle seçti? Eğitim verileri ne kadar temsili? Yanlış pozitif (masum birini haksız yere suçlamak) ve yanlış negatif (gerçek bir sahtekarlığı gözden kaçırmak) oranları nedir ve bu oranlar kabul edilebilir mi? Bu, şeffaflığın ötesinde, aktif bir sorumluluk ve denetim mekanizmasıdır. Bu tür sistemler için bağımsız eşitlik denetimleri (equity audits) zorunlu hale getirilmelidir (Sandvig vd., 2014).

Batı'da geliştirilen bir risk değerlendirme modeli, farklı sosyo-ekonomik ve kültürel yapıya sahip bir bağlamda aynı etik ve verimlilikle çalışmayacaktır. Evrensel teknoloji miti, teknolojik emperyalizmin bir aracı haline gelebilir (Mohamed vd., 2020). Bir Batı ülkesinde, istikrarlı elektrik faturası ödeme geçmişi bir güvenilirlik göstergesi olarak kullanılabilir. Ancak bu model, düzensiz elektrik arzı sorunu yaşayan veya kayıt dışı ekonomide yer alan geniş nüfuslu bir başka ülkede uygulandığında, nüfusun büyük bir kısmını otomatik olarak riskli kategorisine sokacak ve onları kamusal desteklerden mahrum bırakacaktır. Bu, o ülkenin kendi içindeki yapısal eşitsizlikleri görünmez kılan ve derinleştiren bir sonuç doğurur. Bağlama duyarlı bir yönetişim, bu tür modellerin ithal edilmesi yerine, yerel gerçekliklere uygun, adil alternatif göstergelerin (örneğin, topluluk içi dayanışma ağlarının varlığı) geliştirilmesini teşvik eder.

Önerilen yeni aktörler, bu vakada hayati bir rol oynayabilir:

- Bağımsız Algoritik Denetçiler, sosyal yardım algoritmalarının periyodik olarak yukarıda bahsedilen eşitlik denetimlerini yaparak, sistematik ayrımcılık yapıp yapmadıklarını kamuya raporlayabilir.

- Vatandaş Jürileri, bu kadar yüksek riskli bir sistemin kamuya sunulmadan önce kullanımının etik ve toplumsal açıdan uygun olup olmadığını tartışabilir. Bir sosyal yardım algoritmasının hangi amaçlarla kullanılabileceğine, hangi verileri kullanmasına izin verilip verilmeyeceğine dair yönergeler oluşturabilir. Bu, ex-ante (önleyici) bir demokratik meşruiyet mekanizmasıdır.

Otomatik fayda sistemleri örneği, mevcut yönetişim modelinin yapısal adaletsizlikleri nasıl görünmez kıldığını ve derinleştirdiğini açıkça göstermektedir. Önerilen dört bileşenli çerçeve ise, bu sistemleri sadece daha adil algoritmalar tasarlamak için değil, aynı zamanda devletin yurttaşlarına karşı sorumluluğunu, şeffaflığını ve demokratik meşruiyetini yeniden tesis etmek için kullanılabilecek güçlü bir araç seti sunmaktadır.

## 5. Sonuç ve Tartışma

Bu çalışma, dijital dönüşüm çağında teknolojinin yalnızca ekonomik kalkınma, verimlilik ya da inovasyon alanlarıyla sınırlı bir fenomen olmadığını; aynı zamanda toplumsal adalet, eşitlik ve demokratik meşruiyet açısından köklü bir değişim içerdiğini ortaya koymuştur. Algoritmik karar alma sistemleri, özellikle kamu hizmetlerinde ve toplumsal kaynakların dağıtımında giderek daha görünür hale geldikçe, bunların yalnızca teknik araçlar değil, aynı zamanda toplumsal güç ilişkilerinin yeniden üretildiği siyasal aygıtlar olduğu açık biçimde ortaya çıkmaktadır. Bu bağlamda çalışma, algoritmik önyargı sorununu salt teknik bir hata ya da mühendislik eksikliği olarak değil, tarihsel eşitsizliklerin, kurumsal kör noktaların ve toplumsal tahakküm ilişkilerinin bir yansıması olarak ele almıştır.

Mevcut teknoloji yönetişimi yaklaşımlarının, algoritmik önyargıyı çoğunlukla bireysel haklar veya şeffaflık ilkeleri üzerinden ele alması, bu sorunun derinliğini kavramakta yetersiz kalmaktadır. Örneğin adil algoritmalar üretmek ya da açıklanabilir yapay zekâ mekanizmaları kurmak önemli olsa da, bu girişimler genellikle sorunu teknik düzeyde sınırlı tutmakta ve eşitsizliklerin kökenindeki toplumsal ve kurumsal yapıların sorgulanmasını göz ardı etmektedir. Bu da etik ilkeler ile somut düzenlemeler arasında giderek büyüyen bir uçuruma yol açmakta, teknolojinin toplumsal etkilerini yönetme kapasitesini zayıflatmaktadır. Çalışmanın en önemli katkısı bu noktada, adalet sorununu yapısal bir boyuta taşıyarak, yalnızca algoritmaların değil, algoritmaları tasarlayan, işleten ve denetleyen kurumların da sorumluluklarını merkezileştiren bir yaklaşım geliştirmesidir.

Önerilen Yapısal Adalet Odaklı Teknoloji Yönetişimi modeli, dört temel bileşen üzerinden somutlaştırılmıştır. İlk olarak, önleyici adalet ve katılımcı

tasarım ilkesi, algoritmik sistemlerin etkilediği toplulukların karar alma süreçlerine dahil edilmesini ve teknolojik zararın ortaya çıkmadan engellenmesini savunmaktadır. Bu yaklaşım, bireylerin yalnızca sonuçlara maruz kalan pasif özneler değil, tasarımın erken aşamalarından itibaren aktif katılımcılar olarak görülmesi gerektiğini ileri sürmektedir. İkinci olarak, sistemsel hesap verebilirlik boyutu, sorumluluğun tekil algoritmaların kara kutularında değil, bu sistemleri üreten şirketlerde, kamu kurumlarında ve onları denetleyen idari yapılarda olduğunu vurgulamaktadır. Bu, adaletin bireysel hatalara indirgenmesini değil, kurumsal mekanizmaların bütüncül biçimde sorgulanmasını gerektirir. Üçüncü bileşen olan bağlama duyarlı ve dekolonyal perspektif, teknolojinin evrensel ve nötr bir araç olduğu mitini reddederek, farklı kültürel bağlamların, yerel değerlerin ve tarihsel koşulların dikkate alınmasını zorunlu kılar. Böylelikle, teknoloji yönetişimi Batı merkezli bir standart setine indirgenmek yerine çoğulcu ve yerel farklılıklara duyarlı bir anlayışa evrilir. Son olarak, yeni demokratik aktörlerin sürece dahil edilmesi, teknolojik kararların yalnızca teknokratik uzmanların dar çerçevesinde kalmasını engelleyerek, vatandaş jürileri, bağımsız algoritmik denetçiler ve sivil toplum aktörleri gibi yeni mekanizmalarla demokratik meşruiyeti güçlendirmektedir.

Kamu sektöründeki otomatik fayda tahsisi sistemleri üzerine yapılan vaka analizi, bu dört bileşenli çerçevenin pratikte nasıl uygulanabileceğini göstermiştir. Bu inceleme, görünürde tarafsız ve nesnel olan algoritmik süreçlerin, aslında var olan eşitsizlikleri nasıl yeniden üretebildiğini ve kurumsal sorumluluğu nasıl görünmez kıldığını ortaya koymuştur. Böylelikle çalışma, yalnızca teorik bir çerçeve sunmakla kalmamış, aynı zamanda bu çerçevenin uygulamalara nasıl yansıyabileceğine dair ipuçları da geliştirmiştir.

Bununla birlikte, çalışma kendisini kavramsal bir öneriyle sınırlamaktadır. Önerilen çerçevenin operasyonel boyutlarının detaylandırılması, farklı siyasi-idari yapılara uyarlanması ve somut politika araçlarına dönüştürülmesi gelecek araştırmaların gündeminde olmalıdır. Özellikle, yapısal adaletin ölçülmesinde kullanılacak göstergeler, farklı bağlamlarda vatandaş katılımının nasıl etkin kılınacağı ve sistemsel hesap verebilirliği güvence altına alacak kurumsal tasarımlar üzerine ayrıntılı ampirik çalışmalar yapılması gerekmektedir. Bu noktada, disiplinlerarası araştırma yaklaşımı büyük önem taşımaktadır; hukuk, siyaset bilimi, sosyoloji ve bilgisayar mühendisliği gibi farklı alanların ortak katkısı olmaksızın bu paradigmanın pratikte hayata geçirilmesi mümkün görünmemektedir.

Sonuç olarak, dijital dönüşümde adalet arayışı, yalnızca teknoloji politikalarının değil, demokrasi ve toplumsal eşitlik mücadelelerinin de ayrılmaz bir parçasıdır. Yapısal Adalet Odaklı Teknoloji Yönetişimi modeli, algoritmik önyargının neden olduğu sorunları yalnızca teknik çözüm önerileriyle sınırlı

bırakmak yerine, bu sorunları toplumsal güç ilişkilerinin merkezine yerleştirmektedir. Bu yaklaşım, teknolojiyi insanlığın ortak geleceğini daha özgürleştirici ve eşitlikçi kılacak bir araç haline getirmek için paradigmatik bir dönüşüm çağrısıdır.

**Etik bildirim:** sbfdergi@politics.ankara.edu.tr

**Benzerlik denetimi:** Tamamlandı.

**Çıkar çatışması beyanı:** Yazar(lar) bir çıkar çatışması beyanında bulunmamıştır.

## Kaynakça

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities* (Open access ed.). fairmlbook.org.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed, 17*(2), 389–409. https://doi.org/10.2966/scrip.170220.389

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems, 29*, 4349–4357.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. İçinde *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT*)* (ss. 77–91). PMLR.

Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. NYU Press.

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Çev.). Vintage Books.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., & Daumé III, H. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a right to explanation. *AI Magazine, 38*(3), 50–57.

Green, B., & Viljoen, S. (2020). Algorithmic realism: Expanding the boundaries of algorithmic thought. İçinde *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)* (ss. 19–31). ACM. https://doi.org/10.1145/3351095.3372840

Habermas, J. (1987). *The theory of communicative action: Volume 2. Lifeworld and system: A critique of functionalist reason* (T. McCarthy, Çev.). Beacon Press.

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society, 22*(7), 900–915. https://doi.org/10.1080/1369118X.2019.1573912

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. İçinde *Proceedings of the 2016 Innovations in Theoretical Computer Science Conference (ITCS)* (ss. 43:1–43:23). ACM. https://doi.org/10.1145/2840728.2840760

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Metzinger, T. (2019, April 8). *Ethics washing made in Europe*. *Der Tagesspiegel*. https://www.tagesspiegel.de/politik/ethics-washing-made-in-europe-6353211.html

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. İçinde *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT*)* (ss. 220–229). ACM. https://doi.org/10.1145/3287560.3287596

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Mühlhoff, R. (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media philosophy of human-aided computing. *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin, 28*(2), 191–215. https://doi.org/10.1007/s00048-020-00265-3

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. İçinde *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)* (ss. 33–44). ACM. https://doi.org/10.1145/3351095.3372873

Raji, I. D., & Buolamwini, J. (2022). Actionable auditing revisited: Investigating the impact of publicly naming biased performance results of commercial AI products. *Communications of the ACM, 66*(1), 101–108.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. İçinde *Data & discrimination: Collected essays* (ss. 1–23). Open Technology Institute.

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. İçinde *Proceedings of the 2019 Conference on*

*Fairness, Accountability, and Transparency (FAT*)* (ss. 59–68). ACM. https://doi.org/10.1145/3287560.3287598

Sloane, M., Moss, E., & Awomolo, O. (2020). Participation is not a design fix: Limits of participatory design for ethical AI. İçinde *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3313831.3376592

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887.

Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values, 41*(1), 118–132. https://doi.org/10.1177/0162243915605575