Bilişsel Yanlılıktan Algoritmik Yanlılığa: İşe Alım Süreçlerine Nöroorganizasyonel Bir Bakış

From Cognitive Bias to Algorithmic Bias: A Neuro-Organizational Perspective on Recruitment Processes

Selma Kalkavan[1]

AXIS World

Journal of Applied Cross-disciplinary Insights

Cilt / Volume: 2

Sayı / Issue: 2

Ağustos / August 2026

Makale ilk gönderim tarihi / Received (First): 11.06.2026

Revizyon / Revision: 01.07.2026

Makale kabul tarihi /Accepted: 15.07.2026

[1]Corresponding Author: Dr., Bağımsız Araştırmacı, Türkiye, e-mail: selma@kalkavan.net

Anahtar Kelimeler: İnsan Kaynakları Yönetimi, İşe Alım, Nöro-Yanlılık, Yapay Zeka Etiği.
JEL Kodu: M12, D91, O33

Keywords: Human Resource Management, Recruitment, Neurobias, Artificial Intelligence Ethics.

JEL Codes: M12, D91, O33

Citation / Atıf: Kalkavan, S. (2026). Bilişsel Yanlılıktan Algoritmik Yanlılığa: İşe Alım Süreçlerine Nöroorganizasyonel Bir Bakış. *Axis World Journal of Applied Cross-disciplinary Insights*, 2(2), 88-107. https://doi.org/10.71284/axisw.2026222

Axis World Journal is a journal of the Academic Studies Group.

https://journals.academicianstudies.com/axis

Öz

İşe alım kararlarındaki bilişsel yanlılık ile yapay zeka sistemlerindeki algoritmik yanlılık, yönetim ve organizasyon alanında büyük ölçüde birbirinden bağımsız iki araştırma akışı olarak ele alınmaktadır. Bu çalışma, iki akışı birleştiren ve aralarındaki aktarım mekanizmasını modelleyen özgün bir kavramsal çerçeve önermeyi birincil teorik hedefi olarak belirlemektedir. Merkezi argüman, algoritmik yanlılığın sadece teknik bir problem olmadığı, insan nörobilişsel örüntülerinin bireysel kararlar, kurumsal veri yapıları ve algoritmik eğitim süreçleri aracılığıyla dijital sistemlere aktarılıp bu sistemler tarafından yeniden üretilebildiği varsayımına dayanmaktadır. Bu çerçevede çalışma, mevcut literatürde ayrı düzlemlerde ele alınan süreçleri çapraz düzey bir aktarım mekanizması içinde birleştiren "nöro-yanlılık" kavramını önermektedir. Makale, kavramın sınırlarını ve özgünlüğünü sistematik biçimde tanımlamakta, beş sınanabilir formal varsayım ortaya koymakta, kavramın bireysel, kurumsal ve algoritmik düzeylerde nasıl araştırılabileceğine ilişkin çok düzeyli bir çerçeve sunmakta, alternatif teorik yaklaşımlarla karşılaştırmalı bir konumlanma geliştirmekte ve otomasyon yanlılığı korumasını içeren dört katmanlı hibrit nöroorganizasyonel karar mimarisi (HNoKM) modelini önermektedir. Sonuç olarak çalışma, algoritmik yanlılığın teknik bir sistem sorunu olmadığını, bireysel karar örüntüleri, kurumsal veri yapıları ve algoritmik öğrenme süreçleri arasındaki etkileşimlerden beslenen çok düzeyli bir olgu olduğunu ileri sürmektedir. Önerilen nöro-yanlılık yaklaşımı ve Hibrit Nöroorganizasyonel Karar Mimarisi

(HNoKM), işe alım süreçlerinde daha adil, şeffaf ve sürdürülebilir karar sistemlerinin geliştirilmesine yönelik bütünleşik bir teorik çerçeve sunmaktadır.

Abstract

Cognitive bias in hiring decisions and algorithmic bias in artificial intelligence systems are largely treated as two independent research streams in the field of management and organization. This study establishes its primary theoretical objective as proposing an original conceptual framework that integrates these two streams and models the transmission mechanism between them. The central argument is based on the premise that algorithmic bias is not merely a technical problem, rather, human neurocognitive patterns are transferred to digital systems through individual decisions, institutional data structures, and algorithmic training processes, and are subsequently reproduced by these systems. Within this framework, the study proposes the concept of "neuro-bias," which unifies processes treated on separate planes in the current literature within a cross-level transmission mechanism. The article systematically defines the boundaries and originality of the concept, puts forward five testable formal propositions, presents a multi-level framework regarding how the concept can be investigated at individual, institutional, and algorithmic levels, develops a comparative positioning against alternative theoretical approaches, and proposes a four-layered Hybrid Neuro-Organizational Decision Architecture (HNoDA) model that includes automation bias protection. Consequently, the study posits that algorithmic bias is not a technical system flaw, but a multi-level phenomenon fueled by the interactions among individual decision patterns, institutional data structures, and algorithmic learning processes. The proposed neuro-bias approach and the Hybrid Neuro-Organizational Decision Architecture (HNoDA) provide an integrated theoretical framework for developing fairer, more transparent, and sustainable decision systems in recruitment processes.

GİRİŞ

İşe alım kararları, örgütlerin insan sermayesi niteliğini belirleyen en kritik süreçlerden biri olarak değerlendirilmektedir. Buna karşın uzun yıllardır yürütülen araştırmalar, bu kararların önemli ölçüde bilişsel yanlılıklardan etkilenebildiğini göstermektedir. Meta-analitik bulgular, yapılandırılmamış mülakatların iş performansını öngörme gücünün görece düşük düzeylerde seyrettiğini, yüksek yapılandırma içeren değerlendirme süreçlerinin ise daha güvenilir sonuçlar üretebildiğini ortaya koymaktadır (Levashina vd., 2014; Sackett vd., 2022). Bununla birlikte işe alım süreçlerinde sezgisel değerlendirme eğilimleri, kültürel uyum arayışı, benzerlik tercihleri ve değerlendiriciler arası tutarsızlıklar, insan kararlarının sistematik biçimde yanlılaşabildiğine işaret etmektedir. Thomas ve Reimann'ın (2023) araştırması, insan kaynakları uzmanlarının kendi yanlılık düzeylerini diğer değerlendiricilere kıyasla daha düşük algıladığını göstererek, kör nokta önyargısının işe alım süreçlerinde yaygın olabileceğine dair deneysel kanıt sunmaktadır.

Yapay zeka destekli işe alım araçlarının son yıllarda hızla yaygınlaşması, bu tabloya yeni bir boyut katmıştır. Algoritmik sistemler başlangıçta insan kaynaklı yanlılıkları azaltabilecek teknik çözümler olarak değerlendirilmiş olsa da, zaman içinde bu sistemlerin tarihsel veri yapılarındaki eşitsizlikleri yeniden üretebildiği görülmüştür. Raghavan vd. (2020), algoritmik işe alım sistemlerinde kullanılan mevcut yanlılık azaltma yaklaşımlarının sınırlılıklarını ortaya koyarken, Bandara vd. (2025), algoritmik yanlılık yönetim kapasitesinin stratejik insan kaynakları etkinliğiyle doğrudan ilişkili olduğunu göstermektedir. Benzer biçimde, otomasyon yanlılığı üzerine yürütülen araştırmalar, insan değerlendiricilerin algoritmik önerilere eleştirel değerlendirme yapmaksızın uyum gösterebildiğini ortaya koymaktadır (Goddard vd., 2012).

Bu gelişmelere rağmen insan bilişsel yanlılığı ile algoritmik yanlılık literatürleri büyük ölçüde birbirinden bağımsız ilerlemekte, algoritmik sistemlerin söz konusu örüntüleri hangi mekanizmalar aracılığıyla öğrendiği sorusu yeterince açıklanamamaktadır.

Söz konusu teorik boşluğa yanıt olarak nöro-yanlılık kavramı önerilmektedir. Çalışmanın temel argümanı, algoritmik yanlılığın teknik bir problem olarak değerlendirilemeyeceği, insan nörobilişsel örüntülerinin bireysel kararlar, kurumsal veri yapıları ve algoritmik eğitim süreçleri aracılığıyla dijital sistemlere aktarılıp bu sistemlerde yeniden üretilebildiği varsayımına dayanmaktadır. Bu doğrultuda makale, bilişsel yanlılık ile algoritmik yanlılık arasındaki ilişkiyi çapraz düzey bir aktarım mekanizması çerçevesinde ele almakta ve mevcut literatürde ayrı düzlemlerde incelenen süreçleri bütünleşik bir analitik model içinde tartışmayı amaçlamaktadır.

Makale bu kapsamda dört temel katkı sunmaktadır. Öncelikle, nöro-yanlılık kavramının teorik sınırlarını ve mevcut kavramlardan ayrıştığı noktaları sistematik biçimde tanımlamaktadır. Kavramın bireysel, kurumsal ve algoritmik düzeylerde nasıl araştırılabileceğine ilişkin çok düzeyli bir araştırma çerçevesi geliştirmesi de bir diğer katkı alanıdır. Üçüncü olarak, otomasyon yanlılığı riskini de dikkate alan Hibrit Nöroorganizasyonel Karar Mimarisi'ni (HNoKM) önermektedir. Son olarak çalışma, kurumsal örnekler üzerinden algoritmik yanlılığın nasıl güçlenebildiğini ve hangi yapısal müdahalelerle sınırlandırılabileceğini tartışarak teorik çerçevenin uygulama alanlarını değerlendirmektedir.

2. LİTERATÜR

2.1. İşe Alımda Bilişsel Yanlılık ve Karar Sapmaları

İşe alım, örgütlerin ihtiyaç duydukları insan kaynağını belirleme, değerlendirme ve seçme sürecini ifade etmektedir. İnsan sermayesinin örgütsel performans, verimlilik ve sürdürülebilir rekabet avantajı üzerindeki etkisi dikkate alındığında, işe alım kararları örgütler açısından stratejik öneme sahip süreçler arasında yer almaktadır. Bu süreçte adayların teknik yeterlilikleri, eğitim ve deneyim düzeyleri, davranışsal yetkinlikleri ve iş gereklerini karşılama kapasiteleri değerlendirilmektedir.

Bununla birlikte işe alım kararları sadece nesnel ölçütlere dayanmamaktadır. Örgütsel ihtiyaçlar, kültürel uyum beklentileri, ekip dinamikleri ve değerlendiricilerin adaylara ilişkin algıları da karar sürecini etkileyebilmektedir. Özellikle belirsizlik, zaman baskısı ve sınırlı bilgi koşullarında verilen kararlar, değerlendiricilerin sezgisel değerlendirmelere yönelmesine neden olabilmektedir. Bu durum, işe alım süreçlerini bilişsel yanlılıkların ortaya çıkabileceği önemli karar alanlarından biri haline getirmektedir (Highhouse, 2008; Kahneman, 2011; Sackett vd., 2022).

İşe alım kararlarında bilişsel yanlılık konusu, davranışsal karar verme literatürünün en köklü tartışma alanlarından birini oluşturmaktadır. Tversky ve Kahneman'ın (1974) heuristikler ve yanlılıklar yaklaşımı, bireylerin belirsizlik altında tam rasyonel davranmadığını, bilişsel kısa yollar aracılığıyla sistematik karar sapmaları üretebildiğini ortaya koymuştur. Bu yaklaşım, insan kaynakları yönetimi alanında özellikle işe alım, performans değerlendirme ve terfi kararları bağlamında geniş ölçüde uygulanmıştır.

İşe alım araştırmaları, değerlendiricilerin adayları teknik yeterlilik yanı sıra kültürel uyum, benzerlik algısı, sosyal yakınlık ve sezgisel izlenimler üzerinden de değerlendirebildiğini göstermektedir. Highhouse (2008), çalışan seçiminde sezgisel değerlendirmenin bilimsel olarak daha düşük geçerlilik üretmesine rağmen yöneticiler tarafından güçlü biçimde tercih edildiğini belirtmektedir. Benzer biçimde Rivera'nın (2012) elit profesyonel hizmet firmaları üzerine yürüttüğü çalışma, işe alım süreçlerinde kültürel eşleşme arayışının demografik ve

sosyal benzerlik temelli tercihleri güçlendirebildiğini göstermektedir. Bu bulgular, işe alım kararlarının bireysel uzmanlık değerlendirmesi dışında sosyal kategorizasyon süreçleriyle de ilişkili olduğunu düşündürmektedir.

Bilişsel yanlılık literatürü aynı zamanda karar tutarsızlığı problemini de gündeme taşımıştır. Kahneman, Sibony ve Sunstein'ın (2021) geliştirdiği gürültü kavramı, aynı adayın farklı değerlendiriciler tarafından ya da aynı değerlendirici tarafından farklı zamanlarda tutarsız biçimde puanlanabildiğini göstermektedir. Gürültü problemi, sistematik önyargıdan farklı olarak karar sürecindeki rastlantısal değişkenliği ifade etmekte ancak sonuçları bakımından örgütsel adalet ve seçim güvenilirliği üzerinde benzer biçimde olumsuz etkiler yaratabilmektedir.

Bu literatür, işe alım süreçlerinde bilişsel yanlılıkların varlığını güçlü biçimde belgelemekle birlikte, söz konusu örüntülerin dijital sistemlere nasıl aktarılabildiği sorusuna sınırlı ölçüde yanıt vermektedir. Çalışmalar büyük ölçüde bireysel karar verici düzeyinde yoğunlaşmakta, insan kararlarının kurumsal veri yapıları ve algoritmik sistemlerle nasıl ilişkilendiği yeterince açıklanmamaktadır.

2.2. Örgütsel Nörobilim ve Karar Verme Literatürü

Örgütsel nörobilim literatürü, karar verme süreçlerinin davranışsal çıktılarla birlikte bilişsel ve nörofizyolojik mekanizmalar aracılığıyla da incelenebileceğini öne sürmektedir. Becker, Cropanzano ve Sanfey (2011) ile Senior, Lee ve Butler (2011), örgütsel davranış araştırmalarının karar verme süreçlerinin altında yatan bilişsel mekanizmaları dikkate almadan eksik kalabileceğini savunmaktadır. Bu yaklaşım, özellikle belirsizlik, risk, sosyal değerlendirme ve ödül beklentisi gibi süreçlerin örgütsel kararları nasıl etkileyebileceğini anlamaya yönelmektedir.

Sosyal biliş ve nörobilim araştırmaları, bireylerin grup üyeliği, sosyal benzerlik ve tehdit algısı gibi uyaranlara farklı bilişsel ve duygusal tepkiler verebildiğini göstermektedir (Cunningham vd., 2004; Phelps & LeDoux, 2005). Bununla birlikte örgütsel bağlamda bu bulguların doğrudan nedensel çıkarımlar üretmek için kullanılmasının önemli metodolojik sınırlılıkları bulunmaktadır. Poldrack (2018), belirli beyin bölgeleri ile karmaşık sosyal davranışlar arasında doğrudan eşitleme yapılmasının ters çıkarım (reverse inference) riskini artırdığını belirtmektedir. Örgütsel nörobilim alanında elde edilen bulguların, davranışsal süreçlerle ilişkisellik düzeyinde değerlendirilmesi gerektiği vurgulanmaktadır.

Ayrıca, örgütsel nörobilim literatürü karar verme süreçlerinin tamamen rastlantısal olmadığını, belirsizlik, bilişsel yük ve sosyal değerlendirme gibi koşullar altında belirli bilişsel eğilimlerin sistematik biçimde devreye girebildiğini göstermektedir. Karar verme literatüründe ödül beklentisi, belirsizlik azaltma eğilimi ve bilişsel yük gibi süreçlerin sezgisel kararları etkileyebileceğine dair kapsamlı bulgular bulunmaktadır. Kahneman'ın (2011) Sistem 1 ve Sistem 2 yaklaşımı, zaman baskısı ve bilişsel yük altında hızlı ve sezgisel karar süreçlerinin baskın hale gelebildiğini öne sürmektedir. Simon'ın (1955) sınırlı rasyonellik yaklaşımı ise örgütsel kararların çoğu zaman eksik bilgi ve sınırlı bilişsel kapasite altında alındığını göstermektedir. İşe alım süreçleri, yüksek belirsizlik, yoğun aday havuzu ve zaman baskısı nedeniyle bu bilişsel dinamiklerin yoğun biçimde ortaya çıkabildiği alanlardan biri olarak değerlendirilebilir.

Bununla birlikte örgütsel nörobilim literatürü büyük ölçüde bireysel karar süreçlerine odaklanmakta, bu süreçlerin kurumsal veri sistemleri ve algoritmik yapılarla nasıl bütünleştiği sorusunu sınırlı biçimde ele almaktadır. Özellikle insan nörobilişsel örüntülerinin dijital karar sistemlerine nasıl aktarıldığı ve algoritmik yapılarda nasıl yeniden üretildiği konusu, mevcut literatürde bütünleşik bir teorik çerçeve içinde yeterince açıklanmamaktadır.

2.3. Algoritmik Yanlılık ve Yapay Zeka Destekli İK Sistemleri

Yapay zeka destekli işe alım sistemlerinin yaygınlaşmasıyla birlikte algoritmik önyargı konusu insan kaynakları yönetimi alanının temel tartışma alanlarından biri haline gelmiştir. Köchling ve Wehner'ın (2020) sistematik derlemesi, algoritmik karar sistemlerinde ortaya çıkan yanlılıkların veri yapısı, model tasarımı, optimizasyon hedefleri ve temsil problemleri gibi çok sayıda kaynaktan beslenebildiğini göstermektedir. Mehrabi vd. (2021), algoritmik adalet araştırmalarında doğruluk, eşitlik ve temsil arasında çoğu zaman bir denge problemi bulunduğunu ortaya koymaktadır.

Algoritmik işe alım sistemleri başlangıçta insan kaynaklı önyargıları azaltabilecek daha objektif araçlar olarak değerlendirilmiş olsa da, zaman içinde bu sistemlerin tarihsel veri yapılarındaki eşitsizlikleri yeniden üretebildiği görülmüştür. Barocas ve Selbst'in (2016) vekil değişken yaklaşımı, görünüşte nötr değişkenlerin dolaylı biçimde demografik özellikleri temsil ederek ayrımcı sonuçlar üretebildiğini göstermektedir. Benzer biçimde Noble (2018) ve O'Neil (2016), algoritmik sistemlerin toplumsal eşitsizlikleri teknik tarafsızlık görünümü altında yeniden üretebildiğini ileri sürmektedir.

İnsan merkezli yapay zeka yaklaşımı, bu sorunlara karşı insan denetimini ve açıklanabilirliği güçlendirmeyi hedeflemektedir. Shneiderman (2022), güvenilir AI sistemlerinin teknik doğruluk dışında, insan kontrolü, şeffaflık ve hesap verebilirlik ilkeleri üzerinden değerlendirilmesi gerektiğini savunmaktadır. Hunkenschroer ve Luetge (2022) ise AI destekli işe alım süreçlerinde etik risklerin algoritmik modelden kaynaklanmadığını, veri üretim süreçlerinden ve örgütsel karar yapılarından da kaynaklandığını belirtmektedir.

Bununla birlikte algoritmik yanlılık literatürü çoğu zaman teknik sistem düzeyinde yoğunlaşmakta, algoritmik sistemlerin beslendiği insan karar örüntülerinin bilişsel ve örgütsel kökenlerini ikincil planda bırakmaktadır. İnsan kararları ile algoritmik çıktılar arasındaki ilişkinin çoğu araştırmada doğrusal biçimde ele alındığı aktarım sürecinin dinamik ve çok düzeyli yapısının yeterince açıklanmadığı görülmektedir.

2.4. İnsan Kararları ile Algoritmik Sistemler Arasındaki Çapraz Düzey Aktarım Boşluğu

İnsan kaynakları yönetimi, örgütsel davranış ve algoritmik karar verme literatürleri, işe alım süreçlerindeki önyargı problemini farklı analiz düzeylerinde incelemektedir. Bilişsel yanlılık araştırmaları çoğunlukla bireysel karar vericinin sezgisel değerlendirme örüntülerine odaklanırken, algoritmik yanlılık çalışmaları eğitim verilerindeki temsil sorunları ve model çıktılarındaki eşitsizlikler üzerinde yoğunlaşmaktadır. Kurumsal teori ise örgütsel normların ve homojenleşme süreçlerinin karar yapıları üzerindeki etkisini açıklamaktadır. Bununla birlikte bu literatürlerin büyük bölümü, söz konusu süreçleri birbirinden görece bağımsız analiz alanları olarak ele almaktadır.

Mevcut araştırmaların önemli bir kısmı, yanlılığın belirli bir düzeyde nasıl ortaya çıktığını açıklasa da, bu örüntülerin bireysel kararlardan kurumsal veri yapılarına, oradan da algoritmik sistemlere nasıl taşındığını açıklayan bütünleşik bir çerçeve sunmamaktadır. Özellikle insan kararlarının zaman içinde kurumsal veri tabanlarını nasıl biçimlendirdiği, algoritmik sistemlerin bu örüntüleri hangi mekanizmalar aracılığıyla öğrendiği ve öğrenilen yapıların örgütsel süreçlere nasıl geri beslendiği soruları literatürde parçalı biçimde ele alınmaktadır. Bu durum, AI destekli işe alım sistemlerinde ortaya çıkan önyargıların sadece teknik sistem hatalarıyla açıklanmasını güçleştirmektedir. Çünkü algoritmik sistemler çoğu zaman tarihsel insan kararları üzerinde eğitilmekte ve bundan dolayı geçmiş örgütsel tercih örüntülerini yeniden üretme kapasitesi taşımaktadır. Ancak mevcut literatürde bireysel, kurumsal ve algoritmik düzeyleri aynı anda dikkate alan çapraz düzey bir aktarım modeli henüz yeterince geliştirilmiş değildir.

Bu çalışma, söz konusu teorik boşluğa yanıt olarak nöro-yanlılık kavramını önermekte ve bu kavramı bir sonraki bölümde ayrıntılı biçimde tanımlamaktadır.

3. NÖRO-YANLILIK (NEUROBIAS): KAVRAMSALLAŞTIRMA

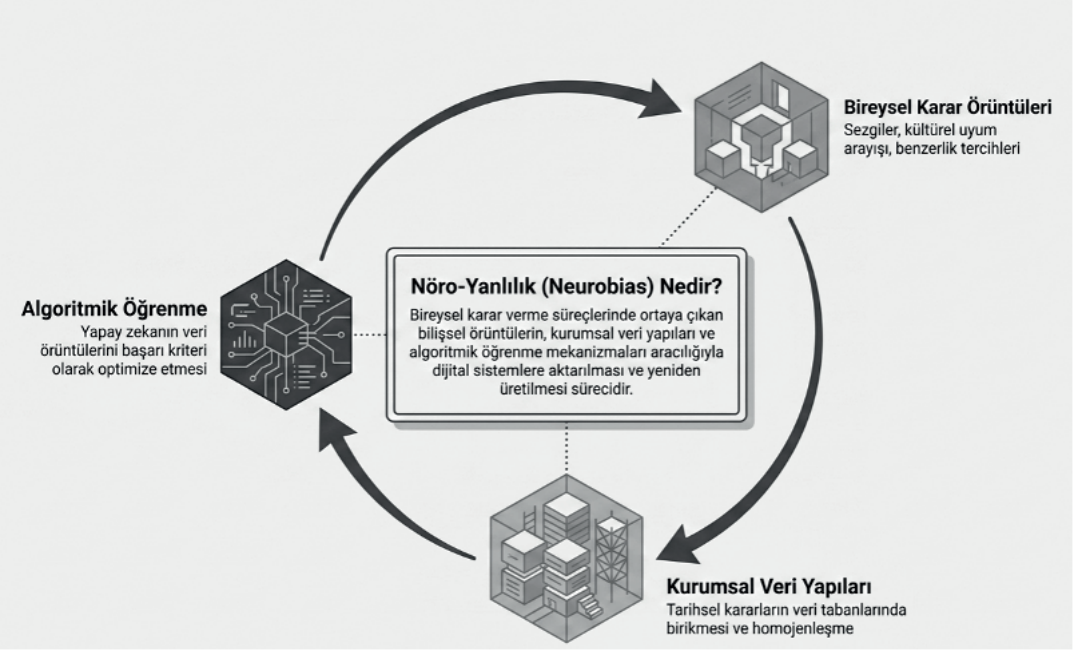
3.1. Tanım ve Teorik Konumlanma

Bu çalışmada nöro-yanlılık (Neurobias), bireysel karar verme süreçlerinde ortaya çıkan bilişsel örüntülerin kurumsal uygulamalar, tarihsel veri yapıları ve algoritmik öğrenme mekanizmaları aracılığıyla farklı analiz düzeylerine aktarılması ve yeniden üretilmesini açıklayan kavramsal bir aktarım mekanizması olarak önerilmektedir. Kavramın temel amacı, bireysel bilişsel yanlılıklar ile algoritmik yanlılık arasında yer alan kurumsal süreçleri görünür kılan bütünleşik bir açıklama çerçevesi sunmaktır.

Bu kavramsallaştırma, bilişsel yanlılık (Kahneman, 2011), örtük yanlılık (Greenwald & Banaji, 1995), algoritmik yanlılık (Barocas & Selbst, 2016) ve örgütsel nörobilim literatüründe geliştirilen insan karar verme yaklaşımlarından (Becker vd., 2011; Senior vd., 2011; Lindebaum, 2014) yararlanılarak geliştirilmiştir. Bununla birlikte, önerilen kavram mevcut yaklaşımların doğrudan birleşimi olmayıp, bu literatürleri çapraz düzey aktarım perspektifi altında bütünleştiren yeni bir teorik öneri niteliği taşımaktadır.

Bu çalışmada kullanılan "nöro" ifadesi, nörobiyolojik bir nedensellik iddiasını içermemekte, karar verme davranışlarının nörobilim tarafından açıklanan bilişsel örüntülerini esas alan metaforik ve analitik bir kavramsallaştırmayı ifade etmektedir. Dolayısıyla çalışma, belirli beyin bölgeleri veya sinir ağları üzerinden davranış açıklaması yapmayı amaçlamamakta, bireysel bilişsel örüntülerin örgütsel uygulamalar ve algoritmik sistemler içinde yeniden üretilebildiği süreçleri kavramsal düzeyde açıklamaktadır.

Şekil 1. Nöro-Yanlılık Kavramının Bileşenleri



Kaynak: Yazar tarafından literatür dikkate alınarak oluşturulmuştur.

Bu yaklaşım özellikle örgütsel nörobilim alanında dile getirilen metodolojik eleştiriler dikkate alınarak geliştirilmiştir. Nörobilim terminolojisinin açıklayıcı gücünün ötesinde kullanılmasının gereğinden fazla bilimsel ikna etkisi oluşturabileceği (Weisberg vd., 2008) ve karmaşık örgütsel davranışların tek başına beyin temelli açıklamalarla yorumlanmasının kavramsal aşırı genişleme ile ters çıkarım riskini artırabileceği vurgulanmaktadır (Lindebaum, 2014; Poldrack, 2018). Bu nedenle nöro-yanlılık, biyolojik belirlenim yaklaşımı olarak kullanılmamış, bilişsel örüntülerin kurumsal ve teknolojik sistemlerde yeniden üretimini açıklayan teorik bir çerçeve olarak kullanılmaktadır.

Alternatif olarak bu süreç "örtük yanlılığın kurumsallaşması", "sosyo-bilişsel aktarım" veya "kurumsal bilişsel aktarım" gibi kavramlarla da açıklanabilir. Ancak bu çalışma, bireysel karar süreçlerinden başlayarak örgütsel uygulamalar, veri yapıları ve algoritmik öğrenme mekanizmalarına uzanan çok düzeyli aktarım sürecini vurguladığı için nöro-yanlılık (Neurobias) kavramını tercih etmektedir.

Mevcut algoritmik yanlılık literatürü çoğunlukla veri setlerindeki temsil sorunları, model tasarımı veya teknik hata kaynakları üzerine odaklanırken, bilişsel yanlılık araştırmaları bireysel karar sapmalarını açıklamaktadır. Buna karşılık bireysel bilişsel örüntülerin kurumsal insan kaynakları uygulamaları aracılığıyla tarihsel veri yapılarına, oradan da yapay zeka sistemlerine aktarılması ve algoritmik çıktılar yoluyla yeniden insan kararlarını etkilemesi biçimindeki döngüsel aktarım süreci sınırlı ölçüde kavramsallaştırılmıştır.

Bu nedenle nöro-yanlılık kavramı, yanlılığı sadece bireysel veya algoritmik bir olgu olarak değerlendirmemekte, insan-örgüt-veri-algoritma etkileşimi içinde sürekli yeniden üretilen dinamik bir süreç olarak ele almaktadır. Çalışmada önerilen HNoKM modeli de bu aktarım mekanizmasını görünür kılarak karar noktalarında gerçekleştirilecek insan müdahaleleriyle yanlılığın azaltılmasını hedefleyen uygulamaya dönük bir çerçeve sunmaktadır.

Aşağıdaki tablo, nöro-yanlılık kavramının mevcut yaklaşımlarla ilişkisini ve hangi noktada ayrıştığını özetlemektedir.

Tablo 1. Nöro-Yanlılığın Mevcut Kavramlarla Karşılaştırmalı Teorik Konumlanması

Kavram	Analiz Düzeyi	Temel Odak	İnsan Kararlarından Algoritmik Süreçlere Aktarımı Açıklama Biçimi
Örtük Yanlılık (Greenwald & Banaji, 1995)	Bireysel	Bilinç dışı tutumlar	Bireysel tutum süreçlerine odaklanır, kurumsal ve algoritmik yeniden üretim süreçlerini doğrudan ele almaz
Bilişsel Yanlılık (Tversky & Kahneman, 1974)	Bireysel	Karar verme süreçlerindeki bilişsel kısayollar ve sistematik sapmalar	Karar süreçlerindeki bilişsel sapmaları açıklar, dijital öğrenme süreçlerine sınırlı odaklanır
Tarihsel Veri Yanlılığı (Barocas & Selbst, 2016)	Algoritmik	Eğitim verilerindeki eşitsizlikler	Veri setlerindeki temsil sorunlarını açıklar, insan karar örüntülerinin oluşum sürecini sınırlı ele alır
Kurumsal Önyargı (DiMaggio & Powell, 1983)	Kurumsal	Normlar ve örgütsel homojenleşme	Kurumsal tekrar ve norm süreçlerini açıklar, bilişsel süreçler ile algoritmik öğrenme arasındaki ilişkiyi doğrudan modellemez
Nöro-Yanlılık (Bu Çalışma)	Bireysel + Kurumsal + Algoritmik	Çapraz düzey aktarım mekanizması	İnsan karar örüntülerinin kurumsal veri yapıları ve algoritmik öğrenme süreçleri aracılığıyla yeniden üretimini birlikte ele almayı amaçlar

Kaynak: Yazar tarafından ilgili literatür temel alınarak oluşturulmuştur.

3.2. Kavramsal Sınırlar ve Kapsam

Bu çalışma kapsamında nöro-yanlılık, insan karar örüntülerinin bireysel düzeyde ortaya çıktıktan sonra kurumsal uygulamalar, tarihsel veri yapıları ve algoritmik öğrenme süreçleri boyunca aktarılması ve yeniden üretilmesini açıklayan kavramsal bir çerçeve olarak önerilmektedir. Bu kavram, özellikle insan kararlarının kurumsal veri üretimini şekillendirdiği ve bu verilerin yapay zeka sistemlerinin öğrenme sürecine kaynak oluşturduğu örgütsel bağlamlarda açıklayıcı bir işlev üstlenmektedir.

Algoritmik yanlılık literatürü, dijital sistemlerde ortaya çıkan eşitsizliklerin insan kararlarının yanısıra veri işleme süreçleri, model mimarisi, optimizasyon tercihleri ve temsil problemleri gibi teknik unsurlardan da kaynaklanabileceğini göstermektedir (Barocas & Selbst, 2016; Mehrabi vd., 2021). Mevcut çalışma, insan karar örüntülerinden bağımsız biçimde ortaya çıkan teknik model hatalarını nöro-yanlılık çerçevesinin dışında değerlendirmektedir.

Benzer biçimde belirli grupları dışlamak amacıyla bilinçli olarak tasarlanan ayrımcı sistemler de nöro-yanlılık yaklaşımının temel inceleme alanı dışında kalmaktadır. Çünkü bu tür sistemler, bilinç dışı bilişsel eğilimlerden çok açık kurumsal tercihlerin ve yönetsel kararların sonucu olarak değerlendirilmektedir. Critical algorithm studies literatürü, algoritmik sistemlerin bazı durumlarda kurumsal güç ilişkileri ve bilinçli politika tercihleri doğrultusunda şekillenebildiğini vurgulamaktadır (Noble, 2018; O'Neil, 2016). Her algoritmik eşitsizliğin otomatik biçimde nöro-yanlılık çerçevesi içinde yorumlanması kavramın kapsamını aşırı genişletme riski taşımaktadır.

Ayrıca algoritmik çıktılarda gözlenen tüm demografik farklılıkların doğrudan insan bilişsel örüntülerine indirgenmesi de metodolojik açıdan sorunlu olabilir. Selbst vd. (2019), algoritmik sistemlerin toplumsal bağlamdan bağımsız değerlendirilemeyeceğini, iş gücü piyasası yapıları, eğitim erişimi ve tarihsel eşitsizlikler gibi daha geniş sosyoteknik dinamiklerin de sonuçlar üzerinde etkili olduğunu belirtmektedir. Nöro-yanlılık yaklaşımı bu nedenle, insan karar örüntülerinin dijital sistemlere aktarım süreçlerini anlamaya yönelik sınırlı fakat bütünleştirici bir teorik çerçeve olarak konumlandırılmaktadır.

Kavramın en anlamlı hale geldiği bağlamlar, geçmiş insan işe alım kararlarının yoğun biçimde veri ürettiği, kurumsal homojenleşmenin tarihsel süreklilik gösterdiği ve algoritmik sistemlerin insan karar örüntülerini optimize etmeye çalıştığı örgütsel yapılardır. İnsan merkezli yapay zeka literatürü, algoritmik sistemlerin etik etkilerinin çoğu zaman insan karar süreçleriyle birlikte değerlendirilmesi gerektiğini vurgulamaktadır (Shneiderman, 2022; Hunkenschroer & Luetge, 2022). İnsan karar süreçlerinin tamamen dışarıda bırakıldığı tam otomasyon senaryolarında nöro-yanlılık yaklaşımının açıklayıcı kapasitesinin daha sınırlı kalabileceği düşünülmektedir.

3.3. Nöro-Yanlılığın Bileşenleri ve Algoritmik Yeniden Üretim Süreci

Nöro-yanlılık yaklaşımı, insan karar süreçlerinde etkili olabilecek bazı bilişsel ve sosyal örüntülerin kurumsal veri yapıları aracılığıyla algoritmik sistemlerde yeniden üretilebildiği varsayımına dayanmaktadır. Bu bileşenler davranışsal eğilimlerle ilişkili teorik örüntüler olarak ele alınmaktadır.

Tablo 2. Nöro-Yanlılığın Teorik Bileşenleri ve Olası Yeniden Üretim Dinamikleri

Bileşen	Teorik Dayanak	İşe Alım Süreçlerine Olası Yansıması	Algoritmik Yeniden Üretim Riski
Tehdit ve Belirsizlik Değerlendirmesi	Sosyal biliş, tehdit algısı ve duygusal değerlendirme araştırmaları (Cunningham vd., 2004; Phelps & LeDoux, 2005)	Belirsizlik içeren durumlarda tanıdık aday profillerine yönelme eğilimi	Benzer aday profillerinin düşük riskli örüntüler olarak öğrenilmesi
Sosyal Benzerlik Tercihi	Sosyal Kimlik Teorisi ve kültürel uyum araştırmaları (Tajfel & Turner, 1979; Rivera, 2012)	Benzer eğitim, kültürel geçmiş veya deneyime sahip adayların daha olumlu değerlendirilmesi	Kurumsal homojenliğin veri aracılığıyla yeniden üretilmesi
Sezgisel Karar Süreçleri	Sınırlı rasyonellik ve Sistem 1 yaklaşımı (Simon, 1955; Kahneman, 2011)	Zaman baskısı altında hızlı ve sezgisel değerlendirmelerin artması	Geçmiş sezgisel tercihlerin algoritmik örüntülere dönüşmesi
Kültürel Şema Yapıları	Şema teorisi ve sosyal kategorizasyon çalışmaları (Fiske & Taylor, 2013)	Üniversite, kariyer geçmişi veya dil yeterliliği gibi göstergelerin başarıyla özdeşleştirilmesi	Görünüşte tarafsız değişkenler üzerinden dolaylı ayrımcılığın yeniden üretilmesi
Sosyal Ödül ve Uyum Eğilimleri	Sosyal ödül, sosyal bağlanma ve olumlu sosyal değerlendirme araştırmaları (Tabibnia & Lieberman, 2007)	Kurum kültürüyle daha uyumlu veya sosyal yakınlık hissi uyandıran adayların daha olumlu değerlendirilmesi	Başarı profillerinin belirli sosyal örüntüler etrafında yoğunlaşması

Kaynak: Cunningham vd. (2004), Phelps ve LeDoux (2005), Tajfel ve Turner (1979), Simon (1955), Kahneman (2011), Rivera (2012), Fiske ve Taylor (2013), Tabibnia ve Lieberman (2007), Barocas ve Selbst (2016) temel alınarak yazar tarafından oluşturulmuştur.

Tablo 2'de yer alan bileşenlerin hiçbiri tek başına algoritmik yanlılığın oluşumunu açıklamamaktadır. Algoritmik yanlılık, bireysel karar örüntülerinin kurumsal uygulamalar aracılığıyla veri yapılarına yansıması ve yapay zeka sistemlerinin bu tarihsel örüntüler üzerinden öğrenmesi sonucunda ortaya çıkan çok düzeyli bir süreçtir. Bu nedenle nöro-yanlılık yaklaşımı, bireysel kararlar, kurumsal veri üretimi ve algoritmik öğrenme süreçleri arasındaki karşılıklı etkileşim sonucunda zaman içinde yeniden üretilen dinamik bir yapı olarak değerlendirmektedir. Bu kavramsallaştırma, çalışmada önerilen HNoKM modelinin teorik temelini oluşturmakta ve yanlılığın hangi aşamalarda ortaya çıkabileceğini ve hangi müdahale noktalarında azaltılabileceğini açıklamayı amaçlamaktadır.

3.4. Nöro-Yanlılığın Çapraz Düzey Yeniden Üretim Mekanizması

Nöro-yanlılık yaklaşımı kapsamında önerilen yeniden üretim döngüsü, bireysel karar örüntülerinin zaman içinde kurumsal veri yapıları ve algoritmik öğrenme süreçleri aracılığıyla nasıl pekişebileceğini kavramsal düzeyde göstermektedir. Bu çerçevede süreç, birbirini besleyen aşamalardan oluşan dinamik bir yapı olarak ele alınmaktadır. Özellikle AI destekli işe alım sistemlerinde geçmiş karar örüntülerinin yeni karar süreçlerini etkileyebilmesi, örgütsel homojenleşmenin dijital sistemler aracılığıyla yeniden güçlenmesine zemin hazırlayabilmektedir.

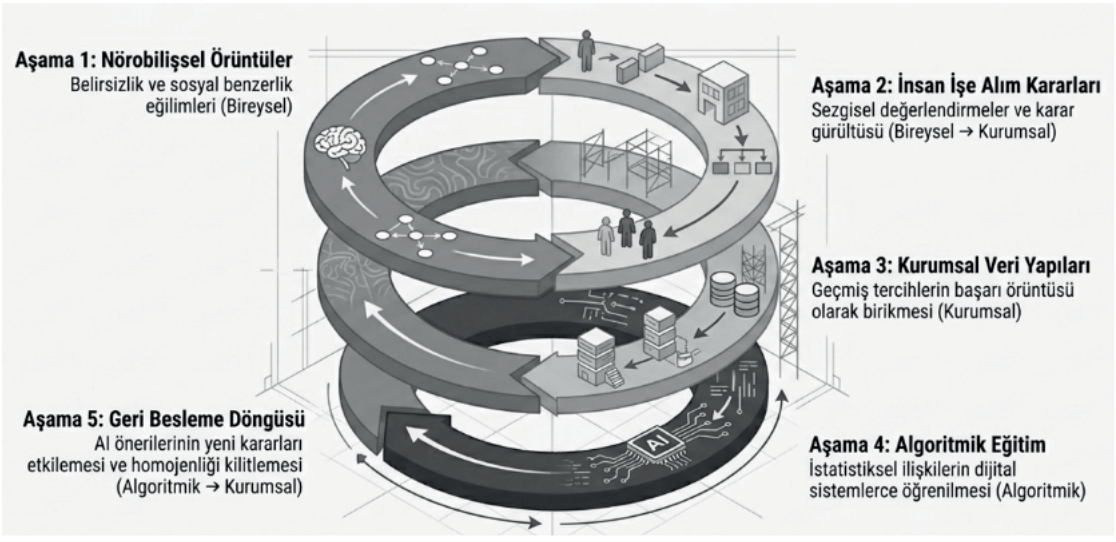
Aşağıdaki çerçeve, nöro-yanlılığın olası yeniden üretim ve güçlenme döngüsünü kavramsal düzeyde özetlemektedir.

Tablo 3. Nöro-Yanlılığın Yeniden Üretim ve Güçlenme Döngüsü

Aşama	Düzey ve Süreç	İçerik ve Mekanizma	Olası Müdahale Noktaları
1	**Nörobilişsel Örüntüler (Bireysel)**	Belirsizlik değerlendirmesi, sosyal benzerlik eğilimleri, sezgisel karar süreçleri ve kültürel şema yapıları işe alım değerlendirmelerini etkileyebilmektedir.	Örtük Çağrışım Testi (IAT) uygulamaları, yapılandırılmış değerlendirme eğitimleri, göz izleme ve yanıt gecikmesi ölçümleri
2	**İnsan İşe Alım Kararları (Bireysel → Kurumsal)**	İşe alım kararlarında kültürel uyum arayışı, benzerlik tercihleri, değerlendirme gürültüsü ve otomasyon yanlılığı gibi örüntüler ortaya çıkabilmektedir.	Yapılandırılmış mülakat sistemleri, kör özgeçmiş uygulamaları, bağımsız değerlendirme protokolleri
3	**Kurumsal Veri Yapılarının Oluşumu (Kurumsal)**	Geçmiş işe alım kararlarının veri tabanlarında birikmesi belirli aday profillerinin başarı örüntüsü olarak görünür hale gelmesine yol açabilmektedir.	Veri temsil dengesi denetimleri, çeşitlilik analizleri, veri temizleme süreçleri
4	**Algoritmik Eğitim ve Çıktılar (Algoritmik)**	Algoritmik sistemler tarihsel veri örüntülerini istatistiksel ilişkiler biçiminde öğrenebilmekte, bu durum geçmiş tercihlerin dijital ölçekte yeniden görünür hale gelmesine neden olabilmektedir.	Ayrımcı etki analizleri, algoritmik adalet metrikleri, açıklanabilir yapay zeka (XAI) uygulamaları, model kalibrasyonu
5	**Kurumsal Geri Besleme ve Güçlenme Döngüsü (Kurumsal → Algoritmik)**	Algoritmik önerilerin yeni işe alım kararlarını etkilemesi, benzer profillerin yeniden seçilmesine ve örgütsel homojenleşmenin güçlenmesine yol açabilmektedir.	HNoKM modeli, etik yönetişim mekanizmaları, çeşitlilik politikaları, insan denetimli karar süreçleri

Kaynak: Kahneman (2011), Greenwald ve Banaji (1995), Simon (1955), Fiske ve Taylor (2013), Barocas ve Selbst (2016), Rivera (2012), Phelps ve LeDoux (2005) ile açıklanabilir yapay zeka ve algoritmik adalet literatürü temel alınarak yazar tarafından geliştirilmiştir.

Şekil 2. Nöro-Yanlılığın 5 Aşaması



Kaynak: Yazar tarafından literatür dikkate alınarak oluşturulmuştur.

Tablo 3 ve Şekil 2'de özetlenen çerçeve, işe alım süreçlerinde ortaya çıkan karar örüntülerinin bireysel düzeyden başlayarak kurumsal veri yapılarına ve algoritmik öğrenme süreçlerine nasıl aktarıldığını ve zaman içinde geri besleme mekanizmaları aracılığıyla nasıl pekişebildiğini kavramsal düzeyde göstermektedir. Bu yaklaşım, algoritmik yanlılığın sadece teknik sistemlerin bir özelliği olmadığını, insan kararları, kurumsal uygulamalar ve yapay

zeka sistemleri arasındaki karşılıklı etkileşim sonucunda yeniden üretilebilen dinamik bir süreç olduğunu ileri sürmektedir. Bu doğrultuda önerilen Hibrit Nöroorganizasyonel Karar Mimarisi (HNoKM), yanlılığın farklı aşamalarda ortaya çıkabileceği müdahale noktalarını görünür kılarak geri besleme döngüsünün sınırlandırılmasına yönelik teorik bir çerçeve sunmaktadır.

4. VARSAYIMSAL ÇERÇEVE VE ARAŞTIRMA TASARIMI

Bu çalışma, nöro-yanlılık kavramını bireysel karar örüntüleri, kurumsal veri yapıları ve algoritmik öğrenme süreçleri arasındaki olası ilişkileri açıklamaya yönelik teorik bir çerçeve olarak ele almaktadır. Bu nedenle aşağıda sunulan ifadeler kesin nedensel ilişkiler öne sürmek yerine, gelecekte ampirik olarak sınanabilecek teorik iddialar olarak değerlendirilmelidir. Özellikle insan kaynakları yönetiminde yapay zeka kullanımına ilişkin güncel çalışmalar, geçmiş insan kararlarının veri setleri aracılığıyla algoritmik sistemlerde yeniden görünür hale gelebildiğini göstermektedir (Barocas & Selbst, 2016; Köchling & Wehner, 2020; Mehrabi vd., 2021).

Bu kapsamda ilk varsayım, bireysel işe alım kararları ile kurumsal veri örüntüleri arasındaki ilişkiye odaklanmaktadır. Yapılandırılmamış işe alım süreçlerinde sezgisel değerlendirmelerin ve sosyal benzerlik eğilimlerinin daha görünür hale gelebildiği bilinmektedir (Kahneman, 2011; Rivera, 2012). Bu çerçevede değerlendiricilerin örtük tercih örüntülerinin, zaman içinde kurumsal işe alım verilerinde belirli aday profillerinin yoğunlaşmasına katkı sağlayabileceği varsayılmaktadır.

Bir diğer varsayım, kurumsal veri yapıları ile algoritmik öğrenme süreçleri arasındaki ilişkiye yönelmektedir. Algoritmik sistemlerin büyük ölçüde tarihsel insan kararları üzerinde eğitilmesi, geçmiş örgütsel tercih örüntülerinin dijital sistemlerde yeniden üretilebilmesi riskini ortaya çıkarmaktadır (Barocas & Selbst, 2016). Özellikle demografik homojenliğin yüksek olduğu kurumsal veri yapılarında, algoritmik sistemlerin belirli aday profillerini başarı göstergesi olarak öğrenme olasılığının artabileceği düşünülmektedir.

İşe alım süreçlerinin yapılandırılma düzeyi de bir diğer varsayım alanıdır. Yapılandırılmış mülakatlar, standart değerlendirme kriterleri ve çoklu değerlendirici sistemlerinin bilişsel yanlılıkları azaltabildiğine ilişkin bulgular bulunmaktadır (Levashina vd., 2014; Sackett vd., 2022). Yapılandırma düzeyi arttıkça bireysel karar örüntülerinin kurumsal veri yapıları üzerindeki etkisinin zayıflayabileceği varsayılmaktadır.

Otomasyon yanlılığı varsayımı kapsamında, insanların algoritmik önerileri çoğu zaman eleştirel değerlendirmeden kabul edebildiği ve bunun insan denetimini zayıflatabildiği gösterilmiştir (Goddard vd., 2012). Algoritmik önerilere dayalı işe alım süreçlerinin, geçmiş veri örüntülerini güçlendirme riski taşıyabileceği düşünülmektedir. Buna karşılık insan denetimi, yapılandırılmış değerlendirme süreçleri ve açıklanabilir yapay zeka uygulamalarını birlikte içeren hibrit karar mimarilerinin bu riski sınırlandırabileceği varsayılmaktadır.

Son varsayım ise açıklanabilirlik ve örgütsel güven ilişkisine odaklanmaktadır. Yapay zeka sistemlerinin karar gerekçelerini şeffaf biçimde sunabilmesinin, kullanıcıların algıladığı adalet ve güven düzeyi üzerinde etkili olabileceği belirtilmektedir (Shneiderman, 2022). Özellikle işe alım süreçlerinde kullanılan açıklanabilir yapay zeka mekanizmalarının, adayların prosedürel adalet algısını güçlendirebileceği değerlendirilmektedir.

Aşağıdaki tablo, çalışmada önerilen varsayımsal ilişkileri ve olası araştırma alanlarını özetlemektedir.

Tablo 4. Varsayımsal İlişkiler ve Olası Araştırma Alanları

Araştırma Boyutu	Temel İlişki	Olası Düzenleyici Faktörler	Araştırma Odağı	Önerilen Örnek Yöntemler
Bireysel karar örüntüleri	Örtük tercihlerin işe alım kararlarına yansıması	Yapılandırılmış mülakat düzeyi	Demografik yoğunlaşma	Deneysel çalışma / IAT
Kurumsal veri yapıları	Tarihsel işe alım örüntülerinin veri tabanında birikmesi	Veri çeşitliliği düzeyi	Algoritmik öğrenme örüntüleri	Arşiv veri analizi
Algoritmik öğrenme süreçleri	Geçmiş tercihlerin dijital sistemlerde yeniden görünür hale gelmesi	Eğitim verisi hacmi ve temsil dengesi	Ayrımcı etki düzeyi	Simülasyon / XAI
Otomasyon yanlılığı	Algoritmik önerilerin eleştirel değerlendirme olmadan kabul edilmesi	İnsan denetimi düzeyi	Karar doğruluğu ve çeşitlilik	İnsan-AI deneyleri
Açıklanabilirlik ve güven	Karar gerekçelerinin şeffaf sunulması	Açıklanabilirlik düzeyi	Algılanan adalet ve örgütsel güven	Anket + deney

Kaynak: Bu çalışma kapsamında geliştirilen kavramsal model çerçevesinde ilgili literatürden yararlanılarak yazar tarafından oluşturulmuştur.

Bu varsayımlar, nöro-yanlılık yaklaşımının ampirik olarak sınanabilmesi için olası araştırma alanları önermektedir. Çalışmanın amacı kesin nedensel ilişkiler ortaya koymaktan çok, insan karar süreçleri, kurumsal veri yapıları ve algoritmik sistemler arasındaki etkileşimleri bütüncül biçimde tartışabilecek bir araştırma çerçevesi geliştirmektir.

5. HİBRİT NÖROORGANİZASYONEL KARAR MİMARİSİ (HNoKM)

Nöro-yanlılık yaklaşımı kapsamında tartışılan yeniden üretim döngüsü, algoritmik sistemlerde ortaya çıkan eşitsizliklerin teknik model sorunlarıyla sınırlı olmadığını bireysel karar örüntüleri, kurumsal veri yapıları ve dijital öğrenme süreçleri arasındaki etkileşimlerden beslendiğini göstermektedir. Bu durum, işe alım süreçlerinde sadece insan kararına ya da tek başına algoritmik sistemlere dayanan yaklaşımların belirli sınırlılıklar taşıyabileceğini düşündürmektedir. İnsan kararları bilişsel yanlılık, değerlendirme gürültüsü ve sezgisel sapmalardan etkilenebilirken, algoritmik sistemler de geçmiş veri örüntülerini yeniden üretme eğilimi gösterebilmektedir (Kahneman vd., 2021; Barocas & Selbst, 2016).

Çalışma kapsamında önerilen Hibrit Nöroorganizasyonel Karar Mimarisi (HNoKM), insan değerlendirmesi ile algoritmik sistemleri birbirinin alternatifi olarak konumlandırılmamakta karşılıklı denetim ve dengeleme mekanizmaları içinde çalışan bütünleşik bir karar yapısı olarak ele almaktadır. Modelin temel amacı, bireysel karar örüntülerinin doğrudan algoritmik yeniden üretime dönüşmesini sınırlandırırken aynı zamanda insan denetimini tamamen ortadan kaldırmayan bir yapı önermektir.

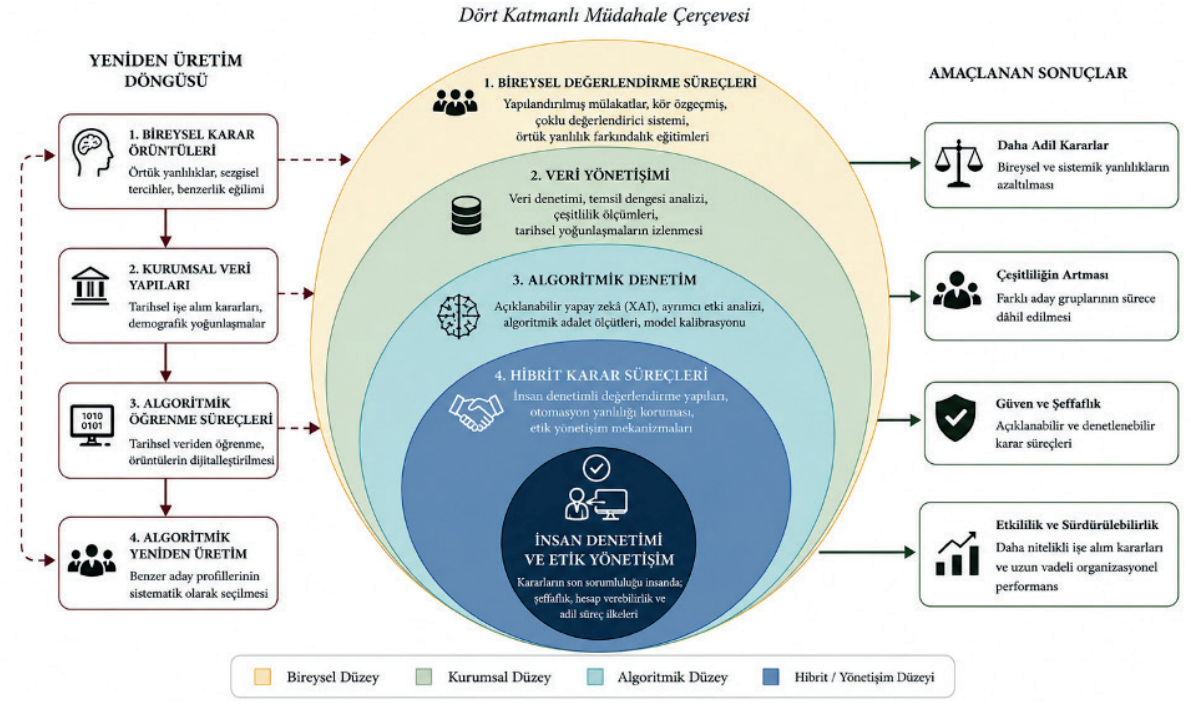
HNoKM yaklaşımı dört temel müdahale alanı üzerine yapılandırılmıştır. İlk katman, bireysel değerlendirme süreçlerine odaklanmaktadır. Yapılandırılmış mülakat sistemleri, kör özgeçmiş uygulamaları, çoklu değerlendirici kullanımı ve örtük yanlılık farkındalık eğitimleri gibi uygulamalar, bireysel karar örüntülerinin kurumsal veri üretimine etkisini azaltmayı amaçlamaktadır. Levashina vd. (2014), yapılandırılmış mülakatların değerlendirme güvenilirliğini artırabildiğini göstermektedir. Bir diğer katman, kurumsal veri yönetişimi ile ilişkilidir. İşe alım süreçlerinde kullanılan veri setlerinin temsil dengesi, çeşitlilik düzeyi ve tarihsel yoğunlaşmaları düzenli biçimde analiz edilmediğinde, geçmiş tercih örüntülerinin başarı modeli olarak yeniden üretilebilme riski artabilmektedir. Bu nedenle veri denetimi, çeşitlilik analizi ve temsil dengesi kontrolleri modelin temel bileşenlerinden biri olarak

değerlendirilmektedir. Algoritmik değerlendirme süreçleri katmanında ise, açıklanabilir yapay zeka uygulamaları, ayrımcı etki analizleri, algoritmik adalet ölçütleri ve model kalibrasyonu gibi mekanizmalar, algoritmik sistemlerin karar süreçlerinin daha görünür hale gelmesini amaçlamaktadır. Özellikle açıklanabilirlik düzeyinin artmasının kullanıcı güveni ve prosedürel adalet algısı üzerinde etkili olabileceği belirtilmektedir (Shneiderman, 2022). Son katman ise insan denetimini koruyan hibrit karar süreçlerini içermektedir. Otomasyon yanlılığı araştırmaları, bireylerin algoritmik önerileri çoğu zaman eleştirel değerlendirmeden kabul edebildiğini göstermektedir (Goddard vd., 2012). Model bu nedenle, algoritmik çıktıları nihai karar mekanizması olarak değerlendirmemekte, insan değerlendirmesini destekleyen karar araçları olarak konumlandırmaktadır. Amaç, insan yanlılığını tamamen algoritmaya devretmek yerine karşılıklı denetim mekanizmalarıyla sınırlandırılmış bir karar yapısı oluşturmaktır.

İlgili araştırmalarda önerilen insan-makine iş birliğine dayalı karar modelleri, genellikle algoritmik çıktının insan tarafından onaylanması ya da açıklanabilirlik mekanizmalarının eklenmesi üzerinde durmaktadır (Shneiderman, 2022). Bu yaklaşımlar, insan denetimini çoğunlukla kararın son aşamasında konumlandırmakta, insan kararlarının kurumsal veri yapılarını nasıl biçimlendirdiği sorusunu kapsam dışında bırakmaktadır. HNoKM ise insan denetimini sadece algoritmik çıktının onaylanması aşamasına eklememekte, bireysel karar örüntülerinin kurumsal veri üretimine dönüştüğü ilk aşamadan itibaren modelin tamamına yayarak konumlandırmaktadır. Bu yönüyle model, mevcut hibrit karar yaklaşımlarından, yanlılığın tek bir noktada değil, bireysel-kurumsal-algoritmik döngünün her aşamasında yeniden üretilebileceği varsayımıyla ayrışmaktadır.

HNoKM modelinin temel bileşenlerini aşağıdaki görsel özetlemektedir.

Şekil 3. Hibrit Nöroorganizasyonel Karar Mimarisi (HNoKM)

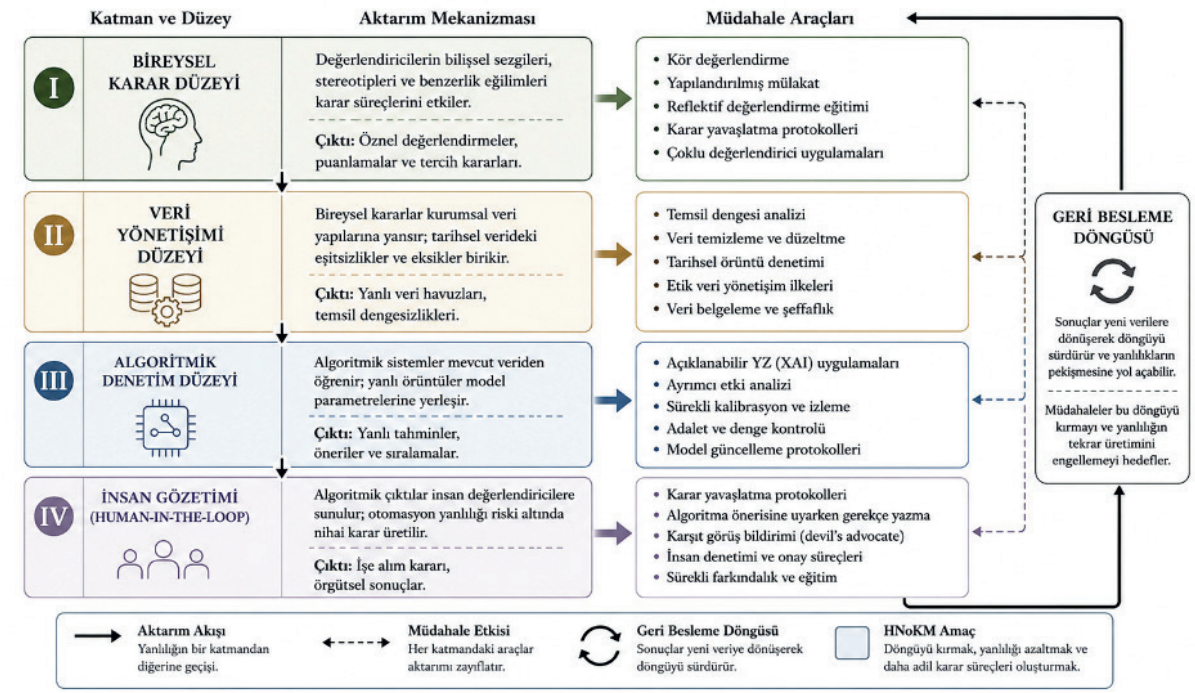


Kaynak: İlgili görsel modelin mimarisine ilişkin yazar tarafından yapay zeka uygulaması kullanılarak oluşturulmuştur.

Şekil 3, bireysel değerlendirme süreçlerinde ortaya çıkan yanlılık eğilimlerinin kurumsal veri yapıları ve algoritmik sistemler aracılığıyla nasıl süreklilik kazanabildiğini göstermektedir.

Karar yavaşlatma protokolleri, yapılandırılmış refleksiyon soruları, çoklu değerlendirici sistemleri ve yanlılık farkındalık uygulamaları, sezgisel değerlendirme süreçlerinin otomatikleşmesini sınırlandırabilecek araçlar arasında değerlendirilmektedir. Özellikle zaman baskısı altında verilen işe alım kararlarında reflektif değerlendirme alanının genişletilmesi, bilişsel yanlılıkların kurumsal veri yapılarına doğrudan yansımasını azaltabilir.

Şekil 4. HNoKM Müdahale Katmanları ve Nöro-Yanlılığın Yeniden Üretim Süreci



Kaynak: İlgili görsel modelin mimarisine ilişkin yazar tarafından yapay zeka uygulaması kullanılarak oluşturulmuştur.

Şekil 4, nöro-yanlılığın çok düzeyli yeniden üretim mekanizmasını ve HNoKM kapsamında önerilen müdahale katmanlarını özetlemektedir. Şekil, nöro-yanlılığın algoritmik sonuçlarla sınırlı olmadığını, bireysel değerlendirme süreçlerinden kurumsal veri yapıları ve insan denetimine kadar uzanan çok düzeyli bir yapı içinde ortaya çıkabildiğini göstermektedir. HNoKM modeli, bu sürecin farklı aşamalarında yapılandırılmış değerlendirme, veri yönetişimi, algoritmik denetim ve reflektif insan gözetimi mekanizmalarının birlikte kullanılmasını önermektedir.

Uygulama düzeyinde değerlendirildiğinde, HNoKM'nin dört katmanı farklı kurumsal aktörler ve düzenli aralıklarla işletilebilecek somut mekanizmalarla ilişkilendirilebilir. Bireysel değerlendirme katmanı, işe alım ekiplerinin yapılandırılmış mülakat protokollerini ve örtük yanlılık farkındalık eğitimlerini periyodik olarak güncellemesiyle sürdürülebilir. Kurumsal veri yönetişimi katmanı, İK analitik birimlerinin düzenli temsil denetimleri ve veri analitiği incelemeleri yürütmesini gerektirir. Algoritmik denetim katmanı, model geliştirme ekipleri ile İK biriminin birlikte yürüttüğü ayrımcı etki analizleri ve model kalibrasyon döngüleriyle işletilebilir. Son olarak insan denetimi katmanı, algoritmik önerilerin nihai karar noktasında bağımsız bir insan onayına tabi tutulmasını ve itiraz mekanizmalarının kurumsallaştırılmasını öngörür. Bu işleyiş biçimi, modelin tek seferlik bir denetim aracı olarak kullanılmaması, sürekli

izlenen ve güncellenen bir yönetişim döngüsü olarak konumlandırılmasını amaçlamaktadır.

HNoKM modeli bu çalışma kapsamında teorik bir çerçeve olarak önerilmektedir. Modelin farklı sektörlerde, kültürel bağlamlarda ve örgütsel yapılarda nasıl sonuçlar üreteceği ise gelecekte gerçekleştirilecek ampirik çalışmalarla değerlendirilebilir.

6. YÖNETİMSEL VE KURUMSAL YANSIMALAR

İnsan kaynakları yönetiminde işe alım süreçleri uzun yıllar büyük ölçüde insan değerlendirmesine dayalı biçimde yürütülmüştür. Son yıllarda algoritmik sistemlerin bu süreçlere giderek daha fazla dahil olması, kararların hızını ve standardizasyonunu artırırken, değerlendirme süreçlerinde hangi örüntülerin görünür hale geldiği sorusunu da gündeme taşımıştır. Özellikle geçmiş işe alım verileri üzerinde eğitilen sistemler, örgütsel çeşitlilik, adalet algısı ve karar şeffaflığı açısından da tartışılmaya başlanmıştır.

İşe alım süreçlerinde tam otomasyona dayalı yaklaşımlar kadar tamamen sezgisel insan kararlarına dayanan sistemlerin de belirli riskler taşıdığı görülmektedir. Kahneman'ın (2011) ikili süreç yaklaşımı ve Kahneman, Sibony ve Sunstein'ın (2021) değerlendirme gürültüsü çalışmaları, yapılandırılmamış karar süreçlerinde tutarsızlık ve bilişsel sapmaların artabildiğini göstermektedir. Diğer taraftan algoritmik sistemler de geçmiş kurumsal tercih örüntülerini veri aracılığıyla yeniden üretebilmekte ve bu örüntüleri daha geniş ölçekte görünür hale getirebilmektedir (Noble, 2018).

Kurumsal uygulamalar incelendiğinde, işe alım süreçlerinde ortaya çıkan problemlerin çoğu zaman nöro-yanlılık döngüsünün sadece belirli aşamalarına müdahale edilmesinden kaynaklandığı görülmektedir. Amazon'un kamuoyuna yansıyan işe alım algoritması örneğinde, tarihsel işe alım verilerindeki erkek ağırlıklı örüntülerin algoritmik sistem tarafından yeniden öğrenildiği ve kadın adayları dolaylı biçimde dışlayan sonuçlar ürettiği rapor edilmiştir (Dastin, 2018). Buna karşılık Unilever'in yapılandırılmış dijital değerlendirme süreçleri ve kör işe alım uygulamaları, özgeçmiş temelli benzerlik sinyallerini azaltmayı hedefleyen daha yapılandırılmış yaklaşımlara örnek oluşturmaktadır. Benzer biçimde LinkedIn'ın algoritmik adalet izleme uygulamaları ve Avrupa Birliği Yapay Zeka Yasası kapsamında geliştirilen insan gözetimi ve şeffaflık yükümlülükleri, sürekli denetim ve açıklanabilirlik ihtiyacının giderek daha merkezi hale geldiğini göstermektedir. HireVue'nun yüz analizi bileşenini bağımsız değerlendirmeler sonrasında devre dışı bırakması ise kullanılan ölçüm araçlarının geçerliliğinin de sürekli denetlenmesi gerektiğini ortaya koymuştur (Langer vd., 2022).

Tablo 5, nöro-yanlılık yaklaşımının farklı kurumsal uygulamalar bağlamında nasıl değerlendirilebileceğini ve önerilen kavramsal çerçevenin bu uygulamalara nasıl açıklama getirdiğini özetlemektedir.

Tablo 5. Nöro-Yanlılık Çerçevesinin Farklı Kurumsal Uygulamalarda Değerlendirilmesi

Kurumsal Uygulama	Temel Risk / Müdahale Alanı	Nöro-Yanlılık Çerçevesi Açısından Değerlendirme
Amazon işe alım algoritması	Tarihsel veri örüntülerinin algoritmik yeniden üretimi	Geçmiş işe alım tercihlerinin dijital ölçekte yeniden görünür hale gelmesi
Unilever yapılandırılmış dijital değerlendirme süreci	Yapılandırılmış ve çok aşamalı değerlendirme süreçleri, İlk izlenim ve benzerlik temelli karar örüntüleri	Karar süreçlerini yavaşlatarak reflektif değerlendirme alanını genişletme
HireVue yüz analizi uygulaması	Ölçüm geçerliliği ve algoritmik denetim problemi	Kullanılan göstergelerin bilimsel geçerliliğinin kritik önemi
LinkedIn algoritmik adalet izleme sistemi	Sürekli algoritmik denetim ve kalibrasyon	Adalet kaymasının izlenmesine yönelik kurumsal müdahale
Kör işe alım uygulamaları	Sosyal benzerlik sinyallerinin azaltılması	Demografik çağrışımların karar sürecine etkisini sınırlama
Avrupa Birliği Yapay Zeka Yasası	İnsan gözetimi ve şeffaflık zorunluluğu	Hibrit yönetişim yaklaşımının düzenleyici düzeyde güçlenmesi
Reflektif değerlendirme ve karar yavaşlatma uygulamaları	Sezgisel ve hızlı karar örüntülerinin baskın hale gelmesi	Sistematik düşünme alanını genişleterek otomatik değerlendirme örüntülerinin etkisini azaltma

Kaynak: Amazon işe alım algoritmasına ilişkin literatür (Dastin, 2018), HireVue uygulamalarına ilişkin çalışmalar, LinkedIn'in algoritmik adalet uygulamaları, Avrupa Birliği Yapay Zeka Yasası (AI Act) ve ilgili literatür temel alınarak yazar tarafından oluşturulmuştur.

Bu gelişmeler, insan kaynakları yönetiminde veri yönetişimi ve algoritmik denetim kapasitesinin stratejik bir yetkinlik alanına dönüştüğünü düşündürmektedir. Özellikle temsil dengesi analizi, veri kökenselliği denetimi, açıklanabilir yapay zeka uygulamaları ve insan denetimli hibrit karar süreçleri, etik unsurların yanısıra örgütsel güveni ve prosedürel adalet algısını etkileyen yapılar olarak değerlendirilebilir. Colquitt'in (2001) örgütsel adalet modeli, karar süreçlerinin nasıl alındığının en az kararın sonucu kadar önemli olduğunu göstermektedir. İşe alım süreçlerinde açıklanabilirlik ve itiraz mekanizmalarının bulunması, adayların sisteme yönelik güven düzeyini artırabilmektedir.

Türkiye bağlamında değerlendirildiğinde ise yapay zeka destekli işe alım araçlarının kullanımının giderek yaygınlaştığı, ancak bu alandaki düzenleyici ve akademik tartışmaların henüz sınırlı olduğu görülmektedir. Özellikle hiyerarşik ilişkilerin ve referans ağlarının güçlü olduğu örgütsel yapılarda, sosyal benzerlik örüntülerinin işe alım süreçlerine etkisi daha görünür hale gelebilmektedir. Bu durum, nöro-yanlılık yaklaşımının farklı kültürel bağlamlarda nasıl işlediğinin incelenmesini önemli bir araştırma alanı haline getirmektedir. Ayrıca Türkiye'de yapay zeka destekli işe alım sistemlerine yönelik standartlaştırılmış etik denetim mekanizmalarının henüz sınırlı olması, veri yönetişimi ve açıklanabilirlik konularını daha kritik hale getirmektedir.

7. SONUÇ, SINIRLILIKLAR VE GELECEK ARAŞTIRMALAR

Bu çalışma, algoritmik yanlılığı teknik bir model problemi olarak ele alan yaklaşımların ötesine geçerek, insan karar örüntüleri, kurumsal veri yapıları ve algoritmik öğrenme süreçleri arasındaki etkileşimi açıklamaya yönelik bütüncül bir çerçeve önermektedir. Çalışma kapsamında geliştirilen nöro-yanlılık kavramsallaştırması, bireysel düzeyde ortaya çıkan bilişsel eğilimlerin kurumsal süreçler ve dijital sistemler aracılığıyla yeniden üretilebileceğini tartışmaya açmaktadır. Bu yönüyle çalışma, insan kaynakları yönetimi, örgütsel davranış ve algoritmik adalet literatürleri arasında çapraz düzey bir bağlantı kurmayı amaçlamaktadır.

Makalenin temel katkılarından biri, algoritmik işe alım sistemlerinde görülen bazı eşitsizliklerin sadece veri kalitesi ya da model tasarımıyla açıklanamayabileceğini göstermesidir. Özellikle geçmiş insan kararları üzerinde eğitilen sistemlerin, tarihsel örgütsel tercih örüntülerini yeniden görünür hale getirebilmesi, insan ve algoritma ilişkisini yeniden düşünmeyi gerekli kılmaktadır. Bu çerçevede önerilen Hibrit Nöroorganizasyonel Karar Mimarisi (HNoKM), tam otomasyon ile tamamen sezgisel insan kararları arasında alternatif bir yaklaşım sunmaktadır. Model, yapılandırılmış değerlendirme süreçleri, veri yönetişimi, algoritmik denetim ve insan gözetimini birlikte ele alarak karar süreçlerinde çok katmanlı bir denge mekanizması önermektedir.

Çalışmanın bir diğer katkısı, nöro-yanlılık tartışmasını örgütsel öğrenme, çeşitlilik ve uzun vadeli performans açısından da değerlendirmesidir. Homojen karar örüntülerinin sürekli yeniden üretilmesi, farklı bakış açılarının örgüte girişini sınırlandırabilir ve bu durum yenilik kapasitesi üzerinde dolaylı etkiler oluşturabilir. Çeşitlilik bu nedenle, sosyal sorumlulukla birlikte stratejik örgütsel kapasite bağlamında da ele alınmalıdır.

Aynı zamanda çalışma bazı önemli sınırlılıklar taşımaktadır. Öncelikle bu araştırma kavramsal niteliktedir ve önerilen ilişkiler ampirik olarak test edilmemiştir. Çalışmada kullanılan nörobilişsel açıklamalar doğrudan biyolojik belirlenimcilik iddiası taşımamakta, ilişkisellik düzeyinde tartışılmaktadır. Ayrıca kültürel bağlam, sektör farklılıkları ve örgütsel yapıların etkisi bu çalışma kapsamında ayrıntılı biçimde incelenmemiştir. Özellikle algoritmik sistemlerin farklı ülkelerdeki düzenleyici çerçeveler altında nasıl sonuçlar üreteceği önemli bir araştırma alanı olarak varlığını korumaktadır.

Gelecekte yapılacak araştırmaların, bireysel değerlendirme süreçleri, kurumsal veri örüntüleri ve algoritmik çıktıları aynı araştırma tasarımında bir araya getiren çok düzeyli çalışmalar geliştirmesi önem taşımaktadır. Boylamsal araştırmalar, belirli işe alım örüntülerinin zaman içinde nasıl güçlendiğini ya da zayıfladığını inceleyebilir. Ayrıca insan değerlendiricilerin algoritmik önerilerle nasıl etkileşime girdiğini analiz eden deneysel çalışmalar, otomasyon yanlılığının örgütsel karar süreçlerindeki etkisini daha görünür hale getirebilir. Açıklanabilir yapay zeka uygulamalarının aday güveni, prosedürel adalet algısı ve örgütsel meşruiyet üzerindeki etkileri de ileriki çalışmalar açısından önemli görünmektedir.

İK liderleri açısından değerlendirildiğinde çalışma, işe alım süreçlerinde yanlılığın tek başına bireysel değerlendiricilerin farkındalığıyla sınırlandırılamayacağını göstermektedir. Özellikle geçmiş işe alım verilerinin düzenli gözden geçirilmesi, yapılandırılmış değerlendirme süreçlerinin yaygınlaştırılması, çoklu değerlendirici uygulamalarının kullanılması ve algoritmik sistemlerin insan gözetimi altında değerlendirilmesi, karar süreçlerindeki tek yönlü örüntülerin güçlenmesini sınırlandırabilecek uygulamalar arasında değerlendirilebilir. Bunun yanında hızlı sezgisel kararların gerekçelendirilmesini destekleyen reflektif değerlendirme uygulamaları, işe alım süreçlerinde daha dengeli karar yapılarının oluşmasına katkı sunabilir. Bu tür uygulamalar, örgütsel çeşitliliğin korunması, aday deneyiminin güçlendirilmesi ve değerlendirme süreçlerinin daha tutarlı yürütülmesi açısından da işlevsel görülebilir.

Sonuç olarak bu çalışma, işe alım süreçlerinde ortaya çıkan yanlılıkların sadece insan ya da algoritma kaynaklı biçimde ele alınmasının yetersiz kalabileceğini ileri sürmektedir. İnsan kararları, kurumsal veri yapıları ve algoritmik sistemler arasındaki ilişkinin birlikte değerlendirilmesi, hem örgütsel adalet hem de sürdürülebilir karar süreçleri açısından giderek daha kritik hale gelmektedir. Nöro-yanlılık yaklaşımı ve HNoKM modeli, bu tartışmalar için kavramsal bir başlangıç noktası sunmayı amaçlamaktadır.

Finansal Destek: *Yazarlar bu makalenin araştırma, yazarlık ve/ veya yayınlanması için herhangi bir finansal destek almamıştır.*

Financial Support: *The authors received no financial support for the research, authorship, and/or publication of this article.*

Çıkar Çatışmaları: *Yazar(lar) bu makaleyle ilgili herhangi bir akademik, finansal, ticari veya kişisel çıkar çatışması olmadığını beyan eder. Bu çalışmanın hazırlanması sırasında sonuçları etkileyebilecek herhangi bir destek alınmamıştır.*

Conflicts of Interest: *The author(s) declare that there is no academic, financial, commercial, or personal conflict of interest related to this article. No support that could influence the results was received during the preparation of this study.*

Teşekkürler: *Yazarların beyan edecek herhangi bir teşekkürü yoktur.*

Acknowledgments: *The authors have no acknowledgments to declare.*

Yazar Katkıları: *Kavramsallaştırma: [SK], Metodoloji ve Yazılım: [SK], Doğrulama: [SK], Biçimsel Analiz: [SK], Araştırma: [SK], Kaynaklar ve Veri Derleme: [SK], Yazma – Orijinal Taslak Hazırlama, İnceleme ve Düzenleme: [SK]. Tüm yazarlar makalenin son halini okumuş ve kabul etmiştir.*

Author Contributions: *Conceptualization: [SK], Methodology ve Software: [SK], Validation: [SK], Formal Analysis: [SK], Investigation: [SK], Resources ve Data Curation: [SK], Writing – Original Draft Preparation, Review ve Editing: [SK], All authors have read and agreed to the final version of the manuscript.*

Etik Onay: *Bu çalışma insan katılımcıları, hayvanları veya herhangi bir kişisel veriyi içermemektedir. Bu nedenle etik kurul onayı gerekmemiştir.*

Ethical Approval: *This study does not involve human participants, animals, or any personal data. Therefore, ethical approval was not required.*

Bilgilendirilmiş Onam: *Bu çalışma insan katılımcılar, anket veya deneysel veri toplama süreçleri içermediğinden bilgilendirilmiş onam alınmasını gerektirmemektedir.*

Informed Consent: *Not applicable. This study does not involve human participants, surveys, or experimental data collection; therefore, informed consent was not required.*

Veri Erişilebilirliği Beyanı: *Bu çalışmanın bulgularını destekleyen veriler, talep üzerine ilgili yazardan temin edilebilir.*

Data Availability Statement: *The data that support the findings of this study are available from the corresponding author upon reasonable request.*

KAYNAKÇA

Bandara, R. J., Biswas, K., Akter, S., Shafique, S., & Rahman, M. (2025). Addressing algorithmic bias in AI-driven HRM systems: Implications for strategic HRM effectiveness. *Human Resource Management Journal, 35*(4), 1047-1063. https://doi.org/10.1111/1748-8583.12609

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671-732. https://doi.org/10.15779/Z38BG31

Becker, W. J., Cropanzano, R., & Sanfey, A. G. (2011). Organizational neuroscience: Taking organizational theory inside the neural black box. *Journal of Management, 37*(4), 933-961. https://doi.org/10.1177/0149206311398955

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386-400. https://doi.org/10.1037/0021-9010.86.3.386

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychological Science, 15*(12), 806-813. https://doi.org/10.1111/j.0956-7976.2004.00760.x

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/world/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-2018-10-10/

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147-160. https://doi.org/10.2307/2095101

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture* (2nd ed.). Sage Publications.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. Psychological Review, 102(1), 4-27. https://doi.org/10.1037/0033-295X.102.1.4

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1*(3), 333-342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics, 178*(4), 977-1007. https://doi.org/10.1007/s10551-022-05049-6

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research, 13*(3), 795-848. https://doi.org/10.1007/s40685-020-00134-w

Langer, M., König, C. J., & Papathanasiou, M. (2022). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment, 30*(1), 1-16. https://doi.org/10.1111/ijsa.12358

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*(1), 241-293. https://doi.org/10.1111/peps.12052

Lindebaum, D. (2014). Ideology in organizational cognitive neuroscience studies and other misleading claims. *Frontiers in Human Neuroscience, 7*, Article 834. https://doi.org/10.3389/fnhum.2013.00834

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1-35. https://doi.org/10.1145/3457607

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 Publishing Group.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48*(2), 175-187. https://doi.org/10.1016/j.neuron.2005.09.025

Poldrack, R. A. (2018). *The new mind readers: What neuroimaging can and cannot reveal about our thoughts*. Princeton University Press.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469-481. https://doi.org/10.1145/3351095.3372828

Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review, 77*(6), 999-1022. https://doi.org/10.1177/0003122412463213

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040-2068. https://doi.org/10.1037/apl0000994

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68). ACM. https://doi.org/10.1145/3287560.3287598

Senior, C., Lee, N., & Butler, M. J. R. (2011). Organizational cognitive neuroscience. *Organization Science, 22*(3), 804-815. https://doi.org/10.1287/orsc.1100.0532

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*(1), 99-118. https://doi.org/10.2307/1884852

Tabibnia, G., & Lieberman, M. D. (2007). Fairness and cooperation are rewarding: Evidence from social cognitive neuroscience. *Annals of the New York Academy of Sciences, 1118*(1), 90-101. https://doi.org/10.1196/annals.1412.001

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Brooks/Cole.

Thomas, O., & Reimann, O. (2023). The bias blind spot among HR employees in hiring decisions. *German Journal of Human Resource Management, 37*(1), 5-22. https://doi.org/10.1177/23970022221094523

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*(3), 470-477. https://doi.org/10.1162/jocn.2008.20040