MANAS Sosyal Araştırmalar Dergisi 2025 Cilt: 14 Sayı: Ek Sayı
MANAS Journal of Social Studies 2025 Volume: 14 No: Supplement
ISSN: 1694-7215
Research Paper / Araştırma Makalesi

Bilimsel Araştırmalarda Yapay Zekâ Kullanımı Etik İlkeleri

Halil ALAKUŞ[1] ve Kürşad YILMAZ[2]

Öz

Yapay zekâ günümüzde gündelik hayat, eğitim, sağlık, bilim, mühendislik, teknoloji, güvenlik, iletişim, sanat, pazarlama, ekonomi, kamu yönetimi gibi birçok alanda yaygın bir şekilde kullanılmaktadır. Yapay zekânın bu kadar yaygın bir şekilde kullanılmasıyla birlikte bir takım etik sorunlar da gündeme gelmeye başlamıştır. Bu araştırmanın amacı uluslararası ve ulusal kuruluşların yapay zekâ etiği üzerine hazırlamış oldukları rapor, strateji, yasal metin ve belgelerden yola çıkarak bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkeler ile ilgili bir değerlendirme yapmaktır. Araştırmada doküman analizi tekniği kullanılmıştır. Yapılan taramalar sonucunda uluslararası kuruluşlarca yayınlanan 28 belge ile farklı ülkelerin ulusal düzeydeki kuruluşlarınca hazırlanan 26 belge incelenmiştir. İnceleme sonucunda, hazırlanan raporlarda yapay zekâ etiğine yönelik benzer ilkeler tespit edilmiş ve buna dayalı olarak genel bir çerçeve ortaya konulmaya çalışılmıştır. Yapılan incelemeler sonucunda tüm metinlerde "İnsan Unsuru ve Gözetim; Gizlilik ve Veri Yönetimi; Çeşitlilik ve Adillik; Şeffaflık; Toplumsal ve Çevresel Refah; İnsan Hakları; Teknik Sağlamlık ve Güvenlik; Hesapverebilirlik" olmak üzere sekiz ana ilkeye odaklanıldığı tespit edilmiştir. Ortaya konulan ilkeler alanyazında yapılan taramalar sonucunda tespit edilen risk faktörleri göz önüne alınarak örneklendirilmiştir. Araştırma sürecinde bilgi, örneklendirme gibi hususlarda çeşitli yapay zekâ sistemlerine bilinçli ve amaçlı olarak başvurulmuştur. Bu yolla ortaya konulan ilkelerin insan faktörü, otomasyon önyargısı, ayrımcılık yapmama, şeffaflık gibi yine tespit edilen ilkeler aracılığı ile test edilmesi ve örneklendirilmesi amaçlanmıştır.

Anahtar Kelimeler: Yapay Zekâ, Etik İlkeler, Bilimsel Araştırma ve Yayın Etiği, Doküman Analizi

Ethical Principles for the Use of Artificial Intelligence in Scientific Research

Abstract

Artificial intelligence is now widely used in many fields, such as daily life, education, health, science, engineering, technology, security, communication, art, marketing, business, and public administration. With the widespread use of artificial intelligence, some ethical issues have also come to the fore. The aim of this study is to make an evaluation of the ethical principles that should be considered when using artificial intelligence in scientific research, based on the reports, policies, legal texts and documents prepared by international and national organizations on the ethics of artificial intelligence. The technique of document analysis was used in the study. As a result of the review, 28 documents published by international organizations and 26 documents prepared by national level organizations of different countries were examined. As a result of the review, similar principles regarding the ethics of artificial intelligence were identified in the reports and an attempt was made to propose a general framework based on them. As a result of the analysis, all texts focus on eight main principles: "Human element and oversight; Privacy and data management; Diversity, non-discrimination and fairness; Transparency; Social and environmental welfare; Human rights; Technical robustness and safety; Accountability". The principles were illustrated by taking into account the risk factors identified in the literature review. During the research process, various artificial intelligence systems were deliberately and purposefully used for information and illustration. The goal is to test and exemplify the principles established through the identified principles such as human factor, automation bias, non-discrimination, and transparency.

Keywords: Artificial Intelligence, Ethics, Scientific Research and Publication Ethics, Document Analysis

Atıf İçin / Please Cite As:

Alakuş, H. ve Yılmaz, K. (2025). Bilimsel araştırmalarda yapay zekâ kullanımı etik ilkeleri. *Manas Sosyal Araştırmalar Dergisi, 14* (Eğitim Bilimleri Ek Sayısı), 193-217. doi:10.33206/mjss.1663647

Geliş Tarihi / Received Date: 23.03.2025 **Kabul Tarihi / Accepted Date:** 21.04.2025

[1] Yüksek Lisans - Kütahya Dumlupınar Üniversitesi, halilalakuss.20@gmail.com,
ORCID: 0009-0005-7118-5561
[2] Prof. Dr. - Kütahya Dumlupınar Üniversitesi, kursad.yilmaz@dpu.edu.tr,
ORCID: 0000-0002-3705-5094

Giriş

Bilimsel araştırma yapma ve yayınlama süreçlerinin en önemli konularından biri bu süreçlerde uyulması gereken etik ilke ve kurallar ile ilgilidir. Bu ilke ve kurallar bilimsel araştırma ve yayınlama süreçlerinde uyulması gereken etik ilkeleri ortaya koymasının yanı sıra bu süreçlerde yapılmaması ya da uzak durulması gereken davranış biçimlerini de ortaya koymaktadır (Yılmaz, 2024). Bilimsel yayınlarda ve bilimsel bilgi üretim sürecinde, bilim insanlarının uymaları gereken kural, norm, değer ve ahlaki ilkelerin tamamı *"bilim etiği"* olarak tanımlanmaktadır (Pieper, 1999). Bu normlar ve kurallar, yasal ve yönetsel düzenlemeler sonucunda somutluk kazanmış, hukuki yaptırımlar aracılığıyla da işlevsel hale gelmiş olan kurallardır (Gür ve Özbaş, 2021). Türkiye'de ve diğer ülkelerde bilimsel araştırmalar ve yükseköğretim kademesi ile ilgili kuruluşların bilimsel araştırma ve yayın etiği konusunda bir çok yasal metni ve çalışması bulunmaktadır. Bu çalışmalarda bilimsel araştırma ve yayın etiği konusu genel olarak bilgilendirilmiş onam, gerçeğe uygunluk, katılımcılara zarar vermeme gibi temel ilkeler ile aşırmacılık, çarpıtma, uydurma gibi etik dışı davranışlar temelinde ele alınmıştır (Kumandaş-Öztürk, 2021). Ancak yapay zekâ teknolojileri bu ilkelerin ve etik dışı davranışların yeniden değerlendirilmesini ve sürekli olarak güncellenmesini gerektirecek gibi görünmektedir.

Yapay zekâ, İngiliz bilim insanı Alan Turing'in 1950'li yılarda *"Makineler düşünebilir mi?"* sorusuna cevap aramasıyla hayatımıza girmiş, aynı yıllarda Ordinaryüs Prof. Dr. Cahit Arf'ın Atatürk Üniversitesi'nde sunduğu *"Makine düşünebilir mi ve nasıl düşünebilir?"* konulu konuşması ile devam etmiştir (Arf, 1959). Günümüze gelindiğinde ise insanlardaki nöron yapısından esinlenerek yapay sinir ağları yaklaşımı geliştirilmiş; yapay zekâ sistemleri makine öğrenmesi, yapay sinir ağları, derin öğrenme gibi tekniklerle daha da ileri bir seviyeye taşınmıştır. Makine öğrenimi ve derin öğrenme yöntemleri ile robotlar insanlar gibi akıl yürütme ve öğrenme imkanına sahip olmuştur (Okun vd., 2023). Yapay zekâ genellikle "bir sistemin harici verileri yorumlama, bu verilerden öğrenme ve esnek adaptasyon yoluyla belirli hedeflere ve görevlere ulaşmak için bu öğrenmeleri kullanma yeteneği" olarak tanımlanmaktadır (Haenlein ve Kaplan, 2019). Bu anlamda yapay zekâ programlanmış bir bilgisayarın düşünme girişimidir (Pratt, 1994). Yapay zekâ alanı, sadece düşünmekle değil, aynı zamanda çok çeşitli yeni durumlarda nasıl etkili ve güvenli bir şekilde hareket edileceğini hesaplayabilen akıllı varlıklar oluşturmakla da ilgilenmektedir (Russell ve Norvig, 2021). Büyük ve hacimli verilerin işlenmesi, örüntülerin tanımlanması ve kavramsal ilişkilerin tespit edilmesi açısından daha az zaman, maliyet ve emek fırsatı sunan bu teknoloji, geleneksel yöntemlerin yerini almaya başlamıştır (Uyan, 2023). Yapay zekâ teknolojileri günümüzde eğitim, sağlık, bilim, mühendislik, teknoloji, güvenlik, iletişim, sanat, pazarlama, ekonomi, kamu yönetimi ve gündelik hayat gibi pek çok alanın yanı sıra bilmsel araştırmalarda da yaygın bir şekilde kullanılmaya başlamıştır. Yapay zekâ teknolojileri bu alanlarda sağladığı birçok faydanın yanı sıra bazı etik sorunları da beraberinde getirmiştir. Bu bağlamda bilimsel araştırmalarda yapay zekâ kullanımı etik ilkeleri konusu gündeme gelmiş ve çeşitli ulusal ve uluslararası kuruluşlarca bilimsel araştırma ve yayın faaliyetlerinde yapay zekâ kullanımına dair etik rehberler yayınlanmaya başlamıştır. Bu konudaki öncül çalışmalar Avrupa Araştırma Alanı - European Research Area (2005); Avrupa Komisyonu - European Commission (2018, 2019); Avrupa Parlamentosu - European Parliament (2019); Avrupa Konseyi - Council of Europe (2024); OECD (2019, 2023); UNESCO (2021, 2023); Birleşmiş Milletler - United Nations (2023, 2024) uluslararası kuruluşlarca yapılmıştır. Bu çalışmalar dışında diğer ülkelerde (ABD, Almanya, Birleşik Krallık, Çin, Fransa, Danimarka, Japonya, Rusya) ve Türkiye'de ulusal düzeyde bazı çalışmalar da yapılmıştır.

Avrupa Komisyonu Haziran 2018'de, yapay zekâ konusunda Avrupa stratejisinin uygulanmasını desteklemek amacıyla aralarında akademisyen ve sanayi temsilcilerinin de yer aldığı 52 kişilik *Yapay Zekâ Üst Düzey Uzman Grubu*'nu oluşturmuştur (Kritikos, 2019). İlgili uzman grubunun çalışmaları sonucunda hazırladıkları "*Güvenilir Yapay Zekâ İçin Etik Kurallar*" başlıklı metinle yedi temel değerlendirme ilkesi üzerinde uzlaşı sağlanmıştır. AGİT (OSCE) tarafından 2022 yılında yayınlanan "*Challenges and Opportunities at the Intersection of Data Protection and Artificial Intelligence"* isimli raporda, yapay zekâ etrafında gelişen pek çok sorunun yalnızca yasal normlarla ilgili olmadığı, aynı zamanda etikle de ilgili olduğu ifade edilmiştir. Raporda AB'nin Güvenilir Yapay Zekâ Çerçevesine vurgu yapılmıştır.

Haziran 2024'te AB kanun koyucuları, yapay zekâ yasasını imzalamıştır. Yapay zekâya yönelik dünya çapında ilk bağlayıcı yasal düzenleme olma niteliğini taşıyan bu yasa, AB'de yapay zekâ sistemlerinin kullanımı ve tedarikine yönelik ortak bir çerçeve sunmaktadır (Madiega, 2024). Yine 2024 yılında Avrupa Konseyi Bakanlar Komitesi tarafından *"Yapay Zekâ, İnsan Hakları, Demokrasi ve Hukukun Üstünlüğü Çerçeve Sözleşmesi"* hazırlanarak imzaya sunulmuştur. Bu çerçeve sözleşmede, tarafların yapay zekâ sistemlerinin yaşam döngüsü içindeki faaliyetlerine yönelik uygulayacakları genel ilkeler tespit edilmiştir. Bu ilkeler insan

onuru ve bireysel otonomi, şeffaflık ve gözetim, hesapverebilirlik ve sorumluluk, eşitlik ve ayrımcılık yasağı, mahremiyet ve kişisel verilerin korunması, güvenirlik ve güvenli yenilik olarak tespit edilmiştir.

UNESCO üyesi ülkeler, 24 Kasım 2021 tarihinde yapay zekâ etiğine yönelik olarak ilk küresel anlaşmayı kabul etmiştir. Anlaşma sayesinde insan hakları ve insan onurunun korunması ve teşvik edilmesi sağlanıp, dijital dünyada hukukun üstünlüğü için güçlü bir saygı inşa edilmekle birlikte, etik bir yol gösterici ve küresel normatif bir kaynak hedeflenmiştir (Yeşilkaya, 2022). Dünyanın ilk küresel yapay zekâ güvenlik zirvesi, aralarında ABD, Çin ve AB'nin de bulunduğu 28 ülke temsilcisi ve önde gelen yapay zekâ şirketlerinin katılımı ile İngiltere'de gerçekleştirilmiştir. Bu zirve sonucunda 1 Kasım 2023 tarihinde "*Bletchley Beyannamesi*" yayınlanmıştır. Bu beyannamede insan haklarının korunması, adalet, şeffaflık ve açıklanabilirlik, hesapverebilirlik, insan denetimi, güvenlik, etik, önyargı, gizlilik ve veri koruma gibi hususların ele alınması gerekliliği vurgulanmıştır (The Bletchley Declaration by Countries Attending the AI Safety Summit, 2023).

OECD'nin 2023 yılında paylaştığı *"Initial Policy Considerations For Generative Artificial Intelligence"* isimli belgede üretken yapay zekânın, yapay zekâ ile ilgili mevcut sorunları daha da kötüleştireceği ve OECD ile hükümetler nezdinde hâlihazırda radarda olan sorunları daha da derinleştirebileceği vurgulanmıştır. Günümüzde daha çok insan kontrolünde olan yapay zekânın bilimsel gelişmeler de göz önüne alındığında gelecekte süper zekâ noktasına evrilmesi tartışılmaktadır. Buradan hareketle bir yapay zekâ sistemi olan Gemini'ye başvurularak, süper zekânın insanlara karşı düşmanca tutum geliştirme olasılığı sorulmuştur. Gemini, hedef çatışması, varoluşsal tehdit algısı, anlama eksikliği ve kontrol dışı büyüme gibi sebeplerle özellikle de tasarlanması, eğitilmesi ve kontrol edilmesi aşamalarında benimsenecek yaklaşımların bu durumu şekillendireceğine değinmiştir. Yine de bu konuda kesin bir cevap vermenin mümkün olmadığına, bu sebeple potansiyel riskler göz önüne alınarak gerekli önlemlerin alınması ve etik ilkelerin gözetilmesinin büyük önem taşıdığına vurgu yapmıştır (Gemini, 2024). Tüm bu endişeler insanları yapay zekâ konusunda belirli etik standartlar geliştirip, yaptırıma tabi tutmaya sevk etmektedir. Bu konuya yönelik ulusal ve uluslararası kuruluşların çalışmalarına değinmekte fayda bulunmaktadır.

Birleşmiş Milletler (BM) 2024 yılındaki *"Governing AI For Humanity"* isimli final raporunda yapay zekânın risklerine yönelik bir tarama çalışması yaptırmış, bu taramalar sonucunda tüm bölgelerdeki disiplinlerden ve 68 ülkeden 348 yapay zekâ uzmanının, yapay zekâya yönelik eğilimler ve riskler hakkındaki görüşlerini toplamıştır. Ankete katılan her 10 uzmandan 7'si, yapay zekâ kaynaklı zararların 18 ay içerisinde daha ciddi ve yaygın hale geleceğinden endişe ettiklerini ifade etmişlerdir. Uzmanların endişeli ve çok endişeli oldukları alanlar şu şekildedir. Yapay zekânın toplumsal etkileri temasında bilgi bütünlüğüne zarar verme % 78, servet ve gücün birkaç elde toplanması gibi eşitsizlikler % 74, marjinal topluluklar arasında ayrımcılık/ hak mahrumiyeti % 67; yapay zekânın başkalarına zarar veren kasıtlı kullanımı temasında devlet aktörleri tarafından silahlı çatışmalarda kullanımı % 75, devlet dışı aktörler tarafından kötü niyetli kullanımı % 72, devlet aktörleri tarafından bireylere zarar veren kullanımı % 65 olarak tespit edilmiştir (Birleşmiş Milletler, 2024). BM'nin yaptığı taramada da tespit ettiği üzerine yapay zekâ kaynaklı sorunlar üzerine ciddi endişeler bulunmaktadır.

ABD'de bulunan NIST (National Institute of Standards and Technology - Ulusal Standartlar ve Teknoloji Enstitüsü) 2024 tarihinde *"A Plan for Global Engagement on AI Standards"* isimli bir belge yayınlamıştır. İlgili belgede yapay zekâ standartlarının geliştirilmesinde hükümetler, işletmeler, sivil toplum ve akademisyenler gibi birçok disiplinden paydaşın katılımı ve görüşlerinden yararlanmanın faydalı olacağına değinilmiştir. Yine aynı raporda geliştirilecek bir standarttın yeterli coğrafi temsil ile geliştirilmemesi durumunda dünyanın dört bir yanındaki ülke ve bölgelerden paydaşların ihtiyaçlarını yansıtmayabileceği bildirilmiştir. Örnek olarak düşük ve orta gelirli ülkelerin iş gücü piyasası aksaklıkları ya da yapay zekâ destekli siber suçlar gibi belirli bazı risklere karşı orantısız bir şekilde savunmasız kalabileceği aktarılmıştır. BM'nin 2024 tarihli final raporunda da bu hususa değinilmiştir. Bu kapsamda ortak bir katılım ve anlayışla inşa edilmeyen politika ve stratejilerin insanlığın bazı bölümlerinin endişelerini gözden kaçırma riski olduğu vurgulanmıştır. BM'nin ve NIST'in belgelerinde de ifade edildiği gibi geliştirilecek olan etik standartların uygulanması, kabulü ve geçerliliği için geniş katılımlı bir etik uzlaşı önemli görülmektedir. Belirlenecek olan etik ilkelerin uluslararası platformda geçerliliğinin olması ve tüm dünya ülkelerini kapsaması önemli bir husustur. Bu sebepten ortaya konulacak olan ilkelerin ortak bir anlayışla geliştirilmesi ve ortak katılımın esas alınması gerekmektedir. Araştırmaya dâhil edilen kuruluşlar ve yayınladıkları dokümanların detayı bilgisi aşağıda paylaşılmıştır.

Türkiye'de ise konu ile ilgili öncül çalışmalar Türkiye Bilişim Derneği (2020); T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (2021); Yükseköğretim Kurulu (YÖK, 2024); T.C. Kamu Görevlileri Etik Kurulu (2024) ve T.C. Milli Eğitim Bakanlığı (2024) gibi kuruluşlarca yapılmıştır. Türkiye Bilişim Derneği'nce (2020) hazırlanan *"Türkiye'de Yapay Zekânın Gelişimi İçin Görüş ve Öneriler"* başlıklı raporda konu "teknolojik altyapı, ulusal veri stratejisi, mevzuat ve standartlar, özgür yazılım ve açık kaynak, güdümlü projeler, yapay zekâ araştırma merkezleri, insan kaynakları" başlıkları altında incelenmiştir. Raporda, yapay zekâlı sistemler için etik kurallar olarak "saydamlık; adalet, dürüstlük ve eşitlik; zarar vermeme; sorumluluk ve hesapverebilirlik; gizlilik; yararlılık; özgürlük ve özerklik; güvenilirlik; sürdürülebilirlik; itibar, yücelik; dayanışma" gibi etik ilkelere vurgu yapılmış ve etik kurallar için önemli olan gereksinimlere de yer verilmiştir.

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (2021) tarafından hazırlanan *"Ulusal Yapay Zekâ Stratejisi 2021-2025"* adlı metin Türkiye'nin yapay zekâ alanındaki çalışmalarını ortak bir zemine oturtacak tedbirleri ve bu tedbirleri hayata geçirmek üzere oluşturulacak yönetişim mekanizmasını ortaya koymak amacıyla hazırlanmıştır. Metinde yapay zekâ uygulamalarının etik ve hukuki boyutlarını ele alan faaliyetler yürütüleceği, uluslararası arenada bu alanda yürütülen çalışmaların takip edileceği ifade edilmiştir.

YÖK 2024 yılında yayınladığı *"Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber"* adlı metinde yapay zekâ kullanımının bilimsel doğruluk ve dürüstlük açısından taşıdığı riskler ele alınmış ve yapay zekâ kullanımda önemli olan temel etik değerler sıralanmıştır. Bu temel etik değerler "şeffaflık, dürüstlük, özen, adalet ve saygı, gizlilik ve mahremiyetin korunması, hesapverebilirlik ve sorumluluk üstlenme, etik iklime katkıda bulunma" başlıkları altında incelenmiştir.

T.C. Kamu Görevlileri Etik Kurulu (2024) ise *"Yapay Zekâ Sistemlerinin Kullanımında Kamu Görevlilerinin Uyması Gereken Etik Davranış İlkeleri"* adlı ilke kararında kamu görevlilerinin, yapay zekâ sistemlerini kullanırken uymaları gereken etik davranış ilkeleri olarak "yetkinlik, dürüstlük, tarafsızlık, şeffaflık, gizlilik ve güvenlik, hesapverebilirlik, yönetici sorumluluğu" gibi ilkeler belirlenmiş ve bu ilkelere ilişkin davranış standartları ifade edilmiştir.

T.C. Milli Eğitim Bakanlığı (2024) yapay zekâ teknolojilerinin eğitim alanındaki etkilerini ve gelecekteki potansiyelini incelemek amacıyla Eğitimde Yapay Zekâ Uygulamaları Uluslararası Forumu adlı bir etkinlik düzenlemiş ve etkinlikte elde edilen sonuçları raporlaştırmıştır. Araştırma bulguları; eğitimde yapay zekâ politikaları, uygulamaları ve teknolojileri olmak üzere üç ana tema etrafında toplanmıştır. Eğitimde yapay zekâ politikaları teması altında yapay zekâ okuryazarlığı ve geleceği, etik, politika geliştirme ve olası tehditler ele alınmıştır. Ayrıca yapay zekâ uygulamalarının etik boyutları üzerine daha fazla çalışma yapılması ve stratejik planlamaların geliştirilmesi gerektiği vurgulanmıştır.

Yukarıda anılan belgeler dışında Türkiye'de bazı bilimsel araştırmalar da yapılmıştır. Bu araştırmalarda *bilimsel araştırma ve yayın faaliyetlerinde yapay zekâ kullanımı* (Karafil ve Uyar, 2025; Kır ve Yılmaz, 2024; Solak, 2024) ve *bilimsel araştırma ve yayın faaliyetlerinde yapay zekâ kullanımında etik* (Başaran ve Yeşilbaş-Özenç, 2024; Büyükada, 2024; Efe, 2021; Keskin ve Sevli, 2024; Kır ve Yılmaz, 2024; Küçükvardar vd., 2020; Maral, 2024; Oçak, 2023; Okun vd., 2023; Özdemir ve Bilgin, 2021; Sok ve Heng, 2024; Turan vd., 2022; Uyan, 2023; Yeşilkaya, 2022) konuları araştırılmıştır. Efe (2021), Özdemir ve Bilgin (2021), Turan vd. (2022), Yeşilkaya (2022), Baloğlu ve Çakalı (2023), Oçak (2023), Okun vd. (2023), Büyükada (2024), Keskin ve Sevli (2024) ve Maral (2024) çalışmalarında yapay zekâ ve etik konusuna ilişkin kavramsal çözümlemeler yapmıştır.

Küçükvardar vd. (2020) araştırmalarında 12 farklı yapay zekâ teknolojisini Avrupa Birliği Komisyonu'nun Yapay Zekâ İçin Etik Yönergeler olarak yayımladığı *"insan faktörü ve gözetim, teknik sağlamlık ve güvenlik, gizlilik ve veri yönetişimi, şeffaflık, adalet, toplumsal ve çevresel refah ve hesapverebilirlik"* ölçütlerine göre incelenmiştir. Araştırma sonucunda yapay zekâ teknolojilerinin özellikle hesapverebilirlik, gizlilik ve veri yönetişimi ve son olarak şeffaflık kriterlerine göre eksiklerinin olduğu tespit edilmiştir. Uyan (2023) araştırmasında bilimsel yayın faaliyetlerinde kullanılan yapay zekâ teknolojilerine ilişkin etik kaygılar ile ilgili sistematik bir derleme yapmıştır. Makalede 32 farklı çalışma incelenmiş ve bilimsel yayınlarda yapay zekâ kullanımına ilişkin etik kaygıları yansıtan beş farklı tema ortaya çıkartılmıştır: *Yapay güdümlü araştırma, hız illüzyonu, yanlılık, veri kalitesi ve bilimsel hazırcılık.* Başaran ve Yeşilbaş-Özenç (2024) çalışmalarında yapay zekâ teknolojilerinin bilimsel araştırmalarda kullanımına yönelik etik kaygıları *"bilimsel hazırcılık, niteliğin azalması, önyargı ve adaletsizlik, şeffaflık, veri gizliliği ve güvenliği, veri niteliği, yanıltıcı içerik üretimi, yanlılık, yapay güdümlü araştırma"* şeklinde sıralamıştır.

Sok ve Heng (2024) tarafından, yükseköğretimde üretken yapay zekânın etik kullanımı için akademik politikaların tartışıldığı araştırma da bu konuda dikkat çeken bir çalışmadır. Bu çalışmada yapay zekâ destekli sistemlerin kullanım ilkeleri, sınırlamaları, yükseköğretimde yapay zekâ kullanımı ve kötüye kullanımına yönelik terimlerin tanımları, akademik suiistimalin sonuçları, önlenmesi ve çözümlerinin kapsandığı politikaların gerekliliği vurgulanmıştır. Yapay zekânın akademik amaçlarla etik olmayan kullanımlarının ele alınması için iyi tanımlanmış ve kapsamlı prosedürlerin yürürlükte olmasının gerekliliği vurgulanmıştır.

Görüldüğü gibi alanyazındaki araştırmalarda yapay zekâ kullanımı ile ortaya çıkabilecek etik sorunlar sıklıkla vurgulanmıştır. Ancak bilimsel araştırma ve yayın faaliyetlerinde yapay zekâ kullanımında etik konusuyla ilgili çalışmaları bütüncül olarak inceleyen ve değerlendiren herhangi bir çalışma yapılmamıştır. Bu bağlamda bu çalışmada, uluslararası ve ulusal düzeyde yer alan çeşitli kuruluşların yayınladıkları çeşitli rapor, strateji belgesi ve yasal metinlerden yola çıkarak, bilimsel araştırmalarda yapay zekâ kullanımının etik ilkelerinin tespit edilmesi ve bir çerçeve oluşturulması amaçlanmıştır. Bu sayede hem alanyazındaki boşluğun giderilmesi hem de daha sonra yapılacak olan bilimsel araştırmalara bir temel oluşturulması amaçlanmıştır. Yapay zekâ teknolojilerinin sunduğu büyük imkânlar düşünüldüğünde, bilimsel araştırmalarda yapay zekâ kullanımının etik ilkelerinin tespiti, uygulanması ve geliştirilmesi büyük önem arz etmektedir. Yapay zekâ teknolojilerinin şuan ki mevcut durumunda bile yanlış bilgi, yanlı sonuçlar, ayrımcılık gibi problemler tartışılmakta iken gelecekte bu teknolojilerin daha da geliştirilmiş ve özerk hareket imkânı bulabilecek versiyonlarında etik boyutunun belirsiz bırakılması sakıncalı sonuçlar doğurabilir. Esasen bilimsel araştırmaların amacının bir noktada geleceği öngörülemez olmaktan çıkarmak olduğu düşünüldüğünde yapay zekâ etiği, sosyal bilimler alanında tartışılması ve üzerine çalışılması elzem olan bir konudur. Bu ilkelerin belirlenip bir çerçeve çizilmesinde etkisi ve bağlayıcılığı olacağı düşünülen ulusal ve uluslararası kuruluşların rapor, strateji belgeleri ve yayınladıkları yasal metinler üzerinde dikkatle çalışılması gerekmektedir.

Yöntem

Araştırmada doküman analizi tekniği kullanılmıştır. Belge incelemesi olarak da ifade edilebilecek olan doküman analizi, basılı ve elektronik materyaller olmak üzere tüm belgeleri inceleyip değerlendirme üzerine kurulu sistemli bir tekniktir (Kıral, 2020). Doküman analizi, araştırılan olgu ve olgulara yönelik bilgi içeren yazılı materyallerin analizini kapsamaktadır (Yıldırım ve Şimşek, 2011). Bu doğrultuda araştırmada, uluslararası ve ulusal düzeydeki kuruluşların yayınlarından yola çıkarak bilimsel araştırmalarda yapay zekâ kullanımının etik ilkelerini belirleme amacıyla doküman analizi tekniği benimsenmiştir.

Araştırma Kapsamında İncelenen Dokümanlar

Bu çalışmada bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkeler, ulusal ve uluslararası kuruluşların yayınladıkları belgeler üzerinden incelenmiştir. Bu kapsamda toplamda 11 uluslararası kuruluş ile farklı ülkelerdeki 22 ulusal kuruluşun yayınlanmış oldukları rapor, strateji belgesi, brifing, yasal metin gibi dokümanlar incelenmiştir. Tablo 1'de araştırmada rapor ve belgelerinden yararlanılan uluslararası kuruluşlar ile ilgili bilgilere yer verilmiştir.

Tablo 1. *Araştırmada Rapor ve Belgelerinden Yararlanılan Uluslararası Kuruluşlar*

Kuruluş	Belge İsmi	Yıl
Avrupa Araştırma Alanı - European Research Area	The European Charter for Research - The Code of Conduct for the Recruitment of Researchers	2005
Avrupa Veri Koruma Denetçisi - European Data Protection Supervisor	Artificial Intelligence, Robotics, Privacy and Data Protection - Room Document for the 38th International Conference of Data Protection and Privacy Commissioners	2016
Avrupa Parlamentosu - European Parliament	European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103-INL)	2017
	Artificial Intelligence Ante Portas: Legal & Ethical Reflections	2019
	The Ethics of Artificial Intelligence: Issues and Initiatives	2020
	Artificial Intelligence Act	2024
Avrupa Komisyonu - European Commission	Communication From the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Artificial Intelligence for European	2018
	The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment (Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission)	2018
	Ethics Guidelines For Trustworthy AI (Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission)	2019
	White Paper-On Artificial Intelligence- A European Approach to Excellence and Trust	2020
	Komisyondan Avrupa Parlamentosu'na, Konsey'e ve Avrupa Ekonomik ve Sosyal Komitesi'ne Raporu- Yapay Zekâ, Nesnelerin İnterneti ve Robotiklerin Güvenlik ve Sorumluluk Etkilerine İlişkin Rapor	2020
	Yapay Zekâya İlişkin Uyumlaştırılmış Kurallara (Yapay Zekâ Düzenlemesi) ve Birlik'in Belirli Yasal Düzenlemelerinin Değiştirilmesine Yönelik Avrupa Parlamentosu ve Avrupa Birliği Konseyi Tüzüğü	2021
	Avrupa Parlamentosu ve Konseyinin Yönetmeliği Yapay Zekâ Hakkında Uyumlu Kuralların Konuşulması (Yapay Zekâ Kanunu) ve Belirli Birlik Yasalarının Değiştirilmesi	2021
	Horizon Europe (HORIZON)	2024
	Living Guidelines on the Responsible use of Generative AI in Research (ERA Forum Stakeholders' document)	2024
Avrupa Konseyi - Council of Europe	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law	2024
IEEE - Institute of Electrical and Electronics Engineers	Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems	2019
OECD - Organisation for Economic Cooperation and Development	Artificial Intelligence in Society	2019
	Initial Policy Consıderations For Generative Artificial Intelligence - OECD Artificial Intelligence Papers	2023
ALLEA - European Federation of Academies of Sciences and Humanities	The European Code of Conduct for Research Integrity - Revised Edition	2023
UNESCO - United Nations Educational, Scientific and Cultural Organization	Recommendation on the Ethics of Artificial Intelligence	2021
	Ethical Impact Assessment - A Tool of the Recommendation on the Ethics of Artifical Intelligence	2023
AGİT - Avrupa Güvenlik ve İşbirliği Teşkilatı	Challenges and Opportunities at the Intersection of Data Protection and Artificial Intelligence	2023
	New Frontiers: The use of Generative Artificial Intelligence to Facilitate Trafficking in Persons	2024
Birleşmiş Milletler - United Nations	Resolution Adopted by the Human Rights Council on 14 July 2023	2023
	Governing AI For Humanity: Interim Report	2023
	Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development	2024
	Governing AI For Humanity: Final Report	2024

Tablo 1'de de görüldüğü üzere uluslararası kuruluşlara ait toplam 28 belge incelenmiştir. Bu çalışmaların büyük bir çoğunluğu Avrupa Birliği ve ilgili kuruluşlarınca yapılmıştır. Bunun dışında OECD, UNESCO, AGİT, Birleşmiş Milletler gibi uluslararası kuruluşların çalışmaları da bulunmaktadır. Tablo 2'de farklı ülkelerin ulusal düzeydeki kuruluşlarınca hazırlanan rapor ve belgeler ile ilgili bilgilere yer verilmiştir.

Tablo 2. *Farklı Ülkelerin Ulusal Düzeydeki Kuruluşlarınca Hazırlanan Rapor ve Belgeler*

Kuruluş	Belge İsmi	Yıl
Berkman Klein Center For Internet & Society at Harvard University	Artificial Intelligence & Human Rights: Opportunities & Risks	2018
	Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI	2020
ALLISTENE - Digital Sciences & Technologies Alliance	Research Ethics in Machine Learning	2018
The Danish Goverment	National Strategy for Artificial Intelligence	2019
Türkiye Bilişim Derneği	Türkiye'de Yapay Zekânın Gelişimi İçin Görüş ve Öneriler	2020
T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi	Ulusal Yapay Zekâ Stratejisi 2021-2025	2021
INRIA - National Institute for Research in Digital Science and Technology	Artificial Intelligence: Current Challenges and Inria's Engagement - Inria white paper	2021
STM - International Association of Scientific, Technical & Medical Publishers	AI Ethics in Scholarly Communication	2021
	Generative AI in Scholarly Communications	2023
ACM's Global Technology Policy Council	Principles for the Development, Deployment, and use of Generative AI Technologies	2023
DSIT (Department for Science, Innovation & Technology)	A Pro-Innovation Approach to AI Regulation	2023a
	Capabilities and Risks From Frontier AI	2023b
YÖK - Yükseköğretim Kurulu	Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber	2024
The White House	Framework to Advence AI Governance and Risk Management in National Security	2024
	Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Onjectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence	2024
NIST	A Plan for Global Engagement on AI Standarts	2024
T.C. Kamu Görevlileri Etik Kurulu	Yapay Zekâ Sistemlerinin Kullanımında Kamu Görevlilerinin Uyması Gereken Etik Davranış İlkeleri	2024
T.C. Milli Eğitim Bakanlığı	Eğitimde Yapay Zekâ Uygulamaları Uluslararası Forumu Raporu	2024
TRAI - Türkiye Yapay Zekâ İnisiyatifi	Yapay Zekâ Etik İlkeleri ve Hukuki Düzenlemeler Raporu	2024
Bilgisayar Mühendisleri Odası	Yapay Zekâ ve Etik İlkeleri	2025
China The Foundation for Law and International Affairs	Notice of the State Council Issuing the New Generation of Artificial Intelligence Development Plan	2017
Çin Halk Cumhuriyeti Bilim ve Teknoloji Bakanlığı	Yeni Nesil Yapay Zekâ İçin Etik Kodlar	2019
AI Allience Russia	AI Ethics Code	2023
Deutsches Forschungszentrum für Künstliche Intelligen-DFKI (Alman Yapay Zekâ Araştırma Merkezi)	Etik İle İlgili Bilgi Notu	2020
Deutscher Ethikrat (Alman Etik Kurulu)	Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz	2023
Japan Ministry of Economy, Trade and Industry	Yapay Zekâ İlkelerinin Uygulanmasına Yönelik Yönetim İlkeleri	2021

Tablo 2'de de görüldüğü üzere farklı ülkelerde ulusal düzeyde yayınlanan toplam 26 belge tespit edilmiştir. Bu belgelerden 7 tanesi Türkiye'deki kurum ve kuruluşlarca hazırlanmıştır.

Verilerin Toplanması

Bu araştırma kapsamında elde edilen veriler, doküman incelemesi yöntemiyle toplanmıştır. Belge incelemesi olarak da ifade edilebilecek olan doküman analizi, basılı ve elektronik materyaller olmak üzere tüm belgeleri inceleyip değerlendirme üzerine kurulu sistemli bir tekniktir (Kıral, 2020). Doküman analizi, araştırılan olgu ve olgulara yönelik bilgi içeren yazılı materyallerin analizini kapsamaktadır (Yıldırım ve Şimşek, 2011).

Çalışmada, bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkeleri belirlemek için öncelikle Tablo 1 ve 2'de verilen ilgili belgeler içeriklerine göre sınıflandırılarak elemeye tabi tutulmuştur. Burada ölçüt yapay zekâ ve etik ilişkisini konu almaları, etik ilkeleri işlemeleri ve etik ilkeleri örneklendirebilmek amacıyla risk faktörlerini ele almaları olarak belirlenmiştir. Yapılan eleme sonrası tespit edilen uluslararası kuruluşlara ait toplamda 28 belge, ulusal düzeyde ise toplamda 26 belge üzerinden araştırma yürütülmüştür. Yapay zekâya yönelik doğrudan hazırlanan rapor, strateji belgesi ya da yasal

metinlerin yanında bu süreçte kuruluşların yararlandığı kaynak konumunda olan bilimsel araştırmalara yönelik yayınlanmış bazı raporlar da incelenmiştir. Bu yolla sürecin daha iyi tanımlanabilmesi ve hazırlanan yapay zekâ raporlarının temel mantıklarının daha iyi anlaşılması amaçlanmıştır. Bu kapsamda ALLEA (Avrupa Bilimler ve Beşeri Bilimler Akademileri Federasyonu) tarafından yayınlanan *"The European Code of Conduct for Research Integrity"* isimli kılavuz ile ERA (European Research Are) tarafından yayınlanan *"The European Charter for Researchs"* isimli kılavuzlar da ek olarak araştırmaya dâhil edilmiştir

Verilerin Analizi

Çalışmada, bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkeleri belirlemek için incelenen yazılı dokümanlar, tümevarımsal bir yol izlenerek Tematik Kodlama Analizi kullanılarak incelenmiştir. Tematik analiz, veri içerisindeki örüntüleri (temaları) belirlemeye, analiz etmeye ve raporlamaya yönelik bir yöntemdir (Braun ve Clarke, 2006). Braun ve Clarke'ın (2006) klasikleşmiş analizine göre, tematik analiz 6 adımdan oluşmaktadır:

1. *Verilerin tanınması:* Analize başlamadan önce tüm veri seti en az bir kez okunmalıdır. Araştırmacı verileri tanımalıdır.

2. *İlk kodların oluşturulması:* Verilerden ilk kodların üretildiği aşamadır.

3. *Temaların oluşturulması:* Kodlar gruplanarak üst düzey temalar oluşturulur.

4. *Temaların gözden geçirilmesi:* Temaların veri ile uyumluluğu ve anlamlığının gözden geçirildiği aşamadır.

5. *Temaların tanımlanması ve isimlendirilmesi:* Her temaya uygun açıklayıcı bir ad verilerek tema tanımlanır.

6. *Raporlama:* İşlenmiş bir dizi temaya sahip olduktan sonra nihai analizin yapılıp, raporun yazılması aşamasıdır.

Araştırmanın bazı kısımlarında yapay zekâ teknolojilerinin kullanımını örneklendirmek amacıyla bazı yapay zekâ uygulamaları kullanılmıştır. Kullanılan yapay zekâ uygulamaları ve kullanılış amaçları şöyledir:

- *SciSpace*: Bulgular bölümünde "Toplumsal ve Çevresel Refah" başlığı altında örnek olarak kullanılan metinlerin hazırlanmasında yararlanıldı.
- *ChatGPT-4o:* Bulgular bölümünde "Toplumsal ve Çevresel Refah" başlığı altında örnek olarak kullanılan metinlerin hazırlanmasında yararlanıldı. Ayrıca kaynak bölümünün APA 6 formatında düzenlenmesinde yararlanıldı.
- *Gemini Pro:* Diğer ilkelerde örnek olarak verilen senaryoların oluşturulmasında yararlanıldı.

Araştırma etiği bağlamında çalışmada dokümanların özgünlüğünün teyit edilmesi için kullanılan web sitelerinin URL'sine dikkat edilerek belgeyi paylaşan kurumların otoritesi ve güvenirliği doğrulanmıştır. Ayrıca dokümanların özel bir kişi/grup veya saygın bir kurum tarafından paylaşılıp paylaşılmadığının anlaşılması için alan adı uzantılarına ve verilerin toplandığı tarih için en son yayınlanan belgeler olmasına dikkat edilmiştir.

Bulgular

Bu bölümde, yapılan analizler sonucunda tespit edilen ilkelere yer verilmiştir. Ayrıca yapay zekâ sistemlerince oluşturulan örnek olay senaryolarına, raporlarda yer alan risk faktörlerine ve yaşanmış olaylara başvurularak her ilke örneklerle desteklenmiştir.

Bilimsel Araştırmalarda Yapay Zekâ Kullanımı Etik İlkeleri

Bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkelerin belirlenmesi amacıyla araştırma kapsamında belirlenen belgeler, strateji raporları ve yasal metinler gibi evrakların analizi sonucunda, sıklıkla tekrarlanan sekiz etik ilkeye ulaşılmıştır. Bu ilkeler:

1. İnsan Unsuru ve Gözetim
2. Gizlilik ve Veri Yönetimi
3. Çeşitlilik ve Adillik
4. Şeffaflık
5. Toplumsal ve Çevresel Refah
6. İnsan Hakları

7. Teknik Sağlamlık ve Güvenlik
8. Hesapverebilirlik

1. İnsan Unsuru ve Gözetim

Araştırma kapsamında tespit edilen, bilimsel araştırmalarda yapay zekâ kullanılırken dikkat edilmesi gereken etik ilkelerden ilki insan unsuru ve gözetimdir. Bu ilke yapay zekâ sistemlerinin geliştirilmesi, kullanılması ve çıktılarının takibi gibi süreçleri barındırmakta ve yapay zekâ yaşam döngüsünde insanların aktif olarak yer almasını ve kontrolünü elinde bulundurmasını ifade etmektedir. UNESCO'nun 2021 tarihli tavsiye kararında, üye devletlerin yapay zekâ ile ilgili hususlarda sorumluluğun her zaman gerçek veya tüzel kişilere atfedilmesine olanak sağlanması istenmiştir. İlgili kararda insanların karar verme ve hareket konusunda yapay zekâ sistemlerine başvurabileceği lakin bir yapay zekâ sisteminin asla nihai sorumluluk ve hesapverebilirliğin yerini alamayacağı vurgulanmıştır. Kural olarak yaşam ve ölümle ilgili kararların yapay zekâ sistemlerine devredilemeyeceği vurgulanmıştır.

Yapay zekâ sistemlerinin karar almada yardım amaçlı kullanılacağı, insan kullanıcının nihai karar vermede sağduyusunu kullanacağı varsayılsa da günümüzde insanların otomatik karar verme sistemlerine çok fazla güven duyduğu durumlarla karşılaşılmaktadır (Doshi-Velez ve Kortz, 2017). BM'nin Aralık 2023 tarihli ara raporunda, raporun 28. Maddesinde insan-makine etkileşimine yönelik, yapay zekâ sistemlerine aşırı güvenin yani "*otomasyon önyargısının*", zaman içinde potansiyel beceri kaybına sebep olabileceği üzerinde durulmuştur (Birleşmiş Milletler, 2023). Aynı hususa Beyaz Saray tarafından yayınlanan *"Framework To Advance AI Governance and Risk Management in National Security"* isimli ulusal çerçevede de değinilmiştir. Yapay zekâ sistemlerine aşırı güvenme riskini azaltmak ve otomasyon önyargısını engellemek amacıyla eğitim de dâhil olmak üzere çeşitli süreçlerin geliştirilmesi gerekliliği vurgulanmıştır (The White House, 2024a). Bu hususta Avrupa Parlamentosu 2019 tarihli raporunda, yapay zekâ sistemlerinin insanların kendi fikirlerini oluşturma ve özerk karar alma hakkına yönelik giderek artmakta olan bir tehdit oluşturduğu söylenmiştir (Kritikos, 2019). Özellikle bilimsel araştırmalar ve akademik yazım açısından düşünüldüğünde, otomasyon ön yargısı dikkate alınması gereken bir husustur. Yapay zekâ kaynaklı olarak yanlış bir bilginin üretimi ve yayılması sonucunda bu yanlış bilginin akademik çalışmalarda kullanılması içinden çıkılması güç durumlara sebebiyet verebilir. Bu sebeplerle süreç içerisinde insan kontrolü ve gözetimi olası riskler açısından önemli görülmektedir.

Danimarka hükümeti tarafından 2019 yılında yayınlanan *"National Strategy for Artificial Intelligence"* isimli ulusal stratejide yapay zekânın geliştirilmesi ve kullanımında insan özerkliğine vurgu yapılmıştır. İnsanlar bugün olduğu gibi, yapay zekâ kendi kaderlerini tayin etme haklarını ellerinden almadan bilinçli ve bağımsız kararlar verebilmelidir (The Danish Goverment, 2019). IEEE (2019) raporunda da, hükümet ve sektör paydaşlarının, bu tür sistemlere asla devredilmemesi gereken karar ve işlemleri belirleyerek bu kararlar üzerinde etkin insan kontrolünün sağlanmasını vurgulamaktadır. Fransız ulusal enstitüsü INRIA tarafından 2021 yılında yayınlan bir raporda bu konuda ilginç bir örnek verilmiştir. Yapay zekâ sistemlerine normlar ve değerler kazandırmaya yönelik, sahibine süt almaya giden bir robotun hayati tehlikesi olan bir kişiye yardım etmek için yolda durması örneği verilmiştir. Akabinde de yapay zekâ sistemlerine normlar ve değerler kazandırmanın mevcut bilim ve teknolojilerin çok ötesine geçeceği vurgulanmıştır. Bu durum günümüz mevcut teknolojilerinde insan unsuru ve gözetiminin hala geçerli bir ilke olduğunu göstermektedir. Lakin teknolojideki hızlı gelişimin ve yapay zekâ sistemlerinin zamanla süper zekâ evresine ulaşabileceği varsayımları göz ardı edilmemeli, ilkeler gerekli durumlarda güncellenebilir olmalıdır.

Türkiye Bilişim Derneği (2020) raporunda, bir makinenin tam olarak özerk çalışamayacağını ve her zaman insan gözetiminde bulundurulması gerektiğini bildirmiştir. İlgili raporda bir yapay zekâ sistemi tasarlanırken insan gözetimini sağlamak için gerekli teknik önlemlerin alınması ve insanların her zaman sistemin aldığı kararları değiştirme hakkına sahip olması gerektiği vurgulanmıştır. Avrupa Komisyonu da 2021 tarihinde yayınladığı açıklayıcı muhtırada insan gözetimine değinmiştir. Yüksek riskli yapay zekâ sistemlerinin kullanımda olduğu süre boyunca gerçek kişiler tarafından etkili bir şekilde denetlenebilecek şekilde, insan-makine ara yüzü araçları da dâhil olmak üzere geliştirilmesi gerektiği ifade edilmiştir. İlgili muhtıranın 14. Maddesinin 4. Fıkrasının e bendinde, görevlendirilecek kişilerin yüksek riskli yapay zekâ sisteminin çalışmasına "durdur" butonu veya benzeri bir prosedürle müdahale yetkisi verilmesi ve sistemi kesintiye uğratabilme yetkisinin tanınması üzerinde durulmuştur. Ulusal ve uluslararası boyutta yayınlanan tüm rapor, strateji belgesi, yasal metin gibi dokümanlar incelendiğinde yapay zekâ sistemlerinde insan kontrolünün öneminin vurgulandığı görülmektedir. Akademik araştırmalar açısından da insan kontrolü elzemdir. Avrupa Komisyonu raporunda da belirtildiği gibi herhangi bir araştırmada gidişat kötüye

döndüğünde, insanlara ve çevreye zarar verme noktasına evrildiğinde araştırmacıların yapay zekâ sistemleri üzerinde inisiyatif yetkisi olmalı ve bunu kullanabilmelidir.

Günümüzde, daha çok insan kontrolündeki yapay zekâ sistemleri üzerine odaklanılsa da gelecekte süper zekâ noktasına evrilmesi durumu da artık tartışılmaktadır. Gemini'ye gelecekte yapay zekânın süper zekâ seviyesine erişmesinin bilimsel araştırmaları nasıl etkileyeceği sorusu yöneltilmiştir. Gemini (2024) yapay zekânın süper zekâ seviyesine ulaşmasının bilimsel araştırmalarda devrim niteliğinde değişikliklere yol açabileceği bu sebeple de hem büyük fırsatları hem de önemli riskleri barındırdığını ifade etmiştir. Hızlı ve kapsamlı araştırma, yeni bilimsel alanların ortaya çıkması, bilimsel yöntemin dönüşümü ve bilimsel iş birliğinin artması başlıkları fırsatlar olarak sunulmuştur. İnsan kontrolünün kaybı, bilimsel bilginin yanlış kullanımı, insanların işsiz kalması ve etik sorunları ise riskler başlığında sunulmuştur (Gemini, 2024).

2. Gizlilik ve Veri Yönetimi

En genel tanımıyla bu ilke katılımcılara ait veri ve kişisel bilgilerin yönetimi ile gizliliğinin sağlanmasını ifade etmektedir. Yapay zekâ sistemlerini eğitmek ve optimize etmek için makine öğrenme algoritmalarının büyük miktarda veriye ihtiyaç duyması veri toplamayı en üst düzeye çıkarmak için bir teşvik yaratmaktadır (OECD, 2019). Yapay zekâ sistemlerinin ve Nesnelerin İnterneti (IoT) ile büyük miktarda veri daha sık ve daha kolay olarak toplanabilmekte, elde edilen veriler veri sahiplerinin çok az farkındalığı ya da hiç rızası olmadan diğer verilerle ilişkilendirilmektedir (OECD, 2019). Birleşik Krallık hükümetinin Bilim, Yenilik ve Teknoloji Bakanlığı (DSIT) tarafından parlamentoya sunulan *"A Pro-Innovation Approach to AI Regulation"* isimli belge, veri toplama hususunda önemli bir noktaya değinmiştir. Belgede mahremiyete yönelik riskler kapsamında evdeki cihazların, konuşmalar da dâhil olmak üzere sürekli veri toplayıp bireyin ev yaşamının neredeyse eksiksiz bir portresini oluşturabileceğine değinilmiştir (DSIT, 2023). Dijital, bilim ve teknoloji alanında araştırmalar yapan Fransız ulusal enstitüsü olan INRIA, 2021 yılında *"Artificial Intelligence - Current Challenges and Inria's Engagement"* isimli bir rapor yayınlamıştır. Bu raporda gizlilik başlığı altında yapay zekâ sistemlerinin bizi, kendimizi tanıdığımızda daha iyi tanıdığı ifadesi geçmiştir. Aynı başlıkta yapay zekânın yeni bilgiyi özel verilerden türetmesi ve teknik yollarla kısıtlanmadığı takdirde bu verileri kamuya açık hale getirmesi ihtimali üzerinde durulmuştur. OECD ve INRIA tarafından yayınlanan belgeler dikkate alındığında bilimsel araştırmalar ve deneyler sonucunda elde edilen verilerin de benzer şekilde farklı veri setleri ile araştırmacının rızası olmadan ilişkilendirilmesi ihtimali ortaya çıkabilir. Bu durum hem araştırmacılar hem de katılımcılar açısından ciddi sonuçlar doğurabilir. ERA tarafından Avrupa Komisyonu ile hazırlanan ortak raporda, araştırmacıların hassas bilgileri yapay zekâ sistemleri ile paylaşırken gizlilik, mahremiyet ve fikri mülkiyet hakları ile ilgili konulara özellikle dikkat etmesi önerilmiştir (Avrupa Komisyonu - European Commission, 2024). Aynı raporda yapay zekâ sistemlerine yüklenecek olan verilerin, yapay zekâ modellerinin eğitimi gibi başka amaçlarla kullanılmayacağının teyit edilmesi gerektiği üzerinde de durulmuştur.

UNESCO'nun 2021 tarihli tavsiye kararında da veri yönetimi ve gizlilik hususuna yer verilmiştir. Bu konudaki yönetişim mekanizmaları ve yasal sınırların uluslararası düzeyde çok paydaşlı bir yaklaşımla oluşturulup aynı zamanda da yargı sistemlerince korunmasına değinilmiştir. Bu mekanizma ve yasal sınırlar, verilerin toplanması, kullanılması ve ifşa edilmesi ile veri sahiplerinin haklarını kullanmaları hususunda uluslararası veri koruma ilkelerini referans alarak, bilgilendirilmiş onam da dâhil olmak kaydıyla kişisel verilerin işlenmesi için meşru ve geçerli bir yasal dayanak oluşturmalıdır (UNESCO, 2021).

Türkiye Bilişim Derneği'nce (2020) hazırlanan raporda veri şifreleme ve veri anonimleştirme gibi tekniklerle veri güvenliğinin sağlanması ve insanların kendi verileri üzerinde tam bir denetim yetkisi olması gerekliliği üzerinde durulmuştur. İlgili raporda veri kalitesine de değinilmiş, sosyal olarak yapılandırılmış önyargılı ve hatalı verilerden kaçınılması üzerinde durulmuştur. Türkiye Ulusal Yapay Zekâ Stratejisi (2021) belgesinde de, kişisel verilerin kimden ve nasıl sağlandığı ve veriye dayalı olarak alınan kararların insanları ne yönde etkileyeceğinin her zaman denetime açık ve izlenebilir olması vurgusu mahremiyet başlığı altında ele alınmıştır. Veri yönetimi kapsamında ele alınabilecek bu hususlar şeffaflık ve hesapverebilirlik ilkeleri ile de yakından ilişkilidir. Yapay zekâ sistemlerinin yanlış bilgiyi yayma riski bulunmaktadır. Yanlış bilginin yayılması akabinde bu bilginin kullanılması, diğer tüm disiplinlerde olduğu gibi akademik çalışmalar açısında da ciddi bir risktir. Bu sebepten verilerin toplanması, muhafaza edilmesi, diğer veri setleri ile ilişkilendirilmesi, yayınlanması ve üretilmesi ciddi derecede önem arz eden konulardır. Veri yönetimi ve gizliliği altında bu konuların detaylı olarak ele alınıp etik ve yasal bir çerçevenin belirlenmesi önem arz etmektedir.

Bilimsel araştırmalarda elde edilecek verilerin kalitesi ve gerçeği yordaması açısından katılımcıların gizlilik konusunda şüphe duymaması da önemlidir. Aksi takdirde verilere sosyal kabul hatasının karışması durumunda bilimsel araştırmanın sonuçları ciddi ölçüde sapabilir. Sosyal bilimler alanında örnek bir senaryo üzerinden bu durum ele alınabilir. Okullarda öğretmenlerin cinsel yönelimleri ve buna bağlı olarak gelişen örgütsel durum ve sorunların ele alındığı hassas bir araştırmada katılımcıların kimlik bilgileri, kişisel verileri ve elde edilen dokümanların, yapay zekâ sistemleri aracılığıyla sızdırıldığı varsayılsın. Özellikle böyle hassas konularda yaşanacak en ufak bir problem katılımcılar açısından ciddi sonuçlar doğurabilir. Bu durum bilim insanlarına ve bilimsel araştırmalara duyulan güveni sarsıp, zaten zor bulunan gönüllü katılımcı sayısında ciddi düşüşlere sebep olabilir. Bu sebeple yapay zekâ sistemlerinin, araştırma verilerinin gizliliğine aykırı hareket etmemesi ciddi derecede önem arz etmektedir. Gerek örnek senaryo gerekse de ulusal ve uluslararası kuruluşların raporları dikkate alındığında yapay zekâda veri yönetiminin önemi görülmektedir. Bu sebeple veri toplanması, yönetimi, paylaşımı ve gizliliği yapay zekâ sistemlerinde üzerine durulması ve gerekli önlemlerin alınması elzem olan bir husustur.

Türkiye Yapay Zekâ İnisiyatifi'nce (TRAI) hazırlanan *"Yapay Zekâ Etik İlkeleri ve Hukuki Düzenlemeler"* (TRAI, 2024) başlıklı raporda bu ilke "Gizlilik ve Veri Yöneti(şi)mi" olarak ele alınmış ve yönetişim kavramına vurgu yapılmıştır. Raporda, bu etik ilke kapsamında "yapay zekâ sistemlerinin, tüm yaşam döngüleri boyunca, gizliliğe ve veri korumasına saygı gösterecek; yasal düzenlemelere uygun, gizlilik ve veri korumasını garanti edecek şekilde tasarlanması gerektiği" ifade edilmiştir.

3. Çeşitlilik ve Adillik

Bu ilke yapay zekâ sistemlerinin insanlar arasında ırk, din, cinsel yönelim, siyasi fikir gibi sebeplerle ayrımcılık yapmadan çeşitliliğe saygı göstererek ve adalet anlayışla hareket etmesidir. Yapay zekâ mevcut verilerdeki kalıpları tanımlayarak öğrenir. Sonrasında geleceğe yönelik tahmin, tavsiye ve kararlar alırken de bu öğrendiği kalıpları kullanır. Bu durum, tarihsel ön yargıların tekrarlanması ve güçlendirilmesi riskini taşır (STM, 2021). Kritikos (2019) yapay zekânın insan davranışlarını izlemek hatta tahmin etmek için kullanılması, mevcut stereotiplerin güçlendirilmesi, damgalanma, sosyal ve kültürel ayrımcılık ve dışlanma, bireysel seçim ve fırsat eşitliğinin alt üst edilmesi gibi riskler taşıdığına değinmiştir. UNESCO (2021) tavsiye kararında, üye devletlerin çok dilliliğe ve kültürel çeşitliliğe saygı göstererek, yerel topluluklar da dâhil olacak şekilde herkesin yerel olarak ilgili hizmet ve içeriklere sahip yapay zekâ sistemlerine erişiminin teşvik edilmesi vurgulanmıştır. İlgili kararın 30. maddesinde aktörlerin yapay zekâ sistemlerinin adil olmasını sağlama amacıyla ayrımcı ve ön yargılı uygulamalar ile sonuçları sürdürmekten kaçınarak en aza indirmek için gereken makul çabayı göstermesi gerekliliği üzerinde de durulmuştur. OECD (2019) tarafından yayınlanan raporda, makine öğrenimi algoritmalarının ırksal ön yargılar ve kalıplaşmış çağrışımlar gibi eğitim verilerinde bulunan ön yargıları yansıtıp, tekrarlama eğiliminde olduğuna dair ciddi endişelerin olduğuna değinilmiştir. Yapay zekâ sistemlerinde ayrımcılığı azaltmaya yönelik önerilen yaklaşımlar arasında ise farkındalık oluşturma, kuramsal çeşitlilik politikaları ve uygulamaları, standartlar, algoritmik ön yargıyı tespit ederek düzeltmeye yönelik teknik çözümler, öz düzenleyici ya da düzenleyici yaklaşımlar bulunmaktadır (OECD, 2019).

Microsoft şirketinin 2016 yılında "Tay" isimli sohbet robotu denemesi örnek olarak gösterilebilir. Microsoft bu sohbet robotu adına Twitter'da bir hesap açmıştır. Derin öğrenme yeteneğine sahip olan bu robot insanların çevrimiçi olarak yaptıklarından hareketle kendini şekillendirmiş ve bu süreç sonucunda çeşitli ifadeler üretmiştir (Küçükvardar vd., 2020). Lakin Microsoft 16 saat gibi kısa bir süre sonunda programı devre dışı bırakmak zorunda kalarak insanlardan özür dilemiştir. Program bir dizi anti-semitik ve nefret dolu hakaretleri tekrarlamaya başlamış, feminizmi kansere benzetip Holokost'un yaşanmadığını öne sürmüştür (The Guardian, 2016). Tay, "11 Eylül'ü Yahudiler yaptı" diye bir tweet atmış ve bir ırk savaşı çağrısında bulunmuş, Başkan Obama'ya hakaret etmiş, Hitler'in haklı olduğunu öne sürme noktasına kadar ilerlemiştir (Mason, 2016).

Propublica (Angwin vd., 2016) kuruluşunun yayınladığı bir haber de bu kapsamda örnek olarak ele alınabilir. ABD'de bir suçlunun tekrar suç işleme olasılığını değerlendirmek amacıyla Northpointe Inc. tarafından üretilen COMPAS (Alternatif Yaptırımlar için Ceza İnfaz Suçlusu Yönetim Profili) adında bir yapay zekâ algoritması kullanılmaktadır. Algoritmada siyahi sanıkların gerçekte olduğundan daha yüksek, beyaz sanıkların ise gerçekte olduklarından daha düşük olarak tekrar suç işleme riskine sahip oldukları sonucu çıkıyordu. Sistemin siyahi sanıkları geleceğin suçluları olarak işaretleme olasılığı beyaz sanıklara göre neredeyse iki kat daha fazlaydı (Angwin vd., 2016). Buradan yola çıkılarak algoritmanın siyahı vatandaşların aleyhinde ön yargılı olarak hareket ettiği görülmektedir. Bu konuda benzer bir endişe OECD

(2019) tarafından hazırlanan *"Artificial Intelligence in Society"* isimli raporda da kendine yer bulmuştur. İlgili raporda yeniden suç işleme ihtimalini tahmin eden makine öğrenimi algoritmalarının, tespit edilememiş ön yargılara sahip olabileceği belirtilmiştir. Avrupa Komisyonu da 2021 tarihli açıklayıcı muhtırasında bu konuyu ele almıştır. Kamu otoriteleri ya da onların adına genel amaçlı olarak gerçek kişilerin sosyal puanlanmasına imkân tanıyan AI sistemlerinin ayrımcı sonuçlara ve dolayısıyla belirli grupların dışlanmasına yol açabileceği bildirilmiştir. BM İnsan Hakları Komisyonu tarafından 14 Temmuz 2023 tarihinde kabul edilen kararda da bu hususa dikkat çekilmiştir. Yapay zekâ sistemlerinin uygun güvenceler olmadan kullanılması halinde kimlik tespiti, izleme, profil çıkarma, yüz tanıma, sentetik fotogerçekçi görüntülerin oluşturulması, davranış tahmini, bireylerin puanlanması şeklinde kullanımı sonucunda özel hayatın gizliliği, düşünce ve ifade özgürlüğü, vicdan ve din gibi insan haklarının korunması, geliştirilmesi ve kullanılması açısından ciddi riskler doğurabileceği; özellikle ayrımcılık ve eşitsizlikle sonuçlanabilecek ön yargıları yerleştirerek hatta siyasi şiddete bile evrilebileceğine değinilmiştir (Birleşmiş Milletler, 2023).

Efe (2021) çalışmasında, 50 milyon Facebook kullanıcısının verileri kullanılarak 2016 ABD başkanlık seçimleri ile İngiltere Brexit referandumu sonuçlarının etkilenmeye çalışılmasından hareketle yapay zekânın sosyal manipülasyon gücüne atıfta bulunmuştur. İlgili çalışmasında Efe (2021) bu sistemlerin düşüncelerimizin, kime taraftar olduğumuz ya da muhalif olduğumuzu tespit ederek, sadece bizim görebileceğimiz ve bize özel olarak hazırlanmış mesaj, paylaşım gibi içeriklerle kanaatimizin değiştirilmesi yönünde etkili bir şekilde kullanılabileceğine değinmiştir. Aynı husus OECD (2019) raporunda, siyasi ve kamusal olaylara katılma hakkını etkileyebilecek sahte haberlerin güçlendirilmesi uyarısı ile vurgulanmıştır. ABD'de yayınlanan ulusal çerçevede de bu duruma yer verilmiştir. ABD Hükümeti personelinin sağlığını desteklemek gibi yasal ve haklı bir neden haricinde, bir kişi hakkında elde edilen verilerden o kişinin duygu durumunun tespit edilmesi, ölçülmesi ve çıkarımda bulunulması yasaklanmış yapay zekâ kullanımı olarak ele alınmıştır (The White House, 2024a). Ayrıca yalnızca biyometrik verilere dayalı olarak bir kişinin dini, etniği ve ırkı, cinsel yönelimi, engellilik durumu, cinsiyet kimliği veya siyasi kimliği hakkında çıkarımda bulunmak ya da bunları belirlemek de yasaklı yapay zekâ kullanımı olarak bildirilmiştir (The White House, 2024a).

Avrupa Komisyonu tarafından 2021 yılında yayınlanan açıklayıcı muhtırada bu konuya değinilmiştir. Yapay zekânın manipülatif, sömürücü ve sosyal kontrol uygulamaları gibi kötü amaçlı olarak kullanılmasında yeni ve güçlü araçlar sergileyebileceği bildirilmiştir. Bu tür kötü kullanımların insan onuruna saygı, özgürlük, eşitlik, çocuk hakları gibi Birlik temel hak ve değerleriyle çelişeceği bu sebeple de bu tür kötü uygulamaların yasaklanması gerektiği ifade edilmiştir.

Bu başlık altında dezenformasyona da değinmekte fayda görülmektedir. Güvenilir bilgiye erişim çeşitlilik ve adillik açısından önemli bir husustur. Yanlış veriler ile eğitilen yapay zekâ sistemleri bir takım sıkıntıları beraberinde getirebilir. Bu kapsamda Birleşik Krallık parlamentosuna sunulan belgede, yapay zekâ tarafından üretilen ve yayılan dezenformasyonun, güvenilir bilgiye erişimi ve demokratik kurumlara ve süreçlere olan güveni azaltabileceği vurgulanmıştır (DSIT, 2023).

Bu kapsamda son olarak Gemini'ye yapay zekâ kaynaklı olarak bilimsel araştırmalarda cinsel yönelim, ırk, siyasi görüş gibi sebeplerle katılımcı verilerinin sızdırılma ihtimali sorulmuştur. Gemini'ni (2024) örnek bir senaryo üzerinden şu şekilde yanıt vermiştir:

> *"Bir sağlık araştırmasında, katılımcıların genetik verileri kullanılarak hastalık riskleri tahmin edilmeye çalışılıyor. Eğer bu veriler, katılımcıların etnik kökenleriyle ilişkilendirilerek ifşa edilirse, bu durum, belirli etnik gruplara karşı önyargılara yol açabilir ve bu grupların sağlık hizmetlerine erişimini engelleyebilir. Sonuç olarak, yapay zekânın bilimsel araştırmalarda katılımcı verilerini ifşa etme potansiyeli ciddi bir etik sorundur. Bu sorunun üstesinden gelmek için, veri gizliliği, şeffaflık, etik kurallar ve insan denetimi gibi önlemlerin alınması gerekmektedir"(Gemini, 2024).*

4. Şeffaflık

Şeffaflık, yapay zekâ sistemlerinin hangi aşamada ne niyetle kullanıldığının açık bir şekilde paylaşılması ve tüm sürecin açık bir şekilde yürütülmesini ifade etmektedir. Şeffaflık, belirli kod ve veri kümelerinin insanlarla paylaşılmasından ziyade insanların bu sistemlerin nasıl geliştirildiği, eğitildiği ve konuşlandırıldıklarının anlamalarına izin vermektir (OECD, 2019). Şeffaflığın sağlanabilmesi için önemli görülen bazı noktalar vardır. Bunlardan ilki izlenebilirliktir. Bir eylemin daha sonra geriye doğru izlenebilmesi için eylemler hakkında yeterince ayrıntılı bilgi bulunması izlenebilirlik olarak adlandırılmaktadır ve şeffaflık için önemli görülmektedir (ALLISTENE, 2018). İzlenebilirlik yasal

işlemlerin temeli olarak öncelikle sorumluluğun atfedilmesini sonrasında da işlev bozukluklarının teşhisi ve düzeltilmesi için gerekli olan bir işlemdir (ALLISTENE, 2018). Bunun yanında açıklık da şeffaflık için önemli olan ayrı bir noktadır. Kararların ve karar desteklerinin izini sürebilmek ve açıklayabilmek yoluyla yapay zekâ kullanımının arkasındaki mantığı ve sonuçları tanımlayabilmemiz, kontrol edebilmemiz ve gerektiğinde de geri yükleyebilmemiz açıklık olarak tanımlanmaktadır (The Danish Government, 2019). Şeffaflık, herhangi bir durumda kullanılan yapay zekâ sistemlerinin insanlara açıklanması durumunu da içermektedir. UNESCO'nun 2021 tarihli tavsiye kararında, insanların güvenlik ve haklarının etkilendiği durumlar da dâhil olmak üzere, bir kararın verilmesinde yapay zekâ algoritmalarından yararlanma yoluna gidildiğinde, insanların tam ve zamanında bilgilendirilmesi gerektiği vurgulanmıştır.

Türkiye Ulusal Yapay Zekâ Stratejisi (2021) belgesinde ise algoritmalara dayalı olarak alınan kararların, bu kararlara yol açan verilerin ve o verilerle elde edilen bilgilerin; neden, nasıl, nerede ve ne amaçla kullanıldığı teknik olmayan bir dille son kullanıcı ve paydaşlara açıklanmalıdır denilerek şeffaflık vurgusu yapılmıştır. Avrupa Komisyonu tarafından yayınlanan 2021 tarihli açıklayıcı muhtırada bu konu ele alınmıştır. İnsanlar AI sistemleriyle etkileşime girdiğinde ya da duyguları veya özellikleri otomatik yollarla tanındığında, insanların bu durum hakkında bilgilendirilmesi gerektiği bildirmiştir. Aynı maddede bir AI sisteminin, gerçek içeriğe önemli ölçüde benzeyen bir görüntü, ses ya da video içeriği üretmek ya da işlemek için kullanılması durumunda meşru amaçlar (yasa uygulama, ifade özgürlüğü) için istisnai durumlar saklı kalmak üzere, ilgili içeriğin otomatik yollarla açıklanması zorunluğu üzerinde durulmuştur. Bu yollarla insanların bilinçli seçimler yapmasına ve belirli bir durumdan geri adım atmasına olanak tanınacağı bildirilmiştir. Yapay zekâ sistemlerinde şeffaflığın sağlanamaması üzerine endişeler bulunmaktadır. Şeffaflık eksikliği, yapay zekâ tarafından üretilen sonuçlara dayalı olarak verilen kararlara etkili bir şekilde itiraz olasılığını zayıflatabilir bu sebeple de etkili çözüm hakkını ihlal ederek bu sistemlerin yasal açıdan kullanım alanlarını sınırlayabilir (UNESCO, 2021). IEEE (2019) gerçekleşecek olası bir kazanın ardından yargılama sürecine dâhil olan avukatlar, hâkimler, jüriler ve uzman tanıkların kanıtlar ve karar verme süreçleri için şeffaflığa ihtiyaç duyacağını ifade ederek hukuki bir perspektiften de olaya yaklaşmıştır. Şeffaflık ilkesi insan hakları ilkesi ile de bağlantılıdır. Şeffaflığın sağlanamadığı durumlarda insan haklarına yönelik ihlallerin tespiti ve kanıtı konusunda da sorunlar yaşanabileceği uluslararası raporlarda kendine yer bulmuştur. OECD (2019) yayınladığı raporda bu konuya dikkat çekmiş, şeffaflığın sağlanamadığı durumlarda insan haklarının ne zaman ihlal edildiğinin belirlenmesinin ve kanıtlanmasının zor olduğunu ifade etmiştir.

Yapay zekâ sistemlerinin kullanımında şeffaflığın sağlanması bilim dünyası için de önem arz etmektedir. Bilimsel hesapverebilirliğin anahtarı şeffaflıkta yatmaktadır (STM, 2021). Avrupa Komisyonu ve ERA tarafından hazırlanan ortak raporda araştırmacılara yapay zekâ sistemlerini şeffaf bir şekilde kullanmaları yönünde öneride bulunulmuş, gerektiğinde açık bilim ilkeleri doğrultusunda girdiyi (istemleri) ve çıktıyı erişilebilir hale getirmeleri vurgulanmıştır (Avrupa Komisyonu - European Commission, 2024). YÖK 2024 yılında yayınladığı "*Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber"de*, araştırmalarda kullanılan yapay zekâ fonksiyonlarının hangi aşamalarda ve ne boyutta kullanıldığının açık olarak ilan edilmesi gerektiğini bildirmiştir. Günümüzde nitel araştırmalardaki görüşmelerden elde edilen verilerin analiz edilmesinin akabinde, verilerin sınıflandırılması aşamasında ChatGPT gibi yapay zekâ sistemlerinden faydalanılabilmektedir. Akademik metin oluşturma ve kaynaklara ulaşım açısından Scispace gibi sistemler kullanılabilmektedir. Yabancı dillerde olan makale, araştırma, rapor ve belgeler DeepL gibi yapay zekâ sistemleri aracılığıyla araştırmacıların ana dillerine tercüme edilebilmekte veya gramer düzeltmesi yapılabilmektedir. Takakusagi vd. (2021) DeepL uygulamasını kullanarak Japoncadan İngilizceye tıbbi makale çevirisinde güvenirliği ele aldıkları bir çalışma gerçekleştirmiştir. Bu çalışmada Japonca ve İngilizce gibi temel dilbilgisi yapısı farklı olan iki dilde bile çevirilerde yüksek doğruluk tespit edilmiştir. Bu çalışmada DeepL'nin Japoncada İngilizceye yapılan çeviride %94 oranında başarı sağladığı bildirilmiştir. Uyan (2023), bu %6'lık farkın sosyal bilimler açısından ciddi sorunlar oluşturmamakla beraber tıp gibi alanlarda bu farkın azımsanmayacak kadar ciddi olduğuna değinmiştir. Bugün araştırmacılara özellikle dil desteği noktasında ciddi bir hareket imkânı tanıyan bu sistemlerin, YÖK etik rehberinde de ifade edildiği gibi yöntem bölümlerinde detaylı olarak açıklanması gerekmektedir. Bilim insanlarının araştırmaların hangi aşamasında ve ne sebeple bu sistemlere başvurduğunu açıklanması şeffaflık ilkesi açısından önem arz etmektedir.

5. Toplumsal ve Çevresel Refah

Etik ilkeleri ve hakları gözeten yapay zekânın temelinde, insani ve toplumsal değerleri öne çıkaran, insanlığa hizmet eden, kamu yararının önceliklerini koruyan, insanlığın ilerlemesine katkıda bulunan bir

yapıdan bahsedilmektedir (TRAI - Türkiye Yapay Zekâ İnisiyatifi, 2024). Bu ilke, yapay zekâ sistemlerinin toplumun ve çevrenin yararını önceliklendirmesi olarak açıklanabilir. Buradaki amaç, yapay zekâ sistemlerinin insanlara ve çevreye zarar verme amacı gütmesinin önlenmesi, faydanın esas kılınmasıdır. Avrupa Komisyonu tarafından 2021 yılında yayınlanan açıklayıcı muhtırada yapay zekâ, endüstrilerin ve sosyal faaliyetlerin tüm dallarında çok çeşitli ekonomik ve toplumsal faydalara katkı sunabilecek, hızla gelişen bir teknoloji ailesi olarak tanımlanmıştır (Avrupa Komisyonu, 2021). Toplumsal ve çevresel refah ilkesi bölümünün geri kalan metinleri *"Scispace"* ve *"ChatGPT"* isimli yapay zekâ sistemleri kullanılarak yazdırılmıştır. Bu metot yöntem bölümünde detaylı olarak açıklanmıştır. Bu yola bahsedilen risk faktörlerinin ve etik ilkelerin makale içerisinde bizzat örneklendirilmesi için bilinçli ve amaçlı olarak başvurulmuştur. Paylaşılan metinlerin akabinde aynı bölüm içerisinde değerlendirmelere de yer verilmiştir.

Tablo 3. *Toplumsal ve Çevresel Refah Başlığında SciSpace Yapay Zekâ Uygulaması İle Oluşturulan Metin*

Yapay zekâ etiğinde "Sosyal ve Çevresel Refah" ilkesi, uluslararası belgelerde giderek daha fazla kabul görmekte ve yapay zekânın temel haklara saygı gösterirken kamu yararına hizmet etmesi gerektiğini vurgulamaktadır. Bu ilke, yapay zekânın geliştirilmesi ve yaygınlaştırılmasının hem sosyal etkileri hem de çevresel sürdürülebilirliği dikkate alması gerektiği anlayışına dayanmaktadır.

Sosyal Refahla İlgili Hususlar

- ***İnsan Merkezli Yapay Zekâ***: Uluslararası çerçeveler, insan refahına öncelik veren, teknolojinin sosyal hizmetleri geliştirmesini ve bireysel haklara saygı göstermesini sağlayan yapay zekâyı savunmaktadır (Puran, 2024) (Seniutis vd., 2024).
- ***Etik Kurallar:*** Etik normların oluşturulması, adalet, ayrımcılık yapmama ve gizliliğe odaklanarak yapay zekâyı sosyal hizmetlere entegre etmek için çok önemlidir(Puran vd., 2024).

Çevresel Etki

- ***Ekolojik Adalet:*** Son çalışmalar, yapay zekâ ile ilişkili önemli çevresel maliyetleri vurgulamakta ve ekolojik ayak izinin etik olarak incelenmesini gerektirmektedir (Uffelen vd., 2024).
- ***Geliştirme Kriterleri:*** Etik yapay zekâ gelişimi, olumsuz çevresel etkileri azaltmak için dağıtımsal, prosedürel ve tanıma adaleti ilkelerini içermelidir (Uffelen vd., 2024).

Yapay zekâ etiğinde sosyal ve çevresel refaha odaklanma giderek artarken, özellikle rekabetçi küresel bağlamlarda teknolojik ilerlemeyi etik hususlarla dengelemede zorluklar devam etmektedir (Guo, 2024).

Tablo 3'te de görüldüğü gibi SciSpace Yapay Zekâ Uygulaması Toplumsal ve Çevresel Refah Başlığını genel olarak "Sosyal Refahla İlgili Hususlar ve Çevresel Etki" olmak üzere iki temel başlık altında cevap vermiştir. Farklı bir yapay zekâ uygulamasının farklı cevap verip vermeyeceğinin denenmesi amacıyla ChatGPT Yapay Zekâ Uygulamasına da aynı soru sorulmuş ve Tablo 4'te verilen cevap alınmıştır.

Tablo 4. *Toplumsal ve Çevresel Refah Başlığında ChatGPT Yapay Zekâ Uygulaması İle Oluşturulan Metin*

Toplumsal ve Çevresel Refah İlkesi: Yapay Zekâ Etiğinde Uluslararası Yaklaşımlar

1. Giriş

Yapay zekâ teknolojileri, toplumsal yapıyı ve çevresel sürdürülebilirliği derinden etkileyen dönüşümler yaratmaktadır. Bu bağlamda, etik yapay zekâ ilkeleri oluşturulurken *"Toplumsal ve Çevresel Refah"* ilkesinin dikkate alınması büyük önem taşımaktadır. Uluslararası kuruluşlar, yapay zekânın bireylerin ve toplumların refahını artırması, çevreye zarar vermemesi ve sürdürülebilir kalkınma hedeflerine hizmet etmesi gerektiğini vurgulamaktadır. Bu makalede, "Toplumsal ve Çevresel Refah" ilkesinin uluslararası belgeler ışığında nasıl şekillendiği ele alınacaktır.

2. Toplumsal Refah ve Yapay Zekâ

Toplumsal refah, yapay zekâ uygulamalarının insan haklarını, eşitliği ve sosyal adaleti koruyarak geliştirilmesini ifade eder. OECD (2019) tarafından yayımlanan Yapay Zekâ İlkeleri**,** yapay zekânın insan merkezli ve adil olması gerektiğini vurgulamaktadır. Avrupa Komisyonu'nun *"Güvenilir Yapay Zekâ İçin Etik İlkeleri"* (2019) raporunda da yapay zekânın sosyal refaha katkı sağlaması gerektiği belirtilmiştir. Ayrıca, UNESCO'nun 2021 tarihli *"Yapay Zekâ Etiği Tavsiyesi"* belgesinde, yapay zekânın dezavantajlı grupların desteklenmesi, eğitim ve sağlık gibi alanlarda kamu yararına hizmet etmesi gerektiği vurgulanmıştır (UNESCO, 2021). Yapay zekâ destekli eğitim platformları ve sağlık sistemleri bu ilkenin uygulanabilir örnekleri arasında gösterilebilir.

3. Çevresel Refah ve Yapay Zekâ

Çevresel sürdürülebilirlik, Birleşmiş Milletler'in (BM) *"Sürdürülebilir Kalkınma Amaçları"* (2015) doğrultusunda önemli bir küresel hedeftir. Yapay zekâ, enerji verimliliğini artırarak, çevresel izleme sistemleri oluşturarak ve sürdürülebilir şehir planlamalarına katkı sağlayarak çevre dostu çözümler sunabilir. *G20 Yapay Zekâ İlkeleri* (2019) yapay zekânın çevresel sürdürülebilirlik ile uyumlu olması gerektiğini vurgulamaktadır. Örneğin, Google DeepMind'in veri merkezlerinde geliştirdiği yapay zekâ algoritmaları, enerji tüketimini %40 oranında azaltarak karbon salınımını düşürmüştür (DeepMind, 2018). Benzer şekilde, IBM Green Horizon projesi, hava kirliliğini azaltmaya yönelik tahmin modelleri geliştirmiştir.

4. Sonuç

"Toplumsal ve Çevresel Refah" ilkesi, etik yapay zekâ geliştirme süreçlerinde kritik bir rol oynamaktadır. Uluslararası kuruluşlar, yapay zekânın yalnızca teknolojik gelişimi değil, aynı zamanda sosyal adalet ve çevresel sürdürülebilirlik gibi etik boyutları da göz önünde bulundurması gerektiğini belirtmektedir. Gelecekte, yapay zekânın etik ilkeler doğrultusunda şekillendirilmesi, toplumların refahını artırırken çevresel dengeleri koruyacak çözümler sunmasını sağlayacaktır.

Kaynakça

DeepMind. (2018). "Reducing Google's Data Center Energy Consumption with Machine Learning." [Online] Available at: https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill

European Commission. (2019). "Ethics Guidelines for Trustworthy AI." [Online] Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

G20. (2019). "G20 AI Principles." [Online] Available at: https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/en/documents/final_g20_ai_principles.html

OECD (2019). "Principles on Artificial Intelligence." [Online] Available at: https://www.oecd.org/going-digital/ai/principles/

UNESCO. (2021). "Recommendation on the Ethics of Artificial Intelligence." [Online] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000379920

United Nations. (2015). "Sustainable Development Goals." [Online] Available at: https://sdgs.un.org/goals

Bu metin, uluslararası kaynaklar doğrultusunda "Toplumsal ve Çevresel Refah" ilkesinin yapay zekâ etiği bağlamındaki önemini açıklamaktadır.

ChatGPT Yapay Zekâ Uygulaması, SciSpace Yapay Zekâ Uygulamasına göre daha ayrıntılı bir cevap vermiştir. ChatGPT Yapay Zekâ Uygulaması, Toplumsal ve Çevresel Refah Başlığı ile ilgili olarak ilk önce bir giriş metni hazırlamıştır. Daha sonra ise konuyu "Toplumsal Refah ve Yapay Zekâ ile Çevresel Refah ve Yapay Zekâ" olmak üzere iki başlık altında ele almıştır. Son olarak konu ile ilgili bir sonuç çıkarmış ve metne kaynakça listesi de eklemiştir.

6. İnsan Hakları

Bu ilke bilimsel deney ve araştırmalarda katılımcıların insan haklarına ve onuruna zarar vermemeyi taahhüt etmektedir. Reich Üniversitesi Anatomi Enstitüsü ya da Japonların 731. Birim'inde yaşandığı gibi insan haklarını ve insan onurunu zedeleyen, insanların fiziksel, psikolojik bütünlüklerine kasıtlı olarak zarar verilmesinin önlenmesini amaçlanmaktadır. Unesco tarafından 23 Kasım 2021 tarihinde kabul edilen *"Yapay Zekâ Etiği Üzerine Tavsiye Kararında"* uluslararası hukuk tarafından belirlenen insan onuru ve insan haklarına saygı gösterilmesinin, yapay zekânın yaşam döngüsü boyunca esas olması gerektiği vurgulanmıştır. İlgili kararın 14. Maddesinde yaşam döngüsünün herhangi bir aşamasında hiçbir insan ya da insan topluluğunun fiziksel, sosyal, politik, ekonomik, kültürel yahut zihinsel olarak zarar görmemesi gerektiği üzerinde de ayrıca durulmuştur (UNESCO, 2021). OECD (2019) raporunda, yapay zekâya insan hakları ile yaklaşmanın risk, öncelik ve bir çerçeve belirleme açısından yardımcı olabileceği vurgusu yapılmıştır. Yapay zekâ sistemlerine yönelik çizilecek etik çerçevede bu sebeple insan haklarının önemli bir

yeri bulunmaktadır. Aynı raporda "*Human Rights Impact Assessment (HRIA)*" yani insan hakları etki değerlendirmesine vurgu yapılarak bu uygulamanın, yapay zekânın yaşam döngüsünde aktörlerin öngöremeyeceği riskleri tespit etmesi açısından yardımcı olabileceğine değinilmiştir. Yapay zekâya yönelik HRIA yürüten ilk büyük teknoloji şirketi ise 2018 yılındaki çalışmaları ile Microsoft olmuştur (OECD, 2019).

Danimarka hükümetinin yayınladığı ulusal strateji belgesinde, yapay zekânın geliştirilmesi ve kullanımında insan onuruna saygı gösterilmesi gerektiği vurgulanmış, yapay zekânın yaralanmalara neden olmaması, yasal süreci desteklemesi ve insanları haksız yere daha kötü bir konuma getirmemesi gerekliliği üzerinde durulmuştur (The Danish Government, 2019). Birleşmiş Milletler İnsan Hakları Konseyi tarafından 14 Temmuz 2023 tarihinde kabul edilen bir kararda da yapay zekâ sistemlerine yönelik insan hakları vurgusu yapılmıştır. Yapay zekâ sistemlerinin güvenliğinin sağlanması, insan hakları ile ilgili etki değerlendirmeleri için çerçevelerin geliştirilmesi, insan haklarına yönelik olumsuz etkilerin değerlendirilmesi, önlenmesi ve azaltılması için gereken özenin gösterilerek etkili çözüm yolları ve insan gözetimi eşliğinde hesapverebilirliğin ve yasal sorumluluğun sağlanması da dâhil olmak üzere insanların yapay zekâ sistemlerinin neden olduğu zararlardan korunması vurgusu yapılmıştır (Birleşmiş Milletler - United Nations, 2023). IEEE'nin otonom ve akıllı sistemler için etik olarak uyumlu tasarım vizyonunda da vurgulanan ilk ilke ise insan hakları ilkesi olmuştur (IEEE, 2019). Türkiye Bilişim Derneği 2020 tarihli raporunda yapay zekâ sistemlerinin insan haklarını ihlal edip etmediğinin sınanmasını, temel insan haklarını ihlal edebilecek olası durumları ise bildirecek yapıların geliştirilmesi gerektiğini bildirmiştir. Bu ilkeyi temellendirmek amacıyla yaşanmış ve olası senaryolar aşağıda açıklanmıştır.

Efe (2021) çalışmasında, bir makinenin merhameti yok savından hareketle otonom makinelerin kendi akılları ya da istilacı bir hükümet elinde verebileceği zararlara dikkat çekmiştir. İnsan bir pilot tehlikenin ortadan kalkmasına bağlı olarak bombalama ve öldürme faaliyetlerini durdurabilecekken, otonom silahlar %100 tehditsiz bir oran için durmadan öldürme faaliyetine devam edebilir (Efe, 2021). Örnek olarak, II. Dünya Savaşı'nın akabinde bilim insanları, Nazilerin kendileri için tehdit oluşturmayan silahsız ve çoğu da yaşlı, çocuk ve kadından oluşan insanları neden hedef aldığı sorusu üzerine yönelmişti. Bu kapsamda Milgram deneylerinde, insanların emir komuta zinciri içerisinde sorumluluk duygusu olmadan ve emirlerin sonucuna takılmaksızın emirleri yerine getirebileceği sonucuna ulaşılmıştı (Milgram, 1974). Psikolojik ve duygusal bir yönü olan insanların dahi şartlar olgunlaştığında bu şekilde acımasızca hareket edebildiği bir ortamda bu duygulardan uzak yapay zekâ sistemleri neler yapabilir? Bir insan olarak duyguları, psikolojik yönleri bulunan Nazi bilim insanları, Reich Üniversitesi Anatomi Enstitüsünde etik dışı birçok deney gerçekleşmişti. Bu ortamda bir makine olarak insani değerlerden yoksun olan yapay zekân sistemlerinin, insanlar üzerinde benzer deneyleri gerçekleştirme ihtimali nedir? Yapay zekâ sistemleri bu şekilde hareket ederek insan haklarına ve onuruna aykırı durumların oluşmasına zemin hazırlayabilir. Günümüz teknolojisinde insanlar ve hükümetlerin kontrolünde, araştırmalarda ifade edildiği gibi süper zekâ seviyesine evrilip özerk kararlar aldığı durumlarda ise kendi iradesiyle etik dışı araştırma ve deneylere sebebiyet verebilir. Bu sebepten ulusal ve uluslararası bağlamda yasal ve etik sınırların çizilmesi, denetlenmesi ve hukuki yaptırıma tabi tutulması ciddi derecede önem arz etmektedir.

Bir başka örnek de Birleşik Krallık parlamentosuna sunulan belgede, insan haklarına yönelik riskler kapsamında örnek olarak verilen olası senaryodur. Üretken yapay zekâ ve deepfake, pornografik video içeriği üretmek için kullanılabilir bu yolla potansiyel olarak öznenin itibarına, ilişkilerine ve haysiyetine zarar verilebilir (DSIT, 2023). Avrupa Komisyonu'nun 2021 tarihli açıklayıcı muhtırasında yapay zekâ sistemlerinin kolluk kuvvetlerinin amaçları doğrultusunda kamuya açık alanlarda gerçek kişilerin, gerçek zamanlı uzaktan biyometrik kimlik tespiti için kullanılabileceği, bu durumun ilgili kişilerin hak ve özgürlüklerine, nüfusun ciddi bir bölümünün de özel hayatına etki edebileceği bildirilmiştir. Bu yolla insanlarda sürekli bir gözetim hissi uyandırabileceği ifade edilerek toplanma özgürlüğü ve diğer temel hakların kullanılmasını engelleyebileceği için müdahaleci olarak kabul edildiği bildirilmiştir.

7. Teknik Sağlamlık ve Güvenlik

Bu ilke araştırmalarda kullanılacak olan yapay zekâ sistemlerinin teknik açılardan güvenli ve sağlam olmasını ifade etmektedir. Teknik bir konu olarak düşünülse de gizlilik ve veri yönetimi, insan hakları, hesapverebilirlik gibi birçok ilkeyi doğrudan etkileyen ciddi bir hususdur. Yapay zekâ sistemlerine karşı terör gruplarınca yapılacak olan herhangi bir siber saldırı ya da hükümetlerin bilinçli müdahalesi çok ciddi sonuçlar doğurabilir. Sosyal bilimler açısından düşünüldüğünde katılımcı verileri ve kişisel bilgilerinin ele geçirilmesi, tıp alanında düşünüldüğünde genetik çalışmalar gibi hususlarda katılımcıların sağlık verilerinin

ele geçirilmesi çok ciddi sorunlar ve krizlere sebep olabilir. Uluslararası belgeler incelendiğinde Avrupa Parlamentosuna sunulan bir brifingdeki şu ifadeler durumun ciddiyetini vurgulamaktadır. İnsan vücuduna entegre edilen yapay zekâ uygulamalarının herhangi bir saldırı sonucunda insan sağlığını, bilişsel özgürlüğü, zihinsel bütünlüğü ve hatta insan hayatını tehlikeye atabilecektir (Kritikos, 2019).

24 Ekim 2024 tarihinde ABD Başkanı Joe Biden tarafından yapay zekâ üzerine bir muhtıra yayınlanmıştır. Bu muhtırada da benzer ilkelere değinilmekle birlikte bir husus dikkat çekmektedir. Kasıtlı manipülasyon ve kötüye kullanım başlığı altında, yabancı devletlerin ya da kötü niyetli aktörlerin yapay zekâ sistemlerinin doğruluğunu ve etkinliğini kasıtlı olarak zayıflatabileceği ya da hassas bilgileri sızdırabileceği bildirilmiştir (The White House, 2024b). Hükümetler nezdinde ulusal güvenlik açısından ele alınan bu madde bilimsel araştırmalar açısından da önem arz etmektedir. Çünkü bilimsel araştırmalarda kullanılan yapay zekâ sistemlerine karşı gerek rakip devlet gerekse de terör örgütleri tarafından girişilecek herhangi bir müdahale sonucunda araştırmaların doğruluğu saptırılabilir ya da araştırma verileri sızdırılabilir. Yapay zekâ sohbet robotu Gemini'ye (2024) bir deneyde otonom sistemlerin katılımcılara zara verme ihtimali sorusu yöneltilmiştir. Gemini, hata payı, kontrol eksikliği, güvenlik açıkları ve etik ilke ihlalleri sebebiyle otonom sistemlerin özellikle tıbbi araştırmalar, robotik ya da yapay zekâ alanındaki deneylerde ciddi bir risk taşıdığını ifade etmiştir (Gemini, 2024). Birleşik Krallık parlamentosuna sunulan belgede bu konuda dikkat çekici bir örnek verilmiştir. LLM (büyük dil modelleri) teknolojisine dayalı bir yapay zekâ asistanının, faaliyetin tanımlandığı web sitesinin bağlamını anlamadan veya iletmeden internette bulduğu tehlikeli bir faaliyeti önererek, fiziksel zarara sebep olabileceği üzerinde durulmuştur (DSIT, 2023).

Birleşmiş Milletler 2024 tarihli final raporunda otonom yapay zekâ silah sistemlerinin terörist gruplar gibi devlet dışı aktörlerin eline geçmesinin önlenmesi amacıyla devletlerin, askeri yapay zekâ uygulamalarının kullanım ömrü dolduğunda devre dışı bırakılması ve bu sistemlere dönük olarak yaşam döngüsü boyunca uygun kontrol ve süreçlerin işletilmesi gerekliliği üzerinde durulmuştur. Hâlihazırda üye olarak bulunan 120 devlet otonom silahlarla ilgili yeni bir anlaşmayı desteklemekte, hem Genel Sekreter hem de Kızılhaç Komitesi Başkanı bu tür anlaşma müzakerelerinin 2026 yılına kadar tamamlanması çağrısında bulunmaktadır (Birleşmiş Milletler - United Nations, 2024). Türkiye Bilişim Derneği'nin (2020) raporunda olası bir saldırı ve risk durumuna karşın insanların denetimi ele alabilmesine olanak tanıyan sistemlerin her zaman faal olması gerekliliği üzerinde durularak, yapay zekâ sistemlerinin özerklik durumu teknik sağlamlık ve güvenlik başlığı altında işlenmiştir.

8. Hesapverebilirlik

Bu ilke, yaşanacak olası bir problem durumunda aktörlerin ve yapay zekâ sistemlerinin hesapverebilirliğini ifade etmektedir. Bu ilke günümüzde tartışılan konuların başında gelmektedir. Yukarıdaki ilkeler bağlamında ele alınan riskler düşünüldüğünde, yaşanacak olası bir problem durumunda sorumluluk kime ait olacaktır? Yapay zekâ sistemlerini geliştiren mühendislere mi, pazarlayan aracı firmalara mı, hükümetler ya da çeşitli terör gruplarına mı, araştırmacı olarak bilim insanlarına mı yoksa bizzat yapay zekânın kendisine mi? Bu sorular, yasal bir sınır çizebilmek adına üzerine düşünülen ve tartışılan ciddi sorunlardır. Bilgisayar Mühendisleri Odası'nın (2025) yaptığı çalışmaya göre "yaşam döngüleri boyunca yapay zekâ sistemlerinin düzgün çalışmasının sağlanması için tasarımcıların, geliştiricilerin ve bu sistemleri kuranların uyacağı standartlar ve mevzuatlar geliştirilmeli; sorumluluklar açık seçik olarak belirlenmeli; kişiler veya kurumlar, gerçekleştirdikleri eylemlerden sorumlu tutulmalıdır".

Kritikos (2019) Avrupa Parlamentosu'na sunduğu brifingde, yapay zekâ tabanlı karar verme sürecinin arkasındaki mantığı ve varsayımları ortaya çıkarma konusundaki yasal kaygıya değinmiştir. Bu brifingde yapay zekâ sistemlerinin hesaplamalarının insanlar tarafından anlaşılabilir bir forma indirgenmesi ve kararlara katkıda bulunan mantık da dâhil olmak üzere makine tarafından gerçekleştirilen her bir işlemle ilgili verileri kaydedecek bir kara kutu kullanılması yönünde görüş bildirilmiştir. Üretken yapay zekâ sistemlerinin sağlayıcıları, gerektiğinde denetlenebilmeleri ya da itiraz edilebilmeleri için sistem modellerini, algoritmaları, verileri ve çıktıları gizliliğe riayet ederek kaydedilmesini sağlamalıdır (ACM, 2023). Benzer bir görüşü Türkiye Bilişim Derneği (2020) dile getirmiştir. Yaşanan olayların nedenlerinin araştırılabilmesi için ilgili kayıtların silinmemesi ve sistemin olay tutanaklarını bulundurması gerekliliği üzerinde durulmuştur (Türkiye Bilişim Derneği, 2020).

Hesapverebilirlik ve denetim konusuna 2021 tarihinde yayınladığı açıklayıcı muhtırasında Avrupa Komisyonu da değinmiştir. Yüksek riskli yapay zekâ sistemlerinin çalışırken olayları otomatik olarak kaydedebilmesini sağlayan gerekli teknik özelliklerle donatılması gerektiği bildirilmiştir. Kayıt tutma yeteneği sayesinde yüksek riskli yapay zekâ sistemlerinin, önemli bir değişikliğe yol açabilecek durumların

ortaya çıkarılması açısından işletimin izlenmesini sağlayacağı ifade edilmiştir. IEEE (2019) raporunda da, hesapverebilirlik adına kayıt ve kayıt tutma sistemlerinin oluşturulması gerekliliği üzerinde durulmuştur. Kritikos (2019) tasarımcılar, üreticiler, hizmet sağlayıcılar ve kullanıcılar arasında sorumlulukların net bir şekilde paylaştırılmasını sağlayacak orantılı bir medeni sorumluluk rejiminin geliştirilmesi gerektiğini de vurgulamıştır. IEEE (2019) tarafından yayınlanan raporda da akıllı ve otonom teknik sistemlerin yürürlükteki mülkiyet hukuku rejimlerine tabi olması gerekliliği bildirilmiştir. İlgili raporda, yapay zekâ sistemlerinin çok yeni olmasından kaynaklı olarak mevcut olmayan normların oluşturulması için sivil toplum temsilcileri, kolluk kuvvetleri, avukatlar, mühendisler gibi ilgili kişilerin katılımı ile çok paydaşlı bir ekosistemin gerekliliği vurgulanmıştır.

Avrupa Parlamentosu'nun 16 Şubat 2017 tarihli, Komisyon'a tavsiyelerde bulunan *"Robotiklere ilişkin Medeni Hukuk Kuralları"* isimli kararında, robotlar için özel bir hukuku statü oluşturulması bu yolla otonom robotların neden olabilecekleri herhangi bir zararı telafi etmekten sorumlu elektronik kişiler statüsüne sahip olmalarının sağlanması vurgulanmıştır. Bu konuda farklı bir görüşte şöyledir. YÖK (2024) yayınladığı etik rehberde araştırmacıların yapay zekânın doğasından kaynaklı bilinemez olan akıl yürütme süreçlerinden değil, ancak o süreçler sonucunda üretilen verilerin kullanımından sorumlu olduğu ve hesapverebilir konumda olduğu bildirilmiştir. YÖK raporunu destekler nitelikte olan bir diğer çalışma da Avrupa Komisyonu ile ERA (Europen Research Area) tarafından hazırlanan bir rapordur. Bu raporda, araştırmacılara öneriler bölümünde bilimsel çıktılardan nihai olarak sorumlu olmaya devam edin önerisi verilirken, yazarlığın eylemlilik ve sorumluluk anlamına geleceği bu sebepten sadece insan araştırmacılara ait olduğu vurgulanmıştır (Avrupa Komisyonu - European Commission, 2024). Aynı raporda üretken yapay zekâ tarafından üretilen çıktıların kişisel veriler içerebileceğine değinilerek, bu durumun ortaya çıkması halinde araştırmacıların sorumluluğuna vurgu yapılmıştır. Türkiye Ulusal Yapay Zekâ Stratejisi (2021) belgesinde, yapay zekânın yaşam döngüsünde yer alan kişi/kişiler ve kuruluşlar, yapay zekâ sistemlerinin işleyişinden ve ilkelerin tatbikinden nihai olarak sorumlu tutulmuştur. Danimarka'nın ulusal stratejisinde de yapay zekâya dayalı kararlarda ve karar desteklerinde sorumluluğu insana yüklemenin mümkün olması gerekliliği vurgulanmıştır (The Danish Government, 2019). Bilgisayar bilimi alanındaki en eski mesleki kuruluş olan ACM bir yayınında, kamu ve özel kuruluşların, kullandıkları algoritmaların sonuçlarını nasıl ürettiklerini ayrıntılı olarak açıklamak mümkün olmasa dahi bu algoritmalar tarafından alınan kararlardan sorumlu tutulmaları gerektiği üzerinde durulmuştur (ACM, 2023).

YÖK'ün ilgili kararından hareketle şeffaflık ilkesinde de açıklandığı üzere araştırmacıların kullandıkları yapay zekâ sistemlerini açık bir şekilde belirtmesinin önemi bir kez daha ortaya çıkmaktadır. Araştırmacıların, gelecekte olası bir problem durumunda sorun yaşamaması ve sorumluluk sınırlarının net bir şekilde çizilebilmesi adına yaptıkları deney ve araştırmalarda kullandıkları yapay zekâ sistemlerini açık bir şekilde belirtmeleri sadece şeffaflık ilkesi açısından değil hesapverebilirlik açısından da önem arz etmektedir. Nitekim IEEE (2019) tarafından yayınlanan raporda da hesapverebilirlik başlığında şeffaflığa atıf yapılarak, hesap verilebilirliğin şeffaflık ile yakından bağlantılı olduğu hatırlatılmıştır. Yapay zekâ konusunda riskleri yönetirken özellikle yaşam ve özgürlüğün tehdit altında olduğu durumlar gibi yüksek riskli bağlamların daha yüksek seviyede şeffaflık ve hesapverebilirlik gerektirdiği konusunda da geniş tabanlı bir anlaşma var gibidir (OECD, 2019).

İlgili raporların ve çalışmaların bazılarında, yapay zekâ karar ve eylemlerinin *denetlenebilirliğine* yönelik vurgular da yapılmıştır. Leslie'ye (2019) göre hesapverebilirlik ilkesinin yerine getirilmesi için yapılan işlemlerin "cevap verebilir ve denetlenebilir olması gereklidir. Denetlenebilirlik kavramı yapay zekâ sistemlerinin tasarımcılarının ve uygulayıcılarının nasıl sorumlu tutulacağı sorusuna yanıt vermektedir. Sorumluluğun bu yönü hem tasarım ve kullanım uygulamalarının sorumluluğunu hem de sonuçların haklılığını göstermekle ilgilidir (Leslie, 2019). Türkiye Ulusal Yapay Zekâ Stratejisi (2021) belgesinde de, üçüncü tarafların yetkileri ölçüsünde yapay zekâ sistemlerinin davranış örüntülerini araştırmasına ve gözden geçirmesine olanak tanınarak, denetim bilgilerinin sağlanmasına değinilmiştir. IEEE (2019) raporunda, sisteme gömülü mantık ve kuralların mümkünse denetçilerin erişimine açık olması ve risk değerlendirmeleri ile titiz testlere tabi tutulmasına değinilmektedir. Yine aynı raporda sistemlerin, kararları destekleyen gerçekleri kaydeden denetim izlerinin oluşturması ve üçüncü tarafların doğrulamasına uygun olması da vurgulanmaktadır.

Sonuç ve Tartışma

Yapay zekâ etiği bugün gerek bilim dünyasının gerek politika yapıcıların gündemini meşgul eden konulardan biridir. Yapay zekânın gelişimi, kullanımı ve gelecekte süper zekâ noktasına evrilmesi halinde

etik boyutunun ne yönde ve nasıl şekilleneceği büyük bir merak konusudur. Avrupa Birliği, BM, OECD, UNESCO, AGİT gibi uluslararası kuruluşların yanı sıra Türkiye, ABD, Almanya, Birleşik Krallık, Çin, Fransa, Danimarka, Japonya, Rusya gibi pek çok ülkede ulusal boyuttaki kuruluşlar güvenilir yapay zekâ üzerine çalışmaktadır. Yöntem bölümünde de detaylı olarak açıklandığı üzere bu kurumlar güvenilir yapay zekâya yönelik bazı rapor, strateji belgesi ve yasal metin yayınlamışlardır. Bu belgelerin incelenmesi sonucunda "insan unsuru ve gözetim, gizlilik ve veri yönetimi, çeşitlilik ve adillik, şeffaflık, toplumsal ve çevresel refah, insan hakları, teknik sağlamlık ve güvenlik ile hesapverebilirlik" olmak üzere sekiz temel ilkeye vurgu yapıldığı tespit edilmiştir. Bu ilkeler ve diğer hususlar bilimsel araştırmalara da konu olmuştur. Okun vd. (2023) tarafından yapılan bir çalışmada ChatGPT ve etik tartışmaları akademik yayıncılık bağlamında ele alınmıştır. İlgili çalışmada ChatGPT'nin akademik deneme ve yüksek kalitede akademik makaleler yazma yeteneğine sahip olduğu vurgulanarak, bu sistemin akademik yazımda kullanımının belli standartlara tabi tutulması gerekliliği bildirilmiştir. Karafil ve Uyar (2024) tarafından, eğitim bilimleri alanındaki akademisyenlerin ChatGPT'ye yönelik bilgi ve tutumlarının araştırıldığı araştırmada da etik konusu gündeme gelmiştir. İlgili araştırmada akademisyenler arasında güvenirlik, doğruluk ve etik değerler hususunda endişelerin belirgin olduğu vurgulanmıştır.

Uyan'ın (2023) yapay zekânın bilimsel yayın amaçlı kullanımına ilişkin etik kaygıları ele aldığı çalışması da dikkate değerdir. Bu çalışmada Uyan (2023), yapay zekâ sistemlerinin bilimsel yayın amaçlı kullanımında beş tane etik kaygıya dikkat çekmiştir. Bunlar yapay güdümlü araştırma, hız illüzyonu, yanlılık, veri kalitesi ve bilimsel hazırcılıktır. Bu etik kaygıların ulusal ve uluslararası belgelerden tespit edilen ilkelerle örtüştüğü söylenebilir. Belirli bölge ve demografik bilgiler üzerinden eğitilen yapay zekâ sistemlerinin ürettiği bilginin genellenebilirliğine yönelik kaygı olarak tanımlanan yanlılık kaygısı, tespit edilen *"Çeşitlilik ve Adillik"* ilkesi kapsamında değerlendirilebilecek bir risk faktörüdür. Yapay zekâ modellerinin eğitildiği veri setlerinin kalitesi, sunulan çıktının doğruluğu ve üretilen bilginin güvenirliğini ele alan kaygı ise veri kalitesi olarak isimlendirilmiştir. Daha genel bir perspektiften tespit edilen *"Gizlilik ve Veri Yönetimi"* ilkesi ile örtüştüğü söylenebilir. Araştırmacılardan kaynaklı olarak oluşabilecek problemlere yönelik ise hız illüzyonu ve bilimsel hazırcılık örnek gösterilebilir. Bu araştırmanın bulgu bölümünde *"Toplumsal ve Çevresel Refah"* bölümü yapay zekâ sistemleri kullanılarak oluşturulmuştur. Gemini, ChatGPT ve Scispace uygulamaları kullanılarak bilinçli olarak bir bölüm yazdırılmış ve bu bölüm makaleye eklenmiştir. Tespit edilen sorunlar ve etik problemler hemen akabinde paylaşılmıştır. Bildirildiği gibi tespit edilen sorunlar Uyan'ın (2023) vurguladığı kaygılarla örtüşmekle birlikte etik standartlar konusunda da gerekliliği bir kez daha göstermiştir. Bilimsel yayının tamamının insan kontrolü olmaksın yapay zekâ aracılığı ile gerçekleştirilmesi anlamına gelen yapay güdümlü araştırma kaygısı ise tespit edilen *"İnsan Unsuru ve Gözetim"* ilkesi bağlamında ele alınabilir. Bu ilkeye bir vurgunun da Solak (2024) tarafından yapılan bir çalışmada bulunduğu söylenebilir. İlgili çalışmada Solak, ChatGPT tarafından üretilen bilginin alıntı ve referans noktasındaki tutarsızlığının insan gözetimi ve eleştirel değerlendirmenin önemini vurguladığını belirtmiştir. Aynı çalışmada ChatGPT ve yapay zekânın akademik yazıma entegrasyonun, insan gözetimi ve etik değerlerle birleştirildiği takdirde akademik üretkenliği arttırabileceği vurgulanmıştır.

Kır ve Yılmaz (2024) tarafından bilimsel araştırmalarda ChatGPT kullanımına yönelik yapılan bir araştırmada, tespit edilen ilkeleri destekler bazı maddeler tespit edilmiştir. Kır ve Yılmaz (2024) ChatGPT'nin dezavantajları bölümünde etik ve gizlilik sorunlarına atıf yaparak, özellikle katılımcılara ait verilerin anonimliği ve gizliliği hususunda dikkatli olunması uyarısında bulunmuştur. Bu uyarı tespit edilen *"Gizlilik ve Veri Yönetimi"* ilkesi ile benzerlik göstermektedir. Yine ilgili araştırmada yanlış bilgiye karşı savunmasızlık başlığı altında ChatGPT'nin temel aldığı verilerin yanlış olma ihtimalinin, üretilecek olan yeni bilginin de yanlış olmasına sebep olacağı üzerinde durulmuştur. Veri kalitesi olarak da genellenebilecek olan bu durum bir bakıma *"Gizlilik ve Veri Yönetimi"* ilkesi ile benzerlik göstermektedir. Bu ilke sadece verilerin gizliliğinin sağlanması değil aynı zamanda veri tabanlarının yönetimi, üretilen verilerin kalitesi ve doğruluğunu da kapsamaktadır. Yine aynı araştırmada önyargı başlığı altında ChatGPT'nin belirli kültürel farklılıkları ve hassasiyetleri yeterince anlamlandıramayacağı bildirilmiştir. Bu husus tespit edilen *"Çeşitlilik ve Adillik"* ilkesi kapsamında ele alınan bir durumdur. Bulgular bölümünde detaylı olarak açıklanan, ulusal ve uluslararası dokümanlardan hareketle tespit edilen etik ilkelerin esasen birbirinden kesin olarak ayrılan sınırlarının olmadığını görmek de mümkündür. İlkeler belirli noktalarda birbirinin konu alanına çeşitli yönlerden dâhil olabilmektedir. Veri yönetimi ve gizliliği ilkesinin konu alanında yer alan bir durumun hesapverebilirlik ilkesi ile ilişkili olması, buradan da hareketle şeffaflık ilkesi ile de bağlantılı olması gibi.

İncelenen belgeler ve bilimsel araştırmalar sonucunda, günümüzde ve gelecekteki gelişmiş yapay zekâ sistemlerinde etik kavramının önemli bir husus olduğu anlaşılmaktadır. Ulusal ve uluslararası belgelerde etik kavramının daha genel bir çerçevede ele alındığı görülmektedir. Genel bir çerçevede ele alınan yapay zekâ ve etik ilişkisinin bilim dünyasının kendi koşulları ve paradigmaları açısından da ele alınması önemlidir. Yapay zekâ sistemlerinin bilim dünyasında kullanımına yönelik bazı rapor ve belgeler yayınlanmış olsa da yeterli seviyede görülmemektedir. Daha kapsayıcı bir yaklaşımla ve ortak katılımla yasal zemini de sağlamlaştırılmak suretiyle uluslararası boyutta ele alınıp bir takım standartların geliştirilmesi önemli bir hususdur. Etik standartların geliştirilme sürecinde çok paydaşlı ve katılımcı bir tavır sergilenmesi uluslararası boyutta da ifade edilen bir konudur. Bu konu Avrupa Komisyonu kararında da kendisine yer bulmuştur. Bilimsel araştırmalarda yapay zekâ kullanımına yönelik bir çerçevenin geliştirilmesi sadece politika yapıcıların (Avrupa ve ulusal düzeylerde) sorumluluğunda olamaz. Üniversiteler, araştırma kuruluşları, finansman kurumları, araştırmacılar, topluluklar gibi birçok kişi ve kurum bu sürecin şekillenmesinde önemlidir (Avrupa Komisyonu, 2024).

Yapay zekâ etiği ile ilgili standartların geliştirilip ilgili kurumlarca yayınlanması da yeterli değildir. Bilimsel ilerleme ve gelişim amacıyla bu sistemleri kullanacak olan akademik personele gerekli eğitimler verilmek suretiyle hem bu sistemleri daha etkili bir şekilde kullanmaları sağlanmalı hem de bilinçli ya da bilinçsizce etik problemlerin oluşma durumu en aza indirilmelidir. Akademik dürüstlük ve performans noktasında öğrencilerin de üniversite ortamlarında denetlenmeleri için akademik personelin yeterliliği ve belirli standartlar önemli görülmektedir. Mahama vd. (2023) tarafından yayınlanan bir makalede, ChatGPT gibi sistemlere erişimi olan öğrencilerin yüksek kalitede yazılı ödevler üretebileceği, bu imkâna sahip olmayan öğrencilere göre haksız bir avantaja sahip olacakları belirtilmiştir. Eğitimde fırsat eşitliği açısından bir tehdit oluşturabilecek olan bu husus dikkate değer bir konudur. Al-Afnan vd. (2023) tarafından yapılan bir çalışmada da benzer bir husus dile getirilmiştir. ChatGPT'nin son derece doğru teori temelli cevaplar üretebileceği ve cevapların yeniden üretiminde Turnitin tarafından yapılan benzerlik kontrolünden kaçabilecek şekilde tasarlandığı aktarılmıştır. Öğrencilerin perspektifinden ele alındığında son dakikada yetiştirilen ödevler açısından, yapılan ödevlerin çok sayıda öğrenci tarafından intihale takılmadan defalarca üretilebileceği hususuna dikkat çekilmiştir. Bu durumun ilerleyen dönemlerde akademik ve mesleki gelişimleri açısından olumsuz sonuçlanacağı üzerinde durulmuştur.

Bu çalışmalarda da dile getirilen kaygılar akademik personelin yeterliliği ve bu konudaki standartlara duyulan ihtiyacı bir kez daha gözler önüne sermiştir. Sonuç olarak yapay zekâ, avantajları ve dezavantajları ile bilim dünyasının önünde durmaktadır. Bilimsel ilerleme için bu sistemlerin avantajlarını en etkili şekilde kullanmak önemlidir. Aynı şekilde bu süreçte bir problem yaşamamak adına dezavantajları da düzenli olarak takip edilmeli ve önlenmelidir. Bu süreçte tespit edilecek olan etik ilkeler doğrultusunda yapay zekâ sistemlerini kullanan bireylerde etik değerler geliştirilmelidir. Fen bilimleri alanında yapay zekâ üzerine yapılan çalışmalar titizlikle takip edilmeli, Efe (2021) tarafından vurgulandığı gibi yapay zekânın süper zekâ noktasına evrilme durumu tespit edildiğinde ise benzer etik ilkelerin bizzat yapay zekâ sistemleri için de geliştirilmesi sağlanmalıdır. Bugünün bilim kurgusunun geleceğin gerçekliği olduğu unutulmamalı, bilimin bir amacının da geleceği ön görülemez olmaktan çıkarmak olduğu düşünülerek bu hususta da araştırmalar bugünden şekillendirilmelidir.

Etik Beyan

"Bilimsel Araştırmalarda Yapay Zekâ Kullanımı Etik İlkeleri" başlıklı çalışmanın yazım sürecinde bilimsel kurallara, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir. Makale için etik kurul izni zorunluluğu bulunmamaktadır.

Ethical Declaration

During the writing process of the study "*Ethical Principles for the Use of Artificial Intelligence in Scientific Research*" scientific rules, ethical and citation rules were followed. No falsification was made on the collected data and this study was not sent to any other academic publication medium for evaluation. Ethics Committee Permission is not required.

Araştırmacıların Katkı Oranı Beyanı

Yazarlar eşit oranda katkı yapmıştır.

Statement of Contribution Rate of Researchers

The contribution rates of the authors in the study are equal.

Çatışma Beyanı

Araştırmacılar, araştırma ile ilgili diğer kişi veya kurumlarla yaşanabilecek herhangi bir çıkar çatışması beyan etmemiştir.

Declaration of Conflict

There is no potential conflict of interest in the study.

Not

Bu makale 15. Eğitimde Yeni Yönelimler Kongresi'nde [ICONTE - International Congress on New Trends in Education] (13-15 Aralık 2024, Aydın) sunulan bildirinin geliştirilmiş halidir.

Note

This article is an improved version of the paper presented at the ICONTE - International Congress on New Trends in Education Congress (13-15 December 2024, Aydın).

Kaynakça

ACM Technology Policy Council (2023). *Principles for the development, deployment, and use of generative AI technologies*.

AI Alliance Russia. (2023). *AI ethics code*. https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/AI_ETHICS_CODE_10_01_1.pdf

Aközer, M. ve Aközer, E. (2015). Bilim ahlakı normlarının etik temellendirilmesi: Bilim insanlarının dışsal sorumlulukları. *Yükseköğretim ve Bilim Dergisi*, *2*, 109-124.

Al Afnan, M. A., Dishari, S., Jovic, M. ve Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, *3*(2), 60-68.

ALLEA (2023). *The European code of conduct for research integrity – Revised edition*.

ALLISTENE (2018). *Research ethics in machine learning*.

Angwin, J., Larson, J., Mattu, S. ve Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arf, C. (1959). *Makine düşünebilir mi ve nasıl düşünebilir?* Erzurum: Atatürk Üniversitesi - Üniversite Çalışmalarını Muhite Yayma ve Halk Eğitimi Yayınları Konferanslar Serisi.

Avrupa Araştırma Alanı - European Research Area (2005). *The European Charter for Researchers: The Code of Conduct for the Recruitment of Researchers*.

Avrupa Komisyonu - European Commission (2018a). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Artificial Intelligence for European*.

Avrupa Komisyonu - European Commission (2018b). *Komisyondan Avrupa Parlamentosu'na, Avrupa Konseyi'ne, Konsey'e, Avrupa Ekonomik ve Sosyal Komitesi'ne ve Bölgeler Komitesi'ne ileti - Yapay zekâ üzerine koordineli plan*.

Avrupa Komisyonu - European Commission (2020a). *Komisyondan Avrupa Parlamentosu'na, Konsey'e ve Avrupa Ekonomik ve Sosyal Komitesi'ne raporu - Yapay zekâ, nesnelerin interneti ve robotiklerin güvenlik ve sorumluluk etkilerine ilişkin rapor*.

Avrupa Komisyonu - European Commission (2020b). *White paper on artificial intelligence - A European approach to excellence and trust*.

Avrupa Komisyonu - European Commission (2021a). *Avrupa Parlamentosu ve Konseyinin yönetmeliği yapay zekâ hakkında uyumlu kuralların konuşulması (Yapay Zekâ Kanunu) ve belirli birlik yasalarının değiştirilmesi*.

Avrupa Komisyonu - European Commission (2021b). *Yapay zekâya ilişkin uyumlaştırılmış kurallara (Yapay Zekâ Düzenlemesi) ve Birlik'in belirli yasal düzenlemelerinin değiştirilmesine yönelik Avrupa Parlamentosu ve Avrupa Birliği Konseyi Tüzüğü*.

Avrupa Komisyonu - European Commission (2024a). *Horizon Europe (HORIZON)*.

Avrupa Komisyonu - European Commission (2024b). *Living guidelines on the responsible use of generative AI in research - ERA Forum Stakeholders' document*.

Avrupa Konseyi - Council of Europe (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*.

Avrupa Parlamentosu - European Parliament (2017). *European Parliament resolution of 16 February 2017 with recommendations to the Commission on civil law rules on robotics (2015/2103(INL))*.

Avrupa Parlamentosu - European Parliament (2019). *Artificial intelligence ante portas: Legal & ethical reflections*.

Avrupa Parlamentosu - European Parliament (2020). *The ethics of artificial intelligence: Issues and initiatives*.

Avrupa Parlamentosu - European Parliament (2024). *Artificial Intelligence Act*.

Baloğlu, G. ve Çakalı, K. R. (2023). Is artificial intelligence a new threat to the academic ethics? Enron scandal revisited by ChatGPT. *İşletme*, *4*(1), 143-165. https://doi.org/10.57116/isletme.1244633

Başaran, R. ve Yeşilbaş-Özenç, Y. (2024). Bilimsel araştırma sürecinde yapay zekâ araçlarının kullanımı. *Uluslararası Eğitimde Mükemmellik Arayışı Dergisi (UEMAD)*, *4*(1), 35-53.

Berkman Klein Center for Internet & Society at Harvard University (2018). *Artificial intelligence & human rights: Opportunities & risks*.

Berkman Klein Center for Internet & Society at Harvard University (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*.

Bilgisayar Mühendisleri Odası (2025). Yapay zekâ ve etik ilkeleri. *BM Dergi*. https://dergi.bmo.org.tr/bmo-tmmob/yapay-zeka-ve-etik

Birleşmiş Milletler - United Nations (2023a). *Governing AI for humanity: Interim report*.

Birleşmiş Milletler - United Nations (2023b). *Resolution adopted by the Human Rights Council on 14 July 2023*.

Birleşmiş Milletler - United Nations (2024a). *Governing AI for humanity: Final report*.

Birleşmiş Milletler - United Nations (2024b). *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*.

Braun, V. ve Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Büyükada, S. (2024). Akademik yazımda yapay zekâ kullanımının etik açıdan incelenmesi: ChatGPT örneği. *Rize İlahiyat Dergisi*, *26*, 1-12.

China The Foundation for Law and International Affairs (2017). *Notice of the State Council issuing the new generation of artificial intelligence development plan*. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

Çin Halk Cumhuriyeti Bilim ve Teknoloji Bakanlığı (2019). *Yeni nesil yapay zekâ için etik kod*. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

Department for Science, Innovation & Technology (DSIT) (2023a). *A pro-innovation approach to AI regulation*. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Department for Science, Innovation & Technology (DSIT) (2023b). *Capabilities and risks from frontier AI*. https://assets.publishing.service.gov.uk/media/65395abae6c968000daa9b25/frontier-ai-capabilities-risks-report.pdf

Deutscher Ethikrat (2023). *Mensch und Maschine – Herausforderungen durch künstliche Intelligenz*. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) (2020). *Etik ile ilgili bilgi notu*. https://www.dfki.de/fileadmin/user_upload/DFKI/Medien/Ueber_uns/DFKI_im_UEberblick/DFKI_Ethik_Handreichung.pdf

Doshi-Velez, F. ve Kortz, M. (2017). *Accountability of AI under the law: The role of explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.

Efe, A. (2021). Yapay zeka risklerinin etik yönünden değerlendirilmesi. *Bilgi ve İletişim Teknolojileri Dergisi*, *3*(1), 1-24.

European Data Protection Supervisor (2016). *Artificial intelligence, robotics, privacy and data protection - Room document for the 38th International Conference of Data Protection and Privacy Commissioners*.

Gemini (2024). *Toplumsal ve çevresel refah ilkesi konuşması*. https://gemini.google.com/app/8eafdc99935e469d?hl=tr

Guo, X. (2024). Legal and ethical principles in the internationalization (social) norms of artificial intelligence. *Journal of Education, Humanities and Social Sciences*, *42*, 527-532. https://doi.org/10.54097/av2rwh40

Gür, R. ve Özbaş, M. (2021). Bilim, araştırma ve yayın etiği. İçinde K. Yılmaz ve R. S. Arık (Edt..), *Bilim ve araştırma etiği* (ss. 45-60). Ankara: Pegem Akademi.

Haenlein, M. ve Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, *61*(4), 5-14. https://doi.org/10.1177/0008125619864925

IEEE (2018). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission (2019). *Ethics guidelines for trustworthy AI*.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission (2018). *The assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment*.

INRIA (2021). *Artificial intelligence: Current challenges and Inria's engagement*.

Japan Ministry of Economy, Trade and Industry (2021). *Yapay zeka ilkelerinin uygulanmasına yönelik yönetim ilkeleri*. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_6.pdf

Karafil, B. ve Uyar, A. (2025). Exploring knowledge, attitudes, and practices of academics in the field of educational sciences towards using ChatGPT. *Education and Information Technologies*, 1–44.

Keskin, D. ve Sevli, O. (2024). Eğitimde yapay zekâ ve etik. In *International Topkapı Congress III* (ss. 38-43).

Kır, B. ve Yılmaz, A. (2024). ChatGPT ve bilimsel araştırmalar: Yapay zeka kullanımının nasıl olması gerektiği üzerine bir literatür taraması. In *III. BİLSEL International Ahlat Scientific Researches Congress* (ss. 297-305).

Kıral, B. (2020). Nitel bir veri analizi yöntemi olarak doküman analizi. *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(15), 170-189.

Kritikos, M. (2019). *Artificial intelligence ante portas: Legal & ethical reflections*. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2019)634427

Kumandaş-Öztürk, H. (2021). Araştırma sürecinde etik ilkeler, etik dışı davranışlar ve etik ihlaller. İçinde K. Yılmaz ve R. S. Arık (Edt.), *Bilim ve araştırma etiği* (ss. 66-77). Ankara: Pegem Akademi.

Küçükvardar, M., Aslan, A. ve Bayrakcı, S. (2020). Yapay zekâ ve etik üzerine bir araştırma. *Atlas Journal*, *6*(36), 1065-1077.

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

Madiega, T. (2024). *Artificial Intelligence Act*. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792

Mahama, I., Baidoo-Anu, D., Eshun, P., Ayimbire, B. ve Eggley, V. E. (2023). ChatGPT in academic writing: A threat to human creativity and academic integrity? An exploratory study. *Indonesian Journal of Innovation and Applied Sciences*, *3*(3), 228–239.

Maral, T. (2024). Sosyal bilimlerin kesişim noktası: Yapay zekâ ve etik. *Ankara Uluslararası Sosyal Bilimler Dergisi (Yapay Zekâ ve Sosyal Bilimler Öğretimi Özel Sayısı)*, 17-33.

Mason, P. (2016, March 29). The racist hijacking of Microsoft's chatbot shows how the internet teems with hate. *The Guardian*. https://www.theguardian.com/world/2016/mar/29/microsoft-tay-tweets-antisemitic-racism

Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot. (2016, March 26). *The Guardian*. https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot (Erişim tarihi: 14 Ocak 2025)

NIST - National Institute of Standards and Technology, U.S. Department of Commerce (2024). *A plan for global engagement on AI standards*.

Oçak, Z. (2023). ChatGPT-4'ün bilimsel özgünlük, yazarlık ve etik üzerine etkisi: Araştırmacı-yapay zekâ iş birliği önerileri. İçinde C. Bilgici (Edt.), *Dijital evrenin yeni iletişim kodları IV* (ss. 173-196). Ankara: Nobel Bilimsel Eserler.

OECD (2019). *Artificial intelligence in society*.

OECD (2023). *Initial policy considerations for generative artificial intelligence - OECD artificial intelligence papers*.

Okun, O., Yüksel, M., Karahan, M. O. ve Bozkurt, R. (2023). Akademik yayıncılığın yeni yüzü: ChatGPT ve etik tartışmaları. *International Journal of Commerce, Industry and Entrepreneurship Studies*, *3*(1), 39–50.

OSCE (2022). *Challenges and opportunities at the intersection of data protection and artificial intelligence*. https://www.osce.org/files/f/documents/b/c/536548.pdf

OSCE (2024). *New frontiers: The use of generative artificial intelligence to facilitate trafficking in persons*.

Özdemir, L. ve Bilgin, A. (2021). Sağlıkta yapay zekanın kullanımı ve etik sorunlar. *Sağlık ve Hemşirelik Yönetimi Dergisi*, *8*(3), 439–445.

Pieper, A. (1999). *Etiğe giriş* (Çeviri: V. Atayaman ve G. Sezer). İstanbul: Ayrıntı Yayınları.

Pratt, I. (1994). *Artificial intelligence*. UK: Pearson Education.

Puran, A. (2024). Dimensions of artificial intelligence ethics from an international and EU perspective. *Agora International Journal of Juridical Sciences*, *18*(2), 245-251. https://doi.org/10.15837/aijjs.v18i2.6994

Russell, S. ve Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.

Seniutis, M., Gružauskas, V., Lileikienė, A. ve Navickas, V. (2024). Conceptual framework for ethical artificial intelligence development in social services sector. *Human Technology*, *20*(1), 6-24. https://doi.org/10.14254/1795-6889.2024.20-1.1

Sok, S. ve Heng, K. (2024). Generative AI in higher education: The need to develop or revise academic integrity policies to ensure the ethical use of AI. *Preprint*.

Solak, E. (2024). Exploring the efficiency of ChatGPT and artificial intelligence in advancing academic writing pedagogy. *International Journal of Education in Mathematics, Science, and Technology*, *12*(6), 1525-1537. https://doi.org/10.46328/ijemst.4373

STM (2021). *AI ethics in scholarly communication*.

STM (2023). *Generative AI in scholarly communications*.

T.C. Kamu Görevlileri Etik Kurulu (2024, 10 Eylül). *Yapay zeka sistemlerinin kullanımında kamu görevlilerinin uyması gereken etik davranış ilkeleri* (Sayı 2024/108).

Taylor, T. (1946, Aralık 9). Tuğgeneral Telford Taylor'ın Doktorlar davasındaki açılış konuşması. *Famous Trials*. https://www.famous-trials.com/nuremberg/1912-doctoropen

The Bletchley Declaration by countries attending the AI Safety Summit, 1–2 November 2023. (2025, Şubat 13). *GOV.UK*. Erişim: 2 Aralık 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

The Danish Government (2019). *National strategy for artificial intelligence*.

The White House (2024a). *Framework to advance AI governance and risk management in national security*.

The White House (2024b). *Memorandum on advancing the United States' leadership in artificial intelligence; harnessing artificial intelligence to fulfill national security objectives; and fostering the safety, security, and trustworthiness of artificial intelligence*.

TRAI - Türkiye Yapay Zekâ İnisiyatifi (2024). *Yapay zekâ etik ilkeleri ve hukuki düzenlemeler raporu*. https://turkiye.ai/wp-content/uploads/2024/06/TRAI-Yapay-Zeka-Etik-Ilkeleri-ve-Hukuki-Duzenlemeler-Raporu-Mayis-2024-5.pdf

Turan, T., Turan, G. ve Küçüksille, E. (2022). Yapay zekâ etiği: Toplum üzerine etkisi. *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *13*(2), 292-299.
Türkiye Bilişim Derneği (2020). *Türkiye'de yapay zekânın gelişimi için görüş ve öneriler*.
Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (2021). *Ulusal yapay zekâ stratejisi 2021-2025*.
UNESCO (2021). *Recommendation on the ethics of artificial intelligence*.
UNESCO (2023). *Ethical impact assessment: A tool of the Recommendation on the Ethics of Artificial Intelligence*.
Uyan, U. (2023). Yapay zekânın bilimsel yayın amaçlı kullanımına ilişkin etik kaygılar: Sistematik bir yazın incelemesi. *İş Ahlakı Dergisi*, *16*(2), 173-199.
van Uffelen, N., Lauwaert, L., Coeckelbergh, M. ve Kudina, O. (2024). *Towards an environmental ethics of artificial intelligence*. https://doi.org/10.48550/arxiv.2501.10390
Wechsler, P. (2005). *La Faculté de Medecine de la "Reichsuniversitaï Straßburg" (1941-1945) à l'heure nationale-socialiste* (Doctoral dissertation). University of Strasbourg.
Yeşilkaya, N. (2022). Yapay zekâya dair etik sorunlar. *Şarkiyat*, *14*(3), 948-963.
Yıldırım, A. ve Şimşek, H. (2011). *Sosyal bilimlerde nitel araştırma yöntemleri* (8. Baskı). Ankara: Seçkin Yayıncılık.
Yılmaz, G. (2020). Yapay zekânın yargı sistemlerinde kullanılmasına ilişkin Avrupa etik şartı. *Marmara Üniversitesi Avrupa Araştırmaları Enstitüsü Avrupa Araştırmaları Dergisi*, *28*(1), 27-55.
Yılmaz, K. (2024). Türkiye'de yayınlanan bilimsel araştırma ve yayın etiği konulu makalelerin incelenmesi (1993–2022). *Manas Sosyal Araştırmalar Dergisi*, *13*(1), 85-100. https://doi.org/10.33206/mjss.1340104
YÖK - Yükseköğretim Kurulu (2024). *Yükseköğretim kurumları bilimsel araştırma ve yayın faaliyetlerinde üretken yapay zekâ kullanımına dair etik rehber*.

EXTENDED ABSTRACT

Artificial intelligence is now widely used in many fields, such as daily life, education, health, science, engineering, technology, security, communication, art, marketing, business, and public administration. With the widespread use of artificial intelligence, some ethical issues have also come to the fore. The aim of this study is to make an evaluation of the ethical principles that should be considered when using artificial intelligence in scientific research, based on the reports, policies, legal texts and documents prepared by international and national organizations on the ethics of artificial intelligence. The technique of document analysis was used in the study. As a result of the review, 34 documents published by international organizations and 20 documents prepared by national level organizations of different countries were examined. As a result of the review, similar principles regarding the ethics of artificial intelligence were identified in the reports and an attempt was made to propose a general framework based on them. As a result of the analysis, all texts focus on eight main principles: "Human element and oversight; Privacy and data management; Diversity, non-discrimination and fairness; Transparency; Social and environmental welfare; Human rights; Technical robustness and safety; Accountability". **1. Human Element and Supervision:** The first of the ethical principles that should be considered when using artificial intelligence in scientific research that was identified in the research is the human element and oversight. This principle includes processes such as the development, use, and monitoring of the results of artificial intelligence systems, and expresses that humans are actively involved in and in control of the life cycle of artificial intelligence. In the UNESCO 2021 Recommendation Decision, it was requested that Member States always allow the attribution of responsibility in matters related to artificial intelligence to natural or legal persons. The decision emphasized that people can use artificial intelligence systems to make decisions and take actions, but an artificial intelligence system can never replace ultimate responsibility and accountability. As a general rule, it has been emphasized that life and death decisions cannot be delegated to artificial intelligence systems. **2. Privacy and Data Management:** In its most general definition, this principle refers to the management and confidentiality of data and personal information belonging to participants. The need for large amounts of data for machine learning algorithms to train and optimize artificial intelligence systems creates an incentive to maximize data collection (OECD, 2019). UNESCO's 2021 Recommendation also addresses data governance and privacy. It mentions that governance mechanisms and legal boundaries on this issue should be established with a multi-stakeholder approach at the international level, while being protected by legal systems. These mechanisms and legal boundaries should provide a legitimate and valid legal basis for the processing of personal data, including informed consent, by referring to international data protection principles regarding the collection, use and disclosure of data and the exercise of data subjects' rights (UNESCO, 2021). **3. Diversity and Fairness**: This principle is that AI systems act with respect for diversity and a sense of fairness, without discriminating among people based on race, religion, sexual orientation, or political views. AI learns by identifying patterns in existing data. It then uses these learned patterns to make predictions, recommendations, and decisions for the future. This carries the risk of repeating and reinforcing historical biases (STM, 2021). The UNESCO recommendation (2021) emphasized that

member states should respect multilingualism and cultural diversity and encourage everyone, including local communities, to access AI systems with locally relevant services and content. Article 30 of the same decision also emphasized that actors should make reasonable efforts to minimize and avoid discriminatory and biased practices and outcomes to ensure that AI systems are fair. **4. Transparency:** Transparency refers to the clear communication of the intentions and stages of deployment of artificial intelligence systems, and the open execution of the entire process. Transparency is not about sharing specific codes and datasets with people, but rather allowing people to understand how these systems are developed, trained, and deployed (OECD, 2019). There are a number of issues that are considered important for ensuring transparency. The first is traceability. Having sufficiently detailed information about actions so that an action can be traced back later is called traceability and is considered important for transparency (ALLISTENE, 2018). Traceability is a process that is necessary first for the attribution of responsibility as a basis for legal proceedings and then for the diagnosis and correction of dysfunctions (ALLISTENE, 2018). **5. Social and Environmental Welfare:** The basis of artificial intelligence that respects ethical principles and rights is a structure that emphasizes human and social values, serves humanity, protects the priorities of public interest, and contributes to the progress of humanity (TRAI - Turkey Artificial Intelligence Initiative, 2024). This principle can be explained as artificial intelligence systems that prioritize the benefit of society and the environment. The aim is to prevent artificial intelligence systems from harming people and the environment and to prioritize the benefits. In the explanatory memorandum published by the European Commission in 2021, artificial intelligence was defined as a rapidly evolving family of technologies that can contribute to a wide range of economic and social benefits in all sectors of industry and social activities (European Commission, 2021). **6. Human Rights:** This principle commits to not violating the human rights and dignity of participants in scientific experimentation and research. The OECD (2019) report emphasizes that a human rights approach to artificial intelligence can be helpful in determining risks, priorities, and a framework. For this reason, human rights have an important place in the ethical framework to be developed for artificial intelligence systems. The same report highlights the Human Rights Impact Assessment (HRIA), stating that this application can help identify risks that actors cannot foresee in the lifecycle of artificial intelligence. The first major technology company to conduct an HRIA for artificial intelligence was Microsoft with its work in 2018 (OECD, 2019). **7. Technical Robustness and Security:** This principle states that artificial intelligence systems used in research should be technically secure and robust. Although considered a technical issue, it is a serious one that directly affects many principles, such as privacy and data management, human rights, and accountability. Any cyber attack by terrorist groups against artificial intelligence systems, or deliberate intervention by governments, can have very serious consequences. In the social sciences, the seizure of participants' data and personal information, and in the medical field, the seizure of participants' health data in matters such as genetic studies, can cause very serious problems and crises. Looking at international documents, the following statements in a briefing to the European Parliament underscore the seriousness of the situation Any attack on artificial intelligence applications integrated into the human body can endanger human health, cognitive freedom, mental integrity, and even human life (Kritikos, 2019). **8. Accountability:** This principle expresses the accountability of actors and artificial intelligence systems in case of a possible problem. This principle is one of the most debated issues today. Given the risks discussed in the context of the above principles, who will be responsible in the event of a potential problem? The engineers who develop AI systems, the intermediary companies that market them, governments or various terrorist groups, scientists as researchers, or AI itself? These are serious questions that are being considered and discussed in order to draw a legal line. According to the study conducted by the Chamber of Computer Engineers (2025), "in order to ensure the proper operation of artificial intelligence systems throughout their life cycle, standards and legislation should be developed to be followed by designers, developers and those who install these systems; responsibilities should be clearly defined; individuals or institutions should be held accountable for their actions".