

Gönderim Tarihi : 27.05.2024

Kabul Tarihi : 29.06.2024

DOI: 10.5281/zenodo.12637303

Nurcan DENİZ*

Yapay Zekânın Sürdürülebilirliği: Sorumlu Yapay Zekâ

The Sustainability of Artificial Intelligence: Responsible Artificial Intelligence

Özet

Abstract

Dijital dönüşümün yıkıcı teknolojileri arasında ilk sırada yer alan yapay zekâ, her geçen gün önemini artırmaktadır. İnsanlar bir taraftan yapay zekâyı kullanmanın nimetlerinden faydalanma yoluna giderken, diğer taraftan işlerini kaybetme korkusu yaşamaktadır. Bir yandan otonom araçların bir kaza anında kime çarpması gerektiğine dair ahlaki deneylerin yapıldığı endişe ile izlenirken, diğer yandan yapay zekâ aracılığıyla hastalık tespitlerinde büyük yol kat edilmesi ümit vermektedir. Bir yandan yapay zekânın halüsinasyon görebileceği tartışılırken, diğer taraftan cinsiyet, ırk vb. konularda ayrımcı kararlar almasını sağlayan yanlılıklara sahip olduğunu gösteren çalışmalar ortaya çıkmaktadır. Tüm bu gelişmeler, son yıllarda diğer bir önemli konu olarak gündemimizde olan sürdürülebilirliğin sosyal boyutu bağlamında ele alındığında, insan ve toplum tarafında çalışmalar yapılması gerektiğini göstermektedir. Yapay zekâ ile ekonomik, çevresel ve sosyal sürdürülebilirliğe katkı sağlanmasının ötesinde, yapay zekânın bizzat kendisinin sürdürülebilir olması için sorumlu yapay zekâ kavramının tartışılması gerekmektedir. Bu noktada etik ve açıklanabilirlik gibi konular öne çıkmakla birlikte, sorumluluk kavramının bu kavramları içermekle birlikte daha öte bir kavram olduğu da dikkatlerden kaçırılmamalıdır. Bu çalışma uluslararası yazında son birkaç yıldır yer almakla birlikte Türkçe yazında bir boşluk olduğu tespit edilen sorumlu yapay zekâ konusunda genel bilgileri vermeye yönelik olarak tasarlanmıştır. Bu bağlamda oldukça sınırlı ve yeni bir çalışma alanı olan sorumlu yapay zekâ konusunda yapılan yazın taraması neticesinde son birkaç yılda çalışmalar gerçekleştirildiği, yazının disiplinlerarası ve farklı kültürlerde yapılan çalışmalarla desteklenmesi gerektiği ortaya çıkmaktadır.

Anahtar Kelimeler: yapay zekâ, sürdürülebilirlik, sorumlu yapay zekâ, açıklanabilirlik

The importance of artificial intelligence is increasing day by day as being the first rank disruptive technologies of digital transformation. While people are trying to benefit from using artificial intelligence, they are also afraid of losing their jobs. While it is being watched with concern that ethical experiments are being conducted on who autonomous vehicles should hit in the event of an accident, it is also promising that great progress has been made in disease detection through artificial intelligence. While it is being discussed that artificial intelligence can hallucinate, studies are emerging about its biases which make discriminatory decisions on issues such as gender, race, etc. All these developments, when considered in the context of the social dimension of sustainability, which has been on our agenda as another important issue in recent years, show that studies should be carried out on the human and society side. Beyond contributing to economic, environmental and social sustainability with artificial intelligence, the concept of responsible artificial intelligence should be discussed in order for artificial intelligence itself to be sustainable. At this point, issues such as ethics and explainability come to the fore, it should not be overlooked that the concept of responsibility is a concept that includes these concepts but goes beyond them. This study was designed to provide general information on responsible artificial intelligence, which has been included in international literature for the last few years, but has been identified as a gap in Turkish literature. In this context, as a result of the literature review conducted on responsible artificial intelligence, which is a relatively limited and new field of study, it is revealed that studies have been carried out in the last few years and it needs to be supported by interdisciplinary and cross-cultural studies.

Keywords: artificial intelligence, sustainability, responsible artificial intelligence, explainability

*Eskişehir Osmangazi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Türkiye, nurcanatikdeniz@gmail.com

Giriş

Yapay zekâ (YZ) her geçen gün gelişmekte ve gündemde daha fazla yer almaktadır. Sağlık gibi birçok farklı alanda başarılı uygulamaları olan YZ, diğer taraftan gizlilik endişeleri, kullanıcı güvensizliği, veri sızıntıları, bilgi şeffaflığı ve etik endişeleri de beraberinde getirmektedir. Algoritmaların şeffaflık eksikliği etik endişelerin artmasına neden olmaktadır (Wang ve diğerleri, 2020). YZ kullanımı kötü niyetli kullanımlar, algoritmik önyargılar ve/veya ayrımcılık, şeffaflık eksikliği ve hesap verebilirlik sistemlerinin eksikliği gibi istenmeyen sonuçlar doğurabilmektedir (Merhi, 2023). Ntoutsis ve diğerleri (2020) tarafından “NoBias - Artificial Intelligence without Bias,” adlı proje kapsamında “algoritmik yanlılık” konusunda ırk, cinsiyet vb. konularda veri güdümlü yapay zekânın yanlılığı olabileceği ifade edilmektedir. Yazın da bazı işe alım algoritmalarının kadınlar yerine erkekleri işe alma yanlılığına sahip olduğunu ortaya koymaktadır (Bozkurt Gümrükçüoğlu ve Yakacak, 2023). Meske ve diğerleri (2022) yapay zekânın en temel potansiyel probleminin “yanlılık (bias)” olduğunu ifade etmekte ve “vasıfsızlaştırma (deskilling)”ya neden olabilecek “otomasyon yanlılığı”na dikkat çekmektedirler. Zira insanlar otomatik karar vermeye aşırı güvenebilmektedir. Bunun bir örneği doktorların kendi teşhisleri doğru olmasına rağmen yapay zekânın teşhisinden farklı olduğu için göz ardı etmeleri şeklinde gerçekleşebilmesidir.

Üretken yapay zekânın gelişimi ile birlikte küresel anlamda yapay zekâ etiği gibi konular daha fazla tartışılmaya başlamıştır. Hatta “veri”nin yeni “petrol” olarak görüldüğü ifade edilmekte ve yapay zekânın ancak sorumlu bir şekilde geliştirilirse bu petrolden değer üretebilecek en yeni elektrik kaynağı olabileceği ifade edilmektedir (Gupta ve diğerleri, 2023).

Sorumlu YZ (responsible artificial intelligence) kavramının sadece “hoş bir şey” değil, başarı için “olmazsa olmaz bir gereklilik” olduğuna vurgu yapan Gupta ve diğerleri (2023), yapay zekânın sorumlu değilse, aslında akıllı olmadığını da iddia etmektedir. 2020 yılında Kanada ile Fransa’nın öncülüğünde kurulan ve sekretaryası OECD tarafından gerçekleştirilen “Yapay Zekâ Konusunda Küresel İş Birliği” (GPAI) uluslararası inisiyatifi de “sorumlu YZ” konusunda araştırmalar yürütülmesini teşvik etmektedir (UYZS). Konunun önemine binaen T.C. Dijital Dönüşüm Ofisi ve T.C. Sanayi ve Teknoloji Bakanlığı tarafından hazırlanan ve 2021-2025 yıllarını kapsayan Ulusal Yapay Zekâ Stratejisi’nde “Güvenilir ve Sorumlu YZ Çalışma Grubu” YZ ekosisteminin bir parçasıdır. Bu raporda yer verilen “Kamu Yönetimi YZ Hazır Olma Endeksi 2020 Raporu” sonuçlarına göre Türkiye, “Sorumlu YZ” alt-endeksinde mahremiyet, şeffaflık, kapsayıcılık ve hesap verebilirlik puanları dikkate alındığında 34 ülke içerisinde 31’inci sırada yer almıştır (UYZS).

Tam da bu noktada Türkiye’de bu konuda çalışmaların artması gereği aşikârdır. Bu bağlamda uluslararası yazın son birkaç yıl içinde sınırlı da olsa gelişmeye başlamışken, Türkçe yazında “Scholar” veritabanında “sorumlu yapay zekâ” anahtar kelimesi ile arama yapıldığında sadece iki makaleye erişilebilmiştir. Bu makalelerden ilkinde Avundukluoğlu (2023) sorumlu yapay zekâ konusuna bir alt başlık olarak kısaca yer vermektedir. Diğer taraftan Atiker (2024) ise güzel sanatlar alanında yapay zekâ ve sorumlu yapay zekâ ilişkisine yer vermekte ve şeffaflık, telif hakları ve veri yönetimi gibi konulara dikkat çekmektedir. Diğer yandan “sorumlu yapay zekâ” anahtar kelimesi ile yapılan arama sonucunda herhangi bir tez de tespit edilememiştir. Türkçe yazında bu konudaki yayın eksikliği, bu makalenin temel motivasyonudur. Konunun önemine binaen bu makalede uluslararası yazında konu ile ilgili güncel çalışmaları ve gelişmeleri paylaşmak amaçlanmaktadır. Böylelikle hem bu konuda araştırma yapmak veya bilgi edinmek isteyenlerin konuyu anlayabilmeleri, hem de Türkçe yazına katkı sağlamak amaçlanmaktadır.

Bu makalede öncelikle yapay zekâ kavramına yer verilecek, sonrasında sürdürülebilirlik konusuna kısaca yer verildikten sonra sorumlu yapay zekânın tanımı ve sorumlu yapay zekâ ile ilgili kavramlara (açıklanabilirlik, riskler, engeller, yönetim, katılımcılık vb.) yer verilecektir. Çalışmada ayrıca uluslararası yazından örnek çalışmalar yer almaktadır. Sonuç bölümü ile makale sonlandırılacaktır.

Yapay Zekâdan Sorumlu Yapay Zekâya

Bu bölümde öncelikle yapay zekâ kavramına yer verildikten sonra, sürdürülebilirlik ile ilgili özet bilgi verilecektir. Sonrasında da sorumlu yapay zekâ kavramı ve bileşenleri ortaya konacaktır.

Yapay Zekâ

Psikologların hala tartıştığı ve bu alanda farklı kuramlar ortaya koyduğu “zekâ”, genellikle neyin doğru neyin yanlış olduğuna ilişkin doğru sonuca varma ve problemlerin nasıl çözüldüğü ile ilişkili olarak görülse de kabul edilmiş tek bir tanımı olmayan çok yönlü bir kavram olarak ifade edilmektedir (Dignum, 2022). “Yapay zekâ” terimini ilk kez ortaya koyan John McCarthy yapay zekâyı “*akıllı ajanların incelenmesi ve tasarımı*” şeklinde tanımlamıştır (Dignum, 2022).

Yapay zekâyı anlamak için belki de ilk anlaşılması gereken “algoritma” kavramıdır. 9. yy âlimlerinden olan El-Harezmi’nin batıda bilinen adı olan algoritmanın ne demek olduğunu anlamanın en kolay yolunun onu bir yemek tarifi olarak düşünmek olduğunu ifade eden (Dignum, 2022); yapay zekânın algoritma olmadığı, algoritmalara dayalı olduğunu vurgulamaktadır.

“Makine öğrenmesi” ve makine öğrenmesinin bir alt kümesi olan “derin öğrenme”nin yanında “doğal dil işleme”, “bilgisayar görüşü”, ve “veri bilimi”de YZ ile ilişkili alanlardır (Ahmed vd., 2022). 1950’lerden beri üzerinde çalışılan YZ (Meske vd. 2022), özellikle yanlış karar verme maliyetinin yüksek olduğu sağlık, finans ve hukuk gibi alanlarda açıklanabilirliklerinin düşük olması nedeniyle sınırlı kullanıma sahiptirler (Deliloğlu ve Çakmak Pehlivanlı, 2021). İç içe doğrusal olmayan yapılara sahip derin öğrenme algoritmaları, her ne kadar yüksek tahmin doğruluğuna sahip olsalar da, şeffaf olmadıkları için kara kutu (black-box) olarak anlandırılmaktadır (Samek vd. 2017).

Sürdürülebilirlik

Türk Dil Kurumu sözlüğünde “*sürdürülebilir olma durumu*” şeklinde tanımlanan “sürdürülebilirlik” (sustainability) kavramı Brundtland Raporu’nda (1987) “*gelecek nesillerin kendi ihtiyaçlarını karşılama yeteneğinden ödün vermeden bugünün ihtiyaçlarını karşılmasını sağlamak*” şeklinde ifade edilmektedir. “Ekonomik” boyutun yanı sıra “çevresel” ve “sosyal” boyutlarının da olduğu (Elkington,1994) sürdürülebilirliğin önemine binaen Birleşmiş Milletler (BM) tarafından 17 adet Sürdürülebilir Kalkınma Amacı (Sustainable Development Goals-SDGs) ortaya konmuştur. YZ gibi yenilikçi teknolojiler, karbon salınımının azaltılması gibi çevresel sürdürülebilirliği artırıcı çalışmalarda kullanım potansiyeline sahiptir. Hatta “ikiz dönüşüm” (twin transition) kavramı ile hem dijital hem de yeşil dönüşümün birlikte ele alınmaya başladığı görülmektedir (Deniz ve Büyük, 2023).

2016 yılında Japon Hükümeti’nin ortaya koyduğu “Toplum 5.0” (Society 5.0) kavramı ile birlikte sosyal sürdürülebilirlik anlayışı çerçevesinde “*teknolojinin bir amaç olmaktan ziyade, toplum yararına gerçekleştirilecek çalışmalar için bir araç olduğu*” anlayışı öne çıkmıştır. Sürdürülebilirlik ile ilgili olarak yapay zekânın insanlık için fayda sağlama gereği “insan merkezli yapay zekâ” veya “sosyal fayda için yapay zekâ” şeklinde de ifade edilmektedir (Arrieta vd. 2020).

Sorumluluk Boşluğu

"Sorumluluk" terimi felsefe ve hukuk alanında belirsiz ve çok anlamlı bir terim olarak bilinmektedir. “*Ya öğrenme sistemlerinin tasarımı ve kullanımına devam edip böylece insanların davranışlarından sorumlu olma olasılığından vazgeçeriz ya da insan sorumluluğunu koruruz ve böylece toplumda öğrenme sistemlerinin tanıtılmasından vazgeçeriz*”. Bu ikilem 2004 yılında Matthias tarafından “sorumluluk açığı (responsibility gap)” sorununu ortaya konulurken ifade edilmiştir. Bu açığın tespit edilmesinin üzerinden tam 20 yıl geçmiş olmakla birlikte ve YZ ile ilgili olarak yeniden gündeme gelmiştir. Üstelik sadece yapay zekânın öğrenme kapasitesi ile ilgili değil, bu sistemlerin genel olarak sergilediği opaklık, karmaşıklık ve öngörülemezlikle ilgili olarak da sorumluluk açığı olduğu ifade edilmektedir (Sio ve Mecacci, 2021).

Sio ve Mecacci (2021) felsefî, hukuki ve sosyolojik sorumluluk kuramlarından yararlanarak dört farklı tür sorumluluk boşluğunu tanımlamışlardır: suçluluk boşluğu (culpability gap), ahlaki hesap verebilirlik boşluğu

(moral accountability gap), kamusal hesap verebilirlik boşluğu (public accountability gap) ve aktif sorumluluk boşluğu (active responsibility gap). Burada yer alan son boşluk (aktif sorumluluk boşluğu) “*yapay zekâyı tasarlayan, kullanan ve onunla etkileşim kuran kişilerin, tasarladıkları, kontrol ettikleri veya kullandıkları sistemlerin davranışlarına yönelik ahlaki yükümlülüklerini görüp bunlara göre hareket etmek için yeterince farkında, yetenekli ve motive olmayabilecekleri riski*”ne yazında ilk kez yer vermesi yönüyle önemlidir.

Tarif edilen sorumluluk boşluklarına kısmi cevaplar ise sorumluluk boşluğunun yeni ve çözümsüz bir sorun olduğu fikri (kadercilik), sorumluluk boşluğunun yeni olmadığı ve bir sorun olmadığı fikri (defasyonelizm) ve sorumluluk boşluğunun, basitçe yeni teknik ve/veya yasal araçlar getirilerek çözülebilecek bir sorun olduğu fikri (çözümçülük)” riskleri şeklinde sıralanmıştır (Sio ve Mecacci, 2021).

Diğer taraftan “açıklanabilir yapay zekâ” savunucularının, yapay zekâ ile etkileşimin psikolojik, sosyal ve politik boyutunu görmezden gelme eğiliminde olduğu da iddia edilmektedir. Bunun sonucu olarak da yapay zekânın davranışına ilişkin tüm ahlaki ve sosyal sorumluluğun, yapay zekâ tekniklerinin işleyişindeki bir gelişmeyle çözülebileceği bir tür “teknik çözümçülük” benimseme riskiyle karşı karşıya olduğu konusunda da uyarılar bulunmaktadır. Bu bağlamda tanımlanan diğer bir risk ise “yasal çözümçülük” riskidir. Bu riskin ise mevcut yasal sorumluluk rejimlerinin revizyonunu öneren avukatlar ve politika yapıcılarının, yapay zekânın davranışı üzerinde bir tür insan ahlaki sorumluluğunu sürdürmenin önemini küçümsemesinden veya ahlaki ve sosyal uygulamaların yapay zekânın kullanımına sorumlu bir geçişi yönetmek için nasıl değişmesi gerektiğini söylemeden bu ihtiyacı kabul etmelerinden ortaya çıkabileceği ifade edilmektedir (Sio ve Mecacci, 2021).

Sorumlu Yapay Zekâ

Tanımı

Günümüzde, yapay zekâ teknolojilerinin güvenli, yararlı ve adil kullanımını sağlamak için yapay zekâyı karşı sorumlu bir yaklaşıma ihtiyaç duyulduğuna dair farkındalık artmaktadır. Yapay zekâyı sorumlu bir şekilde geliştirmek, insan değerlerini ortaya çıkarma ve temsil etme, bu değerleri teknik gereksinimlere dönüştürme, ahlaki ikilemler ve değer tercihleriyle başa çıkma araçlarını geliştirme ve sistemleri insan refahına katkıları açısından değerlendirme araçlarına ihtiyaç duymaktadır (Dignum, 2017).

Sürdürülebilirlik bağlamında önemli bir konu haline gelen “sorumlu yapay zekâ” IBM tarafından, “*model açıklanabilirliğine ilave olarak yapay zekâ yöntemlerinin adil bir biçimde ve hesap verebilirliğe imkân verecek şekilde kuruluşlarda uygulanmasına yönelik bir metodoloji*” şeklinde tanımlanmaktadır. OECD ve AB’nin YZ ile ilgili tanımlamaları esas alan Ulusal Yapay Zekâ Stratejisi (UYZS)’nde ise sorumlu YZ, “*refah, karar hakkına saygı, mahremiyet, dayanışma, demokratik katılım, eşitlik, çeşitlilik, ihtiyat, sorumluluk ve sürdürülebilir kalkınma ilkelerini esas alan YZ*” şeklinde tanımlanmıştır.

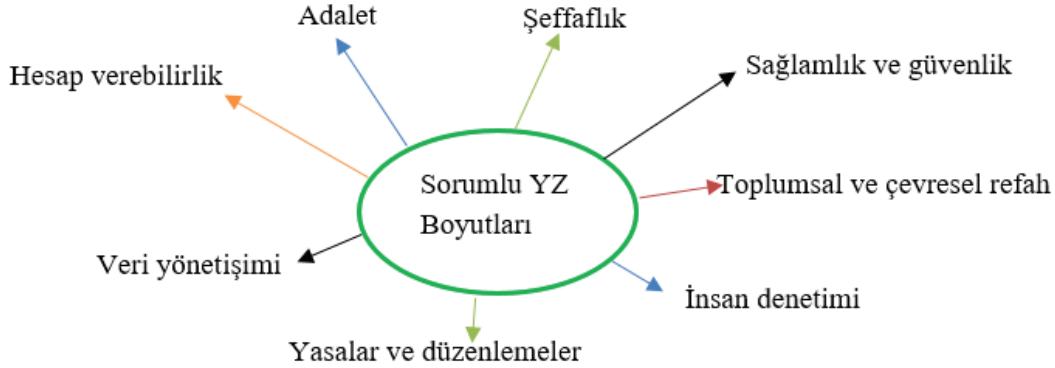
Diğer taraftan Merhi (2023) sorumlu yapay zekânın, güvenlik, önyargılar ve ayrımcılıkla ilgili sorunlardan kaçınmak ve sonuçların yorumlanabilirliğini basitleştirmek için standartlar ve değerler oluşturmakla ilgilendiğini ifade etmektedir. Zira şimdiye kadar bireyci ve rasyonel olarak kabul edilen YZ’nin aslında sınırlı rasyonel olduğu ve kendi algılanan çıkarları doğrultusunda hareket ettiği ifade edilmektedir (Dignum, 2022).

Algoritma kavramını yemek tarifi ile açıklayan Dignum (2022) elmalı turta yaparken organik elmaları seçmek ile adalete, mahremiyete, şeffaflığa ve diğer tüm değerlere saygı duyan ve bunu sağlayan verileri seçen yapay zekâ arasında analogi kurmaktadır. Ayrımcı önyargıları ve adil olmayan sonuçları en aza indirmek veya ortadan kaldırmanın düşük kaliteli verilerin kullanılmaktan daha fazlası olduğunu ifade eden (Dignum, 2022), başarılı bir düzenlemenin, neyin (teknolojinin kendisi mi, etkisi mi, yoksa uygulamanın sonuçları mı) düzenleneceği konusunda net seçimler yapılmasını gerektirdiğini iddia etmektedir. Ayrıca, düzenlemelerin yapay zekânın geliştirildiği ve kullanıldığı girdileri, süreçleri ve koşulları da ele almasının önemine vurgu yapmaktadır. Ayrıca ilave düzenlemelerden önce mevcut yasaların anlaşılmasıyla işe başlanması gerektiği

tavsiye edilmektedir. Standartların bu anlamda önemli olduğu ve ISO ve IEEE gibi standartların özellikle yapay zekânın emniyeti, güvenliği ve sağlamlığı, açıklanabilirlik garantileri ve algoritmik kararlarda önyargıyı azaltma araçlarıyla ilgili olduğunda yapay zekâ politika amaçlarını destekleyebileceği ifade edilmektedir. Bu bağlamda ISO, IEC ile birlikte YZ konusunda SC42 adlı bir komite kurmuştur. IEEE de “Ethically Aligned Design (EAD)” adlı 700 uluslararası araştırmacı ve uygulamacının yer aldığı küresel bir girişim başlatmıştır. P700 çalışma grubu sistem tasarımı, otonom sistemlerde şeffaflık, algoritmik önyargı, kişisel, çocuk, öğrenci ve işveren veri yönetimi, dürtme veya haber kaynaklarının güvenilirliğinin belirlenmesi ve derecelendirilmesi gibi çok çeşitli konular üzerinde hâlihazırda çalışmaktadır (Dignum, 2022).

Boyutları

Şekil 1’de görüleceği üzere Mikalef ve diğerleri 2022 yılında gerçekleştirmiş olduğu çalışma sonucunda sorumlu YZ’nin sekiz boyutunun adalet, şeffaflık, hesap verebilirlik, sağlamlık ve güvenlik, veri yönetimi, yasalar ve düzenlemeler, insan denetimi ile toplumsal ve çevresel refah olduğunu ifade etmektedir (aktaran Merhi, 2023). Sorumlu YZ konusunda eşit derece önemli üç konu olduğunu vurgulayan Dignum (2017) öncelikle toplumun genelinin yapay zekânın etkisinden sorumlu olmaya hazır olması gerekir. Diğer önemli konular ise YZ sistemlerinin etik ve insani değerler hakkında akıl yürütmesini ve bunlara uygun hareket etmesini sağlayacak mekanizmalar ve sorumlu YZ çerçeveleri geliştirecek insanların katılımıdır. Her üç konuda da eğitimin gereğine vurgu yapmaktadır.



Şekil 1. Sorumlu YZ Boyutları

(Kaynak: Mikalef ve diğerlerinden 2022 aktaran Merhi, 2023)

Konuya kurumsal sosyal sorumluluk (Corporate Social Responsibility- CSR) bakış açısıyla yaklaşan Wang ve diğerleri (2020) sorumlu yapay zekânın dört uygulamasını Şekil 2’de görüleceği üzere sırasıyla veri yönetimi, etik çözümler tasarlama, insan merkezli gözetim/risk kontrolü ve Eğitim ve Öğretim şeklinde tanımlamaktadır. Veri Yönetimi başlığı altında Sorumlu YZ yönetimi, “şeffaflık”, “güven” ve “açıklanabilirlik” oluşturmaya odaklanmaktadır. Öncelikle yapay zekânın kurumsal kullanımının paydaşlar için şeffaf olması, bir YZ uygulamasının verilerini nasıl işlediğini tam olarak anlamalarına katkı sağlamasının yanı sıra belirli kararlara varmalarına olanak sağlaması açısından da önemlidir. Diğer taraftan güven inşa etmek için veri kalitesi yönetimi ve veri sahipliği yönetimi konusu önemlidir. Bu bağlamda Güvenilir YZ, yüksek kaliteli veri ve kullanım onayıyla oluşturulmaktadır. Tüketici verilerinin kullanılabilirliği, tüketicilerin gizlilik ihlali, dolandırıcılık, bilgi sızıntısı ve kimlik hırsızlığı gibi ciddi sorunlara yol açma riski nedeniyle önemli bir konudur ve ülkeler bu anlamda kanuni düzenlemeler gerçekleştirmektedir. Avrupa Birliği Genel Veri Koruma Yönetmeliği (GDPR) ve Japonya Kişisel Bilgilerin Korunması Yasası (APPI) bunlara örnek olarak verilebilir. (Wang ve diğerleri 2020). Örneğin GDPR’de, algoritmik hesap verebilirliğe ilişkin hükümlere yer verilmesinin yanı sıra yapay zekâ tarafından alınan bireysel kararlar hakkında bilgi edinme hakkı da bulunmaktadır (Canbay ve Demircioğlu, 2021). Son olarak yapay zekâ modelleri tarafından üretilen sonuçlar hakkında anlamlı ve kişiselleştirilmiş açıklamalar sağlamanın belirsizliği azaltabileceği ve kullanıcılarla güven oluşturabileceği dikkate alındığında “Açıklanabilirlik” de veri yönetimi açısından önemli bir konudur (Wang ve diğerleri 2020).



Şekil 2. Sorumlu YZ uygulamaları için kavram haritası

(Kaynak: Wang ve diğerleri 2020)

Güvenilir (trustworthy) yapay zekâya yönelik tüm gereksinimlerin gerekli, ancak insan merkezli (human-centred) yapay zekâyı geliştirmek için yeterli olmadığını anlamak önemlidir (Dignum, 2022).

Yapay zekânın kullanılan tekniklerin ötesinde sosyal etki ve insan katılımı açısından da incelenmesi ve bu yönüyle sosyo teknik eko-sistem olarak anlaşılması gerektiğini ifade eden (Dignum, 2022) yapay zekâyı anlamak için insan ve teknoloji arasındaki ilişkinin anlaşılması gereğine vurgu yapmaktadır. Ayrıca yapay zekâyı güven ve sorumluluğu sağlamaya yönelik kılavuzlar, ilkeler ve stratejilerin, yapay zekânın geliştirildiği ve kullanıldığı sosyo-teknik ekosisteme yönelik olması gerektiğine dikkat çekilmektedir. Tam da bu noktada etik, güvenilir veya sorumlu olması gerekenin yapay zekâ ürünü veya uygulamasından ziyade, bu ekosistemin sosyal bileşeninin sorumluluk alması ve toplum tarafından güvenilebileceği bir etik çerçeveyi dikkate alarak hareket etmesi gereği ifade edilmektedir. Etik yapay zekânın, makinelere eylemleri ve kararları konusunda bir tür “sorumluluk” vermenin ve bu süreçte insanları ve kuruluşları sorumluluklarından kurtarmanın bir yolu olmadığını dikkat çeken Dignum (2022) aksine, ilgili kişi ve kuruluşlara daha fazla sorumluluk ve hesap verebilirlik sağladığını iddia etmektedir.

Düzenlemeler

Sorumlu YZ kapsamında düzenlemelerin ve sertifikasyonun önemine vurgu yapan Dignum (2020) IEEE tarafından hazırlanan ve otonom ve akıllı sistemlerde şeffaflığı, hesap verebilirliği ve algoritmik önyargının azaltılmasını ilerleyen sertifikasyon ve işaretleme süreçleri için özellikler oluşturmak amacıyla taşıyan Otonom ve Akıllı Sistemler Etik Sertifikasyon Program (The Ethics Certification Program for Autonomous and Intelligent Systems -ECPAIS) örnek olarak vermektedir. Sorumlu Yapay Zekânın Değerlendirmesi, bazı etik “kutuların” işaretlenmesinden veya yapay zeka sistemlerinde bazı eklenti (add-on) özelliklerin geliştirilmesinden daha fazlasıdır (Dignum, 2022).

Wang ve diğerleri (2020) seçmiş oldukları örnek olaylardan hareketle sorumlu yapay zekâ geliştirmek isteyenler için kılavuz niteliğinde beş strateji ortaya koymuşlardır:

1. Baş Sorumlu Yapay Zekâ Görevlilerinin (Chief Responsible AI Officers- CRaiO) Ortaya Çıkışı
2. Yapay zekânın ekonomik ve sosyal sürdürülebilirliğini dengeleme

3. Şeffaf ve müşteri odaklı veri politikası
4. Yapay zekâ ile sosyal sorumluluk sahibi girişimler yaratmak
5. Yapay zekâ kullanımını düzenlemek için havuç (ödül) ve sopa (ceza) mekanizması

Açıklanabilirlik

Sorumlu YZ'nin en önemli bileşenlerinden bir tanesi açıklanabilirliktir. Van Lent ve arkadaşları açıklanabilir yapay zekâyı (Explainable Artificial Intelligence- XAI) 2004 yılında “*kullanıcıya, sistemin bilgisi ve çıkarımı yoluyla, kullanıcının isteğinden ortaya çıkan davranışa kadar kolayca anlaşılabilir bir akıl yürütme zinciri sunma*” olarak ifade etmiştir (Meske vd. 2022:56). Açıklanabilir YZ'nin önemi 2017'de Amerika'da Defense Advanced Research Projects Agency (DARPA) adlı araştırma kurumunun kurulması ile artmıştır (DARPA, 2016). Arrieta ve diğerleri (2020) açıklanabilir YZ için en önemli noktanın “anlaşılabilirlik” olduğunu ve birincil amacının “güvenilirlik” olduğu iddia etmektedir. Diğer amaçları ise nedensellik, transfer edilebilirlik, bilgilendiricilik, güven, adalet, ulaşılabilirlik, etkileşim ve gizlilik farkındalığı şeklinde sıralanmaktadır (Deniz, 2022). Kullanıcı güvenini artırması açısından da açıklanabilirlik önemlidir (Ahmed vd. 2022). DARPA, daha iyi performans göstermesine rağmen derin öğrenme yöntemlerinin en az açıklanabilir (Carabantes, 2020) YZ olduğunu ifade etmektedir. Arrieta ve diğerleri (2020) sorumlu yapay zekânın açıklanabilirliğe ilave olarak “adalet”, “hesap verebilirlik” ve “gizlilik” konularını da içerdiğini ifade ederken; Meske ve diğerleri (2022) açıklanabilirliğin diğerleri için bir ön koşul olduğunu öne sürmektedir.

Yönetişim ve Katılımcılık

Sorumlu YZ alanında bir diğer önemli bileşen de yönetim (governance) olmuştur. Bu bağlamda paydaşların katılımı oldukça önemlidir. Alan uzmanları/model kullanıcıları ile veri bilimcileri/geliştiricileri/ürün sahipleri paydaşlar arasındadır. Diğer paydaşlar ise model kararlarından etkilenen kullanıcıların yanı sıra yöneticiler/yürütme kurulu üyeleri ve düzenleyici kurumlardır (Arrieta ve diğerleri, 2020). “Katılım (inclusion)” ve “çeşitlilik (diversity)” yapay zekâ gelişiminin merkezinde yer alan geniş bir toplumsal zorluk halini almıştır. Söz konusu çeşitlilik sadece cinsiyet değil, kültürel ve etnik anlamda da sağlanmalıdır. Bilişsel çeşitliliğin daha iyi karar almaya katkıda bulunduğuna ilişkin kanıtlar mevcuttur. Yapay zekânın artık bir mühendislik disiplini olarak görülemeyeceğini ifade eden Dignum (2022), YZ geliştirme takımlarına sosyal bilimci ve felsefecileri de dahi etmenin gerekliliğini ifade ederek çok disiplinli ve çok paydaşlı bir katılım olması gerektiğine vurgu yapmaktadır. Sorumlu YZ için yönetim bağlamındaki girişimlerden bazıları AB tarafından güvenilir YZ için hazırlanan kılavuz (Guidelines for Trustworthy AI), IEEE tarafından hazırlanan EAD (Ethically Aligned Design) raporu, Asilomar prensipleri, Barselona ve Montreal Deklerasyonları, Japanese Society for Artificial Intelligence tarafından ortaya konan etik kılavuz şeklinde sıralanabilir. Tüm bu girişimlerin ortak noktası ise insan refahını merkeze almasıdır (Dignum, 2020).

Diğer yandan şirketlerde Yapay Zekâ Etiği Uzmanlığı (AI Ethicist), baş yapay zekâ etik sorumlusu (chief AI ethics officer) gibi kadrolar oluşmaya başlamıştır. Ayrıca YZ etik kurulları veya danışma panelleri de kurularak farklı disiplinleri dâhil etme ve farklı paydaşların açık katılımlarının sağlanması amaçlanmaktadır (Dignum, 2022).

Riskler

Etik çözümler tasarlama kapsamında tasarım mühendislerinin yapay zekâ sistemleri geliştirirken yapay aptallık (artificial stupidity), ırkçı robotlar, veri ve siber güvenlik gibi olası etik zorlukların farkında olması gereğine vurgu yapılırken, farklı disiplinlerden oluşan sorumlu yenilikçilik takımlarının oluşturulması gereğine dikkat çekilmektedir.

Sorumlu yapay zekâ uygulamaları için hayati öneme sahip eğitim ve öğretim kapsamında eğitim programları geliştirilmesinin yanı sıra paydaşlarla işbirliği yapma vurgulanırken, başarılı sorumlu yapay zekâ uygulamaları için tasarım, uygulama ve değerlendirme aşamalarında güvenlik riskleri (siber saldırı riskleri, gizlilik riskleri ve açık kaynaklı yazılım riski), ekonomik riskler (örneğin iş kaybı riskleri) ve performans riskleri (örneğin

hata ve önyargı riski, kara kutu riski ve açıklanabilirlik riski) gibi birtakım risklerin dikkate alınarak en küçüklenmesine yönelik çalışmaların yapılması gerektiği ifade edilmektedir (Wang ve diğerleri 2020).

Gupta ve diğerleri (2023) tarafından dijital sağlık alanında gerçekleştirilen çalışmada sorumlu YZ modelinde performans, güvenlik, kontrol, ekonomik, toplumsal ve etik olmak üzere altı tip riskten söz edilmektedir. Dijital sağlık alanında gerçekleştirmiş oldukları analiz sonucunda performans, kontrol ve güvenlik ile ilişkili riskleri kontrol etmek ile sorumlu YZ geliştirmek arasında istatistiksel olarak anlamlı bir ilişki tespit edilemezken; ekonomik, toplumsal ve etik ile ilişkili riskleri kontrol etmenin sorumlu YZ geliştirmek üzerinde pozitif etkisi olduğu sonucuna varılmıştır. Bu sonuçlar uygulayıcılar açısından önemli bir katkı sağlamakta ve odak noktasını belirlemektedir.

Kültürel Farklılıklar

Dignum'un (2022) dikkat çektiği diğer bir konu da kültürdür. Yapay zekânın ulus ötesi niteliği göz önüne alındığında, yapay zekânın dünyanın çeşitli bölgelerinde toplum üzerinde nasıl bir etki yaratabileceği veya toplum tarafından nasıl kabul görebileceği konularının da ele alınmasının zorunlu olduğunu ifade etmektedir. MIT tarafından gerçekleştirilen "Moral Machine" projesinin sonuçları da söz konusu kültürel farklılığı gözler önüne sermektedir (Canbay ve Demircioğlu, 2021).

Engeller

Yazından hareketle sorumlu yapay zekânın önündeki engellerin teknoloji (veri kalitesi, adalet, gizlilik ve güvenlik, şeffaflık ve açıklanabilirlik), örgüt (örgüt kültürü, sorumluluk ve hesap verebilirlik, yetenekli çalışanlar), çevre (etik konular, yasal konular, direnç, güven ve algı) olduğunu ifade eden Merhi (2023) ayrıca Rakova ve arkadaşları tarafından 2021 yılında yapılan yarı yapılandırılmış mülakatlar neticesinde sorumlu YZ girişimlerinin önündeki en önemli engelin örgütsel kültür olduğunu ve değişimin başarılı bir şekilde gerçekleştirilmesi için gerekli teknik ve analitik yetenekleri kazanması için çalışanların eğitilmesinin önemine vurgu yapmaktadır. Ayrıca çok ölçütlü karar verme tekniklerinden biri olan Analitik Hiyerarşi Süreci (Analytical Hierarchy Process- AHP) aracılığıyla söz konusu engelleri yedi uzmandan aldığı ikili karşılaştırmalara dayanarak önceliklendirmiştir. Sonuç olarak toplam on bir engel arasında en önemli engel "veri kalitesi" olarak ortaya çıkarken, en önemli engel sınıfı "teknoloji" olarak belirlenmiştir.

Örgütsel kategoride en önemli engelin veri ile başa çıkabilen yetenekli çalışan olduğunu ifade eden Merhi (2023), çalışanların veri analitiği, modelleme ve algoritmalar gibi konularda eğitim alarak sertifikalandırılmasını önermektedir. Diğer taraftan direncin sadece çalışanlardan değil, diğer paydaşlardan da gelebileceği Wamba ve Queiroz tarafından 2021 yılında sağlık alanında yapılan çalışma sonucunda hastaların da kültürel olarak direnç gösterebileceği şeklinde ortaya konmuştur. Söz konusu dirençlerin endüstriler, ülkeler, kültürler ve sektörler arasında farklılık gösterebileceği de araştırmacıların dikkat etmeleri gereken diğer bir konudur (aktaran Merhi, 2023).

Sorumlu Yapay Zekâ Konusunda Yazındaki Bazı Çalışmalar

Sorumlu yapay zekâ (Responsible Artificial Intelligence-RAI) konusunda yazın incelendiğinde son yıllarda farklı disiplinlerde konu ilgili makale, bildiri ve kitap/bölümü yazıldığı görülmektedir. Bu çalışma kapsamında sistematik bir yazın taraması yapılmamıştır. Scholar veri tabanında "responsible artificial intelligence" anahtar kelimesi ile yapılan sorgu neticesinden elde edilen çalışmalardan bazılarının örnek olması açısından yer verilmiştir. Bu çalışmalar eskiden yeniye olacak şekilde sıralanmıştır. Böylelikle konunun gelişiminin takip edilmesi de sağlanmış olacaktır.

Sorumlu yapay zekâ konusunda en çok atıf verilen akademisyenlerden biri olan Dignum, 2017 yılında yazmış olduğu makalesinde tasarım yöntemlerinde etik prensiplerin dâhil edilmesi ve sosyal endişelerin cevaplanması gerektiğine dikkat çekmektedir. Bu bağlamda da hesap verilebilirlik-sorumluluk ve şeffaflıktan (accountability, responsibility, transparency-ART) oluşan tasarım prensiplerini önermiştir.

Wang ve diğerlerinin (2020) çalışmasını diğer çalışmalardan ayıran yönü, sorumlu yapay zekânın potansiyel faydalarını ortaya çıkarmak amacıyla değişik endüstrilerden on sorumlu yapay zekâ uygulamasını seçmesidir. Bu çalışmanın sonucunda “yönetişim”, “etik tasarım çözümleri”, “risk kontrolü” ve “eğitim” olmak üzere dört sorumlu yapay zekâ uygulaması tanımlanmış olup; sorumlu yapay zekâ uygulamaları kullanmak isteyen şirketlere de tavsiyeler verilmektedir. Arrieta ve diğerleri (2020) makine öğrenmesi alanında açıklanabilir yapay zekâ üzerine bir yazın taraması yaptıkları çalışmanın bir bölümünde sorumlu yapay zekâyâ giden yoldaki fırsat ve zorlukları tanımlamışlardır. Aynı yıl yayımlanan “The Oxford Handbook of Ethics of AI” kitabında da “Sorumluluk ve Yapay Zekâ” adlı bir bölüm ile katkı sağlayan Dignum (2020), etik ve YZ arasındaki ilişkinin üç düzeyine dikkat çekmektedir: “tasarım ile etik” (ethics by design), “tasarımda etik” (ethics in design) ve “tasarım için etik” (ethics for design).

Sio ve Mecacci (2021) ise araştırmalarında ahlaki ve hukuki felsefe ve teknoloji etiği yazınına dayanarak yapay zekâ ile ilişkili “sorumluluk boşluğu” kavramına yer vermektedir. Yapay zekâ alanında “etik” konusunun popüler olmasına rağmen Afrika’nın akademik tartışmaların radarında olmadığını ifade eden Coeckelbergh (2022), “Responsible AI in Africa” adlı kitabın önsözünde sağlık, tarım ve eğitim gibi alanlarda Afrika’da da yapay zekânın kullanıldığını ifade etmiş, tıpkı batılı ülkelerde olduğu gibi yanlılık ve ayrımcılık, iş hayatının geleceği, iklim değişikliğinin yanı sıra “sorumluluk” konusunun riskler arasında olduğunu ifade etmiştir. Batı ve Afrika arasında kültürel farklılıklara dikkat çeken Coeckelbergh (2022), belki de Afrika etik anlayışının bireyci ve yeteri kadar sosyal olmadıkları için aileye ve yaşlılara yeteri kadar saygı duymayan Batılı etik anlayışının yenilenmesinde bir kaynak olabileceğini ifade etmektedir. Bu kitapta genel anlamda sorumlu yapay zekâ konusuna yer verilmekle birlikte, özel olarak Kenya’da işbirlikçi robotlarla (co-bot) çalışma konusunda ortaya çıkan etik problemler, Nijerya’da sohbet botları (chatbots) kullanımı gibi özel konular da incelenmektedir. Dignum bu kitabın son bölümünde sorumlu yapay zekâ konusunda tavsiyeler ve öğrenilen dersler konusuna yer vermiştir. Bu bağlamda 17 adet Sürdürülebilir Kalkınma Amaçlarına ulaşılmasına katkı sağlama potansiyeli olan yapay zekanın sorumlu bir şekilde geliştirilmesinin yapay zeka alanındaki araştırma ve uygulamaların ana konusu olması gerektiği vurgulanmaktadır. Altı yüzden fazla politika önerisi, strateji raporu vb. yazıldığına dikkat çeken Dignum (2022), beş temel etik prensip konusunda birleştiğini ifade etmektedir: “Şeffaflık”, “Adalet ve Adil Olma”, “Zarar Vermeme”, “Sorumluluk” ve “Gizlilik”.

Gupta ve diğerleri (2023) çalışmalarında dijital sağlık alanındaki yapay zekâ risklerinin sorumlu yapay zekâ ile ilişkili olup olmadığını incelemek amacıyla “Sorumlu YZ uygulamaları nelerdir?” ve “Sorumlu YZ geliştirilmesi ile ilgili algılanan riskler nelerdir?” şeklinde iki araştırma sorusunu cevaplamaya çalışmışlardır. Sağlık alanında diğer bir çalışma da aynı yıl Lukkien (2023) tarafından gerçekleştirilmiş olup, yaşlılar için uzun dönemli bakım için kullanılan zeki teknolojilerin ana bileşeni olarak yapay zekânın önemine yer verilmiş; bu teknolojilerin tasarımı ve uygulanmasının nasıl sorumlu bir şekilde yapılacağı sorumlu yenilikçilik (Responsible Innovation-RI) olarak ifade edilmiştir. Herrmann (2023)’in çalışmasında da sorumlu yapay zekâ, sorumlu yenilikçilik konusu ile birlikte ele alınmıştır. Sorumlu yapay zekâ konusundaki yayınların, daha eski bir geçmişi olan sorumlu yenilikçilik konusundaki yayınlardan üç kat fazla olduğuna dikkat çekilen makale kapsamında önerilen sistematik bilim haritalama (systematic science mapping) adlı yeni yöntem sonucunda ortaya çıkan tematik alanlar ise “etik”, “yönetişim”, “paydaş katılımı” ve “sürdürülebilirlik” olmuştur.

Diğer taraftan Merhi (2023), son yıllarda sorumlu yapay zekâ konusunda artan ilgi nedeniyle yazın taramalarının gerçekleştirildiğini ancak hiçbir çalışmada sorumlu YZ önündeki önemli engellerin ortaya konularak değerlendirilmediğini öne sürmüştür. Buradan yola çıkarak sorumlu yapay zekâ alanındaki engelleri “Teknoloji”, “Örgüt” ve “Çevre” olmak üzere üç sınıfa ayırmıştır. (Bu makalenin “sorumlu yapay zekâ” başlığı altında “engeller” bölümünde çalışmanın detayları yer almaktadır.)

Sonuç

Son zamanlarda insanlar yapay zekânın yaptıkları karşısında şaşkınlıklarını gizleyemez hale gelmiştir. Belki de bu makalenin okunduğu zamanlarda daha neler yapabilir hale gelecektir. Tam da bu noktada sorumlu yapay zekâ konusunda öncü araştırmacılardan biri olan Dignum (2022), yapay zekânın bir sihir olmadığını, tüm sorunlarımızı çözemeyeceğini ve doğal kaynakları kullanmadan ve milyonlarca insanın çalışması

olmadan var olamayacağını ifade etmektedir. Kate Crawford tarafından 2021 yılında kaleme alınan “Yapay Zekâ Atlası (The Atlas of AI)” adlı kitapta da yapay zekânın başarısının saklı kalan tarafına dikkat çekilmekte ve Chaplin’in “Modern Zamanlar” adlı eseri hatırlatılarak, insanların veri etiketleme çiftliklerinde düşük ücretlerle veri sınıflandırma yapacağından bahsedilmektedir (aktaran Dignum, 2022).

Merhi (2023) sorumlu YZ alanındaki çalışmaların artmakla birlikte yeterli olmadığını ve araştırmacıların sorumlu YZ keşfinin erken aşamalarında olduğunu öne sürmektedir. Bu bağlamda gelecekte konu ile ilgili hem kuramsal hem de uygulamalı çalışmaların gerçekleştirilmesi faydalı olacaktır. Sorumlu YZ alanında yazındaki boşluklardan hareketle gelecek çalışmalarda yasal konuların incelenmesi (Merhi, 2023), farklı ülke ve kültürlerden veri toplanarak çalışmalar gerçekleştirilmesi (Gupta ve diğerleri 2023), hesap verilebilirlik ve sorumluluk engellerinin önem derecelerinin (çalışmada on bir engel arasında onuncu sırada yer almıştır) tekrar değerlendirilmesi (Merhi, 2023) önerilmektedir. Bu çalışma kapsamında sadece makale, kitap ve bildiri gibi akademik çalışmalara yer verilmiştir. Yabancı çalışma raporları (white paper) çalışmaya dâhil edilmemiştir. Gelecek çalışmalarda bu raporlar da dâhil edilerek çalışma geliştirilebilir. Ayrıca yazın biraz daha geliştiğinde WoS gibi veri tabanları kullanılarak sistematik bir yazın taraması gerçekleştirilerek, bibliyometrik ve sosyal ağ analizleri ile de geliştirilebilir.

Sonuç olarak, yapay zekâ insanlar için zararlı ve sıkıcı işlerden kurtulmak, kaynakları daha etkin kullanmak vb. için bir yol olabilirken, diğer taraftan insanlık için bir tehdit olarak algılanabilmektedir. Burada yapay zekâyı iyi yönde mi, kötü yönde mi kullanmak konusunda son kararı yine “insan” verecektir. Bu da “insan”ın her zaman olduğu gibi önemini koruyacağını göstermektedir. Bu bağlamda Mugurusu ve Oluca (2021)’nin da ifade ettiği gibi insan ve yapay zekâ birbirlerini “tamamlayıcı” olarak görülmelidir.

Kaynakça

- Ahmed, I., Jeon, G., ve Piccialli, F. (2022). From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where. *Transactions on Industrial Informatics*, 18(8).
- Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., ve Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Information Fusion*, 58, 82-115.
- Atiker, E. Ş. (2024) Güzel Sanatlar Ekseninde Sorumlu Yapay Zekâ: Potansiyel Riskler ve Etik Boyutlar, *Reflektif Journal of Social Sciences*, 5(1) DOI: 10.47613/reflektif.2024.149
- Avundukluoğlu, P. (2023) SAI20 2023 Gündemi: Mavi Ekonomi Ve Sorumlu Yapay Zekâ, *Sayıştay Dergisi*, 34 (128)
- Brundtland, G.H. 1987. “Report of the world commission on environment and development: Our common Future”
- Canbay, P. ve Demircioğlu, Z. (2021). Endüstri 5.0’a Doğru: Zeki Otonom Sistemlerde Etik ve Ahlaki Sorumluluklar. *Academic Journal of Information Technology*, 12 (45).
- Carabantes, M. (2020). Black-box Artificial Intelligence: An Epistemological and Critical Analysis. *AI & Society*, 35, 309–317.
- Coeckelbergh, M. (2022) Foreword, İçinde D.O. Eke · K. Wakunuma & S. Akintoye (Editörler) *Responsible AI in Africa Challenges and Opportunities*(s. v-x). Switzerland: Plagrove McMillan
- DARPA (Defense Advanced Research Projects Agency) (2016). Explainable Artificial Intelligence (XAI). DARPA-BAA-16-53. <https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf>, Erişim Tarihi: 24.08.2022.

- Deniz, N. (2022). Lojistikte Açıklanabilir Yapay Zekâ. İçinde E. Gelmez (Editör), *Lojistikte Güncel Yaklaşımlar*, Ankara: Nobel.
- Deniz, N. Ve Büyük, K. (2023). İkiz Dönüşüm: Sürdürülebilir ve Dijital Dönüşüm, *Dijital Teknolojiler ve Eğitim Dergisi*, 2(1) 57-70.
- Deliloğlu, S., & Çakmak Pehlivanlı, A. (2021). Hibrit Açıklanabilir Yapay Zekâ Tasarımı ve LIME Uygulaması. *Avrupa Bilim ve Teknoloji Dergisi*, 27, 228-236.
- Dignum, V. (2017). Responsible Artificial Intelligence: Designing AI For Human Values, *ITU Journal: ICT Discoveries*, 1.
- Dignum, V. (2020). Responsibility and Artificial Intelligence, İçinde M. D. Dubber F.Pasquale & S. Das (Editörler), *The Oxford Handbook of Ethics of AI* (s 217-219). Oxford: Oxford University Press
- Dignum, V. (2022). Responsible Artificial Intelligence: Recommendations and Lessons Learned İçinde D.O. Eke · K. Wakunuma & S. Akintoye (Editörler) *Responsible AI in Africa Challenges and Opportunities*(s. 195-215). Switzerland: Plagrove McMillan
- Gupta, S. Kamboj S. Ve Bag S. (2023). Role of Risks in the Development of Responsible Artificial Intelligence in the Digital Healthcare Domain, *Information Systems Frontiers*, 25:2257–2274
- Bozkurt Gümrükçüoğlu Y., Yakacak G. A. (2023). Yapay zekânın işe alım süreçlerinde kullanımı ve algoritmik ayrımcılık, *Ankara Üni. Hukuk Fak. Dergisi*, 72 (4),1701-1757.
- Herrmann H. (2023). What’s next for responsible artificial intelligence: a way forward through responsible innovation *Heliyon*, 9, e14379
- IBM, <https://www.ibm.com/tr-tr/watson/explainable-ai>, Erişim Tarihi: 19.07.2022.
- Meske, C., Bunde, E., Schneider, J., ve Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 39(1), 53-63.
- Mugurusi, G. & Oluca, P. N. (2021). Towards Explainable Artificial Intelligence (XAI) in Supply Chain Management: A Typology and Research Agenda. *IFIP Advances in Information and Communication Technology*, 633.
- Samek, W., Wiegand, T., ve Müller, K.R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. <http://arxiv.org/abs/1708.08296v1>.
- Sio, F. S. · Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them *Philosophy & Technology*, 34:1057–1084
- Elkington, J. (1994). “Towards the sustainable corporation: win-win-win business strategies for sustainable development”. *California Management Review*. 36(2), 90-100.
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdil, W., Vidal, M. E. vd. (2020). “Bias in data-driven artificial intelligence systems- An introductory survey”, *WIRES Data Mining Knowledge and Discovery*, 10, 1356.
- Ulusal Yapay Zekâ Stratejisi, <https://cbddo.gov.tr/SharedFolderServer/Genel/File/TR-UlusalYZStratejisi2021-2025.pdf> Son Erişim Tarihi: 7 Aralık 2023.
- Wang, Y., Xiong, M. ve Olya, H. (2020) Toward an understanding of responsible artificial intelligence practices. In: Bui, T.X., (ed.) Proceedings of the 53rd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences (HICSS 2020), 07-10 Jan 2020, Maui, Hawaii, USA. Hawaii International Conference on System Sciences (HICSS) , s. 4962-4971.