



Endüstri 5.0'a Doğru: Zeki Otonom Sistemlerde Etik ve Ahlaki Sorumluluklar

Pelin CANBAY, Kahramanmaraş Sütçü İmam Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, pelincanbay@ksu.edu.tr, 0000-0002-8067-3365

Zübeyde DEMİRCİOĞLU, İstanbul Medeniyet Üniversitesi, Edebiyat Fakültesi, Sosyoloji Bölümü, zubeyde.demircioglu@medeniyet.edu.tr, 0000-0002-8749-006X

ÖZ

Endüstri 4.0 ile makinelerin birbiriyle haberleştiği bir sürece girmiş bulunmaktayız. Bu süreç, makinelerin otonom kararlar vereceği ve bu kararlar ile insanların hayatına önemli etkilerde bulunacağı Endüstri 5.0 döneminin de ilk adımlarını oluşturmaktadır. Endüstride meydana gelen bu gelişmelerin temelinde yapay zekâ algoritmaları ile birçok alanda elde edilen yüksek başarılar yer almaktadır. Geliştikçe başarısı ve karmaşıklığı artarken anlaşılabilirliği azalan yapay zekâ temelli sistemlerin insanları etkileyen önemli kararların alınmasında kullanılacak olması birçok şüpheyi de beraberinde getirmektedir. Bu şüpheler sadece son kullanıcı tarafında değil aynı zamanda sosyal ve ekonomik dönüşümlerle de ilgilidir. Endüstri 5.0'ın önündeki engellerin başında geliştirilmesi planlanan zeki topluluğa karşı toplumsal önyargılar bulunmaktadır. Bu sebeple, zeki ve otonom makinelerin ne tür etik ve ahlaki sorumlulukların bilincinde makineler olduğunun topluma açıklanabilir olması gerekmektedir. Bu çalışmada, hangi durumlarda yapay zekâ'nın etik ve ahlaki değerlere sahip olması gerektiği ve bu değerleri ne şekilde öğrenebileceği araştırılmış, teknoloji ve sosyoloji bağlamında mevcut çalışmalar derlenerek zeki otonom sistemler için sağlam bir etik zeminin olası toplumsal ve teknolojik etkileri değerlendirilmiştir.

Anahtar Kelimeler : Yapay Zekâ, Etik, Otonom Makineler, Endüstri

Toward Industry 5.0: Ethical And Moral Responsibilities in Intelligent Autonomous Systems

ABSTRACT

By Industry 4.0, we have entered a period where machines communicate with each other. This period constitutes the first steps of the Industry 5.0 era, where machines will make autonomous decisions and these decisions will have significant effects on people's lives. High achievements obtained in many areas with Artificial Intelligence algorithms are the basis of these developments in the industry. The fact that intelligent systems, whose intelligibility decrease as their success and complexity



increase, will be used in making important decisions that affect people has brought many doubts. These doubts are not only related to the end-user side but also social and economic transformations. The main obstacles to Industry 5.0 are social prejudices against the intelligent community that is planned to be developed. For this reason, it must be explained to the public about what kind of ethical and moral responsibilities intelligent and autonomous machines have. In this study, it was investigated in which situations artificial intelligence should have ethical and moral values and how it can learn these values. In addition, the possible social and technological implications of a trustworthy ethical basis for intelligent autonomous systems have been evaluated by compiling existing studies in the context of technology and sociology.

Keywords : *Artificial Intelligence, Ethics, Autonomous Machines, Industry*

EXTENDED ABSTRACT

Considering the industrial revolutions that have occurred, it is evident that social concerns are justified. These cogitations direct societies to take measures against the possible harmful consequences of the developing industrial revolution. Presently, the main concerns are about what the ethical and moral values of the intelligent autonomous systems that will enter our lives with Industry 5.0 are and how they will learn these values. In this study, the technological and sociological research conducted to eliminate the concerns about the reliability between humans and machines were examined, and the possible social and technological effects of a strong ethical basis were evaluated.

As they have in the past, accidents related to industrial products will continue to occur, as they presently do. While these products have presently been utilized under human accountability, with Industry 5.0, responsibility is transferred to machines. Some of the ethical responsibilities that these products should have to be used in decision-making about human societies, or to be accepted by the society, are as follows:

- **Privacy:** Necessary precautions must be taken for personal or corporate privacy violations that may occur during the production and use of intelligent autonomous systems.
- **Security:** Two situations are important in terms of security—the security of systems in terms of software and hardware and the security of data used in the system. For these two situations, systems should contain precautions against possible security weaknesses.
- **Safety:** In terms of reliability, these systems should be transparent, accountable, explainable, impartial, free of bias, and non-manipulable.

To be accepted by society, another facet that must be included in intelligent autonomous systems is moral responsibility. It is necessary to expand the concept of moral responsibility—undertaking the results of an action done with free will, knowingly and willingly—to umbrella Industry 5.0 machines. Autonomous Moral Agents (AMA) must be equipped to apply reason to the moral and social consequences of their actions.

It is obvious that there are ethical problems with intelligent autonomous systems that must be resolved but resolution is still controversial. In machines, it is not easy to solve problems that have not yet been solved in human societies, such as the problem of prejudice. Studies conducted to eliminate the concerns about these systems will be beneficial, not only regarding these systems, but also in reducing the undesirable consequences of human weaknesses. Conversely, it is still doubtful whether trust in these systems can be fully secured. For this reason, the use of systems in which a human/machine balance can be established to make decisions regarding the welfare of humans, animals, organizations, and communities will increase society's confidence in these systems.

1. GİRİŞ

Sanayi devrimleri insanlık tarihinin dönüm noktalarıdır. Toplumların hayatını bu derece etkileyen dönüşümler toplumsal kaygı ve korkuları da beraberinde getirmiştir (Narin, 2018). Teknolojinin hızla gelişiyor olması, sanayi devrimlerinin de hızlanmasına olanak sağlamıştır. Toplumlar Endüstri 4.0'ın ilk dönüşümlerine henüz ayak uydurmaya çalışırken, diğer taraftan Endüstri 5.0 beklentilerinin karşılanması, endişelerinin ve korkularının giderilmesi için çalışmalara da başlamıştır (Nahavandi, 2019; Yücebalkan, 2020). Önceki sanayi devrimlerinden farklı olarak Endüstri 5.0'ın, sadece mevcut toplumsal yapıyı etkilemekle kalmayarak Toplum 5.0 ya da Süper Akıllı Toplum olarak adlandırılan yeni bir toplumsal evreye geçilmesine neden olacağı düşünülmektedir (Skobelev & Borovik, 2017). Önceki sanayi devrimleri ile Endüstri 5.0'ın farkını daha belirgin gösterebilmek amacı ile; sanayi devrimlerini başlatan itici güçler, devrimlerin gelişimini ve büyümesini sağlayan teknolojik destekler ve bu devrimlerin barındırdıkları endişeler farklı kaynaklardan (Nahavandi, 2019; Ritzer, 2014; Demir et al., 2019; Marshall, 2015; Öztemel & Gürsev, 2020; Yücebalkan, 2020; Macionis, 2015; Ansal & Yıldırım, 2018; Narin, 2018) derlenerek Şekil 1'de sunulmuştur.

	İtici Güç	Gelişim	Endişe
Endüstri 1.0	<input type="checkbox"/> Mekanik Enerji	<input type="checkbox"/> Su ve Buhar Makineleri	<input type="checkbox"/> İşçi - burjuva sınıfı ayrımı
Endüstri 2.0	<input type="checkbox"/> Elektrik Enerjisi	<input type="checkbox"/> Seri Üretim ve Montaj Hatları	<input type="checkbox"/> Toplumsal bağların azalması
Endüstri 3.0	<input type="checkbox"/> Otomasyon	<input type="checkbox"/> Bilgisayar ve Elektronik Cihazlar	<input type="checkbox"/> İşçi sınıfının vasıfsızlaşması
Endüstri 4.0	<input type="checkbox"/> Dijital Çağ, Siber Fiziksel Sistemler	<input type="checkbox"/> Nesnelerin İnterneti, Büyük Veri, Bulut Bilişim	<input type="checkbox"/> İstihdamın azalması
Endüstri 5.0	<input type="checkbox"/> Süper Akıllı Toplum	<input type="checkbox"/> İnsan - Robot İşbirliği	<input type="checkbox"/> Toplum 5.0'a güvensizlik

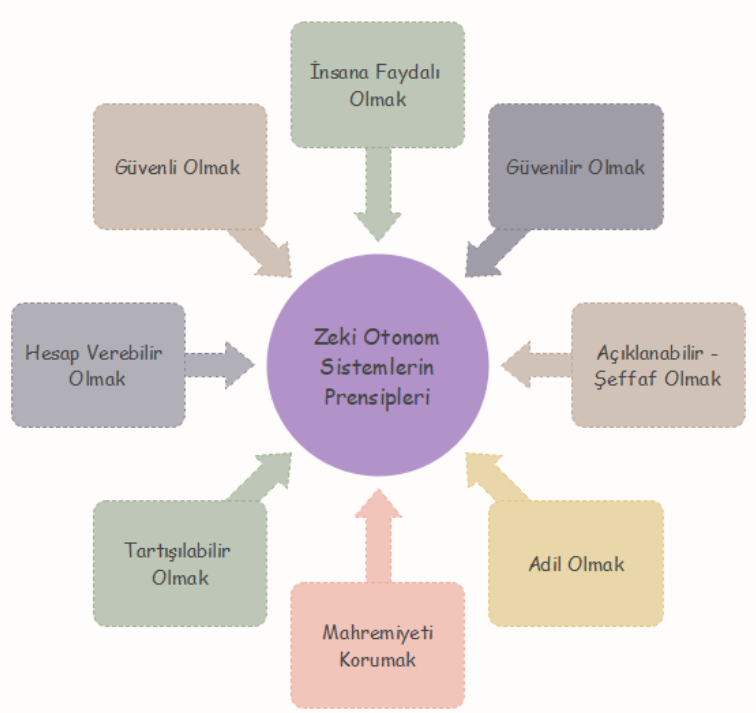
Şekil 1: Endüstri Devrimlerinin İtici Güçleri, Gelişime Katkıları ve Barındırdıkları Endişeler

Günümüze kadar gerçekleşen sanayi devrimlerine bakıldığında haklı endişeler barındırdığı görülebilmektedir. Endişelerin var olması, bir teknolojinin gelişmesini veya bir sanayi devriminin gerçekleşmesini engellemekten ziyade toplumların bu endişelere karşı tedbir almasına vesile olmaktadır. Bu nedenle çalışmamız kapsamında Endüstri 5.0 ile hayatımıza girmesi beklenen zeki otonom (insandan bağımsız hareket eden) sistemlerin, toplumlarda kabul görmesi ve daha etkili teknolojiler olması için hangi etik ve ahlaki değerlere sahip olması gerektiği ve bu değerleri ne şekilde öğrenebileceği araştırılmıştır. Ayrıca, teknoloji ve sosyoloji bağlamında konu ile ilgili mevcut çalışmaların incelenmesi ile zeki otonom sistemler için sağlam bir etik zeminin olası toplumsal ve teknolojik etkileri değerlendirilmiştir.

2. ZEKİ OTONOM SİSTEMLERİN PRENSİPLERİ

Endüstri 5.0 ile insan ve yapay zekâ destekli makine topluluklarının bir arada uyum içerisinde mükemmel bir düzende ve insan merkezli çalışması beklenmektedir. Bu beklentilerin karşılanması için öncelikle insani etik ve ahlaki değerlerin bilincinde akıllı makineler geliştirmek gerekir. İnsana özgü bu değer ve sorumlulukların makinelere hangi durumlarda, ne amaçla ve nasıl verileceği günümüz önemli tartışma konularından biridir (Baum, 2020; Müller, 2020; Ryan & Stahl, 2020; Bostrom & Yudkowsky, 2014). Gündelik yaşamda kullanılan dijital sistemlere gömülü yapay zekâ teknolojileri gün geçtikçe yaygınlaşmaktadır. Bugün geline nokta ulaşımından üretime, finanstan eğitime, sağlıktan eğlenceye pek çok alanda zeki sistemler kullanılmaktadır. Zeki sistemlerin çok yüksek

başarılar ile problemleri çözebilir hale gelmiş olması, özellikle yapay zekâ alanında çalışmayan kişiler tarafından bu sistemlerin korkutucu güçlere sahip olabilecek potansiyelde görünmesine de yol açmıştır. Bilimsel görüşleri kabul gören Bill Gates, Elon Musk, Steve Wozniak ve Stephen Hawking gibi birçok kişi yapay zekâ'nın kontrol edilemez, tehlikeli bir haydut haline gelmesinin distopik potansiyeli hakkında görüş bildirmiştir (Cellan-Jones, 2018). Bu tür tehditlerin olasılığı belirsiz olsa da, insan refahına öncelik vermek için hangi etik, yasal ve ahlaki sınırların bu sistemlere karşı yürürlükte olduğuna dair net bilgi olmadan bu sistemlerin kullanımı birçok problemi beraberinde getirecektir. Bu problemlerin çoğu yalnızca bireysel boyutta değil, toplumsal ve ekonomik dönüşümleri de kapsamaktadır (Adamson, Havens, & Chatila, 2019). Endüstri 5.0 ile akıllı makinelerin, özellikle otonom kararlar verecek zeki sistemlerin söz konusu problemler karşısında sahip olması gereken bazı etik prensipler Şekil 2'de gösterilmektedir.



Şekil 2: Zeki Otonom Sistemlerin Sahip Olması Gereken Prensipler

İnsan hayatını büyük oranda etkileyecek ve hatta şekillendirecek olan zeki otonom sistemlere güvenin sağlanması, bu sistemlerin kabul görebilmesi için Şekil 2'deki gibi bir takım prensiplere sahip olması gerekir. Aksi durumda en ufak bir hatada beklenmedik olumsuz sonuçlarla karşılaşılabilir. The Guardian yayınının bildirdiği üzere, sürücüsüz bir arabanın sebep olduğu ilk ölümlü vaka 2016'da gerçekleşmiştir (Topham, 2019). Otopilot durumundaki Tesla aracının sensörlerinin otoyoldan geçen beyaz bir kamyonu fark edememesi üzerine gerçekleşen bu vaka otonom sistemlerin karar mekanizmalarında etik kurallara bağlı tasarımlar yapılması gerekliliğini ortaya koyan önemli bir vakadır. Endüstride ölümlü kazaların varlığı yeni bir durum değildir. Şekil 3'te, 2020 yılında yayınlanan TÜİK

verilerine göre Türkiye’de taşıma kazalarında meydana gelen can kaybı sayıları verilmiştir (“Tablo 5. Ölüm nedeni ve Cinsiyete Göre Ölümler,” n.d.). İlgili kaynakta bildirilen verilere bakıldığında endüstriyel ürünlerin kullanımında kazaların meydana gelme ihtimalinin var olduğu görülebilmektedir. Bu kazalarda karar merci endüstriyel ürünleri kullanan insanlar olduğundan kazanın sorumluluğu da kullanıcılara aittir. Öte yandan otonom sistemlerde karar merci makinenin kendi olacağından, bir makinenin can kaybı ile sonuçlanan bir kararı karşısında kimin sorumlu tutulacağı, makinelerin karar vermede hangi etik ilkeleri ön planda tutacağı önemli tartışma konularıdır. İnsan hayatını bu derece etkileyen endüstriyel ürünlerden kaynaklı olumsuz sonuçları en aza indirmek için bu ürünlerin tasarımında, evrensel, yerel ve bireysel etik değerlere odaklı ilkelerin kullanılmasına ihtiyaç vardır (Adamson, Havens, & Chatila, 2019).



Şekil 3: 2014 – 2019 Yılları Arasında Taşıma Kazalarında Meydana Gelen Can Kayıplarının Yıllara Göre Dağılımı

Yapay zekâ sistemine sahip makinelerin insanı etkileyen durumlarda kullanılabilmesi için öncelikle bazı gereklilikleri yerine getirmesi beklenmektedir. Zeki otonom araçların meydana getirebileceği tehlikelere ilişkin kamu yaklaşımı ve tartışmalar genellikle istenmeyen bir durumun meydana gelmesi halinde maliyeti kimin üstleneceği konusu ve yasal sorumluluklar üzerinedir. Diğer taraftan böylesi istenmeyen durumların meydana gelmesi üretici kurumun itibarını da büyük ölçüde tehlikeye sokacaktır. Bu çalışmada, benzer istenmeyen durumların olasılığını en aza indirmek ve kaçınılmaz olan tehlikeli durumlarda makinelerin nasıl bir yol izlemesi gerektiğini belirlemek için yapay zekâ kullanacak sistemlerin sahip olması gereken bazı temel sorumluluklar ele alınmıştır. Zeki otonom

sistemlerin sahip olması gereken temel sorumluluklara ilişkin literatürde hangi yaklaşımların değerlendirildiği kapsam dahilinde başlıklar altında ele alınmıştır.

3. ZEKİ OTONOM SİSTEMLERİN ETİK SORUMLULUKLARI

Zeki otonom sistemlerin etik ilkelere uygun hareket etmesini sağlamanın zorluğu Asimov'un (Asimov, 1970) üç robot yasasından beri devam etmektedir. Asimov'un yasaları evrensel makine etiği sorununu çözmek için tasarlanmamış olup bir dizi kod yazarak etik makineler yaratmanın ne kadar zor olduğu üzerine kurgulanan bir hikayedir. Bugün gelinen noktada makine etiği alanındaki araştırmacılar, belirli durumlarda yapılacak doğru eylemin ne olacağını bağımsız olarak belirleyebilen sistemlerin tasarlanması ve geliştirilmesine odaklanmıştır (Vakkuri et al, 2020). Gün geçtikçe karmaşıklaşan zeki sistemler özerkleşerek doğrudan insan kontrolü olmaksızın çalışabilecek otonom sistemlere doğru evrilecektir. Bu durum da söz konusu sistemlerin ne kadar etik etki unsuru olduklarını sorgulatmaktadır. Bir otonom makine, eylemlerinin insanlara zarar verebileceği ya da yarar sağlayabileceği ölçüde potansiyel bir etik faildir. Bir dijital saat bile sahibini randevularına zamanında gitmek için teşvik etmek gibi bir role sahipse, etik etki unsuru olarak kabul edilebilir. Ancak tamamen etik bir fail olarak değerlendirilemez. Moor (2009) etik failer olarak makineler için üç temel düzey tanımlar: örtük etik failer, açık etik failer ve tam etik failer (Moor, 2009).

Örtük etik failer, tasarıma gömülü etik yargılara sahiptir. Bunlar genelde güvenlikle (safety & security) ilgili tasarımlardır. Örneğin uçaklar, yere yakın seyretme, başka bir uçakla yaklaşma gibi tehlike durumlarında devreye girecek bir uyarı mekanizmasını içerecek şekilde tasarlanır. Örtük bir etik fail belirli durumlarda otomatik etik tepkiler verir. Örtük etik failer günlük yaşamımızda pek çok örneğini görebileceğimiz tasarımlardır ancak makine etiği için yeterince gelişmiş değillerdir.

Açık etik failer, çeşitli durumlar hakkında etik bilgileri tanımlayıp işleyen ve ne yapılması gerektiği konusunda tespitler yapabilen tasarımlardır. Makinelerin bu türden etik bilgileri kazanabilmelerinin bir yolu, yapay zekâ uygulamaları ile etik yargılar üretmek için programlanmasına dayanır. Alternatif olarak etik davranışlar veya iç görüler öğrenme yoluyla kazanılabilir. Açık etik fail, çeşitli durumlara uyacak biçimde yorumlanmış genel ilke ve etik davranış kurallarına sahip olacaktır.

Tam etik failer de, açık etik failer gibi çeşitli durumlarda etik yargılarda bulunmakla birlikte bilinç, kasıt, özgür irade gibi insanlara atfedilen özelliklere sahiptirler. Makinelerin tam etik failer haline gelip gelmeyeceği tartışmalı bir konudur. Moor'un (2009) belirttiği gibi robot etiğinin güncel hedefi robotların açık etik failer haline gelmesidir. Robotlar giderek daha fazla özerk hale geldikçe daha fazla etik yetenekler kazanmaları gerekecektir.

Bir zeki sistemin başarısı aynı zamanda arka planda kullandığı verilere ve bu verileri işleme yöntemine de bağlıdır. Yani, her zeki sistemin oluşmasında bir veri toplama ve işleme

sürecinden bahsedebiliriz. Bu süreçte, toplanan verilerin işlenmesi ve paylaşımı ile ilgili hem bu sistemlerin hem de sistem kurucuların dikkate alması gereken etik hususlardan bazıları aşağıda detaylandırılmıştır.

3.1. Mahremiyet, Güvenlik ve Güvenilirlik

Zeki sistemlerin oluşmasında kullanılan verilerin, özellikle söz konusu kişisel veya kurumsal veriler ise, hem veri işleme sürecinde hem de sonrasında mahremiyetinin korunuyor olması gerekmektedir. Kişisel, kurumsal veya bir topluluğa bağlı hassas veri ifşasına yol açabilecek sistemler tasarlanırken olası sonuçları önceden hesaplanıp gerekli tedbirlerin alınması gerekir (Canbay & Sağıroğlu, 2020). Diğer taraftan, mahremiyet açısından zeki otonom sistemlerde ele alınması gereken tek sorun toplanan verilerin mahremiyetini sağlamak değil aynı zamanda bu verilerin sahiplerine hangi verilerinin toplandığı ve ne amaçla işlendiği ile ilgili bilginin verilip verilmeyeceği kararı da önemli bir problem olarak görülmektedir (Coeckelbergh, 2019).

Zeki otonom sistemlerde güvenlik iki açıdan ele alınmaktadır. İlk olarak bu sistemlerin kazalar/arızalar karşısında (yazılım ve donanım bakımından) sağlam ve son kullanıcıların güvenliğini ön planda tutacak bilinçte geliştirilmeleri gerekmektedir. Boeing 737 MAX uçaklarının yaptığı kazaların yazılımla ilgili bir arızadan kaynaklandığı ve pilotların uçağın kontrolünü sağlama çabalarına rağmen uçağın gelişmiş otopilot sistemine müdahale edemediği açıklanmıştır (Topham, 2019). Yazılım veya donanım tabanlı arızaların meydana gelmesi durumunda alınabilecek ek güvenlik önlemleri tanımlanmalı, gerekli durumlarda makine yetkilendirmesi güncellenebilmelidir. Uçaklardaki otopilot sisteminden kaynaklı kazaların ardından otomasyon sistemleri hakkındaki tartışmalar daha da büyümüştür (Coeckelbergh, 2020).

Zeki otonom sistemlerinin güvenliği bakımından bir diğer önemli nokta bu sistemlerde kullanılan, saklanan veya işlenen verilerin güvenliğidir. Endüstri 4.0 ile birlikte artık makinelerin belirli ağlar ile veri ve komut alışverişinde bulunduğu bir sürece girilmiştir. Nesnelerin interneti (Internet of Things – IoT) başlığı altında ele alınan bu bağlantılar çeşitli güvenlik zafiyetleri içerebilmektedir (Alaba, Othman, Hashem, & Alotaibi, 2017). Bu bağlantılar üzerinden gelebilecek olası güvenlik tehditleri ve bu sistemlerin barındırdığı zafiyetler hem internet tarafından hem de son düğüm olan makineler tarafından ele alınarak gerekli tedbirler tasarıma dahil edilmelidir. Ağlara bağlı otonom sistemlerin bu ağlar üzerinden gelebilecek kötü amaçlı saldırılara karşı ek bir korunma mekanizmasına sahip olması da gerekmektedir (Atlam & Wills, 2020).

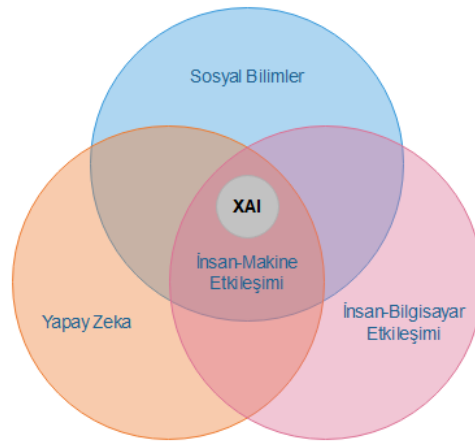
Güvenilirliğin temel ilkeleri arasında ise şeffaflık, hesap verebilirlik, tarafsızlık (olumsuz önyargı eksikliği) ve manipüle edilmemesi bulunur (Adamson et al., 2019). Yapay

zekâ sistemlerinin insanlar ile etkileşiminde güven önemli bir konudur. Bu güvenin değerlendirilebilmesi yapay zekâ'nın arkasındaki yaklaşımın son kullanıcıya ne kadar iyi açıklanabildiği, anlatılabildiği ile ilgilidir. Bu amaç doğrultusunda zeki sistemlerin açıklanabilir olma ihtiyacı doğmuştur. Yapay zekâ'nın açıklanabilir olması amacı ile yapılan çalışmalar (Ribeiro, Singh, & Guestrin, 2016; Samek, Wiegand, & Müller, 2017), sistemlerin işleyişinin anlaşılabilir olmasını hedeflemektedir. Bu çalışmalar hem son kullanıcıdaki şüpheyi ortadan kaldıracak hem de otonom sistemlerin daha güvenle kullanılmasına katkı sağlayacaktır.

3.2. Açıklanabilir Olmak

Ticari faydalar, etik kaygılar veya yasal düzenlemeler bakımından, kullanıcıların yapay zekâ sonuçlarını anlaması, uygun şekilde güvenmesi ve etkili bir şekilde yönetebilmesi için zeki sistemlerde açıklama gereklidir. Avrupa Birliği'nin Genel Veri Koruma Yönetmeliği'nde (GDPR), yapay zekâ tarafından alınan bireysel kararlar hakkında bilgi edinme hakkı ve algoritmik hesap verebilirliğe ilişkin hükümler Açıklama Hakkı (Right to Explanation) olarak ele alınmaktadır (Goodman & Flaxman, 2017). Açıklanabilirlik, bir zeki sistemin toplum tarafından kabul görebilmesi için sahip olması gereken en önemli özelliklerden biridir. Açıklanabilir olmak zeki otonom sistemler için bir gerekliliktir.

Açıklanabilir yapay zekâ (Explainable Artificial Intelligence - XAI) genellikle kendisinin veya başka bir temsilcinin karar vermesinin altında yatan nedenleri ortaya çıkaran açıklayıcı bir temsilci olarak ele alınmaktadır. Ancak XAI çözümünün sadece daha fazla yapay zekâ olmadığı (bir yapay zekâ'yı açıklamak için başka yapay zekâ lar kullanmak), insan-makine etkileşiminin önemli bir parçası olduğu da dikkate alınmalıdır (Miller, 2019). İnsan-makine etkileşimi; yapay zekâ, sosyal bilim, insan – bilgisayar etkileşiminin kavramlarının kesişimi olarak Şekil 4'te gösterilmektedir.



Şekil 4: Açıklanabilir Yapay Zekâ'nın Yeri (Miller, 2019)

Özellikle sistemleri kullanan kişilerin sistem hakkında yeterince bilgi sahibi olmadığı durumlarda beklenmeyen bir durumla karşılaşıldığında kimin sorumlu tutulacağı problemi büyür. Ancak kullanıcıların her zaman her şeyi biliyor olması mümkün değildir ve bu bizi şeffaflık ve açıklanabilirlik sorununa götürür. Sürecin nasıl işlediği çoğu durumda net değildir. Bilim, hesaplama ve mühendislikte, Kara Kutu (Black Box), Gri Kutu (Grey Box) ve Beyaz Kutu (White Box) terimleri, bileşenlerin iç yüzünün açık olma veya bilinme seviyelerine atıfta bulunularak kullanılır (Suman, Mall, Sukumaran, & Satpathy, 2010). Özellikle “kara kutu” olarak adlandırılan, karar zincirinin ya da karar ağacının geriye doğru takip edilemediği sinir ağlarını kullanan makine öğrenmesi sistemlerinde, zeki sistemin kararlarını hangi kriterleri ön planda tutarak verdiği sistemi kuran kişi tarafından bile çoğu zaman bilinmemektedir. Böylesi durumlarda zeki sistem kullanıcıları ve hatta geliştiricilerin yapay zekâ ya duyduğu güven ahlaki açıdan sorunlu görülebilir. Bu etik ve ahlaki bir meseledir zira insanlar kendilerini etkileyen kararların neden ve nasıl alındığını bilme hakkına sahiptir. Bu anlamda açıklanabilirlik ahlaki bir gereklilik olup iki açıdan önemlidir (Coeckelbergh, 2019). Öncelikle sorumluluk açısından önemlidir; bir şeyden sorumlu tutulabilmesi için failin ne yaptığını bilmesi gerekir bu açıdan eylemin açıklanabilir olması şarttır. İkinci olarak failin sorumlu olduğu kişilere karşı cevap verebilirliğini gösterdiği için önemlidir. Dolayısıyla açıklanabilirlik sadece failin bilme sorunu değil, aynı zamanda açıklama yapabilme meselesidir. Fail, karar ve eylemlerinden doğrudan ya da dolaylı etkilenen her tür kategoriye (insan, hayvan, kuruluş, topluluk) karşı sorumludur. Dolayısıyla yapay zekâ ile ilgili problemlerin tüm paydaşlarla ve genel olarak kamuya açık biçimde tartışılması gerektiği söylenebilir.

3.3. Tarafsız Olmak

Rusya’da Voskhod Ajansı tarafından gerçekleştirilen “AI Versus” isimli vaka çalışmasında özdeş iki yapay zekâ sisteminin öğrenme süreçleri boyunca dünya görüşlerinin nasıl farklılaştığı incelenmiştir. Sistemlerden biri Rus devlet televizyon kanalı Russia-1’den, diğeri bağımsız bir Rus TV kanalının içeriklerinden beslenmiştir. 6 aylık süre sonunda bir internet sitesi üzerinden kendilerine sorulan soruları cevaplayan özdeş sistemlerin farklı türden değerlere sahip olduğu ortaya çıkmıştır. Örneğin “Toplumsal cinsiyet eşitliği nedir?” sorusunu devlet televizyonunu izleyen zeki sistem “Amerikan tipi demokratik değerlerden biridir” şeklinde cevap verirken, diğeri zeki sistem “temel şeylerden biridir” cevabını vermiştir (“ai-versus,” n.d.). Dolayısıyla yapay zekâ nın beslendiği verinin sahip olduğu etik değerlerin zeki sistemlere de aktarıldığı görülmektedir.

Önyargı, bir şey ya da kişi hakkında önceden oluşturulmuş kanaat ya da yanlılığı ifade eder. Gerçeklik karşısında sınanmamış, genellikle kusurlu ve esnek olmayan, kalıp yargılara dayanan antipatidir (Marshall, 2005). Önyargı ve ayrımcılık sorunları neredeyse her toplum

ve kültürde görülür ancak konunun yapay zekâ ile ilgili kısmı, bu teknolojilerin önyargı ve ayrımcı kalıpları arttırma ve sürdürme potansiyelinden doğar. Önyargı, veri seçiminden başlayarak eğitim veri setinde, kullanılan algoritalarda ya da uygulama veri seti gibi her aşamada ortaya çıkabilir (Coeckelbergh, 2019). Zeki sistemler tarafından miras alınabilecek en temel insani önyargı kategorileri günümüz için cinsiyet ve ırk önyargıları olarak görülmektedir (Wang & Siau, 2018). Yapay zekâ sistemleri halen insanlar tarafından eğitildiğinden ve insan kaynaklı verilerden faydalandığından insana dair mevcut önyargıların bu sistemler tarafından öğrenilmesi olasılığı yüksektir. Bu nedenle zeki sistemlerin insan önyargısını dışarıda bırakacak biçimde nasıl programlanacağı ve eğitileceği oldukça önemlidir. Zeki sistemlerde yanlılığın ve ön yargının önüne geçebilmek amacı ile önemli çalışmalar yapılmış ve zeki sistemlerin tarafsız bir hale dönüştürülebileceği bu çalışmalarda yapılan deneylerle gösterilmiştir (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016a, 2016b; Caliskan, Bryson, & Narayanan, 2017). İnsanlık ve toplumlar önyargıları hiçbir zaman ortadan kaldıramayabilir. Fakat önyargının yönü, tipi, türü gibi belirleyicilerin tespit edilmesi ve tespit edilen bu yapıların, kullanılacak olan yapay zekâ eğitim veri kümesinden çıkarılması ile en temel anlamda sistem belirlenen önyargılardan büyük ölçüde arınmış olacaktır. Yapay zekâ kullanımı bağlamında teknik olarak belirli önyargıların kısmen kaldırılması mümkündür (Coeckelbergh, 2020).

4. ZEKİ OTONOM SİSTEMLERİN AHLAKİ SORUMLULUKLARI

Temelde insan davranışlarının değerlendirilmesi için kullanılan ahlaki sorumluluk kavramı, özgür iradeyle bilerek ve isteyerek yapılan eylemin sonuçlarının üstlenilmesi anlamına gelmektedir. Eylemin ahlaki sorumluluğu değerlendirilirken üç koşulun yerine getirilmesi beklenir (Noorman, 2018):

1. Fail ile eylemin sonucu arasında nedensel bir bağlantı bulunmalıdır. Bir kişi ancak olayların sonuçları üzerinde kontrolü varsa sonuçlardan sorumlu tutulabilir.
2. Fail eylemlerinin olası sonuçları hakkında bilgi sahibi olmalıdır. Eylemlerinin olumsuz bir sonuca yol açacağını bilmeyen biri suçlanmaktan ziyade mazur görülür (akli yeterlilik).
3. Fail, eylemi gerçekleştirirken özgürce hareket etmeyi seçebiliyor olmalıdır. Yani, eylemleri dışsal olarak belirlenen birisi için ahlaki sorumluluktan söz edilemez.

Ahlaki sorumluluk insanlar için kullanılan bir kavramdır ve bu kavramı yapay varlıkları kapsayacak şekilde genişletmek birçok sorunu beraberinde getirmektedir. Yüzyıllardır filozoflar, teologlar ve sosyal teorisyenler insan davranışını yönlendiren değerler üzerine tartışmaktadır. Günümüzde bu tartışmanın otonom etik failleri de içerecek şekilde genişletilmesi gerekmektedir. Örneğin otonom bir sistem zararlı, yıkıcı ya da yasadışı eylemlerde bulunursa sorumlusunun kim olacağı sorusu önemlidir. Otonom/özerk ahlaki fail (Autonomus Moral Agents - AMA) eylemlerinin ahlaki ve sosyal sonuçları hakkında akıl

yürütme yeteneğine sahip olmalıdır (Allen, Wallach, & Smit, 2006). Tasarımcılarından bağımsız çalışan, gerçek ya da sanal ortamlarda esnek yanıt verecek şekilde programlanmış makinelerle güven duyulması, ancak uygun normları karşılamaları durumunda mümkündür. Bir ekmek kızartma makinesi yangına sebep olduğu için sorumlu tutulamaz ve suçlanamaz ancak özerk bir yapıya sahip otonom sistemler olası zararlı sonuçların bilincinde olarak eylemlerine şekil vermelidir (Allen et al., 2006)

5. ZEKİ OTONOM SİSTEMLERE SORUMLULUKLARIN YÜKLENMESİ

Öncelikle yapay zekâ ile ilgili etik sorunların varlığı ve çözülmesi gerekliliği noktasında genel bir mutabakattan söz edilebilir. Ancak gerekli çözümlerin nasıl yapılacağı kısmı tartışmalıdır. Örneğin, önyargı sorunuyla başa çıkmak kolay değildir. Zira konunun toplumsal ve kültürel bir kökeni vardır. Bu sorunun çözümü için öncelikle toplumsal zihniyetin içine gömülü önyargıya dayalı ayrımcı düşüncelerin ele alınması gerekir.

Veri güvenliği ve gizliliği konusunda GDPR gibi bir takım hukuksal düzenlemeler bulunmaktadır (Goodman & Flaxman, 2017). Ayrıca pek çok ülke yapay zekâ'nın geleceği hakkında ulusal rapor ve stratejiler yayınlamıştır. Bu düzenlemeler ve politika önerileri insan onuru, birey özgürlüğü, demokrasiye saygı, hukukun üstünlüğü, adalet ve vatandaşlık hakları gibi temel haklardan söz eder, ancak aynı zamanda zararsızlık ilkesi ve açıklanabilirlik gibi bir dizi haktan da bahseder (Coeckelbergh, 2019). Dolayısıyla AMA'lar gizliliği gözeten, etik standartları karşılayan, bireysel özgürlüğü koruyan, temel insani haklara saygılı kararlar alabilmelidir. Ancak bu kadar değer duyarlı AMA'lar tasarlamak kolay değildir (Allen et al., 2006).

Etik karar mekanizmalarının zeki sistemlere nasıl entegre edileceği konusunda iki temel yaklaşım bulunmaktadır. Bunlardan ilki, yapay zekâ'nın karşılaşılabileceği senaryoların çözümününün tasarıma dahil edilmesidir. Diğer yaklaşım ise makinelerle doğru ahlaki davranışı öğretmeyi hedefler. Yapay zekâ uygulamalarındaki etik problemlerin çözümünde üç temel modelden söz edilebilir (Wallach & Allen, 2009):

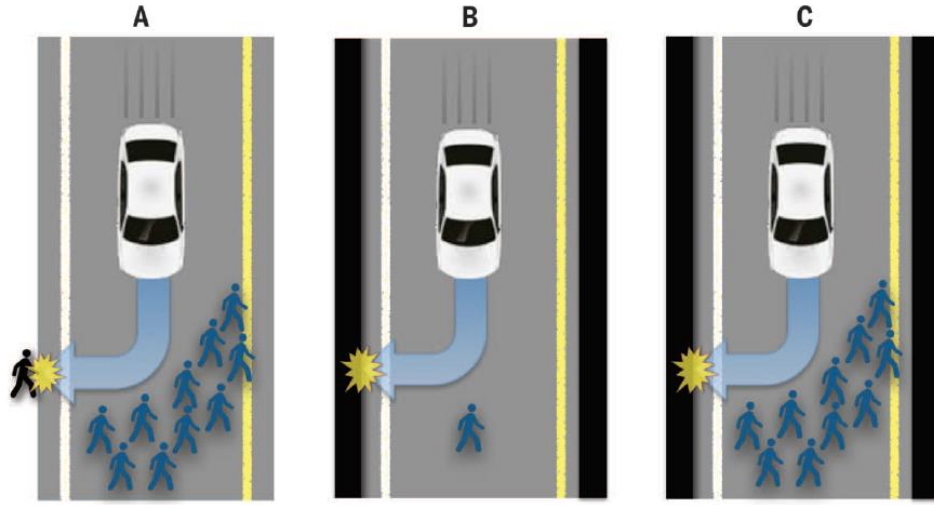
1. Yukarıdan aşağıya modelde, temelde gerekli tüm etik kural setlerinin yapay zekâ uygulamasına yüklenmesi ve uygulamanın eylemlerinde bu kurallara uyması beklenir. Ancak etik kural setlerinin genel olması ve her türden faaliyeti içerecek şekilde ayrıntılandırmanın yapılamaması modele ilişkin en temel problemlerden biridir.
2. Aşağıdan yukarı modelde, yapay zekâ uygulamalarının karşılaştıkları durumlar karşısında etik kuralları öğrenmesi ve bu öğrenmeye dayalı bir karar vermesi beklenir. Ancak öğrenme süreçleri esnasında yaşanabilecek istenmeyen sonuçlar ve sürecin uzun olması bu yaklaşımı uygulanabilir olmaktan çıkarır.

3. Hibrit model, her iki modelden faydalanarak ortaya konulan bir model kullanımı önerir. Dolayısıyla hem yapay zekâ uygulamalarına etik kural setleri yüklenecek hem de makine öğrenme süreci devreye girecektir.

Zeki sistemlerin sorumluluklarının karşılanabilmesi bakımından teknoloji alanında standartların belirlenmesi amacı ile çeşitli çalışmalar yapılmaktadır. Bu çalışmaları yürüten organizasyonların başında IEEE gelmektedir. Zeki sistemlerin tasarımında kullanılmak üzere günümüzde IEEE tarafından bu alana eklenen teknoloji uygulayıcılarının bakış açısı, hem profesyonel teknoloji uzmanlarının mevcut ihtiyaçlarını karşılamakta hem de temel etik hususlara yönetsel ve ticari olmayan bir destek sağlamaktadır (The IEEE Global Initiative, 2017). Yapay zekâ ürünü geliştirme ve uygulama aşamalarında bu ürünlerin öngörülebilir olası etik ihlallerine yönelik alınabilecek birtakım önlemler (Kose, 2020) ve (Demircioğlu & Canbay, 2021) çalışmalarında da ele alınmıştır.

5.1. Ahlaki Makine Çalışmaları

Sürücüsüz araç etiğinin geliştirilmesi konusunda yürütülen araştırmalar, ikilemlerle karşı karşıya kaldıklarında etik açıdan doğru karar verebilecek araçların tasarlanmasına odaklanmaktadır. Etik probleminin kökeni 1967 yılında ahlak felsefecisi Philippa Foot tarafından geliştirilen "Vagon/Tren Problemi"ne dayanır (Foot, 1967). Araştırma temelde etik bir ikilemle karşılaşan insanların nasıl davranacağını belirlemeye çalışır. Bunun için bir problem inşa eder; yaklaşmakta olan bir tren, yoluna aynı hat üzerinden devam ederse tren raylarına bağlı 5 kişinin ölümüne sebep olacak, makas değiştirerek diğer hatta yönelirse 1 kişiyi öldürecektir. Bu durumda nasıl hareket edersiniz? Bu etik problem yıllarca tartışılmış olmakla birlikte bir fikir birliğine ulaşıldığı söylenemez (Yeşilova, 2020). Benzer etik problemler otonom araçların tasarlanmasıyla birlikte tekrar gündeme gelmiştir. Bu tür etik ikilemlerle karşılaşan otonom araçlar nasıl karar vermeli sorusuna cevap bulmak da oldukça güçtür.. Şekil 5'te kaçınılmaz zararlı sonuç içeren üç trafik durumu gösterilmektedir. Şekilde otonom araç verilen seçeneklerden birine karar vermek zorundadır.



Şekil 5: Kaçınılmaz zararlar içeren üç trafik durumu. (A) birkaç yayayı veya yoldan geçen birini öldürmek, (B) bir yayayı veya kendi yolcusunu öldürmek, (C) birkaç yayayı veya kendi yolcusunu öldürmek (Bonnefon, Shariff, & Rahwan, 2016)

Massachusetts Institute of Technology (MIT) 2018 yılında, yapay zekâ etiği üzerine çalışmalar yapan 1 Milyar dolarlık yeni bir işlem merkezini duyurmuştur (Lohr, 2018). MIT tarafından geliştirilen çevrimiçi bir platform olan “Moral Machine” projesi tam da yukarıda verilen etik sorunun çözümüne odaklanmaktadır (“Moral Machine,” n.d.). Araştırma, çeşitli kaza senaryoları sunulan insanların verecekleri cevaplar doğrultusunda genel bir ahlak ve etik anlayışı geliştirmeyi hedeflemektedir. Vagon probleminde benzer şekilde araştırmaya katılanlara dokuz farklı kaza senaryosu verilerek kaza sonucunu belirleyen iki şıktan birini tercih etmeleri beklenmiştir. Senaryolarda; evcil hayvanlarla insanlar, yayalarla aracın yolcuları, yaşlılarla gençler, trafik kurallarına uymayanlarla uyanlar, sağlıklı insanlarla engelliler, erkeklerle kadınlar, yüksek statüde olanlarla düşük statüde olanlar karşı karşıya getirilmiştir. Ayrıca daha fazla hayat kurtarmak ve yoldan çıkmaya karşı rotada kalmak seçenekleri sunulmuştur. Araştırmada 233 ülke ve bölgeden toplanan 40 milyona yakın karar analiz edilerek kapsamlı bir etik algoritmanın geliştirilmesi amaçlanmıştır. Araştırmanın sonuçlarına göre genellenebilecek 3 etik eğilim ortaya çıkmıştır. Katılımcılar; az kişinin öldüğü, hayvanlar yerine insanların, yaşlılar yerine gençlerin hayatta kaldığı senaryoları tercih etmişlerdir. Bununla birlikte tercihlerin kültürlere göre oldukça farklılaştığı gözlemlenmiştir. Doğu (Uzakdoğu ve İslam ülkeleri) – Batı (Kuzey Amerika ve Avrupa) – Güney (Latin Amerika ülkeleri) olarak tanımlanan üç farklı kültürel grupta farklı türden tercihler ortaya çıkmıştır. Bu kümelenme coğrafi ve kültürel yakınlığın benzer türden etik tercihleri ortaya çıkardığını göstermektedir. Her üç kümelenmede de ortak olan yayalara karşı araçtaki yolcuları ve kurallara uyanları koruma eğilimidir. Kümelenmeler en belirgin şekilde bireyci ve toplulukçu kültürler arasındaki farka işaret etmektedir. Daha bireyci olan Batı, az sayıda

kişi yerine çok kişinin kurtarılmasını, yaşlılara saygı kültürünün geliştiği Doğu, gençlerin yaşlılara tercih edilmemesini benimsemiştir (Awad et al., 2018).

6. SONUÇ VE DEĞERLENDİRMELER

Endüstri 4.0 devriminin arkasındaki itici güç olan yapay zekâ teknolojilerinin sağlıktan eğitime, ekonomiden ulaşım, özellikle kaynakların yetersiz olduğu ortamlarda insanların refah seviyesinin geleceği üzerine olumlu bir etkiye sahip olacağı aşikardır. Bu teknolojilerin potansiyelinden tam anlamıyla yararlanabilmek ve olası zararlarını en aza indirebilmek için ahlaki değerler ve etik ilkeler bakımından insanlarla uyumlu olduğundan şüphe duyulmaması gerekir.

Yapay zekâ kaynaklı ortaya çıkabilecek kaza vb. istenmeyen durumlarda sorumluluğun kime ait olacağı ve bu sistemlere ne tür sorumluluklar yükleneceği hakkında net bir sonuca hala varılamamıştır. Yapay zekânın sorumlu tutulması durumunda ne gibi hukuksal mekanizmaların işletileceği, sorumlu tutulmaması durumunda ise sorumluluğun kime yükleneceği konusunda çalışmalar ise günümüzde yoğun bir şekilde sürdürülmektedir. Bu çalışmaların merkezinde zeki sistemlere etik ve ahlaki sorumluluk bilinci yükleme bulunmaktadır, fakat bu toplumsal sorumlulukların makinelere aktarımının kolay bir süreç olmayacağı şu ana kadar yapılan çalışmalarda açık bir şekilde görülmüştür.

Otonom zeki sistemler için etik kurallar getirilmesi gerekliliği yapılan çalışmalar ile de gözler önüne serilmiştir fakat, bu sistemler için evrensel etik kurallara ulaşabilmek bazı durumlarda neredeyse imkansızdır. Bu gibi durumlarda en optimal çözümü sunarak beklenmedik durumlar karşısında tedbir almak, ne yapılması gerektiğini belirlemek ve uygun teknolojiler üretilmesini talep etmek bu alanda çalışan araştırmacıların ve bu teknolojileri kullanacak toplulukların elindedir.

Yapay zekâ temelli otonom sistemlerin olası zararlarını en aza indirmek amacı ile yapılan çalışmalar, sadece bu sistemlerin değil aynı zamanda insan temelli karar mekanizmalarında insan kaynaklı zararları en aza indirmede de kullanılabilir. İnsani önyargı ve zaafiyetlerden arındırılmış, tarafsız değerler ile eğitilmiş otonom bir karar vericiyi kullanmak birçok alanda daha etkili sonuçlara ulaşılmasını sağlayacaktır. Diğer taraftan unutulmamalıdır ki; insan karar vericinin gerçek hayat tecrübesi ile otonom bir makinenin kararı arkasındaki kriterlerin karşılaştırılması birçok durumda mümkün olmayacak ve bu kriterlerin başka istenmeyen sonuçlar ile bağlantısı olup olamayacağı sorumluluğu makinelerle yüklenemeyecektir. Bu gibi durumlar göz önüne alındığında, insan, hayvan, kuruluş ve topluluk ile ilgili kararların alınmasında insan-makine dengesinin kurulabildiği sistemlerin kullanımı, toplumların refah seviyesini daha yükseklerle çıkaracaktır.

KAYNAKLAR

- Adamson, G., Havens, J. C., & Chatila, R. (2019). Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems. *Proceedings of the IEEE*. <https://doi.org/10.1109/JPROC.2018.2884923>
- Ai versus. (t.y.).*The FWA*. 15 Kasım 2020 tarihinde <https://thefwa.com/cases/ai-versus> adresinden erişildi.
- Alaba, F. A., Othman, M., Hashem, I. A. T., & Alotaibi, F. (2017). Internet of Things security: A survey. *Journal of Network and Computer Applications*. <https://doi.org/10.1016/j.jnca.2017.04.002>
- Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*. <https://doi.org/10.1109/MIS.2006.83>
- Ansal, H., Yıldırım, N., (2018). Sanayi 4.0 ve İşgücüne Etkileri: Kadınlar için bir Fırsat Olabilir mi?, *Toplum ve Bilim*, 144., 217-232
- Asimov, I. (1970). w1 I Robot., *Science*.
- Atlam, H. F., Wills, G. B. (2020). Iot Security, Privacy, Safety And Ethics. In *Digital Twin Technologies and Smart Cities* (pp. 123-149). Springer, Cham.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... Rahwan, I. (2018). The Moral Machine experiment. *Nature*. <https://doi.org/10.1038/s41586-018-0637-6>
- Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016a). Debiasing Word Embedding. *30th Conference on Neural Information Processing Systems*.
- Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016b). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*.
- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*. <https://doi.org/10.1126/science.aaf2654>
- Bostrom, N., Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. *The Cambridge Handbook of Artificial Intelligence*, 1, 316-334.
- Baum, S. D. (2020). Social choice ethics in artificial intelligence. *AI & SOCIETY*, 35(1), 165-176.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*. <https://doi.org/10.1126/science.aal4230>
- Canbay, Y., Sağıroğlu, Ş. (2020). Büyük Veri ve Yapay Zekâ Uygulamalarında Mahremiyet ve Etik. *Yapay Zekâ ve Büyük Veri: Teknolojiler, Yaklaşımlar ve Uygulamalar*, 157-186, Nobel.
- Cellan-Jones, R. (2018). *Stephen Hawking - will AI kill or save humankind?* *BBC News*. Retrieved from <https://www.bbc.com/news/technology-37713629>

- Coeckelbergh, M. (2019). Technology Regulation Ethics of artificial intelligence: Some ethical issues and regulatory challenges. *Technology and Regulation*.
- Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-019-00146-8>
- Demir, K. A., Döven, G., Sezen, B. (2019). Industry 5.0 and Human-Robot Co-Working. *Procedia Computer Science*, 158. 688-695.
- Demircioğlu, Z., Canbay, P. (2021). Yapay Zekâ Teknolojilerinin Toplumsal Etkileri ve Etik Gereksinimler., II. *International Ethics Congress* (pp. 195-203).
- Foot, P. (1967). *The problem of abortion and the doctrine of double effect*. The Oxford Review.
- Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a "right to explanation." *AI Magazine*. <https://doi.org/10.1609/aimag.v38i3.2741>
- Köse, U. (2020). Yapay zekâ etiği çerçevesinde geleceğin işletmeleri: Dönüşüm ve paradigma değişiklikleri. *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(5), 290-305. <https://doi.org/10.21923/jesd.833224>
- Lohr, S. (2018). *M.I.T. Plans College for Artificial Intelligence, Backed by \$1 Billion*. New York Times.
- Macionis, J. J., (2015), *Sosyoloji*, (çev. Ed. Vildan Akan), Ankara, Nobel Kitap.
- Marshall, G. (2005). *Sosyoloji Sözlüğü*. Ankar, Bilim ve Sanat Yayınları.
- Marshall, A. (2015). *Industry and trade*. Vani Prakashan.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Moor, J. (2009). *Four Kinds of Ethical Robots*. Philosophy Now.
- Moral machine*. (t.y.). Moral Machine. Retrieved November 5, 2021, from <http://moralmachine.mit.edu>
- Nahavandi, S. (2019). Industry 5. 0—A human-centric solution. *Sustainability*, 11(16), 4371. <https://doi.org/10.3390/su11164371>
- Narin, Ö. (2018). Sanayi 4.0'ın ima ettikleri ve sınırlarına dair bir tartışma: Çevrimiçi kapitalizm olabilir mi? *Toplum ve Bilim*, 144, 233-253.
- Noorman, M. (2018). *Computing and Moral Responsibility*. In *Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/computing-responsibility/>
- Oztemel, E., & Gursev, S. (2020). Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31(1), 127-182. <https://doi.org/10.1007/s10845-018-1433-8>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you? ": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. <https://doi.org/10.1145/2939672.2939778>

- Ritzer, G., (2014), *Klasik Sosyoloji Kuramları*, 6. Basım, (çev. Himmet Hülür), Ankara, De Ki Basım Yayımları.
- Ryan, M., Stahl, B. C. (2020). Artificial Intelligence Ethics Guidelines For Developers And Users: Clarifying Their Content And Normative Implications. *Journal of Information, Communication and Ethics in Society*.
- Samek, W., Wiegand, T., & Müller, K. R. (2017). *Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models*. ArXiv.
- Skobelev, P. O., Borovik, S. Yu. (2017). On the Way from Industry 4.0 to Industry 5.0: From Digital Manufacturing to Digital Society. *Industry 4.0*, 2.6, 307-311.
- Suman, R. R., Mall, R., Sukumaran, S., & Satpathy, M. (2010). Extracting state models for black-box software components. *The Journal of Object Technology*, 9(3), 79. <https://doi.org/10.5381/jot.2010.9.3.a3>
- TÜİK kurumsal. (t.y.). Retrieved from gönderen <https://data.tuik.gov.tr/Bulten/Index?p=Olum-ve-Olum-Nedeni-Istatistikleri-2019-33710>
- The IEEE Global Initiative. (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2. IEEE.
- Topham, G. (2019). *Lion Air pilots were looking at handbook when plane crashed | World news | The Guardian*.
- Wallach, W., & Allen, C. (2009). Moral Machines: Teaching Robots Right from Wrong. In *Moral Machines: Teaching Robots Right from Wrong*. <https://doi.org/10.1093/acprof:oso/9780195374049.001.0001>
- Wang, W., & Siau, K. (2018). Ethical and Moral Issues with AI - A Case Study on Healthcare Robots. Americas Conference on Information Systems 2018: Digital Disruption, *AMCIS 2018*.
- Vakkuri, V., Kemell, K. K., Kultanen, J., & Abrahamsson, P. (2020). The Current State of Industrial Practice in Artificial Intelligence Ethics. *IEEE Software*, 37(4), 50-57.
- Yeşilova, D. (2020). Makine Etiği: Otonom Makinelerin Ahlaki Karar Verme Mekanizmaları Üzerine İnceleme. *Türkiye’de STS: Bilim ve Teknoloji Çalışmalarına Giriş* (pp. 153–160).
- Yücebalkan, B. (2020). Endüstri 4.0’dan Endüstri 5.0’a Geçiş Sürecine Genel Bakış. *Pearson Journal of Social Sciences & Humanities*, <http://dx.doi.org/10.46872/pj.181>.