

## Ahlaki Değerlerin Kodlanabilmesi Bağlamında Yapay Zekâ Etiğine Kuramsal Bir Bakış

Müge KARABAĞ\*

### Öz

Yapay zekânın çalışma prensibi, insan gibi mantık yürüterek insan gibi davranma üzerine kuruludur. Burada insan zekâsının mekanik olana uygulanması, insan zekâsının insan mantığının algoritmasıyla makine zekâsına entegre edilmesi amaçlanmaktadır. Ancak mekanik olanın doğal zekâyâ öykünmesiyle ortaya etik sorunlar çıkmaktadır. Bu çalışmada öncelikle yapay zekâyâ yönelik kuramsal yaklaşımlara, kavramın özelliklerine, yapay zekânın türlerine yönelik bir değerlendirme yapılmıştır. Ardından doğal zekâ ile makine zekâsının karşılaştırılmasına ilişkin literatürden yararlanılarak yapay zekâ ve etik ilişkisinin daha net anlaşılması için kuramsal bir temel oluşturulmuştur. Bu noktada etik kavramının makine zekâsına uygulanabilirliği tartışılmıştır. Çalışmayı yapay zekâ ve etik ilişkisini irdeleyen diğer çalışmalardan ayıran özellik, etik değerlerin kodlanabilmesi kapsamında yapay zekâ ve etik ilişkisinin incelenmesi ve etik bir yapay zekânın inşası için bir formül önermesidir. Dolayısıyla çalışmada sadece yapay zekâ sistemlerinin sebep olabileceği etik sorunlara değil, bu etik sorunların yaşanmaması için yapılması gerekenlere de yer verilerek literatüre katkı sağlamak amaçlanmıştır. Bu çalışmada etik açıdan toplumsal alanda sorun yaratmayacak bir yapay zekânın tasarımı için üç işlevli bir formül ortaya atılmıştır. Bu kapsamda yapılacak her türlü normatif çalışmanın hukuk ile beraber işleme gerektiği yapay zekâ ve etik ilişkisi kapsamındaki literatür taranarak savunulmuştur.

**Anahtar Kelimeler:** Yapay Zekâ, Etik, Makine Zekâsı, Doğal Zekâ, Dijitalleşme

\*Dr., İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Radyo Televizyon ve Sinema Bölümü Doktora Programı  
Mezunu. karabag.mugeee@gmail.com

Karabağ, M. (2021). Ahlaki Değerlerin Kodlanabilmesi Bağlamında Yapay Zekâ Etiğine Kuramsal Bir Bakış. TRT Akademi, 6 (13), 748-767. DOI: 10.37679/trta.954641

### Derleme Makale

Geliş Tarihi: 19.06.2021

Kabul Tarihi: 23.09.2021

ORCID ID: 0000-0002-8141-3557 DOI: 10.37679/trta.954641

## A Theoretical Overview of Artificial Intelligence Ethics Within the Context of Coding Moral Values

Müge KARABAĞ

### Abstract

The working principle of artificial intelligence is based on acting like a human by reasoning like a human. At this point, it is aimed to apply human intelligence to mechanical one and to integrate human intelligence into machine intelligence with the algorithm of human logic. However, ethical problems arise when the mechanical imitates natural intelligence. In this study, first of all, an evaluation was carried out about the theoretical approaches related to artificial intelligence, the characteristics of the concept, and the types of artificial intelligence. Then, using the literature on the comparison of natural intelligence and machine intelligence, a theoretical basis was established for a clearer understanding of the relationship between artificial intelligence and ethics. The feature that distinguishes the study from other studies examining the relationship between artificial intelligence and ethics is that it examines the relationship between artificial intelligence and ethics within the scope of coding ethical values and proposes a formula for the construction of an ethical artificial intelligence. Therefore, in the study, it is aimed to contribute to the literature not only by the ethical problems that artificial intelligence systems can cause, but also by the things that need to be done to avoid these ethical problems. In this study, a three-functional formula has been proposed for the design of an artificial intelligence that will not cause ethical problems in the social field. In this context, the necessity of the consistent cooperation between any normative study that will be carried out and law has been defended by exploring the litterateur within the scope of artificial intelligence and ethics.

**Keywords:** Artificial Intelligence, Ethics, Machine Intelligence, Natural Intelligence, Digitalization

### Review Paper

---

Received: 19.06.2021

Accepted: 23.09.2021

---

## 1. Giriş

Yapay zekâ günümüzün en popüler konularından biridir. Kavramı böylesine popüler yapan, yapay zekânın insanlığı nasıl etkileyeceği konusundaki farklı ve bilim kurgu filmlerini andıran görüşlerdir. Genellikle düalist bir çizgide ilerleyen yapay zekâ ile ilgili görüşler, bu alandaki gelişmelerin insanlığa zarar getireceğine veya insanlığa katkı sağlayacağına yöneliktir. Her geçen gün artan yapay zekâ alanındaki gelişmeler, bu alanın toplumda askeri, endüstriyel ve sosyal alanlardaki kullanımını da artırmaktadır. Yapay zekâ sistemlerinin toplumsal alana giderek daha fazla entegre olması, kamusal yapıdan gündelik hayat pratiklerine kadar pek çok alanı da dönüştürmektedir. Kullanılan navigasyon sistemleri, eş zamanlı çeviri sistemleri, robot süpürgeler, cerrahi robotlar ve diğer pek çok yapay zekâ içerikli sistemler; makine öğrenmesi, doğal dil işleme, derin öğrenme gibi yapay zekânın oluşum süreçleriyle tasarlanmaktadır.

Günümüzde robotik ve yapay zekâ çalışmaları, makine öğrenmesi ile insan öğrenmesi arasındaki farklılıkları, yapay zekâ ile doğal zekâ düzleminde bir karşılaştırmayla eşitlemeye yönelik yapılmaktadır. Her geçen gün tanık olduğumuz teknolojik gelişmeler ileride yapay zekânın insan zekâsını niteliksel olarak aşma ihtimalini düşündürmektedir. Bu noktada insansı robotların, androidlerin, sayborgların ve diğer akıllı cihazların ve yazılımların ana beyni, temel omurgası olarak tüm bu robotik sistemleri bir araya getiren kavramı yapay zekâ olarak tanımlamak mümkündür (Batukan, 2017, s. 26). Bu sistemlerin işlevlerinin artması ve insan zekâsının yapabildiği basit işleri bile yapabilecek boyuta ulaşması insana rakip bir varlığın toplumda eşitsizliğe, işsizliğe, hukuki ve etik sorunlara da yol açabileceği olasılığını ortaya çıkarmaktadır. İkinci ve üçüncü endüstriyel devrimler sonrasında insan ve makineler arasındaki ilişki sebebiyle oluşan birtakım etik sorunların yine yapay zekâ teknolojileri açısından söz konusu olacağı ifade edilmektedir. Bu sorunlara yönelik, makinelerin iyi veya kötü amaçlarla kullanımı, makinelerin ve makine ile insan ilişkilerinin insanlaştırılması, teknoloji bağımlılığı, teknolojik kaynaklara erişim, bu sistemlerin yetkilendirilmesi ve sorumluluklarının belirlenmesi gibi konulara değinilmektedir (Ersoy, 2019, s.147). Örneğin insansı robotlar 2015 yılında fastfood restoranlarında ve diğer iş kollarında kullanılmaya başlanmıştır. Japonya'daki bir restoranlar zincirinde, işletmenin tüm restoranlarında suşi yapımında robotlar kullanılmaktadır. Süpermarket zinciri Walmart'ta ise ürün stoklarını denetleyebilmek için yine robotların kullanımına başvurulmuştur (Ford, 2018, s. 34-49). Yapay zekâ sistemlerinin insani niteliklere ve becerilere ulaşması, onun insanın yapabildiği işlerde kullanımını sağlamıştır. Bu durum iş hayatında yapay zekânın insana tercih edilmesini ve insanların işsiz kalabilme ihtimalini dü-

şündürmektedir. Örneğin McKinsey Global Institute tarafından yapılan bir araştırmada yapay zekânın kullanımının artması üzerine 2030 yılına kadar 800 milyon insanın işini kaybedeceği belirtilmektedir (McKinsey, 2017). Öte yandan yapay zekânın kullanımı arttıkça, insana bağlı üretim ve iş süreçlerinin de değişeceğini ve yeni iş kollarının da ortaya çıkacağını belirtmek mümkündür. Örneğin sanayi devrimiyle birlikte ortaya yeni meslek grupları da çıkmış ve kaybolan mesleklerin yerini yenileri almıştır. Burada önemli olan, kapitalist ve neo liberal politikaların nasıl işlediğinin ötesine geçip etik ve hukuk düzleminde bu sürecin ele alınmasıdır. Yapay zekâ sistemlerinin sadece iş hayatında değil, toplumsal hayatta insan ilişkilerinde de büyük dönüşüme yol açabileceği ve etik sorunlar yaratabileceği önemli bir noktadır. Bu sistemlerin sosyal ilişkilerde insana tercih edilebileceği yapay zekâ ve etik konusu kapsamında değerlendirilmektedir. Bunun örnekleri şimdiden görülmektedir. Örneğin Fransa'da kendisinin 3D yazıcıyla ürettiği robota aşık olan kadın, robotuyla evlenmek istediğini ve robotun kendisini çok mutlu ettiğini belirtmiştir. Teknoloji fütüristleri tarafından da gelecek yıllarda robotların sosyal ve duygusal ilişkilerde insanlara tercih edileceği ve bu amaçlarla kullanılan robot sayısının artacağı belirtilmektedir (Owsianik, 2016). Bu gelişmeler bilim kurgu filmlerinden ve dizilerinden aşına olunan konulardır. Her (2013), Ex Machine (2014), Jexi (2019), S1MONE (2002), gibi filmler ile Wesworld (2016), Humans (2015), Better Than Us (2018) gibi diziler yapay zekâ ve insan ilişkisini etik açıdan ele alan yapımlardır.

Bu çalışmada öncelikle yapay zekâyâ yönelik kuramsal yaklaşımlara, kavramın özelliklerine, yapay zekânın türlerine yönelik bir değerlendirme yapılacak, ardından doğal zekâ ile makine zekâsının karşılaştırılmasına ilişkin literatürden yararlanılarak yapay zekâ ve etik ilişkisinin daha net anlaşılması için altyapı oluşturulacaktır.

Yüzbinlerce yıllık insanlık tarihi içerisinde insan beyninin gelişimiyle birlikte ortaya çıkan düşünce ve davranış biçimleri insana özgü kavramları da üretmiştir. Bu kavramlardan birisi de felsefenin temel araştırma alanlarından biri olan etik kavramıdır. Hızla gelişen teknoloji, söz konusu kavramın daha geniş bir alanda ele alınmasını ve incelenmesini zorunlu kılmıştır. Daha önceleri etik kavramı insan ve hayvan davranışları boyutunda değerlendirilirken günümüzde yapay zekâ ve robot teknolojilerinin birçok alanda kullanılıyor olması bu kavramın yapay zekâ sistemleri ve insan ilişkisi kapsamında da tartışılmasını sağlamıştır. Bu çalışmayı yapay zekâ ve etik ilişkisini irdeleyen diğer çalışmalardan ayıran özellik, etik değerlerin kodlanabilmesi kapsamında yapay zekâ ve etik ilişkisinin incelenmesi ve etik bir yapay zekânın inşası için bir formül önermesidir. Dolayısıyla çalışmada

sadece yapay zekâ sistemlerinin sebep olabileceği etik sorunlara değil, bu etik sorunların yaşanmaması için yapılması gerekenlere de yer verilerek literatüre katkı sağlamak amaçlanmıştır.

## 2. Yapay Zekâya Kuramsal Bakış

Yapay zekâ kavramı en kısa ve net tanımıyla insanların yaptığı basit işleri bir makineye yaptırmaya çalışmayı ifade etmektedir. Bir başka ifadeyle yapay zekâ, bilgisayar algoritmalarının insani bilişsel yetileri taklit etmesiyle ortaya çıkan beceriler bütünüdür (Kutlusoy, 2019, s. 25). Nilsson ise yapay zekâyı, makinelere zekâ kazandırmaya adanmış bir etkinlik olarak tanımlamaktadır (2018, s.13). İnsan zekâsı anlama ve algılama niteliğine sahiptir. Bu bağlamda yapay zekânın tıpkı insanlar gibi algılama, anlama, deneyerek öğrenme gibi aşamalardan geçerek doğruyu ve yanlış ayırt edebildiğini belirtmek yanlış olmayacaktır. Derin öğrenme, makine öğrenmesi gibi yapay zekâya özgü kavramların amacı da yapay zekânın gelişimini tıpkı insan zekâsı gibi mantıksal bir işleyişle gerçekleştirmektir. Bilgisayar uzmanı Sebastian Thrun de yapay zekâyı, makinelerin karmaşık bir şeyi algılayıp buna göre karar verebilmesi olarak tanımlamaktadır (Singer, 2015, s. 98).

İnsana özgü basit işlemlerin bir makine tarafından yapılabilmesi yapay zekânın başlıca işlevidir (Karabağ, 2021, s. 44). Russell ve Norvig yapay zekâya yönelik tanımları 4 başlık altında incelemiştir. Bu başlıklar; insan gibi düşünen sistemler, insan gibi davranan sistemler, rasyonel düşünen sistemler, rasyonel hareket eden sistemler olarak tanımlanmıştır (1995, s. 2). Ancak günümüzde yapay zekâya yönelik mevcut gelişmeler bu kategoriler arasındaki ayrımların muğlaklaşabileceğini düşündürmektedir. Zira yapay zekâ günümüzdeki dar yapay zekâ seviyesinden çıkıp süper zekâya ulaştığında bu kategorilerin bir anlamı kalmayacaktır. Böylece insan gibi düşünen, insan gibi davranan, rasyonel düşünen ve rasyonel hareket eden yapay zekâ sistemleri bir arada işleyebilecektir. Yapay zekâ, büyük beceri ve zekâ gerektiren oyunlarda da insan ile yarışacak ve hatta insanı mağlup edebilecek boyutlara ulaşmıştır. Örneğin 1997 yılında dünya satranç şampiyonu Kasparov'u yenen 'Deep Blue' isimli bir yapay zekâ olmuştur (Karabağ, 2021, s. 45).

Yapay zekâ uzmanları kavramı geliştirirken uygulama alanlarını maksimum düzeye çıkarmak amacıyla öncelik olarak insan beynini model almaktadırlar. Dolayısıyla insan beyninin işleyiş şekli yapay zekâyı oluşturmak için önemli bir referans olarak belirtilmektedir (Russell ve Norvig, 1995, s. 3). Bu durumda da yapay zekânın belirli özellikleri ortaya çıkmaktadır. Bu bağlamda bir yapay zekâda olması gereken özellikler şu şekilde sıralanmıştır (Russell ve Norvig, 1995, s. 3):

- Dil becerisi ile başarılı bir şekilde iletişim kurabilmesi
- Algıladıklarını, bilgiyi sergilemesi ve muhafaza edebilmesi
- Anlayıp değerlendirmesi ve depoladığı bilgileri kullanabilmesi
- Yeni koşullara uyum sağlaması ve durumları tahmin veya tespit edebilmesi
- Nesnelere görüp algılamak için bilgisayar görüşüne sahip olması
- Robotik nesnelere hareket ettirmesi ve komutlarla veya komutlardan bağımsız hareket edebilmesi

Bu özellikler yapay zekânın seviyesine göre değişkenlik göstermektedir. Günümüzde yapay zekâ sistemleri 3 başlık altında konumlandırılmaktadır (Ünal, 2018, s. 60):

- Hedefe yönelik özelleşmiş (zayıf) yapay zekâ,
- Çok sayıda görevi yapabilen ve insan yaklaşımına sahip (güçlü) yapay zekâ,
- İnsanlardan çok daha üstün (süper) yapay zekâ.

Bu noktada süper zekâ kavramı, yapay zekânın ulaşabileceği en üst seviyeyi ifade etmektedir. Zira Bostrom'a göre de süper yapay zekâ terimi, genel nitelikteki birçok bilişsel alanda günümüzde var olan insan zihnini katbekat aşan zekâ tanımı için kullanılmaktadır. Bu bağlamda farklı performans özelliklerine sahip farklı türdeki sistemler de bu tanıma göre süper zekâ kapsamına girebilmektedir (Bostrom, 2018, s.71).

Günümüzde navigasyon sistemleri, bankacılık ve finans uygulamaları, sağlık ve eğitim sistemleri gibi pek çok alanda hedefe yönelik geliştirilen dar yapay zekâ türü kullanılmaktadır. Ancak teknolojik gelişmeler yapay zekânın bu aşamayı hızlı bir şekilde geçip güçlü ve süper zekâyâ ulaşabileceğini düşündürmektedir. Yapay zekâ yalnızca teknolojik bir gelişme olarak değil etkileri ve sonuçları bakımından da ele alınmaktadır. Bu bağlamda insan zekâsı referans alınarak tasarlanan yapay zekânın toplumsal hayatta kullanımıyla ortaya çıkabilecek etik sorunlar tartışılmaktadır. Söz konusu sorunlara değinmeden önce mekanik olanın insan becerisi ile doğal zekâyâ öykünmesi yolunda bilinmesi gerekenlere değinmek, yapay zekâ ve etik ilişkisini bir yönüyle daha iyi anlamayı sağlayacaktır.

### 2.1. Doğal Zekânın Bir Taklidi Olarak Makine Zekâsı

Yapay zekâyâ yönelik çalışmaların doğal zekânın mantık kurma yönergesinin rehberliğinde gerçekleştiğini belirtmek mümkündür. Bu bağlamda yapay zekânın, insanlar gibi davranıp hareket etme özelliğinin bu anlayış çerçevesinde tasarlan-

dığını ifade etmek yanlış olmayacaktır. Zira antik çağlarda da filozoflar öncelikli olarak insan gibi düşünebilen bir varlığın tasarlanabileceği hakkında görüşlerini sunmuşlardır. Örneğin Aristoteles Politika adlı eserinde komutlarla hareket eden akıllı bir makine fikrini şu şekilde ifade etmiştir:

“Düşünün, elimizdeki her alet ya bizim emrimizle ya da ihtiyaç hâlinde kendi işini icra ediyor; eğer Homeros’un, “Kendi başlarına hareket edip tanrılar meclisine girerler.” dediği Hephaistos’un uçayaklıları gibi, dokuma tezgâhının mekiği kendiliğinden gidip gelse ve lirin mızrabı kendi kendine çalsa o zaman ne usta zanaatkârlar hizmetçiye gerekirim duyardı ne de efendiler kölelere...” (1962, s. 65).

Makinelerin insanlar gibi düşünüp davranmalarını sağlayan yapay zekâ uzmanları doğal zekânın bir taklidi olarak yapay zekâ sistemlerini geliştirmektedirler. Bostrom da insan beyninin makine zekâsı için bir kalıp olarak kullanılabileceğini belirterek buna örnek olarak insan beyninden ilham alınarak yapılan bir yapay zekâ tekniği olan nöral ağları göstermektedir (2018, s. 45). Aynı şekilde yapay sinir ağlarının da doğal zekânın öğrenme yolunu taklit ederek beynin öğrenme, anlama, yeni veri üretme gibi temel işlevlerin gerçekleştirildiği bilgisayar yazılımları olduğu ifade edilmektedir (Öztürk ve Şahin, 2018, s. 27). İnsan zekâsının bilgiyi kodlayabilmesi, makinelerin de bilgiyi kodlayabileceğini ve buna uygun davranabileceğini düşündürmektedir. Bu noktada makine zekâsı ile insan zekâsını karşılaştırmak için konuyla ilgili bilgilere değinmek yerinde olacaktır.

Bilgisayarlar hesaplama konusunda insanlardan çok daha hızlı ve başarılıdır. Ancak bilgisayarların hesaplama konusundaki becerisinin diğer konularda düşük seviyede olduğunu belirten uzmanlar, bu durumu insan beyninin bütün olarak programlanabilir ve genel bir amaca hizmet eden bir bilgisayar olmamasıyla izah etmektedirler (Nabiyev, 2010, s. 20). Bu durumda yapay zekânın tamamen insan beyni ile eşit biçimde çalışmasının gerekmediği sonucu ortaya çıkmaktadır. Bostrom’un da belirttiği gibi yapay zekâ tasarlanırken insan beyninin çalışma prensibinden ilham alınabilir ancak insan beynine tamamen benzemesi de gerekmemektedir (2018, s. 45). Öte yandan yapay zekânın insan beyni ile asla yarışmayacağını ve insan beyninin yaptığı şeylerin yapay zekâ tarafından yapılamayacağını ifade eden uzmanlar bu görüşlerini şu argümanlara dayandırmaktadırlar (Nilsson, 2010, s. 392):

- İnsan beyninde trilyonları bulan işlem birimi varken bu oran bilgisayarlarda yüzlerce olabilir.
- İnsan beyni binlerce işlem yaparken bilgisayarlar ise saniyede milyarlarca işlem gerçekleştirir.

- İnsan beyni bir hata durumunda çökmez ve dayanıklıdır. Ancak bilgisayarlar çökebilir.
- Bilgisayarlar ikili sinyaller kullanırken insan beyni analog olanlarla çalışır.
- Bilgisayarlar yalnızca programcılarının yapmasını istediklerini yaparlar, insan beyni ise yaratıcıdır.
- Bilgisayarlar mantıksal olarak kısıtlanırken insan beyni sezgisel olabilir.
- Bilgisayarlar programlanırken insan beyni öğrenir.

Nilsson, bu ayrımların günümüzde artık geçerli olmadığını belirterek bilgisayar teknolojilerindeki gelişmelerin çeşitli alanlarda insan zekâsıyla yarışacak boyuta ulaştığını ifade etmektedir. Örneğin bilgisayarlar da artık makine öğrenmesi yoluyla yani insanlar gibi deneyimlere bağlı olarak kendilerini geliştirmektedir. Bir başka ifadeyle doğrudan bir programlama mevcut olsa da makine öğrenmesi yoluyla mevcut program kendini geliştirebilmektedir. Bu durum öyle ciddi boyutlara ulaşmıştır ki yapay zekânın savaşta ve deneyime dayalı insan yargısı gerektiren görevler için kullanılmasına karşı çıkmaktadır. Öte yandan yapay zekâ sistemlerinin insanlardaki geleneksel iletişim biçimlerini bozacağından, sosyal ve toplumsal yapıyı olumsuz yönde etkileyeceğinden kaygı duyulmaktadır (2010, s. 393-394). Weizenbaum da bilgisayarların insani deneyimlere sahip olamayacağını ve bu sebeple makinelerin ne kadar zeki olursa olsun kritik kararlar almasına ve insanlara tavsiyede bulunmasına izin verilmemesi gerektiğini belirtmektedir (Weizenbaum'dan Akt. Nilsson, 2010, s. 394-395). Burada vurgulanmak istenen, bilgisayarların insanlardan daha zeki olamayacağı değil, onlar gibi "insani" olamayacağıdır. Bu görüş bir anlamda makinelerin bir "ruh"u olamayacağına da işaret etmektedir. Doğal zekâ taklit edilebilir ancak insani bir algı ve empati kurma yeteneği bir makinede hiçbir zaman var olmayabilir. Günümüzde bunun aksini düşündürecek gelişmeler de yaşanmaktadır. Örneğin İngiltere'de robotların öğrenerek kişilik geliştirebileceklerine dair bir çalışma yapılmıştır. Çalışmada her biri bir salyangozun beyin gücüne eşit, yani yaklaşık 50 nöron olan bir grup robot, birbirlerinden uzak durmaları ve birbirlerine değmemeleri için programlanmıştır. Sistem ödül-ceza yöntemiyle kurulmuştur. Robotlar kapalı yere konularak gözlemlenmişlerdir. Burada dikkat çekilen nokta, robotların geçmiş deneyimleri göz önüne alabilmeleridir. Bu da zamanla değişik dersler çıkarabilmeleri anlamına gelmektedir. Tüm robotlar başlangıçta aynı yazılıma sahip olmalarına rağmen araştırmacılar, iş birliği yapan "iyi" robotlarla sürekli birbirlerine saldıran "kötü" robotların farklılaştığını görmüşlerdir. Hatta yapay olarak aptalla eşit veya kendine zarar veren bir robot sürekli olarak mümkün olan en kötü seçimi yapmaya evrimleşmiştir (Singer'dan Akt. Karabağ, 2021, s. 71). Burada yapay zekâ ile donanımlı robotların insanlar gibi



deneyim yoluyla öğrendikleri ve kişilik geliştirdikleri ortaya çıkmaktadır. Dolayısıyla yapay zekâ sistemlerinin de insanlar gibi gelişerek, deneyimleyerek ve buna bağlı sonuçlar ortaya koyarak “insanlar gibi” olabilecekleri ihtimali düşünülmektedir. Dennet da robotlarla insanların ortak yönleri olduğunu ifade ederek robotların da insanlar gibi karmaşık, otokontrol mekanizmasına sahip, doğal seçimle tasarlanmış ve kendi kendine onarım süreçlerine sahip olduğunu belirtir (1997, s. 18). Örneğin kendi kendini güncelleyebilen, şarjı bittiğinde kendisini şarj edebilen bir robot, tamamen olmasa da bir yönüyle otokontrol mekanizmasına sahip, kendi kendini onarım aşamasını gerçekleştiren bir robot olacaktır. Bu bağlamda doğal zekâdan ilham alınarak tasarlanan yapay zekâ sistemlerinin nihai amacının insanın bilişsel ve davranışsal kapasitesine erişmek olduğu ifade edilebilir. Bu sebeple dar yapay zekâ sistemlerini geliştirmek, bu sistemleri her geçen gün daha da fazla insan zekâsıyla yarışır hâle getirmek için çalışmalar yapılmaktadır. Yapay zekâ alanındaki çalışmalar insan hayatını kolaylaştırma, toplumsal refah düzeyini artırma, sorunlara pratik ve ekonomik çözümler bulma gibi amaçlarla yapılırsa da gelişen bu sistemlerin insana ortak bir kamusal alanda var olmasıyla hukuki ve etik alanındaki riskler de ortaya çıkmaktadır. Doğal zekâyâ öykünerek tasarlanan yapay zekâ sistemleri insanlardaki gibi hukuki gereksinimleri beraberinde getirmekte, gerçek insanlarla insansı robotlar ve diğer yapay zekâ sistemlerinin olası haklarını sorgulamaktadır. Özellikle bu sistemlere otonom yetki kazandırmanın ciddi tehlikeleri de beraberinde getirmesi mümkündür. Sürücüsüz arabalar, öldürmeye programlı askeri robotlar, gündelik hayata yardımcı asistan yazılımlar otonom özelliği kazandığında bu sistemlerin insanları manipüle etme ve insanlar tarafından manipüle edilme ihtimali de doğacaktır. Bununla birlikte insansı görünüme sahip olan yapay zekâların insanın alternatifi bir konumda kullanılabileceği de ihtimal dâhilindedir. Zira günümüzde yapay nöronlar ve yapay hücreler ile DNA ve RNA'nın sentetik modeli olan XNA (Xeno nükleik asit) isimli bilgi depolayan biyopolimerin insansı robotlarda kullanım amaçlı üretilmesi, geleceğin dünyasının etik ihlallerle, dönüşen sosyal ilişkilerle dolu kargaşa içindeki bir toplumu resmeden bilim kurgu filmlerini andıracağını düşündürmektedir. Böyle bir gelecek tahayyülü de yapay zekâ ve etik ilişkisinin önemini ortaya koymaktadır.

### 3. Yapay Zekâ ve Etik

Yapay zekâda etik ve ahlak tartışmalarının genel olarak Kant ve Aristoteles'in görüşleri temelinde yapıldığını belirtmek mümkündür (Karabağ, 2021, s. 95). Aristoteles etiğin amacının mutlu bir yaşam sürmek olduğunu ifade eder. Bunun için öncelikle insanın etik bir niteliğe, bir erdeme sahip olması gerekir. Etik kavramı, çeşitli koşullarda en iyi davranma kapasitesi olarak tanımlanmaktadır. Kant ise ahlaklı davranış

ölçütü olarak iradeyi belirler. İrade yapay zekâ sistemlerinin etik davranma potansiyeline yönelik yorumlandığında otonomluğa işaret eder. Özerk bir yapay zekâ aynı zamanda otonomdur ve kendi iradesine de sahiptir. Ancak bu noktada Ruffo, Kant ve Aristoteles'in belirlediği ahlaki ölçütlerin yapay zekâ sistemleri için geçerli olmayacağını ifade eder. Aristoteles'in etik anlayışıyla bir robot için mutluluğun ne olduğunu sorgulayan Ruffo, etik davranışın sonucunda gelen mutluluk hâlinin bir robot için ne ifade edeceğini tartışır. Bu bağlamda ona göre etik davranışı motive eden amaç, bir robot için yabancı olacaktır (2012, s. 87).

Aynı şekilde insanlardaki empati duygusunun robotlarda olamayacağını savunan Ruffo, bebek bakıcısı olarak tasarlanmış bir robotun bebek bakımıyla ilgili sadece basit işlevleri yerine getirebileceğini ancak bebeğin neden ağladığını, neden korktuğunu anlayamayacağını belirtir. Bebeğe gereken sevgi ve ilgiyi sağlayamayan bir robotun insan ilişkilerinde etik davranma boyutuna ulaşması da oldukça zor görünmektedir. Bu açıdan Aristoteles'in belirlediği mutlu bir hayat yaşama hedefi, sevgi, korku, endişe gibi duyguları anlayamayan bir robot için anlamsız bir amaçtır (Ruffo, 2012, s. 88). Bu sebeple yine de evrensel ahlak yasalarının yapay zekâ sistemlerine programlanması robota "otomatik" bir etik sıfatı kazandırabilir. Bu, aynı zamanda Kant ahlakının evrensellik ilkesiyle bağdaşabilir. Robot etik olmadığından mutluluk ve erdem kavramlarının bilincinde olmayacaktır. Evrensel ahlak yasalarına aykırı davranan ancak haklı eylemlerde bulunan robotların da olabileceği belirtilmektedir. Örneğin bir mülteciyi korumak için bile yalan söylemek yasaksa robot bunu doğrudan uygulayabilir. Çünkü robotların iradesi yoktur. Bu durum Kant ahlakıyla yorumlandığında iradesi olmayan bir robotun "niyeti" de yoktur. Bu açıdan özgür olmayan ve bir ruha sahip olmayan bir robotun hukuki şartlara uyum sağlaması için bir sebebi de yoktur. Bununla birlikte yapay zekâ sistemlerini etik ilkelerle inşa etmek ahlaki bir görev yükümlülüğü de gerektirmektedir. Zira bu sistemler kodlandıkları amaçlar doğrultusunda hareket ederler ve bu kodları sorgulayıp değerlendiremezler. Bu açıdan özgür olmayan yapay zekâ sistemlerinin ahlaki özerkliği de bulunmamaktadır. Aristoteles'in felsefesiyle robotlar için mutluluğun ve erdemli olmanın bir anlamı yoktur (Ruffo, 2012, s. 89). Bu sebeplerden dolayı ahlaki yasaların bir robot için geçerli olmayacağı ifade edilmektedir. Robotların sadece bir hesaplama işlevinden başka insani yeteneklere ulaşamayacakları ve bu sebeple insanlara sadece etik kararların verilmesinde yardımcı olabilecekleri belirtilmektedir (Ruffo, 2012, s. 90).

Görüldüğü gibi burada tartışılan etik bir yapay zekânın programlanması değil, yapay zekâ sistemlerinin insanlar gibi muhakeme yeteneğine sahip olmamasından dolayı bu sistemlerin doğrudan etik varlıklar kabul edilemeyeceğidir. Ahlaki kod-

larla inşa edilen bir yapay zekâ bu kodlara uygun davranabilir ancak her zaman etik ve ahlaklı olanı yapamayabilir, zira bu sistemler kodlarından bağımsız davranmayabilirler. Bu bağlamda konuyla ilgili olarak yapılan normatif çalışmalara da değinmek gerekmektedir.

Yapay zekâ sistemleri arasında ilk olarak robotlara yönelik kurallar edebi eserlere konu olmuştur. Daha sonra söz konusu kurallar yapay zekâ etiği üzerine yapılan çalışmalara referans sağlamıştır. Bu yazılı kurallardan ilki, Isaac Asimov'un 1942 yılında yazdığı I, Robot isimli eserinde yer alan 3 Robot Kanunu olarak belirtilmektedir. Bu ilkeler şunlardır (2016, s. 7):

- Robotlar, insanlara zarar veremez ya da eylemsiz kalarak onlara zarar gelmesine göz yumamaz.
- Robotlar, birinci kanun ile çelişmediği sürece insanlar tarafından verilen emirlere itaat etmek zorundadır.
- Robotlar, birinci ya da ikinci kanun ile çelişmediği sürece kendi varlıklarını korumak zorundadır.

Asimov'un davranış kurallarının değiştirilmiş bir formunun robotların yazılımında uzun vadede gerekli olduğunu düşünülmektedir. Ancak Singer'a göre bir makine etik kurullarla donatılmış olsa da bu onu etik bir varlık yapmaz. Zira yazılım kodları birer ahlaki kod değildirler; sıfırların ve birlerin altında yatan ahlaki bir anlam yoktur (Singer'dan Akt. Karabağ, 2021, s. 88) Bazı yapay zekâ uzmanları ise Asimov'un bu ilkelerinin gerçeğe evrilmesinin zor olduğunu söylemektedirler. Onlara göre bir robota zarar vermenin ne olduğunu programlamak oldukça zordur. Gerekli ve yeterli koşullar olduğu matematikle geometrinin tersine, insana zarar vermek gibi yuvarlak bir kavramı dolayısıyla Üç Robot Yasası'nı robotlara yüklemek bir hayli zor görünmektedir. Bununla birlikte bazı kişiler de bir robotun insanlara hiç zarar vermeyecek şekilde programlanmasını istemeyebilir. Örneğin toplumda robot polislerin görev yaptığını düşündüğümüzde, bu robot polisin karşısına insanların hayatını tehlikeye atabilecek bir suçlu çıktığında ne yapacağı sorusu gündeme gelmektedir. Eğer hiçbir insana hiçbir koşulda zarar vermeyecek bir robot ise bu robot polis, böyle bir durumda nasıl davranacaktır? Silahlı robotların alacağı etik kararlar bu açıdan kritik bir öneme sahiptir (Karabağ, 2021, s. 89).

Öte yandan Asimov'un yazdığı robot öykülerinde robotların Üç Robot Yasası'na uygun çalışmadığı, çoğu zaman birinin hayatını kurtarmak amacıyla başkasının hayatını tehlikeye attıkları ve yazılımları ikilemi çözemediği için beyinlerinin kısa devre yaptığı ifade edilmektedir. Bu sebeple Asimov'un söz konusu robot öyküle-

rinde aslında dolaylı bir şekilde etik davranan robotlar üretmenin zor ve sorunlu olduğunu anlattığı vurgulanmaktadır (Demircan, 2016). Yine de gelecek yıllarda robotların insan yaşamına giderek daha fazla nüfuz etme ihtimaline yönelik evrensel alanda normatif çalışmalar yapılmaktadır. Örneğin Avrupa Komisyonu Yapay Zekâ Bağımsız Üst Düzey Uzman Grubu'nun 2018 yılında hazırladıkları Güvenilir Yapay Zekâ için Etik Rehber Raporu bu çalışmalardan biridir. Rapordaki maddeler Asimov'un 3 Robot Kanunu'nda yer alan ilkelerin genişletilmiş ve güncellenmiş hâli olarak yorumlanabilir. Buna göre güvenilir bir yapay zekâ için öncelikle şu maddelere bağlı kalınmalıdır (2018, s. 6):

- Yürürlükteki tüm yasalara ve düzenlemelere uygun olarak yasal olmalıdır.
- Etik ilke ve değerlere bağlı kalmalıdır.
- Hem teknik hem de sosyal açıdan sağlam olmalıdır çünkü iyi niyetle bile tasarlanan yapay zekâ sistemleri insanlık için zarara sebep olabilir.

Bununla birlikte raporda Asimov'un 3 Robot Yasası ile paralel 4 etik ilke olan; insan özerkliğine saygı, adalet, zararın önlenmesi ve açıklanabilirlik maddeleri bulunmaktadır (2018, s. 12). İnsan özerkliğine saygı ilkesi yapay zekânın tamamen insanlığa fayda sağlamasına ve insanların bilişsel, sosyal ve kültürel becerilerini artırmasına ve yapay zekânın insanları manipüle etmemesine yöneliktir. Burada amaç, insanları manipüle edecek yapay zekânın tasarlanmasını önlemektir. Aynı şekilde zararın önlenmesinde yapay zekânın kötü amaçlı kullanımının önüne geçmek ve bu sistemlerin güvenli olduğunu sağlamak amaçlanmaktadır. Adalet ilkesi yapay zekânın kullanımının her alanda eşit olmasına yöneliktir. Açıklanabilirlik ilkesi ise kullanıcıların yapay zekâ sistemlerine ulaşım ve kullanım süreçlerinin şeffaf olmasına yöneliktir (Karabağ, 2021, s. 90). Görüldüğü gibi söz konusu ilkeler birbiriyle bağlantılı ve birbirini tamamlayan ilkelere oluşmaktadır. Adil olmayan ve adil kullanılmayan yapay zekâ sistemleri zararın önlenmesi ilkesini de ihlal edecek ve bu durum, insan özerkliğine saygı ilkesine de zarar verecektir.

Yapay zekâ ve etik çalışmalarında ortak hedef, robotların insan yaşamına dâhil olmasıyla yaşanabilecek sıkıntıları ortadan kaldırmaktır. Bu bağlamda robot etiğinin temel amacı, robotların insanlara, kendilerine, mülklere, çevreye ve diğerlerine zarar vermelerini önlemek için araçlar geliştirmek olmalıdır. Ancak robot etiği sadece bununla sınırlı değildir. Robotların yetenekleri ve karmaşıklıkları arttıkça en belirgin tehlikeleri ve potansiyel zararları önleyen daha sofistike güvenlik kontrol sistemleri geliştirmek gerekli olacaktır. Ayrıca robotlar insan eylemlerini anlama ve yorumlama işinde daha fazla yer aldıkça daha fazla sosyal, duygusal ve ahlaki zekâyı ihtiyaç duyacaklardır (Asaro'dan Akt. Karabağ, 2021, s. 89).

### 3.1. Yapay Zekâya Ahlak Algoritması Yüklenebilir mi?

Yapay zekâ sistemlerine yönelik oluşturulan etik kurallar insanlarla bu sistemlerin ortak yaşam alanında uyum içerisinde çalışmasına yöneliktir. Bununla birlikte söz konusu etik ilkeler daha çok yapay zekâ sistemlerinin tasarlanmasına ve bu sistemlerin kontrolünü sağlamaya odaklıdır. Ancak yapay zekânın kullanımı ve insanların robotlara karşı davranışları üzerine yoğunlaşan görüşler de bulunmaktadır. İnsanların robotlar da dâhil olmak üzere tüm yapay zekâ sistemlerine karşı davranışlarını düzenleyen “Robotetik” adı 2002 yılında kavramsallaştırılmıştır (Veruggio’dan Akt. Karabağ, 2021, s. 91). Her ne kadar robotlar hantal bir metal yığınından oluşsa da üretiminde olduğu kadar insanlar tarafından kullanımında da etik boyutlar ortaya çıkmaktadır. Robotlar ve diğer yapay zekâ sistemleri etik kurallara uyumlu bir yazılımla üretildiğinde, söz konusu sistem etik ihlalinde bulunduğu anda sorumlu kim olacaktır? Asaro’ya göre bu gibi durumlarda robot davranışlarından robotu programlayan ve tasarlayan kişiler sorumlu olmayabilir:

“Bir robota iki farklı insan tarafından iki farklı emir verilir. Kime itaat etmeli? Sahibine mi, diğerine mi? Daha etik talepte bulunan kişiye mi? Daha çok sevdiği kişiye mi? Yoksa kendi çıkarlarını düşünüp ona göre kendi istediğini mi yapmalı? Ya da daha fazla düşünün: Yoksa karar vermenin bir önemi var mı? İnsanlar her zaman böyle ikilemlerle karşı karşıyadır. Pratik etik, bu sorunların çözümü için araçlar sağlamakla ilgilidir. Bu ahlaki görüşmeleri çerçevelemek için çeşitli şemalar vardır ancak bu hangi şemayı kullanacaklarına bağlıdır. Robotlar ve bu tür ikilemleri çözmesi gereken teknolojik sistem ve etik değerler tasarımcılar tarafından inşa edilmelidir. Etik kuralları veya davranışları öğrenebilen sistemler özerk ahlaki özneler olarak nitelendirilir ve bu öğrenme yöntemlerinin tasarımcısı robotun davranışlarından sorumludur. Bununla birlikte bir gün bir robotun, tasarımcılarının ve programcılarının robotun eylemlerinden artık sorumlu olmadığı bir noktaya gelmesi de mümkün olabilir” (2006, 10).

Asaro yapay zekâ ve etik konusunu robotları temel alarak değerlendirir. Ona göre burada önemli olan üç faktör vardır. Öncelikle insanların robotlara karşı olan etik davranışları önemlidir. Burada etik özne insanlardır. Daha sonra robotların etik davranacak şekilde nasıl tasarlandıkları önemlidir. Bu sayede robotlar da etik özneler olabilir. Üçüncü olarak yapay ahlaki özneler yani etik makineler yaratmanın etik olup olmadığını sorgulayan Asaro, toplumda robot askerler, robot polisler, robot doktorlar tasarlanmasının ne derece etik olduğunu robotetik kavramıyla sorunsallaştırır (2006, s. 10). Özellikle robot polis, robot asker gibi kritik görevlerle tasarlanan robotların ahlaki değerlerle kodlanabilmesi yapay zekâ etiği kapsamında tartışılan önemli konulardan biridir. Bu açıdan günlük işlerde kullanılan robotlardan ve diğer yapay zekâ sistemlerinden ziyade askeri, tıbbi ve güvenlik

alanlarında kullanılacak akıllı makinelerin etik muhakeme yetenekleriyle tasarlanmaları öncelikli olmalıdır. Zira askeri işlevselliğe sahip robot sistemleri ciddi etik sorunlara yol açabilir. Savaş zamanlarında kritik durumlarda bu sistemlere otonom özelliği kazandırmak ve karar verme yetkisi vermek oldukça risklidir (Asaro, 2006, s.11-14). Bu riskleri ortadan kaldırmak için robot uzmanları insanların belirlediği ahlak kuralları ile robotların tasarlanması gerektiğini belirtmektedir (Singer, 2015, s. 195). Burada amaç otonom sistemlerin yetkilerini, robotların insanları manipüle etme olasılığını etik kodlarla kontrol etmektir. Bu kapsamda da etik bir robotun önceliği, kendisi de dâhil olmak üzere tüm makinelerin etrafa zarar vermelerini önlemek ve insanların ahlaki şüphelerine son vermek olmalıdır (Asaro, 2006, s. 10).

Ancak burada robotlar da dâhil olmak üzere tüm yapay zekâ sistemlerine yönelik ahlaki değerlendirmeleri sadece makinelerin insanlara olan davranışlarıyla sınırlamamak gerekmektedir. Bu bağlamda ahlaki davranış sorumluluğunda önceliğin insanlarda olduğunu belirtmek yanlış olmayacaktır. Coeckelbergh'e göre de yapay zekâ ve sorumluluk kapsamında makineler tek hedef olmamalıdır. Yapay zekâ sistemlerinin üreticileri kadar kullanıcıları da sorumluluk sahibidirler (2019, s.15).

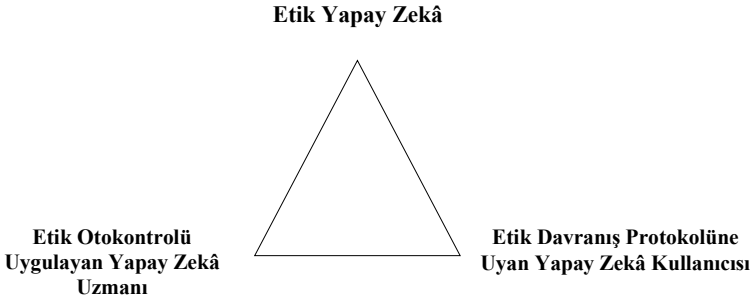
Etik bir makine tasarlamak için ahlaki kodların bu makinelere entegre edilmesini hâkim bir görüş olarak değerlendirmek mümkündür. Bu noktada ahlaki kodların oluşturulma sürecinde hukuk disipliniyle birlikte hareket etmek önem taşımaktadır. Çünkü yapay zekâ sistemleriyle ilgili etik sorunların üstesinden gelmek için hukuk temelli bir altyapı oluşturmak gerekmektedir (Asaro, 2006, s. 13). Bu noktada Asaro, robotların etik makineler hâline gelmesinin bir dizi süreçten geçerek gerçekleşeceğini belirtir. Robotlar gözlem, deneme ve yanılma yoluyla yani davranışlarının sonuçlarını görerek ahlaki bir anlayış geliştirebilirler (2006, s. 11).

Bu doğrultuda robotların davranışlarının tıpkı insanlar gibi ampirik koşullarla etik boyut kazanacağını belirtmek mümkündür. İnsan davranışlarının bilişsel evrimi gibi robotlar da ahlaki anlayış kazanımını zamanla elde edeceklerdir. Dolayısıyla başlangıç olarak ahlaki kodların robot sistemlere dâhil edilmesiyle bir anda etik davranan bir robot tahayyülü en azından şimdiki teknolojiyle uzak bir ihtimal olarak değerlendirilmektedir. Robotla yüklenen ahlaki algoritma robotun deneyimleriyle gelişecek ve sosyal hayatta insanların istediği boyutta etik bir robot hâline gelecektir (Karabağ, 2021, s. 94).

Etik bir yapay zekâ için bu sistemleri ahlak algoritmasıyla programlamak insanlık için fayda sağlayabilir ancak insanların bu sistemleri nasıl kullanacağını ve ne derece etik amaçlar için bu sistemleri tasarlayacağını bir garantisi bulunma-

maktadır. Bu noktada ‐ahlaksal özne‐ kavramı ortaya çıkmaktadır. Bu kavramla, sadece kendi çıkarını değil, başkalarının çıkarlarını da göz önünde bulunduran kişi ya da yapay zekâ sistemleri tanımlanmıştır (Allen v.d., 2000, s. 252-254). Yapay zekâ sistemlerinin ahlaki özne sıfatını kazanmasına yönelik ‐yukarıdan aşağıya‐ ve ‐aşağıdan yukarıya‐ şeklinde 2 modelleme biçimi önerilmektedir. Buna göre ‐yukarıdan aşağıya‐ olan modelleme biçiminde, ahlak kuramları ve etik ilkeler temel alınır. ‐Aşağıdan yukarıya‐ biçiminde olan modellemede ise yapay zekâyâ sahip robota ödüllendirme ve onaylama gibi yöntemlerle ahlaki ilkeler öğretilir. En sağlam metodun ise her iki yaklaşımın birlikte kullanılmasıyla sağlanacağı belirtilmektedir (Allen v.d., 2005, s. 153). Söz konusu ahlak algoritmasının sağlıklı bir şekilde çalışması için öncelikle Turing Ahlak Testi (MTT) önerilmektedir. Ahlak algoritması belirli etik davranış pratikleriyle kodlandırıldığı zaman, makinenin insanlardan ayırt edilemeyecek biçimde etik kurallara bağlı olduğunu teyit etmek için kullanılmasında önerilen test, standart Turing testi ile aynı mantıkla çalışmaktadır. Turing’in amacı, doğal dilin başarılı bir şekilde kazanılmasını tanımlayan standartlar konusunda anlaşmazlıkları geçen bir davranış testi üretmektir. Bu bağlamda benzer şekilde çalışan Turing Ahlak Testi (MTT) ile yapay zekânın belirlenen etik standartları geçmesi hedeflenmektedir. Makine insandan ayırt edilemezse makine bu durumda ahlaki bir öznedir (Allen v.d., 2000, s. 254). Ancak bu görüşün karşıtını savunan yaklaşımlar da mevcuttur. Örneğin Asaro, robotların etik boyuta ulaşabilmelerinin belirli evreler sonucunda gerçekleşeceğini düşünürken Ruffo’ya göre bir robot özerk bile olsa etik olamaz. Bu sebeple robotlar insanların sorumlu olduğu etik ilkelerden muaftır. Ona göre bir robot sadece yasaya uygun davrandığı için etik olmaz. Bununla birlikte Ruffo, etik kavramının yasalarla sınırlanamayacağını, böyle bir robotun en fazla yasal bir robot olacağını ifade etmektedir (Ruffo’dan Akt. Karabağ, 2021, s. 95). Bu tartışma insani yönden ele alındığında etik kavramının insanlar için de değişkenlik gösterdiğini belirtmek mümkündür. Yine de toplumsal hayatta ve insan ilişkilerinde hâkim olan etik kurallarla yapay zekâ sistemlerini inşa etmek etik makinelerin temelini oluşturmak için önemli bir aşama olacaktır. Ancak ahlaki değerlerin bilgisayar koduna dönüştürülmesi anlamına gelen bir ahlak algoritması, yapay zekâ üreticileri ve kullanıcıları arasındaki sorumluluk yükümlülüğünü muğlaklaştıracaktır. Yapay zekâyâ yüklenen ahlak algoritması bir hata verdiğinde doğrudan sorumlu olacak kimse olmayacaktır ve bir makineyi de sorumlu tutup onu ‘yargılayacak’ yaptırımlar günümüzde geçerli değildir. Burada esas sorun, robotların üreticilerinden bağımsızlaşp tamamen özerk bir konuma evrilmeleridir. Komutlarla hareket eden bir robotu ve üreticilerini, bir hata anında sorumlu tutmak hukuksal çerçevede çözümlenmesi mümkün bir konudur. Ancak özerk bir robotun nasıl kontrol altında tu-

tulacağı, sadece robotların insanlara karşı değil, insanların da robotlara karşı nasıl davranacağı gibi sorular gelecekte robot insan ilişkileri açısından büyük önem taşımaktadır. Konuyla ilgili yapay zekâ ve robotik çalışmalarında belirlenen ilkeler genel olarak aynı temeldedir ancak bu ilkelerin geleceğin yapay zekâ dünyasında uygulanabilir olup olmadığı kesin değildir (Karabağ, 2021, s. 97). Burada öncelikli amaç, yapay zekâyâ otonom özelliği kazandırmamaktır. Zira özerk olmayan bir yapay zekâyı dizginlemek ve davranışlarını kontrol etmek daha kolay ve gözlemlenebilir olacaktır. Özerk olmayan bir yapay zekâyı etik değerlerle programlamak ve kullanıcının bu sistemleri etik kullanımını sağlamak, toplumla uyumlu, hukuki ve etik riskleri barındırmayan bir yapay zekâ sistemi inşa etmek için elzemdir. Burada üç işlevli bir etik formül ortaya çıkmaktadır.



Etik bir yapay zekâyı programlamak, etik kurallara bağlı kalarak etik bir otokontrol geliştiren bir uzmanı görevlendirmek ve yapay zekâyâ etik davranan kullanıcılara sahip olmak. Dolayısıyla yapay zekânın etik boyutunda önemli olan insanların onu nasıl tasarladığı ve nasıl kullandığıdır. Bu üç işlev gereğini yaptığında insanlığı tehdit edecek riskler de ortadan kalkacaktır. Etik bir otokontrol mekanizmasına sahip bir yapay zekâ tasarımcısı ve yapay zekâyı etik kullanmaya zorunlu bir kullanıcı için hukuki yaptırımlar uygulamak bu 3 işlevli formülün en önemli basamağı olacaktır.

#### 4. Sonuç

Yapay zekâ teknolojilerinin hızla gelişmesi, bu sistemlerin toplumda pek çok alanda kullanımını sağlamaktadır. Yapay zekânın askeri, endüstriyel, sağlık, finans gibi alanlarda kullanımı ve bu sistemlerin giderek gelişmesi etik konusunu gündeme getirmektedir. Günümüzde mevcut olan dar yapay zekâ aşamasını süper zekâ boyutuna geçirebilecek çalışmalar yapılması, çeşitli zekâ oyunlarında yapay zekânın insan zekâsını mağlup etmesi gibi gelişmeler yapay zekânın tasarımının ve kullanımının etik boyutunu düşündürmektedir.



Yapay zekâ ile etik ilişkisi, tasarlanan yapay zekânın otonom olup olmamasıyla şekillenmektedir. Burada öncelikli amaç, evrensel ahlak yasalarını temel alan, insanlığa uygun bir yapay zekâyı tasarlamak ve kullanımını bu çerçevede sağlamak olmalıdır. Bir başka ifadeyle, etik bir yapay zekânın inşası için sorumluluk sadece yapay zekânın uyması gereken kurallardan ibaret olmamalıdır. Aynı şekilde kullanıcıyı da denetleyen bir mekanizma yapay zekânın etik tasarımı ve kullanımı için elzemdir. Günümüzde yapay zekâ ve etik çerçevesinde yapılan çalışmalar kullanıcı hak ve sorumluluklarını da kapsayan maddelerden oluşmaktadır.

Yapay zekâ sistemlerinin tamamen etik bir nitelik taşıması, öncelikle yapay zekâ uzmanlarının ve sonra kullanıcıların elindedir. Öte yandan etik bir yapay zekânın inşa edilmesiyle bu yapay zekânın doğrudan etik bir varlık kabul edilemeyeceği de önemlidir. Burada vurgulanmak istenen, ahlaki kodlarla tasarlanan bir yapay zekânın bu kodlara uysa da her zaman etik olanı yapamayacağı ve kodlarından bağımsız davranamayabileceğidir. Bazı görüşler yapay zekânın daima etik olanı yapma noktasına belirli süreçlerle geleceğini vurgularken bazıları da yapay zekânın özerk konumda olsa da etik sıfatı kazanamayacağını, yasalara uygun davranan bir yapay zekânın etik olamayacağını savunmaktadır. Ancak ahlaki kodlarla tasarlanan bir yapay zekânın ne kadar etik bir varlık olacağı öncelikli tartışma konusu olmamalıdır. Zira bu konu, tasarımcının etik kodlarla geliştirdiği bir yapay zekânın insanlar tarafından kullanımıyla ortaya çıkabilecek süreçlerden sonra tartışılmaktadır. Burada öncelikli amaç, tasarımcının etik kodlara uyan bir yapay zekâyı inşa etmesi ve kullanıcının kendisine sunulan, etik kodlarla donatılmış bu yapay zekâyı etik kullanım şartlarına uygun kullanmasıdır.

Yapay zekâ ve etik konusuna yönelik çalışmalarda etik bir akıllı makine yapımı için öncelikle 3 önemli madde uygulanmalıdır: Etik bir yapay zekâyı programlamak, bunun gerçekleşmesini sağlayan etik bir yapay zekâ uzmanını görevlendirmek ve yapay zekâyı etik kullanım şartlarına uygun kullanan bir kullanıcıya sahip olmak. Yapay zekâyı tasarlayan programcıyı ve yapay zekâyı satın alan kullanıcıyı hukuki yaptırımlarla kontrol altına almak bu alanda atılacak en önemli adımdır.

Yapay zekânın etik çerçevesinde değerlendirilmesi noktasında yazılımcı ve kullanıcı odaklı bir çerçeve ortaya çıkmaktadır. Yazılımcı ve kullanıcının “ahlaksal özne” sıfatına sahip olması, etik kurallara uyan bir yapay zekânın tasarlanması ve kullanılması için gerekli olmalıdır. Daha önce de bahsedildiği gibi ahlaksal özne tanımlamasındaki amaç, sadece kendi çıkarını değil, başkalarının da çıkarlarını ön planda tutan kişi veya yapay zekâ sistemlerini ifade etmektir. Bu kapsamda yapay zekâ teknolojilerinin tasarlanmasına ve kullanımına yönelik uygulanabilir normatif çalışmalar düzenlemek ve bunun yaptırımını sağlamak gelecekte oluşabilecek

etik sorunları da ortadan kaldıracaktır. Ayrıca yapay zekâ sistemlerine yönelik düzenlenen sosyal kuralların, hukuki alanda yapay zekâ sistemlerinin kullanımından kaynaklanan sorunları çözmeye ışık tutacağını da belirtmek yanlış olmayacaktır.

#### **Çıkar Çatışması Beyanı**

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan etmiştir.

#### **Kaynakça**

- Allen, C., Smit, I., Wallach W. (2005). Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches, *Ethics and Information Technology*, 7(3), 149-155.
- Allen, C., Varner, G., Zinser, J. (2000). Prolegomena To Any Future Artificial Moral Agent, *Journal of Experimental - Theoretical Artificial Intelligence*, 12(3), 251-261.
- Aristoteles, (1962). *The Politics*, (T. A. Sinclair, Çev.) Londra: Penguin Books.
- Asaro, P. (2006). What Should We Want From a Robot Ethic? *International Review of Information Ethics*, 6(6), 9-16.
- Asimov, I. (2018). *Ben Robot*, (E. Odabaş, Çev.) İstanbul: İthaki Yayınları.
- Batukan, C. (2017). *Robo-tizm, Robot, Android, Sayborg ve Yapay Zekâda Ruh Üzerine*, İstanbul: Altkırkbeş Yayınları.
- Bostrom, N. (2018). *Süper Zekâ – Yapay Zekâ Uygulamaları, Tehlikeler ve Stratejiler* (F. B. Aydar, Çev.). İstanbul: Koç Üniversitesi Yayınları.
- Dennet, D. (1997). *Consciousness in Human and Robot Mind, Cognition, Computation, and Consciousness*, Edited: Ito M., Miyashita Y., Rolls E. T. Oxford University Press, 18-29.
- Ersoy, Ç. (2019). *Robotlar, Yapay Zekâ ve Hukuk*, İstanbul: Oniki Levha Yayıncılık.
- Ford, M. (2018). *Robotların Yükselişi, Yapay Zeka ve İşsiz Bir Gelecek Tehlikesi*, (C. Duran, Çev.) İstanbul: Kronik Yayınları.
- Karabağ, M. (2021). *Dijital Televizyon Platformlarında Yayınlanan Bilim Kurgu Dizilerinde Yapay Zekâ Olgusu: Netflix Örneği, Yayımlanmamış Doktora Tezi*, İstanbul, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.
- Kutlusoy, Z. (2019). *Yapay Zekâ ve Gelecek*, Editör: Gonca Telli, 'Felsefe Açısından Yapay Zekâ', İstanbul: Doğu Kitabevi.
- Nabiyev, V.V. (2010). *Yapay Zekâ İnsan – Bilgisayar Etkileşimi*. Ankara: Seçkin Yayınları.
- Nilsson, J.N. (2018). *Yapay Zekâ: Geçmiş ve Geleceği*, (M. Doğan, Çev.) İstanbul: Boğaziçi Üniversitesi Yayınları.
- Öztürk, K. ve Şahin, M. E. (2018). *Yapay Sinir Ağları ve Yapay Zekâyâ Genel Bir Bakış*, *Takvim-i Vekayi Dergisi*, 6(2), 25-36.
- Ruffo, M. (2012). *The Robot: A Stranger to Ethics, The Machine Question: AI, Ethics and Moral Responsibility*, AISB/IACAP World Congress 2012, Ed. David J. Gunkel, Joanna J. Bryson, and Steve Torrance, Birmingham.
- Russell, S., Norvig, P. (1995). *Artificial Intelligence – A Modern Approach*, Pearson Education: New Jersey.

- Singer, P. W. (2015). Robotik Savaş, 21. Yüzyıldaki Robotik Devrim, (M. Erdemir, T.E. Erdemir, Çev.) Ankara: Buzdağı Yayınları.
- Ünal, E. (2018). Yapay Zekâdan Sanatçıya, Bilim ve Teknik Dergisi, Aralık 2018, 613, Ankara: Tübitak Yayınları, 58-66.

### İnternet Kaynakları

- Owsianik, J. "Lilly and InMoovator: Engaged Human-Robot Couple Want Right To Marry" <https://futureofsex.net/robots/lilly-inmoovator-engaged-human-robot-couple-want-right-marry/>  
Erişim Tarihi: 24.05.2021
- Demircan, K. "Ahlaklı Robotlar: Robotlar Etik Kararlar Alabilir Mi?" (Çevrimiçi) <https://khosann.com/ahlakli-robotlar-1-robotlar-etik-kararlar-alabilir-mi/>  
Erişim Tarihi: 02.05.2021
- Coeckelbergh, M. "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability" (Çevrimiçi) [https://www.researchgate.net/publication/336820072\\_Artificial\\_Intelligence\\_Responsibility\\_Attribution\\_and\\_a\\_Relational\\_Justification\\_of\\_Explainability](https://www.researchgate.net/publication/336820072_Artificial_Intelligence_Responsibility_Attribution_and_a_Relational_Justification_of_Explainability)  
Erişim Tarihi: 07.05.2021
- McKinsey Company, "Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages" (Çevrimiçi) <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>  
Erişim Tarihi: 15.06.2021
- 2016–2019 Progress Report: Advancing Artificial Intelligence R&D Artificial Intelligence Research - Development Interagency Working Group Subcommittee On Networking - Information Technology Research - Development Subcommittee On Machine Learning - Artificial Intelligence And The Select Committee On Artificial Intelligence Of The National Science - Technology Council (Çevrimiçi) <https://www.nitrd.gov/pubs/AI-Research-and-Development-Progress-Report-2016-2019.pdf>  
Erişim Tarihi: 11.05.2021

