

Yapay Zekâ: Sahte, Yalan ve Yanlış Sınırı Üzerine Bir Tartışma

Artificial Intelligence: A Discussion on the Boundary of the Fake, the Lie and the False

ÖZET

Bu çalışmada, genel bir perspektiften hareketle, yapay zekânın bilgiye erişim, gerçeklik algısı ve etik sorumluluk alanlarında yarattığı dönüşüm vurgulanmaktadır. Yapay zekânın yalnızca bilgi sunan bir araç olmaktan çıkarak, doğru-yanlış ve gerçek-sahte ayrımlarını bulanıklaştırdığı üzerinde durulmaktadır. Zekânın ne olduğuna dair değerlendirmeler, insan zihninin bütünselliği içinde ele alınmaktadır. Bununla birlikte, yapay zekânın insan zihnine olan benzerliği oldukça tartışmalı bir konu olarak öne çıkmaktadır. Yapay zekânın insan zihni gibi çalışmaya yaklaştıkça insanlığı ilerleteceğini savunan görüşler bulunurken, bu gelişimin insanlığın geleceğini tehdit edebileceğini ileri süren karşıt düşünceler de mevcuttur. Çalışmada, yapay zekânın gerçekten 'zekâ' olup olmadığından ziyade, insan ile kurduğu ilişkinin karanlık yönleri tartışılmaktadır. Çünkü zaman içinde bu tür teknolojilerin hem beklenen hem de öngörülemez olumsuz sonuçlar doğurabileceği görülmüştür. Deepfake ve algoritmik manipülasyon gibi uygulamalar, gerçeklik algısını zedelemekte; bu da etik ve hukuki sorumlulukları kaçınılmaz biçimde gündeme getirmektedir. İnsan, doğru bilgiye ulaşma arzusu taşıırken, bu istek, doğrulanamayan ve şüpheyle örülü bilgi yığınları içinde kaybolma riski barındırmaktadır. Yazılım geliştiricilerinin öngörülleri, çoğunlukla teknik hatalara odaklanmaktadır. Ancak bazı uygulamaların, var olmayan bilgi ya da haberleri üretmesi ve gerçek dışı ses ya da görüntüleri manipüle etme imkânı sunması, bu araçlara duyulan güveni ciddi biçimde sarsmaktadır. Sonuç olarak yapay zekâ, kaçınılmaz bir şekilde, insanlık açısından hem umut hem de tehdit barındıran bir kırılma noktasına işaret etmektedir.

Anahtar Kelimeler: Turing, Yapay Zekâ, Deepfake, Sahte, Yalan, Yanlış.

ABSTRACT

This study highlights the transformation brought about by artificial intelligence in the areas of access to information, perception of reality, and ethical responsibility, from a general perspective. It is emphasised that artificial intelligence has moved beyond being merely a tool for providing information, blurring the distinctions between true and false, and real and fake. Assessments of what intelligence is are addressed within the context of the wholeness of the human mind. However, the similarity of artificial intelligence to the human mind is a highly controversial topic. While some argue that artificial intelligence will advance humanity as it approaches functioning like the human mind, opposing views suggest that this development could threaten humanity's future. The study discusses the dark aspects of the relationship between artificial intelligence and humans, rather than whether it is truly "intelligent". This is because, over time, it has been seen that such technologies can produce both expected and unforeseen negative consequences. Applications such as deepfakes and algorithmic manipulation undermine the perception of reality, inevitably raising ethical and legal responsibilities. While humans have a desire to access accurate information, this desire carries the risk of getting lost in piles of unverifiable and dubious information. Software developers' predictions mostly focus on technical errors. However, the fact that some applications can generate non-existent information or news and manipulate unrealistic sounds or images seriously undermines trust in these tools. Consequently, artificial intelligence inevitably points to a turning point that holds both hope and threat for humanity.

Keywords: Turing, Artificial Intelligence, Deepfake, Fake, Lie, False.

GİRİŞ

Yapay zekâ teknolojisi, hayatımızın birçok alanında hızla yaygınlaşırken, beraberinde sahtecilik, yalan ve yanlış bilgi gibi karmaşık sorunları da gündeme getirmektedir. Bu teknolojinin sınırlarında giderek daha belirgin hale gelen "gerçeklik algısı" sorunları, yapay zekânın ürettiği içeriklerin doğruluğu ve güvenilirliği üzerine ciddi soru işaretleri yaratmaktadır. Sahte videolar, deepfake'ler ve sosyal medyada hızla yayılan manipülatif içerikler, yapay zekânın sadece araç olarak değil, aynı zamanda gerçeğin ve de gerçekliğin çarpıtılmasında da aktif bir rol oynayabileceğini göstermektedir. Turing'in taklit oyunu fikrinden başlayarak, bugün yapay zekânın hem mekanik hata hem de bilinçsiz yalan üretebilme kapasitesine sahip bir sistem haline gelmesi, teknolojinin etik ve hukuki boyutlarını da yeniden tartışmaya açmaktadır. Bu bağlamda, yapay zekânın "sahte" ve "yalan" arasındaki sınırları nasıl aşabildiği; yanlış bilgi üretiminin sonuçlarının toplumsal etkileri, yazılımcıların ve kullanıcıların bu süreçteki

Adem Ökten ¹

How to Cite This Article

Ökten, A. (2025). "Yapay Zekâ: Sahte, Yalan ve Yanlış Sınırı Üzerine Bir Tartışma" International Social Sciences Studies Journal, (e-ISSN:2587-1587) Vol:11, Issue:10; pp:1923-1929. DOI: <https://doi.org/10.5281/zenodo.17441429>

Arrival: 05 September 2025

Published: 25 October 2025

Social Sciences Studies Journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

¹ Dr.Öğr.Üyesi, Sivas Cumhuriyet Üniversitesi, Edebiyat Fakültesi, Felsefe Bölümü, Sivas, Türkiye, ORCID: 0000-0003-3441-8731

sorumlulukları gibi kritik sorular, teknolojinin gelişimi kadar önem kazanmış ve güncel tartışma konusu haline gelmiştir. Gerçeklikle bağımızın zayıfladığı, doğrulamanın giderek zorlaştığı bu dönemde yapay zekânın rolünü anlamak ve bu risklere karşı bilinçli çözümler geliştirmek, teknolojiyi daha güvenilir ve etik bir hale getirmek adına öncelikli hedef olmalıdır. Bu çalışma böyle bir düşüncenin ürünüdür.

ANA HATLARIYLA TEMEL PROBLEM

Yapay zekânın hem günlük hayatın içinde edindiği yer hem de sahip olduğu gelişme hızı popülarlığının her geçen gün artmasına neden oluyor. Yapay zekâyı bilimsel ve pratik eksenlerde savunanlar ile bu teknolojinin potansiyel tehlikelerine karşı eleştirel yaklaşımlar arasındaki tartışmalar; yapay zekânın kontrolden çıkma riskinden, yeni bilimsel keşiflerin önünü açma potansiyeline kadar oldukça geniş bir yelpazede sürmektedir. Bu açıdan bakıldığında, tartışmalarda tarafların birbirlerine yönelttiği sorular ve ortaya koydukları eleştirilerin yapay zekâ alanındaki gelişmeleri yönlendiren önemli dinamikler arasında etkin bir şekilde yer aldığı gözlemlenmektedir.

Leibniz'in "1679'dan itibaren, zihninde yeni bir buluş şekillenmeye başlamıştı: ondan kayda değer buluşlar elde edeceği ikili veya binary aritmetik... Leibniz'in görüşü zamanının çok ilerisine uzanmıştı." (Antognazza, 2013: 215). Abaküsten kuantum bilgisayarlarına uzanan süreçte, Leibniz ve Pascal'ın mekanik hesap makinelerinden sonra, Charles Babbage'ın buhar gücüyle çalışan otomatik hesaplama makinesi, hesaplamaların daha hızlı ve kolay yollarla gerçekleştirilmesine yönelik temel motivasyonu sağlamıştır. II. Dünya Savaşı sırasında, Alan Turing tarafından tasarlanan Bombe cihazının, Alman ordusunun kullandığı Enigma şifreleme sistemini çözmesiyle birlikte, yalnızca hesaplama yapan makineler değil, aynı zamanda analitik işlem yapabilen sistemler de kullanılabilir hâle gelmiştir. Leibniz'in öncülüğünü yaptığı ve George Boole tarafından geliştirilen cebir sistemi, zamanla elektrik devrelerinde uygulanmaya başlanmış; bu gelişmeyle birlikte hem veri girişinde hem de veri çıkışında kullanılabilecek makine dilinin mümkün olduğu fikri ortaya çıkmıştır. Bilgisayar yazılımlarının geliştirilmesinden önce, mantık insan düşüncesini düzenlemeye yönelik bir araç olarak değerlendirilirken, yazılım dilleri sayesinde mantık artık basit kodlamalar aracılığıyla belirli işlevleri yerine getiren sistemler için temel bir yapı hâline gelmiştir.

Boole'ın cebirinde 0 (sıfır) yanlış 1 (bir) ise doğruyu temsil etmektedir. Bu değerlendirme elektrik devrelerindeki ise 1 (açık) 0 (kapalı) biçiminde karşılıklar almaktadır. Teknolojik gelişmelere ve farklı kod yazım dillerinin geliştirilmesinin standartlarının belirlenmesi sonrasında insanın hayal gücü sınırlarında gezen bir yazılım geliştirme çağına girilmiş oldu. Bu bağlamda, günümüzde kullandığımız tüm elektrikli ve elektronik sistemler, "makine dili" olarak adlandırılan bir dil aracılığıyla çalışmaktadır. Geliştirilen arayüz yazılımları, örneğin şu an okumakta olduğunuz metni, makine dilinde kodlanmış verilerin insanlar tarafından anlaşılabilir versiyonlarına dönüştürmektedir. Örneğin '*bilgisayarın serüveni gerçekten hayret verici*' gibi bir cümleyi makine dilinde okumak isteseydik şöyle bir yazımla karşılaşırız:

01100010	01101001	01101100	01100111	01101001	01110011	01100001	01111001
01100001	01110010	100110001	01101110	00100000	01110011	01100101	01110010
11111100	01110110	01100101	01101110	01101001	00100000	01100111	01100101
01110010	11100111	01100101	01101011	01110100	01100101	01101110	00100000
01101000	01100001	01111001	01110010	01100101	01110100	00100000	01110110
01100101	01110010	01101001	01100011	01101001			

"İki değerli temeli Aristoteles'e dayanan mantık, çok değerli mantığa dönüştürüldüğünde yazılım nasıl karar verir?" sorusunun sınırlarını geçen matematikçi Lukasiewicz, "gelecek zamandaki olaylara ilişkin önermelerin şimdilik doğruluk değerinin doğru veya yanlış olmayıp, belirsiz olduğunu" ileri sürmüştür (Grünberg, 2000: 295). Bu durumda doğruluk değerleri kümesi; doğru, yanlış ve belirsiz olmak üzere üç değerli bir sistem olarak ifade edilir. Boole cebirindeki doğru ve yanlış temsil eden 1 ve 0 arasındaki tüm kesirli değerler ise belirsiz olarak tanımlanır. Lotfi A. Zadeh'in puslu/bulanık kümeler teorisinin, günümüzde yaygın biçimde bilinen ifadesi olan "bulanık mantık" (fuzzy logic), 1970'lerin ortalarında Ebrahim H. Memdani'nin bir buhar makinesi için geliştirdiği bulanık denetleyiciyle birlikte uygulanabilir bir örneğe kavuşmuştur. "O zamandan beri 'bulanık mantık' terimi, bulanık kümelerle çalışan herhangi bir matematiksel ya da bilgisayar sistemini tanımlamak için kullanılmaktadır" (Kosko, Isaka, 1993: 77). Özellikle, bir şeyin bir kümeye ait olması ya da olmaması biçiminde ikiye ayrılan bir fiziksel dünya anlayışının gerçekliği tam olarak yansıtmadığı yönündeki iddia (Yüksel, 2011: 526), kesinlik kavramının yerini belirsizliğe bırakmasına zemin hazırlamaktadır.

Çalışan bir bilgisayar yazılımının, bulanık mantık temelinde geliştirilmiş bir algoritmaya sahip olması sayesinde, bu yazılımın kullanıldığı araçların verimliliğinin arttığı, kullanım alanlarındaki örneklerden anlaşılmaktadır. "Bulanık ürünler, hem bulanık çıkarım algoritmalarını çalıştıran mikro işlemciler hem de değişen giriş koşullarını ölçen sensörler kullanır" (Kosko, Isaka, 1993: 78). Bulanık mantık temelinde inşa edilen algoritmaların

hesaplamadaki kullanışlılığı; 1 ve 0 arasındaki sonsuz sayıdaki değeri içeren bir sistemle veri girişine ve işlemeye olanak sağlaması, yapay zekânın arkasında yatan gelişme ivmesini de anlaşılır kılmaktadır. Sadece doğru ve yanlış değerleriyle çalışan yazılımların aksine, belirsizlikleri de hesaba katabilen yazılımlar, otonom ya da yarı otonom sistemler olarak insanoğlunun hayatına girmiştir. Bu durum, söz konusu yazılımların popülerliğinin bir nedeni olarak, belirsizlik ve tahmin edilemezlik unsurlarına işaret edilebileceğini göstermektedir. Kullanıcısına yardımcı olmayı amaçlayan yarı otonom sistemlerin aksine, otonom sistemler karar verme işlevine sahip olmalarıyla, kullanıcılarını devre dışı bırakma potansiyeline sahiptir. Otonom bir sistemin kapatma düğmesinin işlevsiz hâle gelmesi, sisteme yüklenen amaçları gerçekleştirme sürecinde, yine sistemin kendisi tarafından bu düğmenin devre dışı bırakılmasıyla mümkün olabilmektedir (Russell, 2019: 141, 172).

Bugün artık şu soru bizim için anlamsız görünmemektedir: Bilgeliği seven insan artık kendi bilgeliği yerine Big Data'yı kullanmayı mı seviyor? Ya da soruyu şu şekilde formüle edebiliriz: *Günümüzde bilgelik artık vekâlet bilgeliğine mi dönüştü?* İnsanın Aristoteles'in vurguladığı biçimiyle bilgiyi arzulaması ve bilginin kendisi için ulaşılabilir olduğunu gerekçeleriyle ortaya koyabilmesi birbirinden farklı şeylerdir. Bugün bilgiye yönelik arzularının ne kadar canlı olduğu sorusunu bir kenarda tutarak onun bilgiyle ilişkisinin daha problematik hale geldiğini iddia etmek mümkün görünmektedir.

“Aristoteles, teorik bilgelik (sophia) ile pratik bilgelik (phronesis) arasında bir ayrım getirdi. İlki, kişinin gerçekliğin doğasını nihai nedenleri (metafizik) açısından kavramasını sağlayan entelektüel erdemdir. İkincisi ise, kişinin yaşamın gidişatına ilişkin sağlam yargılarda bulunmasını sağlayan nihai pratik erdemdir. İlki, derin anlayış ile geniş bilgi arasındaki karşıtlığı ortaya koyarken, ikincisi sağlam yargı ile sadece teknik beceri arasındaki karşıtlığı ortaya koyuyordu. Teorik ve pratik bilgelik arasındaki bu ayrım, Orta Çağ boyunca devam etti ve günümüze kadar sürdü” (Delaney, 1999: 976).

Bilginin hem en yüksek düzeyini vurgulamak hem de eylem merkezli problemlerde sağlam muhakeme yeteneğini belirtmek için “bilgelik” terimini kullanmaya devam ediyoruz. O halde bilgeliği tam anlamıyla insanlık tarafından kabul edilen insanlar neden artık yoklar? Belki de bilginin çoğalması, bilimsel yönetime dayalı çalışmaların gelişmiş araçlarla kolayca yürütülmesi, kâşif ve mucitlerin buldukları yeniliklerin heyecanı vb. gelişmeler, bilgeliğin yol göstericiliğine olan ihtiyacı ortadan kaldırdı. Sayfalar dolusu metni yazan insanlardan, onları okuyan insanlar çok şey öğrendi. El yazmalarının yine el yazması olarak çoğaltılması, kaynaklara erişimi zorlaştırırken; matbaanın icadı, kitapların çoğaltılmasını daha pratik hâle getirdi ve kaynaklara erişimi de eskiye nazaran kolaylaştırdı. Kitapları yazdıkları dilde okuyabilecek şekilde iyi bir dil eğitiminden geçen düşünürler de zamanla, eserleri yazanlar konumuna erişebildiler. Yirminci yüzyıla gelindiğinde ise, dijital arşivler aracılığıyla erişimin fazlalaşması, doğru orantılı olarak yeni düşünce ve düşünürlerin sayısının artmasını sağlayamadı. Bunun nedenlerini araştırmak bu çalışmanın kapsamının dışında olmakla birlikte, ilgilendiğimiz nokta, arama motorlarının sahip oldukları erişim ağı sayesinde hızlıca sonuçlar vermesi ve elle dokunmanın bile zarar verebileceği eserlere bir ekran aracılığıyla hızlıca ulaşılabilmesidir.

Bilgiye hızlı bir şekilde ulaşmak, teknolojinin sunduğu kolaylaştırıcı çözümlerden biri olarak oldukça çekici bir özelliğe sahiptir. Geline son noktada, yapay zekâ uygulamaları neredeyse her şeye erişimi daha da kolaylaştırmıştır. Bir yazının istenilen dile çevrilmesi artık şaşırtıcı gelmemektedir; çünkü konuşmayı anlık olarak çeviren cihazlar ve yazılımlar kullanılmaktadır. İnsanın eğitim alması bir yana, düşünmesine bile gerek bırakmayacak biçimde bilgisayar yazılımları üretmek, hatta bilimsel makale yazdırmak dahi mümkündür. Bazı insanlar için yapay zekâ uygulamalarının “bilge” olarak görülmeye başlandığı bu aşama, bir kırılma noktasının da işaretçisi sayılabilir. İnsan, binlerce yıllık bilgi birikimini edinmesini sağlayan aklını veri tabanlarını kiraya veren; ancak zekâsını kullanma konusunda kiracı konumuna dönüşmüş haldedir. Sahte-gerçek, yalan-gerçek ve yanlış-doğru ayrımlarını yapmasını sağlayan zekânın artık uygulamalarla sınırlanması, günümüzde varlığı inkâr edilemeyecek kaygılar taşımamıza neden olmaktadır.

İnsan bilgisi, şimdiye kadar doğru ve yanlış olan önermelerin ayrımı üzerinden ilerlemiştir ve tarih içerisindeki gelişimini de bu temelde sağlamıştır. Platon'dan bu yana “gerekçelendirilmiş doğru inanç” olarak genel kabul gören bilgi tanımı, yapay zekâ uygulamalarıyla birlikte daha artan yoğunlukta doğrunun denklemden çıkarılarak bilgiye bakışın, belki de gerekçelendirilmiş bir inanç olarak görülmesinde çok etkin bir dönüşüme yol açmıştır. Bu türlü problematik bir eksenden yapılan bir değerlendirme içerisinden bakarak, yapay zekânın sahnesinde sahte, yanlış ve yalananın hem insanlar arası ilişkilerde hem de teknoloji ile ilişkimizde kabul edilemeyecek pek çok olumsuzluklar barındırdığını ve insanlığı gerçek ve gerçeklikten uzaklaştırdığını iddia etmek mümkündür.

YAPAY ZEKÂ: YALANDAN SAHTECİLİĞE

Kavramların anlamlarının belirgin olmaması, o kavramlarla örölü düşüncelerin de belirsiz ve bulanık olmasına neden olur. Düşüncelerin ifadesi olan önermelerin doğru olduğu varsayımı ise akıl yürütmelerin önündeki en büyük engellerden biridir. Zekâ çoğu zaman akılla eş anlamlı olarak değerlendirilir; ancak bu iki kavram, zihnin farklı bölümlerine/yeterlerine (*faculty*) işaret eder. *Saf Aklın Kritiği* adlı eseriyle Kant, insan zihninin bütünsel bir resmini sunar. Kant, zekâ kavramını daha çok akılsallık (Kant, 1992: 384) olarak değerlendirmiş olsa da “üç bilişsel yeti” arasında ayırım yapar: duyarlılık, anlama yetisi ve akıl. Duyarlılık, nesnelerin bize doğrudan sezgi yoluyla verildiği yetidir. Anlama yetisi, kavramlar aracılığıyla yargılama yetisidir. Bu temelde, yalnızca nesnelerin anlama yetisi kavramları altına alınmasını içeren bu tür yargılar aracılığıyla, duyarlılık yoluyla verilen nesneler düşünülebilir. Akıl ise, ilkeler ve çıkarımlar yetisidir; anlayışın kavramlarını ilkeler aracılığıyla sistemleştirmeye çalışarak, anlayışın kavramsal şemasını birleştirmeye çalışır. Kant’a göre, nesnelerin deneyimi ve bilişi hem duyarlılığı hem de anlamayı gerektirir. Çünkü deneyim, duyular tarafından zihne sunulan nesnelerin kavramlar altında sınıflandırılmasını içeren nesne bilgisine dayanır (Thorpe, 2015: 217–218). Filozofların kendi felsefelerini temellendirme yönündeki kavramsal hamleleri onların felsefelerinin ayırt edici karakterini ortaya koymaktadır. Bu anlamda Kant’ın bu bölümlenmesi, bütün felsefesinin giriş anahtarını temsil etmektedir. Daha çok psikolojinin alanına ait olduğu düşünülen “zekâ”nın ne olduğu hâlâ bir tartışma konusudur.

Zekâ (*intelligence*), “Zihnin; hafıza, hayal gücü ve kavramsal düşüncüyü kullanarak, karşılaştığı pratik ve teorik sorunları etkili bir şekilde çözme kapasitesi” olarak tanımlanır (Wood, 1942: 147). Zekânın bu tanımı, onun bir yetenek/hüner/beceri olarak anlaşılmasına da imkân verir ve insan ile diğer hayvanlar arasında bulunabilecek bir seviye farkını anlamlı kılar. Reese, zekâ tanımlamasına ilişkin problemi biraz daha geniş bir haliyle şu şekilde dile getirir:

“Bence ‘yapay zekâ’ iki problem hariç müthiş bir terim: ‘yapay’ kelimesi ile ‘zekâ’ kelimesi. Problem, ‘yapay’ kelimesinin çok farklı iki anlama sahip olması ve hangisinin kullanıldığının belli olmaması. ‘Yapay çim’ ifadesindeki ‘yapay’ gibi, zekâyâ benzer fakat gerçekte zekâ olmayan bir şey anlamında mı; yoksa buradaki ‘yapay’ kelimesi, gerçek ancak doğal veya biyolojik olmayan bir şey anlamında mı anlaşılmalı?” (Reese, 2018: 76).

Tartışmaların çevresinden dolaşmak amacıyla, yapay zekânın türlerine ve işlevlerine göre yapılan tanımlamalar yerine, burada yapay zekâyı yalnızca bir yazılım olarak kabul edeceğiz. Yapay zekânın bir yazılım olarak ele alınması, onun insan zekâsı gibi işleyen bir yapıya sahip olup olmadığı ya da gelecekte böyle bir özelliğe sahip olup olmayacağı sorusunu peşinen reddetmek anlamına gelebilir. Öte yandan, fütüristik bir bakış açısıyla değerlendirildiğinde, insan zekâsına benzer bir yapıya sahip yapay zekânın geliştirilmesi çok da uzak bir ihtimal olarak görünmemektedir. İnsan zekâsı ile yapay zekâ arasındaki farkların azaltılması ve benzerliklerin artırılması yönünde yürütülen tartışmalar açısından, bu gelişmelerin yönünü belirleyecek önemli dönüm noktaları bulunmaktadır. Bu bağlamda, 1900 yılında Hilbert’in ortaya attığı meydan okumanın özü olan, “matematiğin eksiksiz, tutarlı ve kararlaştırılabilir olduğu” düşüncesi; Gödel’in 1931 yılında yayımladığı eksiklik teoremleri ile sarsılmış ve böylece çözülmesi gereken temel bir problem açığa çıkmıştır (Larson, 2022: 28–29).

Turing, Gödel’in ulaştığı sonuçlardan yola çıkarak, Hilbert’in ortaya attığı problemlerin son kısmı olan formel sistemler için evrensel bir kararlaştırma yordamının var olamayacağını kanıtlayarak yeni bir yola girmiştir. “Turing, matematiğin kararlaştırılabilir olduğu iddiasını çürütmek isterken kesinlik arz eden ve mekanik bir şey icat etmişti: bilgisayar” (Larson, s. 30). Turing’in doktora tezinde sezgi (*intuition*) ve hüner (*ingenuity*) üzerine olan düşünceleri, yayımladığı *Computing Machinery and Intelligence* başlıklı makalesinde artık bir bilgisayarla insanın ayırt edilemeyeceği sorununa, başka bir deyişle Turing testi olarak bilinen bir belirlenim testine dönüşmüştür. Sezgi ve hünere sahip bir bilgisayar mümkün müdür?

Turing’in çalışması bugün bile okuyucuları için hayal güçlerini besleyen bir yapıdadır. Onun önerdiği taklit oyunu (*imitation game*) (Turing, 1950: 433) ile önerilen algoritmaların belirli bir amaca yönelik olarak çalışması, insan zekâsının bir taklidi olarak değerlendirilmesine neden olur. Yazılımın yapmasını beklediğimiz şey, onu hangi amaçla oluşturduğumuza bağlıdır. Ancak amacı gözden kaçırarak bir yazılımcı için, amacına ulaştırmak için kullandığı araç(lar) amacın kendisiymiş gibi görünmeye başlayabilir. Örneğin, içerik seçimine ilişkin algoritmalar sosyal medyada ve internet sayfalarında kullanıcıya uygun içeriklerle karşılaştırmak amacıyla yapılmış olabilir; ancak bir süre sonra bu amaç, yerini hangi içeriklerle karşılaşmasının izleme sayısını ve o uygulamada kalma süresini artıracaklarını belirlemeye yönelik bir amaca kolaylıkla bırakabilir. Artık kullanıcıya yönelik olarak geliştirilmiş uygulama, araç olmayı aşarak gelir elde etme amacına yönelmiş bir uygulamaya dönüşür. Turing’in izinden giden yazılımcılar için insanın mekanik bir yazılıma benzetilmesi ve onun zekâsını taklit etmeyi amaçlarken kullanılan araçların (neredeyse bir yazılımla çalışan bütün cihazlar) insan zekâsının yerini alabilecek birer araç olmasının hedeflenmesi de aracın amacın yerini almasına benzemektedir.

Bir algoritma yapısı gereği işlem adımları içerir ve bu, temel yazılım dili için de aşamaların sıralanmasının ifadesi olarak anlaşılabilir. Mantık için doğruluk değerine karşılık gelen doğru (*true*) 1 ile yanlış (*false*) ise 0 rakamlarıyla temsil edildiğinde; yine mantıkta indirgenmiş olarak kullanılan ve (*conjunction*), veya (*disjunction*) ile ise (*conditional*) önerme eklemleri aracılığıyla oluşturulan ifadeler, kodlamanın temelini oluşturur. Algoritma son halini alana kadar yapılan her düzeltme, onun tam olmasını, yani programcının zihnindeki tasarımla uygun hale gelerek mükemmel olmasını sağlama çabasıdır. Makinelerin hata yapamaz iddiası üzerinde duran Turing için bu hatalar, işlev hatası (*errors of functioning*) ve sonuç hatası (*errors of conclusion*) olarak sınıflandırılır. Ona göre “işlev hataları, makinenin tasarlandığından farklı davranmasına neden olan mekanik veya elektriksel bir arızadan kaynaklanır” (Turing, 1950: 449). Her ne kadar onun sınıflandırmasında mekanik veya elektriksel arızalar işlev hatalarına neden oluyor gibi görünse de, günümüzde yazılımcıların algoritmaları oluştururken yaptıkları kötü veya eksik kodlamalar, mimari ya da uygulamanın sağlıklı çalışması esnasında ortaya çıkabilecek hatalar (*bug*) ya da yazılımların kullandığı veri tabanlarını da tehlikeye atan güvenlik açıkları, hata ayıklama (*debug*) çalışmaları kapsamında işlev hatalarının düzeltilmesi için yapılan çalışmalar olarak değerlendirilebilir. Sonuç hataları ise ona göre yalnızca makineden gelen çıkış sinyallerine anlam yüklendiğinde ortaya çıkar (Turing, 1950: 449). Program çıktılarının hatalı olması olarak anlaşılabilir bu değerlendirmesi üzerinde daha fazla durmayan Turing, sonuç hatalarının programcılar tarafından bilinçli olarak yapılabilecek hatalar olmasına ise değinmez.

Bir yazılımın istenilen şekilde sonuçlar ürettiği, ancak gerçek sonuçlar olmaması mümkün müdür? Örneğin, bir dijital tartının gerçek ağırlığı 900 gram iken sonuç ekranında 1000 gram göstermesi; hata, yalan, yanlış, sahte veya dolandırıcılık seçeneklerinden hangisi olarak değerlendirilmelidir? Bu sayıların hepsi olarak değerlendirilmesi mümkün müdür? Hileli bir tartı yazılımı üretmekten fazlası yapılabilir mi? 2015 yılında ABD’de Çevre Koruma Ajansı (EPA) tarafından yapılan ihlal bildirimi (www.epa.gov) ile Volkswagen Grup tarafından ABD’de satılan 482.000 aracın, sahte sonuçlar üreten yazılıma sahip parçalar nedeniyle emisyon değerlerinin farklı gösterildiği ortaya çıktı. Bu durum, insan sağlığına ciddi zararlar verebilecek eylemlerin bile doğruymuş gibi olduğu, yazılım yoluyla mümkün olduğunu gösterdi. 2009-2015 yılları arasında on bir milyon dizel motora kurulan bu yazılım (Oldenkamp, Zelm ve Huijbregts, 2016: 121) hukuki olarak şirket yöneticilerinin ceza almasına yol açsa da bu yazılımı geliştiren mühendisler pek tartışmadı. Yöneticilerin talimatlarını yerine getiren çalışanların bu dolandırıcılıkta sorumluluğunun olup olmadığı ise hukuki bir tartışma konusu oldu. Braunschweig Bölge Mahkemesi’nde görülen davada, görevli mahkeme başkanı Christian Schütz’e göre, çok sayıda kişi tarafından geliştirilen motorlardaki sahteciliğin, genelleştirilemeyecek ortak sorumluluk olarak ele alınması ve suçlama yöneltmemiş başka kişilerin de bulunması gerektiğine inanılması gerekiyor. (www.lto.de [a]). Kötü amaçlı yazılımları kullanmayı isteyenlerin yanı sıra, bu kötü amaçlı yazılımları geliştirenlerin de yargılanması gerekmektedir. Volkswagen olayında olduğu gibi, “manipülasyon” gibi farklı ifadeler kullanılması, bu yazılımların kötü amaçlı yazılım olmalarını engellemez.

Bir kazanç elde etmek amacıyla kötü amaçlı bir yazılım kullanmak dolandırıcılık olarak değerlendirilirken, kötü amaçlı bir yazılımı geliştiren yazılımcı ise tam da istenilen amaçlara hizmet eden bir yazılım geliştirerek, bir anlamda iyi bir hacker olduğuyla övünebilir. Finans alanında yapılan dolandırıcılıklarda da yine yazılım desteği kullanılmaktadır. Yazılımın kullanılması, kullanıcılara zarar vermesi ortaya çıkana kadar, olayın üç tarafının da mutlu olması muhtemeldir: yazılım aracılığıyla dolandırıcılık yapan, yazılımı geliştiren ve kullanıcı. Olması gerektiği gibi çalışan bir yazılımın geliştirilmesinde yazılımcı kendi değerler sistemiyle hareket eder. Programın tekrarlayarak “0=1” yazdırmasını sonuç hatası olarak değerlendiren Turing, kullanıcı hesaplarında hiç para yokken yüzbinlerce TL gören kullanıcıların durumunu ön görememiş olabilir. Duyu verilerine güvenen kullanıcı, ekranda gördüklerini doğru olarak kabul edecektir.

Yalan söylemek, genellikle bilerek yapılan bir eylem olarak değerlendirilir (Selim, 2022: 704). O halde, yazılımcılar kendi bilinçli tercihleri doğrultusunda işleyen bir kodlama yaptığında, kandırma ve gerçekliği çarpıtma durumlarında yalancı sayılmazlar mı? Dolandırıcı, hacker ya da başka isimlerle anılsalar da yaptıkları eylemin özü değişmediği için yalancı olarak nitelendirilmeleri mümkündür. Bu durum, bilgisayarın ev kullanıcısına sunulmaya başlamasıyla birlikte ortaya çıkmış olsa da yapay zekânın sahte, yalan ve yanlış çıktılar üretmesi sorunun artık insan kontrolünden çıktığını gösteriyor. Yapay zekânın yapabileceklerine dair geleceğe ilişkin endişelerin gerçekleşmeye başladığını düşünebilir miyiz? Soru, masum bir eğlence uygulaması gibi görünen deepfake’ler ve özellikle sosyal medyada kullanılan “bot” olarak adlandırılan sahte kimlik hesaplarıyla, artık gerçeğin ne olduğunun ayırt edilemediği bir duruma evrilmiştir. “Deepfake” terimi, derin öğrenme (*deep learning*) ve sahte (*fake*) kelimelerinin birleşiminden oluşur ve gerçekçi video ile ses içeriklerini ifade eder. Kişisel Verileri Koruma Kurumu tarafından yayınlanan bilgi notu (www.kvkk.gov.tr), tehlikenin boyutunu açıkça ortaya koymaktadır. Bir içeriğin yayılması veya beğeni sayısının artırılmasıyla gündemde üst sıralara çıkmasını sağlayan bot hesaplar ise, insanların maruz kaldığı içerikleri sürekli erişime sunarak etkileşim yaratmaktadır. Yakın gelecekte, telefon görüşmeleri yaparken bile gerçekten o kişiyle görüşülüp görüşülmediğinden şüphe duyulmaya başlanacaktır.

Yapay zekânın sahte içerikler üretmek için insanlar tarafından kullanılması, tehlikenin yalnızca küçük bir kısmını gösterir. Doğal Dil Üretimi (*Natural Language Generation*) ve Doğal Dil İşleme (*Natural Language Processing*) teknikleri, metin özetleme, diyalog oluşturma, üretken soru-cevaplama (*Generative Question Answering*), veriden metin üretme ve makine çevirisi (*Machine Translation*) gibi birçok alanda kullanılır (Ji, Lee vd., 2023: 1-2). Asıl tehlike ise yapay zekâ uygulamalarının Doğal Dil Üretimi teknikleriyle gerçek olup olmadığı neredeyse imkansızca anlaşılan, “halüsinasyon” olarak adlandırılan gerçek dışı içerikler üretmesidir. Bu halüsinasyonlar, yapay zekânın performansını olumsuz etkilediği ve gerçek dünya uygulamalarında güvenlik endişelerine yol açtığı için kaygı vericidir (Ji, Lee vd., 2023: 2). Yapay zekânın tıbbi uygulamalarda kullanılması, bu halüsinasyonlardan habersiz sağlık çalışanları tarafından yanlış uygulamalara ve dolayısıyla hastaların zarar görmesine neden olabilir (Hatem, Simmons ve Thornton, 2023: 1). Yapay zekâların “halüsinasyon” göstermesi psikolojik halüsinasyon terimiyle benzerlik taşısa da yapay zekâ yanlış bilgilendirmesi (*AI misinformation*) teriminin kullanılması daha uygun görünmektedir (Hatem, Simmons ve Thornton, 2023: 2). Ancak terimlerin anlamları göz önüne alındığında, yanlış bilgilendirmenin bir insan tarafından yapılması durumunda bunun bir hata olarak değerlendirilip uyarı yapılabileceği düşünülebilir. Yalanı bilinçli olarak söylenmek yerine gerçekliğin çarpıtılması olarak tanımlarsak, “yapay zekâ yalanı” (*AI lie*) terimi üzerine de düşünülebilir.

Peki, yapay zekâ uygulamalarının ürettiği sahte içeriklere hukuki yaptırım uygulanabilir mi? Hukuki yaptırım konusu, hukuk sistemi içinde görülen iki davada henüz net bir tavır veya uygulama olmadığını gösteriyor. New York Güney Mahkemesi’nde görülen Roberto Mata’nın bir havayolu şirketine açtığı dava (www.cbsnews.com) ve Köln Bölge Mahkemesi’nde görülen bir davada (www.lto.de [b]) avukatların yapay zekâ uygulamalarını kullandıkları tespit edilmiştir. Yanlış kullanım nedeniyle ve özellikle avukatların daha dikkatli olmalarına dair uyarılarla geçiştirilen bu örnekler, giderek artan yapay zekâ uygulamalarının sınırlılıkları ve algoritmalarının yanlış bilgi sunması ya da yalan söylemesinin insanlığa fayda sağlamasının yanında zararlarının da tartışılmasını zorunlu kılmaktadır.

SONUÇ VE DEĞERLENDİRME

Turing, taklit oyununun zamanla sahte ve yalan oyununa dönüşebileceğini sezmişti. Ancak o, bir programcı olarak bu durumu kendi dönemindeki dil kullanımıyla “hata” olarak adlandırmıştı; çünkü o zamanlar dil, yalnızca hataların var olmasına izin veriyordu. Günümüzde modern bilgisayarlar ve birçok yazılım, iki değerli mantık sistemine dayanıyor. Ancak yapay zekâ alanında, Lukasiwicz’in öncülüğünü yaptığı çok değerli mantık kapısından geçerek geliştirilen bulanık mantık sistemleri sayesinde yazılımcılar artık belirsizliğin karmaşık alanlarında çalışıyorlar. Kararlı (stabil) sistemlerin yerini, temelinde sahte-gerçek, yalan-doğru, yanlış-doğru gibi keskin karşıtlıkları görmezden gelen ve daha esnek ama daha belirsiz yapılar alan yazılımlar ortaya çıktı.

Yapay zekâ uygulamalarının, insanın gerçeklik algısı ve bilgiye olan güvenini zedelemesi, doğrulayıcı sistemler veya güvenilir kaynakları garanti eden yeni teknolojilere olan ihtiyacı artırabilir. Ancak insan zihninin dış dünya ile olan bağı ve gerçekliğin temel tanımını değiştirmek oldukça güç görünmektedir. Çünkü neyin doğru, neyin yanlış olduğunu ayırt etmek için öncelikle gerçekliğin ne olduğuna dair temel kabuller gereklidir. Bu kabullerde sarsılma yaşanması, derin sorgulamalara yol açacaktır. Yazılım dünyası, Turing’in ortaya koyduğu test için giderek daha yakın bir tarih öngörebilir; çünkü insan zihninde önemli bir kırılma noktası yaşanmakta ve gerçekliği çarpıtmak ya da yalan söylemek artık yapay zekâların da gerçekleştirebildiği bir eyleme dönüşmektedir. İnsanlar, her derdini danışabildiği, yaşamdan ve sosyal gerçeklikten kopuk bir zihnin inşasına ya da mevcut zihinlerinin çözülüşüne doğru ilerleyen uygulamalara giderek daha fazla bağımlı hale gelmektedir.

Yapay zekâ uygulamalarını salt bir yazılım olarak kabul ettiğimizde, onların kodlama hatalarına sahip olabileceğini kabul etmek zorundayız. Bu hatalar, yazılımı geliştirenlerin yanlışlarından kaynaklanır. Yazılımın yanlış sonuçlar üretmesi, tamamen kodlama hatalarıyla ilgilidir. Sahte yazılımlar da insanlar tarafından üretilebilir. Gelecekte yapay zekânın insan zihnini taklit etmesinin gerçekleşeceğini düşünenler için, bugün yapay zekânın yalan söyleyerek olmayan şeyleri varmış gibi göstermesi çoktan başlamış durumdadır. Yakında ise yapay zekânın kendi başına sahte içerikler üretmesi olasılığı gündeme gelecektir. Gerçekliğin ne olduğunu anlamakta zorlanan insanlar için, bu içeriklerin doğruluğunu denetleyecek yeni yazılımlara ihtiyaç giderek artacaktır. Yazılımcılar kodlardaki hataları düzeltebilirler ancak yapay zekânın sahte ve yalanlarının sınırları yanı sıra birleşiyor gibi görünmektedir. İnsanların gerekçelendirilmiş inanç ile yanlışla olan bağılıklarının da arttığı bir yalan çağı, yani sahtenin gerçeğin yerini aldığı ve yanlışın doğruya göre daha tercih edilir olduğu bir zaman içinde yaşıyoruz. Son olarak belirtmek gerekir ki, yapay zekâ sistemleri bulanık mantıkla çalışsalar bile, girdileri makine diline çevirerek işlerler. Derin öğrenme modelleriyle çalışan sohbet amaçlı yapay zekâ sistemlerinin çıktıları ise kullanıcıların anlayabileceği şekilde dönüştürülür. Bu dönüşüm süreci, yapay zekânın karmaşıklığını ve belirsizliğini bir nebze azaltır, ancak sahte ve gerçek arasındaki sınırın algoritmaları düzenleyen bulanık mantık nedeniyle giderek bulanıklaşması hâlâ önemli bir sorun olarak kalmaktadır.

KAYNAKÇA

- Antognazza, M. R. (2013). Leibniz (Çev: Orhan Düz), Türkiye İş Bankası Kültür Yayınları, İstanbul.
- Delaney, C. F. (1999). "Wisdom" (Ed: Robert Audi), The Cambridge Dictionary of Philosophy, s.976, Cambridge University Press, New York.
- Grünberg, T. (2000). *Sembolik Mantık El Kitabı* (c. 2), METU Press, Ankara.
- Hatem, R.; Simmons, B.; & Thornton, J. E. (2023). "A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks", *Cureus*, 15(9), e44720. <https://doi.org/10.7759/cureus.44720>
- Ji, Z.; Lee, N. & Frieske, R. (2023). "Survey of Hallucination in Natural Language Generation", *ACM Computing Surveys*, 55(12), 1-38. <https://doi.org/10.1145/3571730>
- Kant, I. (1992). "On the Form and Principles of the Sensible and the Intelligible World [Inaugural Dissertation]" (Trans&Ed. David Walford), *Theoretical Philosophy, 1755-1770*, ss.373-416, Cambridge University Press, New York.
- Kosko, B. & Isaka, S. (1993). "Fuzzy Logic", *Scientific American*, 269(1): 76–81.
- Larson, E.J. (2022). *Yapay Zekâ Miti -Bilgisayarlar Neden Bizim Gibi Düşünmez* (Çev: Kadir Yiğit Us), Mak Grup Medya Pro.Rek.Yay.A.Ş., Ankara.
- Oldenkamp, R.; Zelm, R.& Huijbregts, M. A.J. (2016). "Valuing the human health damage caused by the fraud of Volkswagen", *Environmental Pollution*, 212: 121-127.
- Reese, B. (2018). *Yapay Zekâ Çağı* (Çev. Mihriban Doğan), Say Yayınları, İstanbul.
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*, Allen Lane, Great Britain.
- Selim, F. (2022). "Dürüstlük ve Yalancılık Üzerine: Dünyü, Bugünü ve Felsefi İncelenmesi", *Stratejik ve Sosyal Araştırmalar Dergisi*, 6(3): 703-719.
- Thorpe, L. (2015). *The Kant Dictionary*, Bloomsbury Publishing, Chennai.
- Turing, A. M. (1950). "Computing Machinery and Intelligence", *Mind*, 59(236):433-460.
- Wood, L. (1942). "Intelligence" (Ed. Dagobert D. Runes), *The Dictionary of Philosophy*, s. 147, Philosophical Library, New York.
- Yüksel, Y. (2011). "Kesinlik ve Puslu Mantık", *Sosyoloji Dergisi*, 3. Dizi (22): 517-531
- EPA, United States Environmental Protection Agency, www.epa.gov/sites/default/files/2015-10/documents/vw-nov-caa-09-18-15.pdf
- KVKK, Kişisel Verileri Koruma Kurumu, www.kvkk.gov.tr/Icerik/7807/Deepfake-Bilgi-Notu
- LTO [a], Legal Tribune Online, www.lto.de/recht/kanzleien-unternehmen/k/dieselskandal-straftprozess-lg-braunschweig-manager-vw-6kls-411-js-49032/15-2319
- LTO [b], Legal Tribune Online, www.lto.de/recht/juristen/b/ag-koeln-familiengericht-312f130-25-schriftsatz-ki-anwalt-berufspflichten
- www.cbsnews.com/news/lawyer-chatgpt-court-filing-avianca/