

FELSEFE DÜNYASI

2024 YAZ/SUMMER Sayı/Issue: 79

FELSEFE / DÜŞÜNCE DERGİSİ

Yerel, süreli ve hakemli bir dergidir.

ISSN 1301-0875

Sahibi/Publisher

Türk Felsefe Derneği Adına Başkan
Prof. Dr. Murtaza Korlaelçi

Türk Felsefe Derneği mensubu tüm Öğretim üyeleri (Prof. Dr., Doç. Dr., Dr. Öğr. Üyesi) Felsefe Dünyası'nın Danışma Kurulu/Hakem Heyetinin doğal üyesidir.

Felsefe Dünyası, her yıl Temmuz ve Aralık aylarında yayımlanır. 2004 yılından itibaren Philosopher's Index ve TÜBİTAK ULAKBİM/TR DİZİN tarafından dizinlenmektedir.

Felsefe Dünyası is a refereed journal and is published biannually. It is indexed by Philosopher's Index and TUBITAK ULAKBİM/TR DİZİN since 2004.

Editör/Editor

Prof. Dr. Hasan Yücel Başdemir (Ankara Üniversitesi)

Yazı Kurulu/Editorial Board

Prof. Dr. Murtaza Korlaelçi (Ankara Üniversitesi)

Prof. Dr. Sema Önal (Kırıkkale Üniversitesi)

Doç. Dr. Fatih Özkan (Ankara Hacı Bayram Veli Üniversitesi)

Doç. Dr. Muhammet Enes Kala (Ankara Yıldırım Beyazıt Üniversitesi)

Dr. Öğr. Üyesi Aynur Tunç (Ankara Yıldırım Beyazıt Üniversitesi)

Arş. Gör. Ahmet Hamdi İşcan (Ankara Üniversitesi)

Alan Editörleri/Section Editors

Prof. Dr. Ahmet Emre Dağtaşoğlu (Trakya Üniversitesi)

Doç. Dr. Fatih Özkan (Ankara Hacı Bayram Veli Üniversitesi)

Doç. Dr. Mehmet Ata Az (Ankara Yıldırım Beyazıt Üniversitesi)

Doç. Dr. Sebile Başok Diş (Necmettin Erbakan Üniversitesi)

Doç. Dr. Nihat Durmaz (Selçuk Üniversitesi)

Dr. Mehtap Doğan (Ankara Yıldırım Beyazıt Üniversitesi)

Dr. Muhammet Çelik (Ankara Sosyal Bilimler Üniversitesi)

Dr. Kenan Tekin (Boğaziçi Üniversitesi)

Dr. Nazan Yeşilkaya (Şırnak Üniversitesi)

Yazım ve Dil Editörleri/Spelling and Language Editors

Zehra Eroğlu (Ankara Üniversitesi)

Abdussamet Şimşek (Ankara Sosyal Bilimler Üniversitesi)

Hatice İpek Keskin (Ankara Sosyal Bilimler Üniversitesi)

Fiyatı/Price: 300,00 TL | **Basım Tarihi :** Temmuz 2024, 300 Adet

Adres/Address

Necatibey Caddesi No: 8/122 Çankaya/ANKARA

Tel: 0 (312) 231 54 40

<https://dergipark.org.tr/pub/felsefedunyasi>

Hesap No / Account No: Vakıf Bank Kızılay Şubesi

IBAN: TR82 0001 5001 5800 7288 3364 51

Tasarım / Design: Turku Ajans

Baskı / Printed: Uzun Dijital

Zübeyde Hanım, İstanbul Çarşısı, İstanbul Cd. No:48 D:48,

06070 Altındağ/Ankara

Tel: (0312) 341 36 67 | **Sertifika No:** 47865

DİL, DÜŞÜNCE VE TURING: DAVIDSON'IN MAKİNE ZEKASININ SINIRLARINA İLİŞKİN GÖRÜŞLERİ

LANGUAGE, THOUGHT AND TURING: DAVIDSON'S VIEWS ON THE
LIMITS OF MACHINE INTELLIGENCE

Cansu AKOĞLAN

Dr. Arş. Gör., Ankara Hacı Bayram Veli Üniversitesi Felsefe Bölümü, ORCID: 0000-0002-4111-3613,
e-mail: cansu.akoglan@hbv.edu.tr

Felsefe Dünyası Dergisi, Sayı: 79, 2024, ss. 147-164.

Geliş Tarihi: 10.05.2024 | Kabul Tarihi: 29.06.2024

DOI: 10.58634/felsefedunyasi.1481945

Araştırma Makalesi

Öz

Bu makale, ilk olarak 1950 yılında Alan Turing tarafından önerilen ve makine zekasını dilsel iletişimde insan davranışını taklit etme becerisine göre değerlendiren Turing Testi'ni incelemektedir. Turing Testi yapay zeka araştırmalarının temelini oluştururken, önemli felsefi itirazlarla da karşı karşıya kalmıştır. Bu makalenin odak noktası, Turing Testi'nin, temelde sosyal ve iletişimsel olan düşünce ve önermesel tutumların karmaşıklığını açıklamakta başarısız olduğunu savunan Donald Davidson'ın eleştirisidir. Tartışma, sofistike dil modelleri gibi bazı yapay zeka teknolojilerinin bu felsefi itirazlara ne şekilde yenik düştüğünün bir analizini içermektedir. Makalede, Turing Testi'nin makine "zekası" için pragmatik bir ölçüt sunmasına rağmen, Davidson tarafından öngörülen insan düşüncesinin nüanslı yeteneklerini tam olarak yakalayamadığı, dolayısıyla makine düşüncesini değerlendirmek için daha geniş kriterlere ihtiyaç duyulduğu sonucuna varılmakta ve makine bilişini değerlendirirken etkileşimsel ve bağlamsal dinamikleri göz önünde bulunduran bir yaklaşıma duyulan ihtiyacın altı çizilmektedir.

Anahtar kelimeler: Turing Testi, Önermesel Tutumlar, Donald Davidson, İnsan-Makine Etkileşimi, Üçgenleme

Abstract

This article examines the Turing Test, initially proposed by Alan Turing in 1950, to assess machine intelligence by its ability to mimic human behavior in linguistic communication. While Turing's Test has been foundational in artificial intelligence research, it has faced substantial philosophical scrutiny. The focus of this paper is on Donald Davidson's critique, which argues that the Turing Test fails to account for the complexities of thought and propositional attitudes that are fundamentally social and communicative. The discussion includes an analysis of how some AI technologies, such as sophisticated language models, challenge and yet fail to overcome these philosophical objections. The paper concludes that while Turing's Test provides a pragmatic measure for machine "intelligence," it does not fully capture the nuanced capabilities of human thought as envisioned by Davidson, thereby necessitating broader criteria for evaluating machine thinking. The implications for future AI development and testing are considered, underscoring the need for an approach that integrates interactional and contextual dynamics in assessing machine cognition.

Keywords: Turing's Test, Propositional Attitudes, Donald Davidson, Human-Machine Interaction, Triangulation

Giriş

Makinelerin düşünüp düşünemediğinin belirlenebilmesi ve hangi kıstasa göre makinelere zeka atfedebileceğimiz konusu halen bir tartışma konusudur. Makinelerin düşünüp düşünmediğini belirlemeye yönelik bir girişim olarak Turing, “Bilgiişlem Makineleri ve Zeka” (Computing Machinery and Intelligence) adlı ünlü makalesinde, bir makinenin zeki davranışlar sergileyip sergileyemediğinin değerlendirilmesi için daha sonra Turing Testi olarak anılacak bir yöntem öne sürmüştür. Bir makinenin konuşma ortamında bir insanla ayırt edilemez davranışlar sergileyip sergileyemediğini değerlendirmeyi hedefleyen test, yapay zekanın gelişimi için pratik bir ölçüt sunar. Turing Testi, sadece yapay zeka gelişimini yönlendirmekle kalmamış, aynı zamanda düşünenin kriterleri üzerine de uzun süreli tartışmalara neden olmuştur.

Turing’in, makinelerin zekasını, insanı ikna etme becerisiyle ölçme girişimini içeren işlevselci perspektifine, bilhassa zihin felsefesi alanındaki ünlü isimlerden pek çok itiraz gelmiştir. Bu itirazlardan biri de analitik zihin felsefesi ekolünün mensuplarından olan Donald Davidson’a aittir. Düşünmeye dair holistik bir perspektif benimseyen Davidson’a göre “düşünmek sosyal bir özelliktir ve ona yalnızca iletişim kuranlar sahiptir” (Davidson, 1982, s. 327). Davidson’ın düşüncenin doğası üzerine yaptığı inceleme ve özellikle önermesel tutumlarla ilgili çalışmaları, düşüncelerin, onlara sahip olan varlıkların ortak yaşam dünyasına sıkı sıkıya bağlı olduğunu öne sürer. Davidson’a göre, insan düşünce süreçlerinin temel taşı olan, inanmak, arzulamak ve niyet etmek gibi önermesel tutumlar ortak yaşam dünyasında birbiriyle iletişimde olmayan varlıklara atfedilemez. Davidson’ın Turing Testine yönelik bu yönde geliştirdiği itiraz, günümüz yapay zekasının gerçekten düşünüp düşünmediğini belirleme hususundaki ihtilafları anlamak bakımından önemini korur.

Bu makale Davidson’un, Turing Testi’ne yönelik itirazını inceleyerek bu itirazın yapay zeka ve makine düşünmesine ilişkin süregelen zorluklara ne şekilde katkıda bulunabileceğine yönelik bir bakış açısı geliştirmeyi hedeflemektedir. Davidson’un eleştirilerinin nüanslarını inceleyerek, düşünmenin ne şekilde belirlenebileceğine dair diyalogu zenginleştirmeyi ve bu felsefi düşünceler ışığında yapay zekadan beklentilerimizi gözden geçirmeyi amaçlamaktadır. Bu amaç doğrultusunda ilkin, Turing Testi’nin temel prensipleri, yapay zekâ alanındaki etkileri ve testin bilhassa sohbet robotlarının başarısını değerlendirmedeki rolü ele alınacaktır. Ayrıca, Donald Davidson’ın düşünceye ilişkin bütüncül bakış açısı eşliğinde önermesel tutumlar

hakkındaki fikirleri ve Turing Testi'ne yönelik eleştirileri incelenecek ve bu eleştirilerin modern yapay zekâ teknolojileriyle olan ilişkisine, Davidson'ın Turing Testi'ne yönelik itirazı çerçevesinde değinilecektir. Son olarak bu itirazın, genel olarak yapay zeka, özel olarak ise dil işleme modellerinin geldiği son noktada ne şekilde yol gösterici olabileceği üzerinde durulacaktır.

1. Turing Testi

Alan Turing'in 1950 yılında 'Bilgişlem Makineleri ve Zeka' isimli makalesinde sırayla iki soru sorar, bunlardan ilki, 'makinelere düşünebilir mi?' sorusuyken, diğeri 'taklit oyunu' olarak adlandırdığı oyunun, makineye karşı oynandığında elde edilen sonuçların, insan oyuncularla elde edilen sonuçlardan farklı olup olmayacağı sorusudur. Sorulan sorular arasındaki bu sıçrama, Turing'in ilk soruya verilecek yanıtın, makine ve düşünme kelimelerinin ayrı ayrı tanımlarını gerektirmesi ve bu tanımları, bu kelimelere yönelik genel anlayış (halk oylaması) göz önüne alınarak yapmayı 'tehlikeli' bir tutum olarak görmesinden kaynaklanır. Tek başına 'absürt' bir çözüm üretmekle kalmayacak bir soru sormak gereği böylece doğar ve Turing, taklit oyununun makineye¹ karşı oynandığında sonucun ne olacağı sorusunun peşine düşer (Turing, 1950, s. 433).

Taklit oyununun orijinal formu üç insanla oynanır. Ayrı odalarda bulunan bir kadın, bir erkek ve herhangi bir cinsiyetten bir sorgucu birbirleriyle yalnızca yazılı olarak iletişim kurmaktadır. Oyunun amacı, her iki oyuncunun da kendilerine sorulan sorulara cevap vererek sorgucuyu kadın olduklarına ikna etmeye çalışmasıdır. Bu sırada sorgucu, yine sadece sorular sorarak diğer iki oyuncudan hangisinin kadın, hangisinin erkek olduğunu doğru tahmin etmeye çalışır. Doğru tahmin oranı oyunun başarısının belirleyicisidir. Turing oyununun bu ilk formunu tasvir ettikten sonra oyuna bir değişiklik önerir ve oyunculardan erkek olanın bir makineyle yer değiştirmesi halinde sorgucunun kimin kadın olduğunu doğru belirleme oranında bir değişiklik olup olmayacağı sorusunu sorar, bugün Turing Testi olarak adlandırdığımız taklit oyununun ikinci versiyonu böylece ortaya çıkar (Turing, 1950, s. 434). Orijinal oyunda, Turing'in cinsiyet rollerini kullanım tercihinin metodolojik bir sebebe dayanıp dayanmadığını tam olarak bilemesek ve bu kullanım bugün modası geçmiş gibi görünse de, amacın fiziksel görünüşten bağımsız bir değerlendirme yapılmasını sağlamak ve yalnızca konuşma yeteneklerine

1 Turing'in makineden kastı düşünen makinelerdir ve düşünen makinelere yalnızca dijital bilgisayarları dahil eder. Bu türden makinelerin gelecekte ortaya çıkabilecek türden dijital bilgisayarlar olabileceğini ifade eder (Turing, 1950, s. 436). Bu nedenle, Turing imitasyon oyunun kaç oyuncu, kaç sorgulama girişimi vb olması gerektiği hakkında detay vermez. Bu anlamda Turing Testi bir düşünce deneyinin sahip nitelikleri taşıır ve empirik bir test olarak değerlendirilmemektedir.

odaklanmaya hizmet etmek olduğu söylenebilir. Yapısal benzerlik ortadan kaldırılıp, işlevsel benzerlik ön plana çıkacaktır (Harnad, 2003, s. 429). Makinelerin insanları taklit becerisine odaklanan testin amacının, bir makinenin işleyiş süreçlerine odaklanarak düşünüp düşünmediğini tesis etmek değil, makinenin, insan oyuncuyu kandırabilecek kadar maharetli cevaplar üretip üretmeyeceğini belirlemek olduğu söylenebilir. Makine, insanı ne ölçüde isabetli taklit ederse o ölçüde zeki olacaktır.

Makinelerin düşünüp düşünemeyeceği ile makinelerin insanları, insan olduklarına ikna edip edemeyeceği, yani insanı ne kadar iyi taklit edebilecekleri, birbirinden farklı iki sorudur. Turing bu iki soruyu birbiriyle eş tutmaz ancak ilk sorunun aksine, ikinci sorunun, terimleri tanımlama güçlüğüne düşmeyecek olmasını cazip bulur. Zira ilk soru, ‘makinelere düşünebilir mi?’, “üzerine tartışmayı hak etmeyecek kadar anlamsızdır” (Turing, 1950, s. 442). Turing’in ilk soruyu anlamsız bulup terk ederek ikinci soruya geçmesi bir tür işlemsel tanım girişimidir² (Dennett, 2003, s. 300). İşlemsel tanımlar, kavramsal tanımların aksine, bir fenomeni tanımlarken, o fenomenin kapsadığı her şeyi tanıma dahil edecek ve kapsamadığı her şeyi dışarıda bırakacak türden bir küme tespit etme çabasına girilmeden yapılan ve temelde pratik bir ölçüt üretmeyi sağlayan tanımlardır. Örneğin, müşterilerinizin sattığınız üründen memnun kalıp kalmadığını öğrenmek istiyorsanız, memnuniyetin neyi içerdiği ve çeşitli insanların memnuniyetten ne anladığı gibi bir kavramsal belirlemeye girmeksizin, ürünün sahip olduğunu düşündüğünüz birkaç temel niteliğini belirleyerek bu nitelikleri gerçekte ne kadar taşıdığını görmek üzere bir ölçek oluşturup, ölçekten elde ettiğiniz sonuçlara göre ürününüzün söz konusu nitelikler açısından memnuniyet ekseninde başarılı olup olmadığını belirleyebilirsiniz. Sonuçlar size derinlemesine bir bilgi vermez, karmaşık kavramları açığa kavuşturmaz ve eğer ölçeğinizdeki soruların sınırları iyi belirlenmemişse elde ettiğiniz sonuçlar yanıltıcı da olabilir. Bütün bunlara karşın size ölçülebilir, tekrar edilebilir ve kavramsal tanımlama yapmaktan görece daha az muğlak ya da daha az karmaşık olan ve yine görece nesnel sonuçlar elde edebileceğiniz bir araç sunmuş olur. Turing Testi de böylece, araştırmacıların ulaşmak üzere hedefleyebileceği,

2 Copeland (2000) bu anlamda Turing’in düşünce deneyinin yanlış anlaşıldığını ve 1950 yılındaki makalesinde aslında hiçbir zaman bir tanım (işlemsel bir tanım da dahil olmak üzere) öne sürmediğini ifade eder. 1952 yılında yayınlanmış bir radyo programında Turing’in “Düşünenin bir tanımını yapmak istemiyorum, ama yapmak zorunda kalsaydım muhtemelen kafamın içinde devam eden bir tür uğultudan başka bir şey söyleyemedim. Ancak bir tanım üzerinde anlaşmamız gerektiğini de düşünmüyorum. Önemli olan, bir beynin ya da bir insanın tartışmak istediğimiz ve istemediğimiz özellikleri arasında bir çizgi çekmeye çalışmaktır” ifadesini bir dayanak olarak sunar (Copeland, 2000, s. 524). Turing’in bu ifadesinde de zekaya ilişkin kavramsal bir tanım yapmaktan kaçındığı açıktır ancak ifadesinin işlemsel bir tanım yapma niyetiyle yine de örtüşüğünü düşünüyorum.

tartışmalı da olsa net bir ölçüt sağlamıştır ve bu da yapay zeka teknolojilerinin geleceğini belirlemiş ve bilhassa insan konuşma kalıplarını taklit etmeyi amaçlayan doğal dil işleme araçlarının gelişimini şekillendirmiştir.

2. Turing Testi Uygulamaları

Turing Testi'nin etkisi, bilhassa sohbet robotlarının (chatbot) ve doğal dil işleme (natural language processing) sistemlerinin geliştirilmesinde açıkça görülmektedir ki bu alandaki ilk gelişmeler, yapay zeka uygulamalarının günümüzdeki kullanımında da belirleyici olmuştur. Sohbet robotları, doğal dil işleme sistemi ve duyumsama (sentiment) analizi kullanarak bir sohbeti simüle eden bilgisayar programlarıdır (Adamopoulou & Moussiades, 2020).

1964-1966 yılları arasında MIT'de Joseph Weizenbaum tarafından geliştirilmiş olan ELIZA tarihteki ilk işlevsel sohbet robotu olarak genel kabul görür (Natale, 2021, s. 50). ELIZA temelde, bir örüntü eşleştirme metodolojisi kullanarak konuşmayı; bir psikoterapistin konuşmasını taklit etmek üzere tasarlanmıştır. Sohbete dahil olan kişinin girdisini anahtar kelimeler veya ifadeler için taradıktan sonra girdiyi yönetilebilir parçalara ayırır. Ayırıştırılan parçalar daha sonra, önceden tesis edilmiş kurallara uygun olarak yeniden birleştirilir ve yine daha önceden programlanmış olan kelimeler bu birleşime eklenerek anlamlı bir cevap oluşturulmuş olur (Natale, 2021, s. 52) Taklit oyununa uygun olarak geliştirilmiş olan ELIZA'nın tasarımı, insanların etkileşime anlam atfetmeye meyyal doğasından akıllıca yararlanarak, kullanıcıların sisteme gerçekte sahip olduğundan çok daha fazla anlayış atfetmesini sağlamayı başarmıştır. Natale, insanların makinelere karşı bu cömert tutumuna ilişkin olarak 1968 yılında New York Times'da yayınlanmış bir gazete yazısına referansta bulunur (Natale, 2021, s. 60). Yazı, Weizenbaum'un ELIZA isimli projesinde bir bilgisayarı İngilizce gibi doğal bir dili kullanarak iletişim kurabileceği şekilde programlanmış olmasından 'Bilgisayar İngilizce Anlamak Üzere Eğitiliyor' başlığıyla bahseder (Wilford, 1968). Sohbet robotlarının sansasyonel haberlere konu oluşu da bu şekilde başlamış olur. İşleyiş biçiminin basitliğine karşın ELIZA bazı kullanıcıları insan olduğuna ikna edecek performansı sergileyebilmiştir.

Psikiyatrist Kenneth Mark Colby tarafından geliştirilen ve 1972 yılında ortaya çıkan PARRY, bu alandaki gelişimin dönüm noktalarından biri olarak değerlendirilebilecek bir diğer sohbet robotudur. ELIZA'nın psikoterapist kişiliğinin aksine PARRY, bir paranoid şizofreni hastasının iletişim biçimini taklit eder (Adamopoulou & Moussiades, 2020, s. 2). Temelinde herhangi bir psikolojik model olmadan basit örüntü eşleştirme kullanan ELIZA'nın aksine PARRY, kavramsal bir paranoya modeli üzerine inşa edilmiştir. ELI-

ZA'ya benzer şekilde PARRY de kullanıcının girdisindeki önemli unsurları belirlemek için anahtar kelime odaklı bir yaklaşım kullanmıştır. Bununla birlikte, yanıtlarının daha çeşitli oluşu ve konuşmanın gidişatına göre daha dinamik bir tutum sergileme becerisiyle ELIZA'dan farklılık göstermiştir (Colby vd., 1971). 1979 yılında beş psikiyatristten oluşan bir jüri; bir insan ve PARRY ile sırayla görüşme yapmıştır, ilk psikiyatrist kimin insan kimin sohbet robotu olduğunu doğru teşhis ederken; ikincisi, insanı sohbet robotu, sohbet robotunu insan sanmıştır; üçüncü psikiyatrist her iki görüşmesinde de muhatabının insanla olduğunu söylerken, son iki psikiyatrist her ikisinin de sohbet robotu olduğunu ifade etmiştir (Adamopoulou & Moussiades, 2020, s. 2). PARRY, özellikle doğal dil işleme alanında ileri bir adımı temsil etmiştir. Kendisinden önceki modellere kıyasla daha ilgi çekici ve gerçekçi bir fail olarak makine zekası ve insanın psikolojik koşullarını taklit etmenin içerdiği potansiyel karmaşıklıklar hakkındaki tartışmalara önemli ölçüde katkıda bulunmuştur.

ELIZA ve PARRY'den sonra sohbet robotlarına karşı gelişen aşinalık ve dönemin getirdiği teknolojik gelişmelerle, alana artan ilginin sonucunda Amerikalı girişimci Hugh Loebner ve bilişsel psikolog Robert Epstein'in önderliğinde, ilki 1991 yılında düzenlenen toplantıyla birlikte Loebner Ödülü dönemi başlamıştır (Natale, 2021, s. 87). Loebner Ödülü, programcıları bir insanınkinden ayırt edilemeyen konuşmalar yapabilen bilgisayar programlarının yarıştırdığı, yapay zeka alanında yıllık olarak düzenlenen ve sonucusu 2019 yılında gerçekleştirilmiş olan bir yarışmadır. Loebner Ödülü, Turing'in teorik testinin gerçek dünyadaki bir uygulaması olduğu gibi, insan düşüncesini ve dilini sorunsuz bir şekilde taklit edebilen makineler yaratmanın karmaşıklıklarını ve zorluklarını yansıtan doğal dil işleme becerilerinin gelişimi ve ilerlemesi açısından bir vitrin görevi görmektedir. Ancak, Turing Testi'nin pratikte sağladığı tüm avantajlara rağmen bir makinenin düşünüp düşünmediğini belirlemek için kullanılıp kullanılamayacağı bir tartışma konusudur. Düzenlenen ilk yarışmada PC Therapist isimli sohbet robotu, rakiplerini gölgede bırakacak bir performans sergilemiş ve on jüriden beşini insan olduğuna ikna edebilmişse de bu başarısı, onu geliştiren Joseph Weintraub'un programa eklediği 'uçarıklık' özelliği sayesinde olmuştur: sohbet robotu tutarlı ve devamlı bir sohbeti sürdürmek ve dolayısıyla semantik tutarlılık sergilemek yerine uçarı yanıtlarla 'şakacı' bir kişilik yansıtmayı başarmış; sohbetin genel tutarsızlığına ve kurduğu cümlelerin anlamsız oluşuna rağmen jürinin yarısını insan olduğuna ikna edebilmiştir (Natale, 2021, s. 90). PC Therapist, karşısındakini kandırmayı hedefleyen tek sohbet robotu değildir. ELIZA'dan bu yana Turing Testi'ni geçmeyi hedefle-

yen sohbet robotlarının bu tarz hamleler yoluyla, aslında anlamlı olmayan fakat insan katılımcıların anlam atfettiği konuşmalar üretmesi, Turing Testi'nin makinelerin düşünüp düşünmediğini belirleme için yeterli olmadığına dair güçlü bir işaret olarak değerlendirilmiştir.

3. Düşünmenin Yeniden Değerlendirilmesi

Turing'in makinenin yalnızca çıktısını göz önüne alacak şekilde bir düzenlemeye izin veren, bir düşünce deneyi geliştirmiş olması ve bunun sonucunda karşısındakini ikna hedefli sohbet robotlarının gelişip, yapay zeka çalışmalarının büyük ölçüde taklit etmeye yönelmesi, bilhassa testin kandırmaya teşvik edici olduğu yönünde çokça eleştiri almıştır (Natale, 2021, s. 1-2). Ancak bu eleştirilerin Turing Testi'nin kendisinden ziyade, testin alışlagelmiş yorumlarına yöneltildiği isabetlidir. Turing, makalesinin 'Öğrenen Makineler' kısmında, makinenin testten önce "özel bir koçluk yapılmaksızın" (1950, s. 459) eğitime ve öğrenmeye tabi tutulmasını, insan yanılabilirliğinden doğal olarak kurtulmanın bir yolu olarak sunar; makinenin testin gereklerini karşılayıp karşılamadığının anlaşılabilmesi için, kandırmayı içerecek şekilde dizayn edilmemiş olması bir önkoşul gibidir. Dolayısıyla, Turing'in zekaya ilişkin kavramsal bir tanım vermek yerine işlemsel bir tanımlama yoluna girmiş olması, bu yolun daha pratik bir çözüm getirebilecek olmasından başka bir fikre işaret eder ki bu da düşünmenin ne anlama geldiğine dair kanaatiyle ilgilidir: "Takribi elli yıl içinde, ortalama bir sorgulayıcının beş dakikalık bir sorgulama sonrasında doğru tanımlamayı yapma şansının yüzde yetmiş geçmeyeceği kadar iyi bir şekilde taklit oyunu oynamaları için, yaklaşık 10⁹ depolama kapasitesine sahip bilgisayarları programlamanın mümkün olacağına inanıyorum" (Turing, 1950, s. 442). O halde kriter bellidir, belli bir zaman sonra³ inşa edilecek dijital bilgisayarlar sayesinde, taklit oyununu beş dakika boyunca oynayan bir kişi, makinenin taklit becerisi sayesinde en azından yüzde 30 ihtimalle yanlış tahminde bulunacaktır. Yani bilgisayar, taklit becerisi sayesinde, sorgulayıcı ile yaptığı her on yazışmanın en az üçünde sorgulayıcıyı insan olduğuna inandırmayı başarabilecektir. Turing'in bu icadın gerçekleşeceği zaman aralığı için oldukça iyimser bir tutum takınmış olduğu gerçeğinden başka bu ifadede dikkati çeken bir diğer nokta, 'ortalama bir sorgulayıcı'yı konu etmesidir. Ortalama sorgulayıcı, uzman olmayan bir kişiyi işaret eder. Taklit oyununu oynayan kişi, makinenin işleyiş süreçlerini bilen ve onu tuzağa çekip, makine oldu-

3 Turing takribi 50 yıl içinde böyle bir bilgisayarın programlanabileceğini ifade etse de, kastettiği sürenin tam olarak 50 yıl olmadığı da iddia edilmiştir. 1952 yılında vermiş olduğu bir röportajda Turing taklit oyununu insanları kandırabilecek kadar iyi oynayabilecek bir bilgisayarın inşa edilmesi için "en azından bir yüzyıl" geçmesi gerektiğini söyler (Copeland, 2000, s. 527).

ğunu ortaya çıkaracak türden sorular sormayı başarabilecek bir kişi olmalıdır. Bu türden bir oyuncu belirlemesi, Turing'in 'makinelere düşünebilir mi' sorusundan, 'dijital bilgisayarlar bir insanı ne kadar iyi taklit edebilir' sorusuna geçişini anlamamız için de bir ipucudur.

Turing'e göre, bir makinenin düşünüp düşünmemesinin kriteri, gözlemcinin, makinenin düşünüp düşünemediğine dair kanaatinde yatar. Belli bir kavramın ya da niteliğin varlığının, o kavram ya da niteliğin gözlemcisine göre belirleniyor olması fikri, yanıt bağımlılık (response-dependence) olarak adlandırılır (Proudfoot, 2020, s. 494). Locke'un ikincil nitelikler tarifine kavramsal olarak benzer şekilde, Turing'in düşünebilme yetisinin varlığını, gözlemcinin o niteliğin var olduğunu düşünmesine bağlamasıyla, testin direkt olarak ölçtüğü şey makinenin yanıtı değil, gözlemcinin yanıtı olacaktır. Bu nedenle bir makinenin yapısı ya da işleyiş biçimi makineyi zeki yapmaya yeter değildir (Proudfoot, 2020, s. 494). Dolayısıyla makinenin karşısındaki insanı, insan olduğuna ikna etme becerisi, bir varlığa düşünce atfetmenin bir insan eylemi olmasından ötürü, onu zeki kılmaktadır. Turing'in hedefi, çeşitli numaralar yoluyla anlamlı gibi görünen bir etkileşim kurarak zekayı karşıdakini aldatmaya indirgemek değildir.

Öte yandan, Turing'in zeka anlayışını merkeze alan bu tip bir yorum, Turing Testi'nin çokça maruz kaldığı davranışçılık suçlamasını da bertaraf etmektedir. Turing Testi davranışçılık üzerine kurulu olsaydı zihinsel olay ya da nitelikleri ancak belli davranışların gözlemlenmesine indirgerdi ki Turing'in durumunda böyle bir şey söz konusu değildir (Davidson, 2004, s. 80). Ancak, test davranışçılığa dayalı olmasa da, makinelere zeki olup olmadığını belirlemede doğrudan bir bilgi vermiyor olacaktır ve ölçülen şey ortalama - uzman olmayan - bir kişinin makineyi zeki bulup bulmamasıysa, kişinin kanaati dolaylı olarak da olsa makinenin düşünüp düşünemediğini yine de gösterir mi?

Donald Davidson, bu soruya hayır yanıtını verir; gözlemcinin, makinenin yapısal süreçlerinden bağımsız, sadece verdiği tepkiye göre düşünüp düşünmediğini belirlemesi mümkün değildir: "Açıkçası herhangi bir nesne bu teste tabi tutulabilir. Hiçbir tepki vermeyecek şekilde bağlanmış bir çakıl taşı bile bir puan alacaktır (ve özellikle huysuz bir kadın olduğu düşünülürse belki de kötü bir puan almayacaktır)" (Davidson, 2004, s. 78). Davidson'ın testi başarısız bulmasının nedeni testin, cinsiyetçi ya da yanıt bağımlı yapılarından ziyade gözlemciye, makinenin kurduğu cümlelerin bağlamını bilme şansı vermeyişi ve dolayısıyla makinenin düşünüp düşünmediğini anlama imkanı sunmayışıdır; makine, önceden düşünen bir insan tarafından kurul-

muş hazır cümleleri, gözlemcinin sorduğu sorularla tutarlı bir yanıt olabilecek biçimde verdiğiğinde gözlemci makineyi insan sanabilir ancak bu yine de makinenin düşünebildiğini göstermez (Davidson, 2004, s. 80). Makinenin, insan sorularına verdiği yanıtlar, ilgili sorulara yönelimselliğiyle ortaya çıkmış değildir; bu yanıtlar düşünen bir insan tarafından belli sorulara yönelimsellikle önceden hazırlandığından, hali hazırda 'akıllı' cevaplardır. Bu anlamda makine düşünmemektedir ve dolayısıyla anlam üretmemiştir.

Davidson'ın Turing Testine yönelik bu itirazı, bu minvalde yöneltmiş tek itiraz değildir. Makinelerin ya da bugün anladığımız biçimiyle programların tek başlarına bir anlam üretemeyeceği, ürettikleri anlamlı gibi görünen ifadelerin de daha önceden insanlar tarafından hazırlanmış anlamlı ifadelerin geri çağrılmasından başka bir şey olmadığı itirazı, bilhassa Searle'ün 'Çince Odası Argümanı' ile aşına olduğumuz türden bir itirazdır. Searle'ün argümanı ve karşı argümanlar tek başına ayrı bir makalenin konusu olacak kadar geniş bir yer kaplıyor ve bu makalenin kapsamı dışında kalıyor. Yine de, Davidson'ın Turing Testi'ne yönelttiği itiraz bakımından kendine has pozisyonunu anlamak için, Davidson'ın itirazının Searle'ün Çince Odası Argümanı ile birlikte düşünmenin faydalı olduğuna inanıyorum.

Çince Odası Argümanında Searle, Çince bilmediği halde, kendisine verilen Çince cümlelere göre, Çince karakterleri bir kılavuz yardımıyla bir araya getirdiği, dışarıya kapalı bir oda içinde olduğunu hayal eder. Dışarıdan bakan birine göre, doğru yanıtlar verdiği için Çince'yi anlıyor gibi görünür, ancak Searle sadece kılavuzdaki kurallara uyarak sembolleri manipüle etmektedir. Argüman, benzer şekilde, bilgisayarların sembolleri işleyerek insan anlayışını taklit edebileceğini ancak gerçekte işledikleri şeye karşı bir yönelimselliğe sahip olmayıp, ürettikleri şeyi anlamadıklarını, böylece sözdizimsel işleme ile semantik anlayış arasında bir ayrım olduğunu ifade eder (Searle, 1980, s. 417-418). Searle bu düşünce deneyini, programların sözdizimi (syntax) yapabilese bile, bunun anlam (*semantics*) ürettiklerine dair yeterli kanıt olmadığını göstermek için kullanmıştır. Anlamlı görünen bir sözdiziminin gerçekten anlamlı olabilmesi için, sözdizimini oluşturan makinenin insan beynine denk nedensel güçlere sahip olması gerekir; eğer bir makine ya da program bu kriteri karşılayamıyorsa ya da başka bir deyişle insan beyninin sahip olduğu türden nedensel güçlere sahip değilse, yönelimi olmayaacağı için o makinenin düşünebildiğinden söz etmek mümkün olmayacaktır (Searle, 1980, s. 417). Dolayısıyla Turing Testi Searle'e göre bir makinenin düşünüp düşünmediğini ölçmek için uygun bir test olmayacaktır çünkü test sadece makinenin çıktısını, dışa yansıyan davranışını gözlemlemeye imkan

vermektedir. Elde ettiğimiz gözlem ise makinenin üretiminin semantiğe sahip olduğunu düşünmeye yeterli değildir.

Davidson'ın makinelerin düşünmesine yönelik itirazı Searle'ün düşüncesine benzer olarak söz dizimi ve semantik ayrımını içerir ancak Davidson'a göre testin düşünmeyi ölçmek için yetersiz oluşu testin sadece dıştan gözleme imkan vererek semantik opaklığı ortadan kaldırmayı değil dıştan gözleme yeterince imkan vermemesindedir. Davidson'a göre bir makinenin düşünebildiğini iddia edebilmek için makineyle anlamlı bir yazışma yapabiliyor olmak yeterli olmadığı gibi bir makinenin insanın zihin yapısına sahip olması da yeterli değildir; Davidson'a göre düşünebilmenin koşulu önermesel tutumlara sahip olmaktır (Davidson, 1982, s. 318).

4. Önermesel Tutumlar

Literatüre Bertrand Russell'in kazandırdığı (Crane, 2001, s. 108) bir terim olan önermesel tutumlar genel anlamda zihin ve dış dünya arasındaki ilişkiyle ilgilidirler; zihinsel durumların ya da olayların ne şekilde dış dünyayla ilgili olduğu ya da ne şekilde dış dünyaya yönelmiş olduğunu ortaya koyan tutumlardır. Söz konusu tutumlar, bir şeyle ilgili oldukları için yönelimsel durumları içerir. İnanmak, korkmak, istemek, umut etmek, hayal kurmak, hatırlamak gibi zihinsel durum ve olaylar, bir şeyle ilgili olmaları bakımından yönelimsel durumlardır. Bu durumların önermeselliği, dilsel ifadelerinde bir önermeye yönelik olarak kullanılmalardan gelmektedir. Bir failin, bir önermeye yönelik sahip olduğu zihinsel durum bir önermesel tutumdur. Dolayısıyla yönelimim, zihinsel tutumumla bu tutumun ilgili olduğu önerme arasındaki ilişkidir. Yarın havanın güneşli olmasını umuyorsam, 'yarın hava güneşlidir' önermesine yönelik bir umudum var demektir ve bu umut bir önermesel tutumdur. Her tutum bir önermeye yöneliktir ya da onunla ilgilidir. Önermesel tutumların Davidson'ın düşüncesindeki özel yerinin kaynağı da, bu tutumların doğasından gelir. Davidson düşünceye genel olarak tüm önermesel tutumları dahil eder: "Düşünce ile belirli bir içeriğe sahip zihinsel bir durumu kastediyorum. Örneğin: bunun bir parça kağıt olduğu inancı, yavaş ve tane tane konuşma niyeti, yarın havanın güneşli olup olmayacağına dair şüphe" (Davidson, 1989, s. 165). Davidson'a göre genel olarak düşünme ya da bir düşünceye sahip olma, tek tek düşüncelere sahip olmak değildir; düşünmenin bütüncül bir doğası vardır ve bu bütüncül doğanın ortaya çıkabilmesinin zeminini uçgenleme oluşturur (Davidson, 1991, s. 160).

Davidson'ın düşünmeye ilişkin bütüncül bakış açısına göre, zihinselliğin bütüncül olması; düşünmenin tarif edilmesi ya da bir diğer deyişle, konuşabilmek, inanmak, şüphe etmek için birçok kavramın bir arada bulunmasının

gerekliliğinden gelir (Davidson, 1999, s. 7). Buna göre bir önermesel tutum kişiye paket olarak bulunur (Davidson, 1982, s. 318). Bir faile bir zihinsel tutum atfedilebilmesi için pek çok başka zihinsel durum ya da olayın o zihinsel tutuma eşlik ediyor olması gerekir çünkü tüm önermesel tutumlar birbiriyle ilişkilidir ve bir tanesinin anlamlı olabilmesi ya da gerekçelendirilebilmesi için bir diğersinin varlığına ihtiyaç duyulur. Sinemada Wes Anderson'ın yeni filmini izlemeyi planlıyorsam aslında düşündüğüm tek şey film izlemek değildir. Bu planım, Wes Anderson'ın yeni bir film çektiğine ve filmin sinemada gösterime girdiğine dair inancımı, filmi görmeyi isteğimi ve sinema salonuna gitme niyetimi içerir. Üstelik tüm bu yönelimsel tutumlarım birbiriyle ilişki içindeyken anlamlı olacaktır. Sinema salonuna gitme niyetim, filmin sinemada gösterime girmiş olduğu inancım ve filmi görme isteğimle uyumludur. Filmin sinemada gösterime girmiş olduğuna dair inancım olmasaydı, sinemaya gitme niyetimin bir odağı olmayacaktı; filmi görme isteğim olmasaydı, filmin sinemada gösterime girmiş olması belki de bir eylemle sonuçlanmayacaktı. Davidson, bu tür önermesel tutum kümelerinin ya da yığınlarının, bir bireyin eylemlerini ve düşüncelerini yorumladığımız çerçeveyi sağladığını savunur; bu tip bir bütüncül bakış açısı, bir kişinin zihinsel durumunun bir parçasını anlamanın, sahip olduğu inançlar, arzular, niyetler ve diğer tutumlar ağıının tamamını dikkate almayı gerektirir (Davidson, 1982, s. 326).

Düşünmenin bu bütüncül yapısına paralel olarak önermesel tutumlar Davidson'a göre tek bir zihinde var olabilecek şeyler değildir; bir zihinde düşüncenin olabilmesi için, o zihnin doğal bir dünyayı paylaştığı başka düşünce sahibi yaratıkların olması gerekir (Davidson, 1989, s. 170). Başka zihinlerin varlığı, kişinin kendi zihninin önermesel içeriklerinin bilgisi için şarttır. Kişi başka tip bilgilere sahip olmadan kendi zihninin önermesel içeriklerini bilemez çünkü iletişim olmaksızın önermesel düşüncenin olması mümkün değildir (Davidson, 1991, s. 160). Bunun sonucu olarak, bir zihinde bir önermesel tutum oluşması için üç katmanlı bir koşul gerektirir; en azından iki fail olması, bu faillerin ortak bir dünya ile eşzamanlı etkileşim kurmaları ve faillerin birbiriyle etkileşim kurması (Davidson, 1999, s. 12). Üçgenleme olarak adlandırdığı bu fikir, Davidson'ın dilin ve düşünmenin sosyal bir unsur olarak görmesinin nedeni ve Turing Testi'ne itirazının da kaynağıdır. Davidson'a göre düşüncenin içlemsel bir doğası vardır ve bu içlemsel yapı nörolojik, fizyolojik ya da davranışsal herhangi bir şeye indirgenemez (Davidson, 2004, s. 152). Bu indirgenemezlik düşüncüyü semantik olarak opak kılar: "...kedimin dışarıda olduğunu düşündüğüm doğru olsa da mahalledeki en saldırgan kedinin dışarıda olduğunu düşündüğüm doğru olmayabilir; mahalledeki en saldırgan kedi benim kedi olsa bile" (Myers & Verheggen, 2016, s.

14). Söz konusu opaklıktan kasıt şudur; bir varlığın zihninde bir nesneye ilişkin bir kavramın olması için o varlığın söz konusu kavramı yanlış nesne ile ilişkilendirebileceğini anlaması gerekir (Davidson, 1999, s. 13). Kedi diye bir kavramımın olabilmesi için kedi olmayan varlıklara ilişkin kavramlarımın da olması gerekir aksi halde herhangi bir şey kedi olabileceğinden düşüncem belirgin bir şekilde kedilerle ilgili olamayacak; bu nedenle de kedileri düşündüğümde spesifik bir şey düşünemeyeceğim için aslında düşüncemin herhangi bir içeriği olamayacaktır (Davidson, 1991, s. 159). Nesnelere ve kavramları doğru biçimde eşleştirebilmem için başka bir düşünce sahibi varlığın aynı nesne üzerindeki düşüncesine erişebilmem gerekir.

Üçgenleme sayesinde her bir fail iki katmanlı bakış açısına sahip olur; her biri hem bir diğeriyle hem de dünya ile eş zamanlı etkileşim halindedir (Davidson, 1999, s. 12). Birbiriyle etkileşim halinde olan iki failin eş zamanlı olarak yaşadıkları dünya deneyimleri faillerin ortak dışsal çağrıya da maruziyetini gerektirecektir (Davidson, 2004, s. 143). Eş zamanlı olarak aynı uyarana maruz kalan iki fail, etkileşim yoluyla kavramlar ve nesnelere doğru eşleştirmesini yapabilecektir. Bunun doğal sonucu ise şudur: Önermesel tutumların birbiri üstüne binen ve ortak yaşam getiren yapısı ve yönelimsel tutumların bütüncüllüğü nedeniyle, bir varlığın bir diğer varlığa tek bir düşünce atfetmesi için bile oldukça karmaşık davranış örüntülerinin gözlemlenmesi gerekecektir; ortak yaşam alanı içinde ortak uyaranlara yönelik her türden etkileşim, düşüncenin atfında rol oynayacaktır. Davidson bu tip bir örüntünün yokluğunda düşüncenin olmayacağını, bu örüntünün varlığı içinse dilin olması gerektiğini söyler (Davidson, 1982, s. 322). Failler arası etkileşim ancak dil sayesinde önermesel tutumlara dönüşebilecektir.

5. Davidson'ın İtirazı

Davidson'ın Turing Testi'ne yönelik itirazı, düşünmeye dair fikirlerinin ışığında değerlendirildiğinde, itirazının, makinelerin hiçbir zaman düşünemeyeceği gibi bir iddiayı içermediği gibi, makinelerin düşünüp düşünmediğini bilmemizin mümkün olmayacağı iddiasını da içermediği görülecektir. Davidson, yeterli zaman ve doğru şartlar altında kişilerin bir nesnenin düşünüp düşünmediğine karar vermelerinin mümkün olduğunu düşünür (2004, s. 136). Buna ek olarak, Turing'in makinenin düşünüp düşünmediğini belirleyecek olan şeyin onunla etkileşime giren gözlemcinin yargısı olduğu konusunda Turing ile hemfikirdir: Düşüncenin ve anlamın varlığı, yorumcunun yorumlama gücüyle ilintilidir (Davidson, 2004, s. 86). Testin yetersizliği Turing'in işlemselci girişiminden kaynaklanmaz. Davidson'ın itirazı testin, üçgenleme nezdindeki işlevsizliğine yöneliktir.

Davidson Turing'in, 'makinelere düşünebilir mi?' sorusundan taklit oyununa geçişini 'makinenin neyi düşündüğünü belirlemeksizin düşünüp düşünmediğini belirlemeye' geçmek olarak görür ki bu anlamlı cümlelerin varlığını düşünmenin varlığına işaret olarak görmektir (Davidson, 2004, s. 81). Ancak bir varlığın ne düşündüğünü söylemeksizin düşünüp düşünmediğini söylemek mümkün değildir (Davidson, 2004, s. 80). Anlamlı cümlelerin varlığı ya da önermesel ifadeler içlemsel bir doğaya sahip olduğundan ve bu içlemsel yapının herhangi bir kapsamsal bakış açısına indirgenemez olduğundan ötürü, yalnızca failin perspektifinden yola çıkarak düşüncelere ya da dilsel ifadelere anlam atfetmek imkansızdır. Bu nedenle Turing Testi'ne uygun biçimde, makineyle dilsel olarak gerçekleştirilen etkileşim semantik opaklığa takılacak ve makinelerin düşünüp düşünmediğini anlamak mümkün olmayacaktır. Makineler insanlarla ortak yaşam dünyasını paylaşana kadar, düşünüp düşünmediklerine karar verecek bilgiye sahip olmamız mümkün değildir (Davidson, 2004, s. 86). Dolayısıyla makinelerin hangi soruya hangi yanıt ürettiğini belirlemenin bir önemi yoktur, makineyle ortak yaşam dünyasında iletişime geçecek bir insan olmadan makinelerin ne istediği ya da neyi kastettiği anlaşılamayacağı için makineye düşünme atfetmenin anlamı olmayacaktır (Davidson, 1989, s. 171). Semantik opaklığın aşılabilmesi için ise etkileşimin üçgenlemeye imkan vermesi gerekecektir.

Turing Testi üçgenlemeyi içerebilecek bir tasarıma sahip değildir. Bir failin makineyle kurduğu dilsel iletişime dayanan Turing Testi, makinelerin düşünebildiğini göstermeyeceği gibi düşünemediğini de göstermeyecektir (Davidson, 2004, s. 83). Amaç, makinelerin düşünüp düşünmediğine karar verememenin bir yolunu bulmak ise, testin bir kez daha değiştirilerek, sorgunun, makinelerin sadece kendisi ile değil, dünya ile olan etkileşimini gözlemlemesine de imkan tanımak gerekir (Davidson, 2004, s. 84). Bu yolla gözlemcinin bilgisi, makinelerin kendi sorularına cevaplarıyla sınırlı kalmayacak, aynı uyaranlar (olaylar, değişimler ve nesnelere (Davidson, 2004, s. 83)) karşısında ürettiği cevaplar ve bu cevaplarla uyaranların ilişkisini de gözlemeleme fırsatı olacaktır. Testin bu versiyonu, üçgenlemeye imkan vereceğinden gözlemci, kendi düşüncelerini meydana getiren ortamda, makinelerin ne tip ilişkiler içinde olduğunu gözlemleyebilecek ve makineye sorular sorarak aynı koşullar ve ortamda makinelerin düşüncelerinin ne şekilde ortamla ilişkide olduğunu anlayabilecektir. Davidson'ın bu koşulu makinelerin aynı ortam ve koşullarda insan ile birebir aynı kelimeleri kullanmasına ilişkin bir gereklilik değildir: "aynı koşullar altında aynı cümleleri doğru kabul etmeye eğilimli oldukları anlamına gelir" (Davidson, 2004, s. 91). Aynı koşullar altında aynı cümlelere doğruluk atfetme, aynı türden semantik ilişki-

ler kurulduğunun göstergesidir. Bu açıdan, Davidson'ın Turing Testi'ne itirazı, makinenin anlamlı görünen cümleler üretmesinin ötesinde bir kriterin gerekliliğini ortaya koyar.

Davidson'ın itirazının günümüz yapay zeka gelişmeleri çerçevesinde halen bir karşılığının olmasının nedeni de, söz konusu itirazın, anlamlı cümleler kurmaya ya da sadece söz dizimi-semantik arasındaki farka dayanmıyor oluşundan kaynaklanır. Davidson'ın düşüncenin ve dolayısıyla dilin varoluşuna ilişkin ortak doğal dünya paylaşımını şart koşmasının nedeni, insan düşüncesinin büyük ölçüde dil yoluyla kazanılması ve aktarılmasına karşın sadece dilsel olmayışıdır. Bir düşünceye sahip olmak, dile sahip olmakla mümkündür ancak dil de o düşüncenin bağlamını meydana getiren bir ortamın içinde bulunmayı gerektirir. Bu noktada bir soru beliriyor: ortak yaşam alanı içinde bulunulması için makinelerin insana benzemesi gerekli midir? Turing için makinenin insanı taklit etmede olabildiğince başarılı olmasının yolu, yapı bakımından insana en yakın formda olmasıydı ve düşünmenin, bir çocukta olduğu gibi, kademe kademe ilerleyebileceğini iddia ediyordu (Turing, 1950, s. 456). Davidson da makinelerin düşünüp düşünmediğinin belirlenebilmesi için makinelerin dünyada büyük ölçüde insanlarla benzer ilişkiler kurabilmelerini öngördüğünden (Davidson, 2004, s. 86) Turing ile paralel düşündüğü iddia edilebilir. Ancak Turing'in aklında olan şey makinenin, insanın bir kopyası olması değildi (Dennett, 2003, s. 304). Davidson'a göre Turing'in makinelerin insana benzemesi yönündeki tercihi ekonomik bir tercihtir (Davidson, 2004, s. 86). İnsanın zihin yapısı ve zihninin işleyiş biçimi düşünceyi sağladığına göre düşünen bir makine inşa edebilmek için de makinenin insana benzeyecek şekilde tasarlanması en kestirme yol olacaktır. Davidson'ın düşüncesinde ise insana benzerlik, ortak yaşam alanında benzer etkileşime girmeyi içermektedir: "Elbette, haklı olarak, sadece belirli bir biyolojik yapıya sahip canlıların gerçekten düşündüğüne inanıyoruz, ancak (bunca yıldan sonra) arkadaşımın silikondan yapılmış olduğu ortaya çıkarsa, bir insanın hangi maddelerden yapılmış olabileceği konusundaki fikrimi değiştiririm, onun bir insan olduğuna dair kanaatimi değil" (Davidson, 2004, s. 79). Dolayısıyla ne Davidson ne de Turing makinenin ontolojik olarak insan olması gerektiğini düşünmez. Bu bakış açısından, Davidson ve Turing'in insan benzeri özelliklere sahip olması gerekmeyen makineler üzerine yaptıkları tartışmalar, günümüz yapay zeka gelişmeleri ile bağlantılıdır. Turing'in insan gibi düşünen makinelerin olası evrimi ve Davidson'ın bu makinelerin insan toplumlarıyla etkileşime girme potansiyeli üzerine vurguları, sohbet robotları ve dil modellerindeki son ilerlemelerle daha somut bir şekilde ele alınmaktadır.

Son yıllarda yapay zeka alanı, makine yeteneklerine ilişkin anlayışımızı değiştiren önemli atılımlara tanık oldu. Bilhassa sohbet robotlarının oldukça gelişmiş modellerini kullanarak deneyimlediğimiz alanda, Chat GPT'nin teşhis ve tavsiyede uzman göz doktorlarından daha başarılı olduğu (Nacchiappan, 2024), LLM temelli sistemin kimyasal reaksiyonlar tasarlayarak kendi başına ilaç üretmesi (Sanderson, 2023) gibi sansasyonel gelişmeler sıklıkla gündemi meşgul ediyor. İnsan konuşmalarını taklit eden sofistike dil modellerinden karmaşık sorunları çözebilen sinir ağlarına kadar bu gelişmeler, Turing Testi'nin kriterlerinin ve uygunluğunun yeniden değerlendirilmesini gündeme getirdi. Bilhassa, OpenAI'nin üçüncü nesil dil tahmin modeli olan ve geniş dil modellerinin (LLM) üyesi GPT-3 ve Mart 2024'te ChatGPT yoluyla erişime açılan GPT-4 doğal dil işlemede ileriye doğru bir sıçramayı temsil ediyor (Floridi & Chiriatti, 2020, s. 681). Doğal dil işleme modeli, ELIZA'dan bu yana hem kabiliyet hem de yaygınlık noktasında büyük bir başarı elde etti. GPT'lerin verilen bir ipucuna dayalı olarak tutarlı ve bağlama uygun metin üretme yeteneği, makinelerin düşünemediğine dair bir gösterge olarak gündeme geldiyse de bu konuda henüz bir görüş birliğine varılabilmemiş değil. Bir taraftan GPT'lerin, dili kullanma becerisine ek olarak matematik, kodlama, hukuk, tıp, psikoloji ve daha birçok alanda insana yakın performans sergilediği iddia edilirken (Bubeck et al., 2023), diğer taraftan henüz Turing Testi'ni geçemediği (Jones & Bergen, 2024) ve GPT'lerin genel bir yapay zeka özelliği gösterdiği iddiaları "bilgisiz bilimkurgu" (Floridi & Chiriatti, 2020, s. 681) olarak nitelendiriliyor.

Geniş dil modellerini kullanan yapay zeka araçlarının akıllı olup olmadıklarına, başka bir ifadeyle, düşünüp düşünmediklerine ilişkin görüş birliğine varılamaması Davidson'ın itirazını yeniden gündeme getiriyor. LLM'lerin yalnızca dilsel ifadeler yoluyla öğrenmeleri; fiziksel dünyanın bir parçası olan bir donanıma sahip olmamaları nedeniyle dilin kullanımı için gerekli olan dış dünya deneyiminden yoksundurlar. Buradan yola çıkarak bu modellerin düşünemediğine ya da bugün değilse bile bir gün düşünemeyeceğine değil, düşünüp düşünmediğini anlamanın zorluğuna vurgu yapıyorum. Dolayısıyla mesele insanın, makinenin zekasını anlamaması ya da makinenin gösterdiği kabiliyetlerin insanın alışageldiği zekadan daha başka bir formda olması değil, makinece bir zekaya sahip olsa da bunu tespit etmenin güçlüğüdür. Davidson'ın Turing Testi'ne bu minvaldeki itirazı, yapay zeka ve makine öğrenmesinin günümüzde katettiği yola rağmen geçerliliğini koruması açısından dikkat çekicidir. Geniş dil işleme modellerinin önümüzdeki yıllarda Turing Testi'ni herkesin hemfikir olacağı şekilde geçmeyi başarabilse bile gerçekten dili anlayıp anlamadıklarına ya da düşünüp düşünmediklerine iliş-

kin soru, bu modeller ve insanın, ortak yaşam dünyasında aynı anda gözlemlenebilir bir formu bulunmadığı sürece yanıtlanmadan kalacaktır.

Bu gözlemlenebilirliği sağlayacak olan şeyin ne olabileceği tahmin becerilerimizi aşıyor. “Sembolik etkileşimlerin yanı sıra dünya ile sembolik olmayan, duyu-motor etkileşimler de kurabilen, bedenlenmiş bir robot” (Harnad, 2003, s. 486) olabileceği gibi; insan beyninin dil ve dil dışı becerileri keskin şekilde ayıran yapısına benzer bir mimaride üretilmiş, çoklu modaliteye (görsel, duyuusal, akıl yürütme vb.) sahip LLM modelleri de söz konusu kriteri sağlayabilir (Mahowald et al., 2023). Henüz bu tip bir teknolojiyle hangi formda karşılaşacağımızı bilemesek de Davidson’ın bütüncül görüşüne uygun biçimde makinelerin ya da programların düşünüp düşünmediğinin belirlenmesi için herhangi bir modelin dil işlemenin ötesinde, gerçek dünya ile gerek algıya gerekse eyleme dayalı etkileşimlerinin olması ve akıl yürütme ve dolayısıyla dili üretme süreçlerinin gözlemlenebilir olması gerekmektedir.

Sonuç

Turing Testi, yapay zekânın insan zekasıyla kıyaslanması için bir ölçüt olarak ortaya koyulmuş olup, bu testin yapay zeka teknolojilerinin gelişimine önemli katkılarda bulunduğu açıktır. Bu katkıların yararı konusunda ortak bir görüş olmasa bile Turing Testi halen çokça ilgi görmektedir. Testin getirdiği tüm avantaja rağmen Davidson’ın eleştirisiyle, Turing Testi’nin bazı temel kavramsal zaafalarını görüyoruz.

Davidson’ın önermesel tutumlar teorisi, düşüncenin yalnızca dil ile sınırlı olmadığını ve gerçek anlam üretiminin sosyal ve etkileşimli bir bağlam gerektirdiğini vurgular. Davidson’ın düşünceleri, düşünmenin doğasının yalnızca sembolik işlemlerle sınırlandırılmayacağını; zihinsel süreçlerin, sosyal etkileşim ve ortak bir dünya deneyimiyle iç içe geçtiğini savunmaktadır. Bu bakış açısı, Turing Testi’nin yapay zekâyı tam anlamıyla test etmede yetersiz kalabileceğine işaret eder.

Davidson’ın itirazı, yapay zeka teknolojilerinin değerlendirilmesi için daha geniş kapsamlı metodolojilerin geliştirilmesinin önemini göstermektedir. Turing Testi’nin getirdiği dil odaklı sınırlamalar, yapay zekânın insan benzeri zeka gösterme yeteneğini gözleme yönünde engel teşkil edebilir. Davidson’ın eleştirisi yapay zeka araştırmalarının, dil işleme yeteneklerinin ötesine geçerek, makinelerin gerçek dünya ile etkileşimde bulunma kapasitelerini ve sosyal bağlamlarda nasıl işlediklerini anlamamıza yardımcı olacak yeni ölçütler geliştirilmesine olan ihtiyacı vurgulamaktadır.

Sonuç olarak, Turing Testi yapay zekâ araştırmalarında önemli bir kilometre taşı olmuştur, ancak bu testin sınırlılıkları, testin neyi ölçtüğüne ya da ne şekilde anlaşılması gerektiğine yönelik genel bir kanaatin oluşmamış olması ve filozofların itirazları, yapay zekâ teknolojilerinin değerlendirilmesi için daha kapsamlı ve bütüncül yaklaşımlar geliştirilmesinin gerekliliğini göstermektedir. Günümüzde bilhassa LLM'ler, onlara düşünce atfetmemizi teşvik eden başarılarla sahip. Ancak, Davidson'ın eleştirisi ışığında, bu modellerin gerçekten “düşünüp düşünmedikleri” sorusuna cevap vermek, yalnızca teknik bir başarıdan daha fazlasını gerektirmektedir. Makinelerin insanlarla ortak bir yaşam dünyasını paylaşması ve bu dünyada anlamlı etkileşimlerde bulunabilmesi, makine düşünmesi ile ilgili kanaatimizdeki belirsizliği ortadan kaldırmada etkili olacaktır.

Kaynakça

- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*. <https://doi.org/10.1016/j.mlwa.2020.100006>
- Bubeck et al., S. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *digital preprint, arXiv:2303.12712*
- Colby, K. M., Weber, S., & Hif, F. D. (1971). Artificial Paranoia. *Artificial Intelligence*, 1-25.
- Copeland, B. J. (2000). The Turing Test. *Minds and Machines*, 519-539. <https://doi.org/10.1023/A:1011285919106>
- Crane, T. (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford: Oxford University Press.
- Davidson, D. (1982). Rational Animals. *Dialectica*, 317-327. <https://doi.org/10.1111/j.1746-8361.1982.tb01546.x>
- Davidson, D. (1989). The Conditions of Thought. *Le Cahier (Collège international de philosophie)*, 165-171. <https://doi.org/10.5840/gps19893627>
- Davidson, D. (1991). Three Varieties of Knowledge. *Royal Institute of Philosophy Supplement*, 153-166. <https://doi.org/10.1017/s1358246100007748>
- Davidson, D. (1999). The Emergence of Thought. *Erkenntnis*, 51(1), 7-17. <http://www.jstor.org/stable/20012936>
- Davidson, D. (2004). *Problems of Rationality*. New York: Oxford University Press.
- Dennett, D. C. (2003). Can Machines Think? C. Teuscher (Ed.), *Alan Turing: Life and Legacy of a Great Thinker* (ss. 295-316). Heidelberg: Springer Berlin.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 681-694. <https://doi.org/10.1007/s11023-020-09548-1>

- Harnad, S. (2003). Minds, Machines and Turing. H. J. Moor (ed.), *The Turing Test: The Elusive Standard of Artificial Intelligence* (ss. 253-273). Springer Dordrecht.
- Jones, C. R., & Bergen, B. K. (2024). Does GPT-4 pass the Turing test? *arXiv:2310.20216*.
- Mahowald, K. I. (2023). Dissociating language and thought in large language models: a cognitive perspective. *arXiv preprint arXiv:2301.06627*.
- Myers, R. H., & Verheggen, C. (2016). *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*. New York: Routledge.
- Nachiappan, A. (2024, 04 18). *Technology behind ChatGPT better with eye problem advice than non-specialist doctors, study test finds*. <https://news.sky.com/>: <https://news.sky.com/story/technology-behind-chatgpt-better-with-eye-problem-advice-than-non-specialist-doctors-study-test-finds-13117259>
- Natale, S. (2021). *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*. Oxford: Oxford University Press.
- Proudfoot, D. (2020). Rethinking Turing's Test and the Philosophical Implications. *Minds and Machines*, 487-512. <https://doi.org/10.1007/s11023-020-09534-7>
- Sanderson, K. (2023, 12 20). *This GPT-powered robot chemist designs reactions and makes drugs — on its own*. [nature.com: https://www.nature.com/articles/d41586-023-04073-4](https://www.nature.com/articles/d41586-023-04073-4)
- Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 417-457. <https://doi.org/10.1017/S0140525X00005756>
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 443-460. <https://doi.org/10.1093/mind/LIX.236.433>
- Wilford, J. N. (1968, Haziran 15). Computer Is Being Taught to Understand English. *New York Times*, p. 58.