Değini

Yapay Zekâ Tartışmaları ve 'Tersine Modelleme': İnsanlar Gelecekte Makineler Gibi Düşünebilecek mi?

Halil Nalçaoğlu*

Zekâ: Yirmibirinci Yüzyıl'da yapayı doğalından daha makbul olan tek şey. Son yıllarda yapay zekâ araştırma ve uygulamalarının artan bir hızla yaygınlaştığına tanık oluyoruz. Yapay zekâ (YZ) konusunda çok sayıda görüş ortaya atılıyor ve teknoloji alanında bugüne kadar olduğu gibi net bir biçimde gözlemlenebilen keskin karşıtlıklar, bu alan hakkında da kafa karışıklığı yaratıyor. Elon Musk ya da Stephen Hawking gibi popüler figürler yapay zekâyı "insanlığın sonu" olarak tanımlarken, popüler basında yapay zekâyı adapte etmeyen şirketlerin, hatta devletlerin, büyük teknoloji yarışında gerilerde kalmaya mahkûm olduğu yazılıyor. Öyle görünüyor ki, herkesçe anlaşılır argümanlarla tartışmada bir yer kapma eğilimi, konunun karmaşıklığıyla ters orantılı. Bütün bunlar olurken, Alexa ya da Siri gibi kişisel yardımcılar, dijital oyunlar, yapay nöron ağları (YNA), satın alma davranışıyla ilgili öngörü algoritmaları, akıllı sistemler, büyük veri analizleriyle sahteciliğin tespiti ve önlenmesi, satış ve satış sonrası destek birimleri, haber yazımı, kişiselleşmiş çevrimiçi müzik ve video önerileri, güvenlik sistemleri ve benzeri uygulamalar milyonlarca insanın hayatına girdi bile. Modern kent yaşamında YZ sistemlerinin hayatı kolaylaştıran ve satış-Pazarlama ağırlıklı uygulamaları neredeyse hiç sorgulanmadan yasal onay ve toplumsal kabul görürken, bir yandan da bu sistemlerin insanlığın geleceği için nasıl bir tehdit oluşturduğu ciddi, yarı-ciddi ve gayrı ciddi ortamlarda mütemadiyen tartışılıyor.

Yapay zekâ insan zekâsının yerine geçebilir mi? YZ insan türünün geleceği bakımından tehlikeli midir? YZ araştırmaları bir otorite tarafından denetlenmeli ve kontrol altına alınmalı mıdır? Yapay zekâ önünde sonunda "yapay süper zekâya" yol açacak mı? Tüm beyin emülasyonu mümkün ve etik mi?

* İstanbul Bilgi Üniversitesi, İletişim Fakültesi. hnalcaoglu@bilgi.edu.tr

Yapay zekâ yarattığı istihdamdan daha fazlasını ortadan kaldıracak mı? Bir YZ algoritması bir kişiye ya da bir şeye zarar verdiğinde bundan kim(ler) sorumlu tutulacak? Yapay zekânın genel bir otomatik karar verici olmasına izin vermeli miyiz? Bunlar, girişte tarif ettiğim tartışmaya yön veren temel sorulardan bazıları. Bu soruların bir kısmı veya tümü ciddi biçimde incelenmeli ve araştırmacılar, politika yapıcılar, yasa koyucular bu sorular hakkında sonuç getirecek girişimlere hızlıca başlamalılar. Bu yazıda benim üzerinde düşünmeyi hedeflediğim soru yukarıda sıraladığım sorulardan biri değil.

Bu çalışmanın merkezinde duran soru, YZ araştırmalarının kökeninde bulabileceğimiz bir kavramdan (hesaplanabilirlik; *computability*) hareketle -fakat bu kavramdan tamamen habersiz bir şekilde- günümüz popüler söyleminde kendine yer edinmiş görünen bir söylemsel jest ve bu jestin arkasında yattığını iddia edeceğim örtük bir varsayımla ilgili olacak. Bu varsayıma "tersine modelleme varsayımı" adını vermek istiyorum. Hesaplanabilirlik kavramı ve tersine modelleme varsayımını kavramsal olarak tartışmadan önce konunun daha iyi anlaşılabilmesi için somut bir örnekle uğraştığım konuyu açıklığa kavuşturmak yerinde olacak.

Bob ve Alice

Facebook, Bob ve Alice adı verilen iki konuşma robotuyla (*chatbot*) 2017 yazında yapılan bir deneyin, robotların çığrından çıkmaları sonucunda sonlandırıldığını duyurdu. YZ ile donatılmış robotlar pazarlık yapma içeren bir deneyde kendi aralarında araştırmacıların anlamadıkları bir dil geliştirmiş, araştırma bu nedenle sonlandırılmıştı. Her ne kadar konudan anlayan yazarlar bunun ilk kez karşılaşılan, hatta istenmeyen bir durum olmadığını söyleseler de (Robertson, 2017) iki robotun insanların anlamadığı bir dil geliştirmiş olmaları büyük ilgi uyandırdı. Kanımca bu ilginin arkasında YZ ile ilgili oluşmuş olumlu ve olumsuz çok yaygın kanaatlerin ve elbette Facebook PR yetkililerinin malumu uygun bir dille ilan etmelerinin önemli payı var.

Bilimsel deneyler planlandıkları gibi ilerlemeyebilir. Bu durumda araştırmacıların sorun üzerine çalışmaları, sorunu çözebilirlerse deneye devam etmeleri, çözemezlerse deneyi sonlandırmaları doğaldır. Bu vak'ada araştırmacıların nasıl ilerlediklerini bilmiyorum. Ancak benzer pek çok durumun aksine, YZ içeren bu deney büyük bir tartışmanın fitilini ateşlemiş oldu. Yalnızca *The Atlantic* kısa süre zarfında konuyla ilgili üç makale yayınladı. Bunlardan birinin başlığı çok ilginç: "YZ İnsan Zekâsını Nasıl Yeniden Tanımlıyor." Makalede yazar, YZ'nın artık her tür yaratıcı alanda boy gösterdiğini ifade ettikten sonra, makinelerin yaptığı sanatın yaratıcılık ve hayal gücü bakımından sorgulanabilir olmakla birlikte "bizleri insan sanat üretiminin sırlarının derinliklerine taşıdığından" söz ediyor (Lafrance, 2017a). Aynı ya-

zar daha sonraki bir sayıda robotların dil ile ilgili başarılarından bahisle, "araştırmacılar robotların kendi aralarında nasıl iletişim kurduklarını görmenin yanı sıra insanlar arasında sözdizimi ve (dilsel) yapısal düzenin ilk olarak nasıl ortaya çıktığının da ipuçlarını yakalıyor olabilirler" diyecektir (Lafrance, 2017b). Bu ve benzeri fikirleri çoğu kere iddiaların büyüklüğüne gölge düşüren kısıtlayıcı argümanlar takip etse de (kendi dillerini yaratan robotlardan söz ederken *dilden ne anladığımızı* da belirtmeliyiz!), iddiaların kendisi orta yerde durmakta ve hiç kuşkusuz kısıtlar ortaya koyduğu görüşlerden daha fazla ilgi çekmektedir. Evet, insan yaratıcılığı, hayal gücü ya da doğal dillerin sözdiziminin tarihte ilk olarak nasıl şekillendiğini robotlardan öğrenmemiz uzak bir ihtimal olabilir ama gene de dilin yapısal düzeninin laboratuvar ortamında oluşumunu görmek çok ilginç olmaz mıydı?

Bu söylemsel jest, yapay nöron ağı (YNA; *Artificial Neural Networks; ANN*), öğrenen makineler ve yapay zekâ uygulamaları evreninde mevcut bir düzeni tam tersine çeviriyor. Örneğin zihin fonksiyonlarının mekanik emülasyonuyla ilgili ilk fikirlerin 1930'lu yıllarda dev mürekkepbalıklarına ait doğal nöron ağı mekanizmasının keşfiyle gelişmeye başladığı biliniyor (Young, 1938). Yirminci yüzyılın ilk yarısında henüz yapay zekâdan ve YNA'lardan söz edilmezken Alan M. Turing'in dev mürekkep balığı nöronundan hareketle geliştirdiği matematiksel model (1952) 1970'lerde aktif nöron iğciğinin elektronik analoğunun üretilmesinde (*fiber*) önemli rol oynamıştır (Scott, 2002: 5). Tarihsel dizilimden ziyade benim özellikle vurgulamak istediğim nokta, bilimsel buluşların, doğal süreç ve fonksiyonları çözümlenen (açıklanan) fenomenlerden yola çıktıkları olacak (örneğin, havadan ağır uçan makinelerin icadında kuşların nasıl uçtuğunun gözlemlenmesinden sonra nesnelerin hareketi -ve durağanlığı- ile ilgili çok sayıda sağlıklı bilimsel gözlem ve açıklama yapılabilmesi gibi). Yani, doğal olanı yapay evrende tekrar etmek için kurgulanan kavramsal model özünde doğa kaynaklıdır. Bu model kullanılarak geliştirilen teknoloji temel alınarak doğal olan daha iyi (daha *kesin* ve *yetkin*) açıklanabilir mi? Sorulması gereken soru budur.

Tarih boyunca kuşları gözlemleyen insanlar uçma hayalleri kurmuşlardır. Uçma fonksiyonunun iki temel fizik süreciyle (aerodinamik ve güç) gerçekleştiğinin keşfi, kanat çırparak güç üreten ve sabit kanatlarla havada kalmayı başaran kuşların gözlemlenmesi neticesinde olmuştur. Bu keşifle, uçmak isteyen insanların türlü mekanik düzeneklerle kanat çırpan -ve bugün bize komik gelen- makineler üretmek yerine, güç üretmeyi (motor) ve sabit kanatları iki ayrı fonksiyon olarak görmeye başladıktan sonra uçak benzeri ilk aletleri icat ettiklerini biliyoruz (Ackroyd, 2011).

Şimdi ilkel bir soru soralım: Uçaklar baz alınarak oluşturulan kavramsal modelleri kuşları anla-

mak/açıklamak için kullanabilir miyiz? Bu soruya gönül rahatlığıyla olumsuz yanıt verebiliriz. Aynı yanıtı güncel YZ uygulamalarına uyarlayacak olursak, insanın doğal zihinsel süreçlerini anlamak için yapay olanın model olarak kullanılamayacağını söylememiz gerekir. Öte yandan, uçaklar ve kuşlar gibi kaba fiziksel ve kinetik özellikleri olan nesnelerin epistemolojisi dijital çağın süper sofistike nesneleri (YZ, yapay nöron ağları, öğrenen makineler, vs.) için geçerli midir? Bu soru özellikle nasıl işlediğini anlamadığımız yapay sistemlerin hayatımızın önemli bir kısmına hâkim olduğu koşullarda ayrı bir değer kazanıyor. Bu soruya "gönül rahatlığıyla" olumlu ya da olumsuz yanıt vermek kolay değil. Üstelik bu soruya verilebilecek ikiden fazla yanıt olabilir. Belki de şöyle ilerlemeliyiz: Sorulabilecek tüm sorulara yalnızca iki yanıt vermek mümkün müdür? Bu yazının bundan sonraki kısmında aslında bu soruya verilen bazı yanıtların tartışması yer alacak. Daha teknik bir terim kullanacak olursak, kurguladığımız problemi bundan sonra "hesaplanabilirlik" (*computability*) kavramı çerçevesinde irdelemeye çalışacağım.

Hesaplanabilirlik

"Hesaplanabilirlik" kavramını İngilizce "*computable*" karşılığı olarak kullanıyorum. Aynı Türkçe kelimeyle karşılanan "*calculable*"dan farklı olarak *computable*, hem algoritmik hesaplanabilirliğe işaret ediyor hem de tarihsel anlamda önemli bir tartışmanın odağında duruyor. Bu tartışma yakın zamanda MIT tarafından yayınlanan *Computability: Turing, Gödel, Church and Beyond* (2015) başlıklı derlemeden tüm yönleri ile takip edilebilir. Algoritmik hesaplanabilirlikle ilgili kavramsal modellerin oluştuğu 1930'lu yıllarda bugün Enformasyon Çağı olarak adlandırdığımız yazılım ve donanım altyapısının temelleri matematiksel olarak atılıyordu. Bir başka ifadeyle, matematiksel mantık alanında çalışan (bu alanı kuran) büyük isimler (Turing, Church, Gödel, vd.), günümüz yazılım ve donanım teknolojisi henüz ortada yokken öğrenebilen ve düşünebilen makineler hayalini kurmaya başlamıştı bile. Bu hayalin kavramsal izlerini sürmeye çalışalım.

Matematiksel mantık alanında her nedense Almanca ifadesiyle kullanılagelen bir önemli sorun alanından başlayabiliriz: *Entscheidungproblem*. David Hilbert'in 1928 yılında ortaya attığı bu kavram-soru "karar problemi" olarak tercüme edilebilir. Kökenleri Leibniz'e kadar uzanan bu problem özünde bize şu soruyu soruyor: Bir algoritma, belli aksiyomlara bağlı olarak ortaya atılmış bir önermenin doğru veya yanlış olduğuna sadece mantık kurallarına bağlı olarak karar verebilir mi? Bir başka şekilde ifade edecek olursak, sezgilere hiç yer vermeden bir matematiksel fonksiyonun geçerliliği aksiyomların aynı kaldığı her durumda evrensel olarak kanıtlanabilir mi? Leibniz, 17. Yüzyıl sonlarına doğru mekanik hesap makinesi tasarımını geliştirir-

ken, bu soruyu mantık kurallarına tam uyumlu olabilmesi bakımından mutlak bir formel dile ihtiyaç duyulacağı şeklinde formülleştirmişti. Verili aksiyomlara bağlı kalınarak ortaya atılacak herhangi bir problemin evrensel olarak kanıtlanabilmesi, aynı aksiyomların altında önceden belirlenmiş terim ve kurallar dışında hiç bir *yeni* önermenin doğmaması anlamına geliyor. Bu ise düşünebilen makinelerin icadından söz edilen günümüzde karşımıza çıkan en temel problemlerden biri aslında. Basit bir dille soruyu dönüştürecek olursak: Makineye sunulan *her şey* formelleştirilebilir mi?

"Her şey" elbette çok iddialı bir kelime. Leibniz başka pek çok dâhice buluşun yanı sıra diferansiyel integral hesabın da mucidi olarak bilir. Bilindiği gibi bu hesap $y = f(x)$ fonsiyonun iki boyutlu düzlemde çizdiği eğrinin altında kalan alanın hesaplanmasıyla ilgilidir. Alan hesaplama işleminde bir sınırı eğri olan geometrik şekli, sınırlı sayıda kolay hesaplanabilen alanlara (örneğin, karelere veya üçgenlere) bölmek imkânsızdır. İntegral hesabı matematik dünyasına öyle bir kapı aralamıştır ki, biraz eğretilemeye başvurarak ifade edelim, özünde "analog" olan *her şeyin* hesaplanabilirliği umudu doğmuştur. Zira aynı eğretileme seviyesinde durarak konuşacak olursak, doğa doğru çizgilerden değil eğrilerden oluşur. Bunu başka bir şekilde söyleyelim. Hayat sonsuz olasılıklar içerir. Sonsuz olasılıkların kesine *yaklaşan* bir integral hesabı olabilir mi? Evinizden çıkıp işyerinize gidene kadar her gün aynı rotayı takip ediyor olsanız da, akıp giden her saniye size yeni bir şeyler sunacaktır -ırmak aynı gibi görünse de aynı suda iki kere yıkanamazsınız- İşte sözünü ettiğimiz hesaplanabilirlik kavramı bize teorik olarak sonsuza giden olasılıkların bir noktada bağlanıp bağlanamayacağını anlatıyor.

Yirmibirinci Yüzyıl kimilerine göre bir yapay zekâ devri olacak. Bu yazının girişinde belirttiğim çalışma alanları ve bunlara bağlı olarak ortaya atılan çok sayıda soru, her şeyden önce YZ'nın mümkün olup olmaması, Hilbert'in *Entscheidungproblem*'i ile yakından ilgili. YZ, özet olarak ifade edecek olursak, bir makineye sunulan verilerin işlenmesi sonucunda ortaya atılan bir problemin çözülmesi olarak tanımlanabilir. Bir ders kitabı tanımı verecek olursak, YZ "akıllı davranan hesaplayıcı (*computational*) eyleyenlerin (*agents*) sentez ve analizini çalışan alandır." Tanımı yapan yazarlara göre, bu cümlede geçen "hesaplanabilirlik" zekâ için basit bir metafor değildir:

> Bir eyleyenin eylemleri onun yetileri, tarihi, amaç ve tercihlerinin bir işlevleridir. Bu hesaplayıcılığın zekâ için sadece bir metafor olmaktan çıktığı argümanını sunar; akıl yürütme (*reasoning*) hesaplamadır ve hesaplama işi bir bilgisayar tarafından yürütülebilir (Poole ve Macworth, 2011: 7)

Bütün zihin felsefesi ve aşkın felsefe alanlarını bir kalemde kenara

iten bu özgüven bence çok öğretici. Üstelik yeni de değil. Bilişsel bilimlerin bilgisayar bilimleriyle ilişki içinde gelişmeye başladığı 1970'li yıllarda benzer bir tersine okumanın bilim insanlarının zihinlerinde yerleşmeye başladığını görebiliyoruz: "Sıralı (sequential) mimariye yol açan mantığa dayalı hesaplanabilirlik modeli bilişsel bilimlerde insan aklı için de bir model olma hizmeti sundu" (Newell ve Simon, 1976). Bu fikrin dayandığı sıralı bilgisayar mimarisinin mucidi John von Neumann'a (1958, 2000) baktığımızda (*The Computer and the Brain* kitabının bütününde) beynin analog sinyal işleme özelliğinin mekanik hesaplama işi için çok önemli bir model teşkil ettiği yaklaşımını buluruz -tam tersini değil-. Bu "tersine modelleme" artık sorgulanmadan kabul görmüş durumdadır. Bilgisayar bilimleri ve bilişsel bilimlerin son derece verimli işbirliğini göz ardı etmek anlamında değil ama bu alanda yazıp çizen herkesin uzlaştığı, beynin enformasyon işleme özellik ve prensiplerinin hâlâ karanlıkta kalmış peç çok bölgesi bulunduğu da unutulmaması gereken bir gerçektir (Sejnowski, 1989; Hoffman, 2016, 2018).

Tersine modellemeyi sorgulanamaz bir yöntem olarak kabul etme özgüveninin arka planına baktığımızda, Hilbert problemini çözümsüz ilan ederek çözerken hesaplanabilirlik kavramına yeni bir boyut katan Turing'i bulabiliriz (1936). Nitekim yukarıda David L. Poole ve Akan K. Macworth'dan alıntıladığım, zihinsel faaliyetleri hesaplanabilirliğe indirgeyen alıntının başlangıcında Turing-Church hipotezi olarak bilinen önermenin mealini görmek bizi şaşırtmıyor: "Turing-hesaplanabilir olmayan fonksiyonların insanlar tarafından hesaplanabileceğine dair bir kanıt yoktur" (2011: 7).

Turing-Church hipotezi, Alan Turing ve Alonzo Church'ün eş zamanlı olarak karar problemine farklı yollardan giderek çözüm üretmeleri üzerine bu adı almıştır. Literatüre baktığımızda aynı hipotezin birden fazla farklı ifade biçimlerine rastlıyoruz. Detayları bir kenara bırakarak incelediğimizde Turing-Church hipotezinin bize şunu söylediğini görüyoruz: Nümerik bir fonksiyon ancak ve ancak bir Turing Makinesiyle hesaplanabilirse efektif olarak hesaplanabilir. Turing'in adıyla anılan Turing Makinesi aslında hipotetik bir bilgi işlemcidir ve günümüz bilgisayarlarının hala kullandığı bir modeli yansıtır. Bu modelde sonsuz uzunlukta ve üzerine tek bir sembol yazılabilen hücrelere bölünmüş bir bant, bu hücrelerdeki sembolleri okuyabilen, buralara semboller yazabilen ve silebilen bir "kafa," ve bantı ileri geri hareket ettiren bir mekanizma tahayyül edilmiştir. Normal bir insan hesaplayıcının (*computor*) hesap yaparken kullanması gereken zihinsel aşamalar bir algoritma olarak makineye aktarıldığında makinenin hesaplama kapasitesi hesaplanabilirliğin üst sınırı olarak görüldüğünden, makine tarafından hesaplanabilen fonksiyonlar insan hesaplayıcı tarafından

hesaplanabilen fonksiyonların (insanın hesaplayabilme kapasitesinin) teorik olarak eşiti ya da daha üstü olacaktır. Bir başka ifadeyle, bir fonksiyonun hesaplanabilirliğinin kıstası bu fonksiyonun makine tarafından hesaplanabilir olmasıdır.

İnsanın zihinsel işlevlerini hesaplanabilirliğe indirgeyen YZ perspektifinden baktığımızda, bu saptama insan zekâsının (aklının, akıl yürütme kapasitesinin) sınırlarının makinenin hesaplayabilme sınırlarıyla çizildiği gibi bir anlamı ortaya çıkarmaktadır. Kurt Gödel'in efektif hesaplama tanımı anlamında Turing'in başarısını takdir ederken aynı eserde "felsefî zafiyet" bulmasının arkasında işte bu kavramlaştırma var. Her ne kadar Gödel'in Turing'i yanlış okuduğu sonucuna ulaşan çok sayıda makale varsa da (ör., Sieg 2015) gerçek bir matematik bilgisi gerektiren bu tip analizlerin aşkın felsefe ya da zihin felsefesi alanlarında ortaya koyulan yüzlerce yıllık birikimi kaale almayan sonuçlara ulaşılmasına engel olamamaktadır.[1]

Bir yandan Turing makinesini "icad ederken" diğer yandan Hilbert problemine olumsuz yanıt sunan Turing aslında 1930'lu yılların ikinci yarısında Princeton Üniversitesi'nde Church'ün öğrencisi iken tanıştığı Gödel'in matematiksel mantık alanında çığır açan iki teoremini daha sağlam temeller üzerine oturtuyordu. Bu iki teorem "eksiklik" ve "tutarsızlık" teoremleri (*incompleteness and inconsistency theorems*) olarak bilinir. Bu teoremler kabaca şöyle ifade edilebilir: 1) Her zaman önceden tanımlı aksiyomlara dayanmayan doğru bir teorem bulunabilir; 2) Aksiyomatik bir sistem altındaki teoremler sonsuza kadar tutarlı olacak şekilde ispatlanamazlar.

Bunlar ne anlama geliyor? Teoremler mutlaka verili bir aksiyomatik sisteme bağlı olarak ispatlanabilir. Örneğin içinde matematiksel eşitlik bulunan bir teoremi ispatlamadan önce ispata ihtiyaç duymayacak kesinlikte doğruluğunu kabul ettiğiniz ve eşitliği tanımlayan bir aksiyomunuz olması gerekir ("eğer A = B ise, B = A" gibi). Konuyu biraz saptırarak ifade edecek olursak, anlamlı bir cümle kurabilmeniz için önce kullandığınız dilin herkesçe kabul edilmiş gramer kurallarının, kelime anlamlarının, vb. varolması gerektiği gibi. Gödel teoremleriyle bize şunu söylüyor: Her zaman teoremi çerçeveleyen aksiyomlardan indirgenmemiş bir doğru önerme olacaktır ve hiçbir mantıksal sistemin tutarlılığı garanti altına alınamaz çünkü hiçbir mantıksal sistemin sonsuza kadar tamamlanacağı gösterilemez. Zamanında (1931) pek de hoş karşılanmayan bu iki teorem Jacop Boronowski'ye göre (1966) bence günümüz YZ uygulamalarını ve "düşünebilen makinelerini" içerecek kapsamda değer taşıyor. Boronowski'ye göre "bir aksiyomatik sistem dünyayı tümüyle, noktası noktasına betimleyecek şekilde kurulamaz; bazı noktalarda indirgeme ile doldurulamayacak delikler doğacaktır, başka bazı noktalarda ise birbi-

[1] Bu konuda çok kapsamlı bir inceleme için bkz. Gasparyan, 2016.

rine zıt iki indirgeme zuhur edecektir" (1966: 4).

Bu teoremler günümüzde ispatlandıkları 1931 yılına göre daha da değerli. Bir örnekle tamamlayalım.

Tersinden Düşünmek ya da Bilinç Gerçekçiliği

Boronowski, Gödel'in teoremlerini Turing'in *Entscheidungproblem*'e verdiği olumsuz yanıtla beraber okuyarak önemli sonuçlara ulaşıyor. Boronowski'ye göre mantık ve mantıksallığın sınırları vardır; hayal edilebilecek her türden formel dilin de. Gödel ve Turing'den daha güçlü bir şekilde bu sınırların varlığına dair önermeler Alfred Tarski' den geliyor: "(E)vrensel olan bir kesin dil olamaz; en az aritmetik kadar zengin olan her formel dil ne doğru ne de yanlış olarak gösterilebilecek anlamlı cümleler içerecektir" (akt. Boronowski, 1966: 4). Basit bir dille ifade edecek olursak, ne kadar zengin bir aksiyomatik çerçeve çizersek çizelim, "her şey dahil" bir sonlu formel model icat edemeyiz. Boronowski matematiksel mantığın çizdiği çerçevenin dışını bilim ve edebiyatın ortak itici gücü olan hayal gücünde buluyor. En azından "şimdilik" (Boronowski, 1966) hiçbir makine insan bilgisinin tüm kiplerini formelleştiremez. Çünkü makine bir insan icadıdır ve ancak bizim doğayı (doğal olanı) anlama biçimimizi taklit edebilir ve kullanabilir (14).

O halde, hayal gücü ve kazaen ortaya çıkan (niyetlenilmemiş/olumsal) durumlar evrenin formel bir dile tercümesini imkansız kılıyor. Acaba evrenin (doğanın) mutlak bir dışsal varlık olduğundan ne kadar emin olabiliriz? Doğal olanı anlamak için sunî olanın matematiksel modeline başvurma jestine "tersine modelleme" adını vermiştik. Bilgisayar bilimlerinin bilişsel bilimlere model sunması örneğinde olduğu gibi bu eğilim giderek yaygınlık kazanıyor. Beynimiz bize "dış dünyanın" modelini *bir şekilde* sunuyor. Bu işin nasıl olduğunu tam olarak bilemesek de dış dünya olarak adlandırdığımız gerçekliğin somut, dışsal, insan iradesinden bağımsız varlığından kuşku duymak filozofik eksantrikliğin ötesinde bir anlam ifade edebilir mi? Farklı bir biçimde ifade edecek olursak, Gödel, Turing ve Tarski gibi bilim insanlarının gösterdiği matematiksel/mantıksal imkansızlık, matematik-dışı dil ve oluşumlara (ör., evrenin bütününe) teşmil edildiğinde söz konusu kısıtlılığın daha da güçlü argümanlarla savunuluyor olması gerekmez mi? Bazılarına göre hayır, gerekmez.

Bu yazıyı bilişsel bilimci Donald Hoffman'ın "bilinç gerçekçiliği" (*conscious realism*) adını verdiği yaklaşımla bitirebiliriz. Hoffman 2008 yılında yayınladığı makalesinde, kadim zihin-beden ikiliğine, bu iki unsur arasındaki ilişkiye yeni (kuantum) bir boyut katarak yaklaşıyor. Hoffman'a göre tüm gerçeklik bilinçli eyleyenlerin (*agents*) algı ve deneyimlerinden ibarettir. Dış dünyanın insan beynine giriş kapıları belli olduğuna göre bu kapılardan giren her tür enformasyonun matematiksel bir modeli çıkar-

tılabilir. Nihayetinde ortaya çıkartılan bu model bilinen dış dünyanın modeline eşit olacaktır. Yakın zamanda *Quanta Magazin*'e verdiği mülakatta (2016) bu büyük iddianın Turing'den esinlenmiş bir modelleme olduğundan söz eden Hoffman, anılan matematiksel modeli oluşturmak için altı formel unsurun yeterli olacağını belirtiyor. Bilinç gerçekçiliğine göre dış dünya olarak adlandırdığımız varlıklar âlemi her bilinçli varlığın kendi algılarıyla zihninde oluşturduğu evrimsel (türün devamlılığına odaklı) bir kurgudan, bir yanılsamadan başka bir şey olamaz.

Bu noktada YZ üzerine giriş kısmında dile getirdiğim soruların yanıtlanması için kritik bir eşiğin henüz aşılmadığını söyleyebiliriz. Bu eşik teknolojik bir eşik değildir. Kartezyen aşırı şüpheciliğin kuantum fiziği ve evrim argümanlarıyla yeniden gündeme geldiği bu nokta, hiç kuşkusuz, birinci set soru yanıtlanmadan ortaya ikinci ve farklı düzlemde bir soru seti koyuyor. Dış dünyadaki nesnelerin gerçekten ne olduklarından çok, evrimsel anlamda ne işe yaradıklarına göre gerçeklik tanımlarının yapıldığı bir bilişsel evrenden söz ediyoruz. Özünde "her şey" olan bu bilişsel evrenin varlığını mutlak kabul ettiğimiz anda zekânın zaten yapay olduğunu ya da yapay ve doğal zihinsel işlevler arasında bir fark olmadığını söylemek durumunda kalabiliriz.

1960'ların romantik sezgisellik tasavvuru (Boronowski) hayal gücünü ve sanatı ön plana çıkartan bir duygulanımın altını çiziyor. Bazılarımız bunun gerçekten çok önemli ve hatta zarurî olduğunu düşünebilir. Ben kendimi bu grupta görüyorum. Öte yandan YZ araştırmalarının gittiği yöne bakılacak olursa, kendimize ait zannettiğimiz fikir ve görüşlerin büyük veri setinin optimizasyonuyla elde edilecek olasılıklara indirgendiğini kabul etmek durumunda kalabiliriz. Bunun bir alternatifi olamaz mı?

Elbette olabilir. İlginç bir biçimde YZ araştırmalarının başladığı ve günümüzdeki noktasına ulaşmasında büyük rol oynadığı bilinen MIT' den bir ses, Joichi Ito, kendi araştırma programını açıklarken karşımıza yeni bir kavramla çıkıyor: yaygın zekâ (*extended intelligence* -EI). Bu kavram okuduğunuz yazının sınırlarında derinlemesine açıklanamayacak kadar çok boyutlu. Bu kısıt çerçevesinde kaba bir tarif yapmış olma riskini alarak söylersek, yaygın zekâ, bilinçli davranan tekil eyleyenlerin değil, ortak hareket eden ve zekâyı tek bir insanın kafatası içine hapsetmek yerine yaygın, çoklu, ağ-tipi yapılanmış bir varlık olarak tanımlayan bir kavram. Böyle tanımlanmış bir yaygın zekâ temelinde gerçekleştirilecek YZ ve bilişsel bilim uygulamaları, Ito'ya göre, belki de, etik, efektif ve aklıselim olacaktır. İnsan ve makineyi iki ayrı ve rakip varlık gibi gören tartışmaların tersine, yaygın zekâ perspektifi her iki grup zihnî etkinliğin uyum içinde bir ağ oluşturma olasılığından da ciddi biçimde söz ediyor.

İnsanlar gelecekte makineler gibi düşünebilecek mi? Şimdilik bilemiyoruz…

Kaynakça

Ackroyd, J. A. D. (2011). "Sir George Cayley: The Invention of the Aeroplane near Scarborough at the Time of Trafalgar." *Journal of Aeronautical History* 2011/6:130-181.

Boronowski, Jacop (1966). "The Logic of the Mind." *American Scientist* 54/1:1-14.

Gasparyan, Diana (2016). "Artificial Intelligence and Semantics through the Prism of Structural, Post-Structural and Transcendental Approaches." *Integrative Psychological and Behavioral Science* 50(4): 704-743.

Hoffmann, Donald D. (2008). "Conscious Realism and the Mind-Body Problem." *Mind and Matter* 6(1):87-121.

Hoffman, Donald D. (2016). "The Evolutionary Argument Against Reality." *Quanta Magazine* Nisan 2016, https://www.quantamagazine.org/the-evolutionary-argument-against-reality-20160421/ Erişim tarihi: 05.08.2018.

Ito, Joichi (2016). "Extended Intelligence." https://pubpub.ito.com/pub/extended-intelligence Erişim tarihi: 05.08.2018.

Lafrance, Adrienne (2017a). "How AI Will Redefine Human Intelligence." *The Atlantic*. 11 Nisan 2017.

Lafrance, Adrienne (2017b). "What an AI's Non-Human Language Actually Looks Like?" *The Atlantic*. 20 Haziran 2017.

Newell, Allen ve Herbert. A. Simon (1976). "Computer Science as Empirical Enquiry." *CACM* 19:113-126.

Poole, David L. ve Alan K. Mackworth (2011). *Artificial Intelligence: Foundations of Computational Agents*. New York: Cambridge University.

Robertson, Douglas (2017). "This Is How Facebook's Shut-down AI Robots Developed Their Own Language -And Why It's More Common Than You Think." *The Independent*. 1 Ağustos 2017.

Scott, Alwyn (2002). *Neuroscience: A Mathematical Primer*. New York: Springer-Verlag.

Sejnowski, Terrence. J. (1989). "The Computer and the Brain Revisited." *Annals of the History of Computing* 11:197-201.

Sieg, Wilfried (2015). "Gödel's Philosophical Challenge (to Turing)." *Computability: Turing, Gödel, Church and Beyond*. J. Copeland, C. J. Posy ve O. Shagrir (der) içinde. Cambridge ve MA: The MIT. 183-202.

Turing, Alan M. (1936). "On Computable Numbers, with an Application to Entscheidungsproblem." *Proceedings London Mathematical Society* 42(2): 230-265.

Turing, Alan. M. (1952). "The Chemical Basis of Morphogenesis." *Philosophical Transactions of the Royal Society of London. Sries B, Biological Sciences* 237(641) 37-72.

von Neumann, John (2000, orj.1958) *The Computer and the Brain*. New Heaven: Yale University.

Young, John. Z. (1938). "The Functionaing of the Giant Nerve Fibres of teh Squid." *Journal of Experimental Biology* 15(2):170-185.