ÇELİK İRADE: GÜVEN BAĞLAMINDA YAPAY ZEKÂ ÜZERİNE BİR DEĞERLENDİRME

Seda İŞGÜZAR[1]

"İnsanlık, bir zamanlar evrende zorluklarla tek başına mücadele etti. Yalnızdı; ama şimdi yardımcıları var. Hem de kendilerinden daha sadık ve güçlü, daha inançlı ve kullanışlı yaratıklara sahipler. İnsanlar artık tek başına değil. Hiç bu açıdan düşündünüz mü?" (Asimov, 1950)

Masal anlatıp basketbol oynayan, yemek yapan, çocuk bakan, evle ilgili bütün sorumlulukları üstlenip eksiksiz yerine getiren, üzülen, sevinen, sorunlara pratik çözümler üretebilen, ailenin metal üyesi: Rosie. Neredeyse yarım yüzyıl önce "bir gelecek zaman ütopyası" (Wikipedia, 2021) olarak kurgulanıp yayınlanan Jetson's çizgifilmindeki yaşlı ve tecrübeli robot karakter. On yıllar boyunca, Rosie gibi "insansı makinelerin" tasavvuru, yalnızca güçlü bir hayal gücünün ürünüydü: en azından denk gelen yıllarda köylerinin sadece %7'si elektrikle tanışabilmiş Türkiye gibi ülkelerde yaşayan insanlar için. Şimdi ise gerçekler, hayalleri bile şaşırtıyor.

Sürücüsüz araçlar, hastalık teşhisi koyan akıllı algoritmalar, bir makaleyi bir dakikada özetleyen sinir ağları, ev süpüren robotlar… Yapay zekâ, olanca hızıyla hayatlarımıza nüfuz ediyor. Ancak Turing'in (1950) "Makineler öğrenebilir mi?" sorusunun bugün bizi getirdiği nokta, Asimov'un dediği gibi "insanı evren karşısında daha güçlü kılan" teknolojilerin çok daha ötesinde: "dünya, yeniden doğum sancısı çekiyor".

İnsan ırkını yok mu edecek, işsiz mi kalacağız, ekolojik sorunlar sona mı eriyor, kanser bitiyor mu? gibi korku ya da umut dolu birçok bakış açısıyla tartışılan yapay zekâ kavramına biraz daha yakından bakmak faydalı olacaktır.

Yapay zekâ, akıllı makineleri mümkün kılan bir bilim ve mühendislik dalıdır. Akıllı sıfatı ise "çevresini algılayabilme, amacına yönelik olarak en iyi kararları alıp bunları kendi kendine gerçekleştirebilme" yetisini ifade etmektedir (McCarthy, Minsky, Rochester, Shannon 1955). Yani amaç, hafızalama, anlama, çıkarsama, öneride bulunma (Civelek, 2003) gibi insanlar yaptığında zekâ olarak nitelenecek davranışların makineler tarafından da yapılabilmesini sağlamaktır (Tektaş, Akbaş, Topuz, 2002). Bu bağlamda yapay zekâ aynı zamanda beynimizin çalışma şeklini gösteren bir kuramdır (Tektaş, Akbaş, Topuz 2002); nöronlarımızdaki gizli şifrelerin bir dışavurum şeklidir (Aydın, Değirmenci 2018).

Makinelere zeka kazandırabilmek, ancak filozofların, psikologların, biyologların, sinirbilimcilerin, istatistikçilerin, mühendislerin çalışmaları ve sağladıkları bilgi ile mümkündür. Bu nedenle elbette bu bilimin doğuşuna dair izler ilk çağa kadar sürülebilir. Ancak bugün yapay zekâ üzerine konuşuyor olabilmemize neden olarak Alan Turing'in (1950) "Computing Machine and Intelligence" isimli makalesi referans gösterilir. Çünkü Turing, bu ihtimali ele alan ilk modern makaleyi yazan kişidir (Nilsson, 2019).

"Yapay zekâ (Artificial Intelligence)" terimi ise ilk defa J. McCarthy'nin liderlik ettiği bir grup bilim adamının 1955 yılında "zeki makinelerin yapılabilme olasılığını" araştıracakları projelerine destek bulmak için Rockfeller Vakfı'na yazdıkları ve 1956'da DarthMouth Koleji'nde gerçekleştirilen seminerde sundukları metinde kullanılmıştır (Say, 2018 ; İTÜ Vakfı Dergisi 2018). Yine aynı döneme denk gelen bir çok toplantı, yapay zekânın bir bilim olarak doğmasına neden olmuştur (Nilsson, 2019). O zamandan bu zamana pek çok parlama ve duraksama dönemi yaşanmış ve oldukça yol katedilmiştir. Geçtiğimiz on yılda ise yaşanan gelişmeler büyük bir ivme ile gerçekleşmiş; öyleki yapay zekâlı sistemler bazı spesifik işleri insanlardan daha iyi yapar hale gelmiştir. Hatta daha fazlası vardır: dünya, yapay zekânın diktiği yeni gömleği giyinmek üzeredir.

Harari (2018)'nin ifadesiyle yapay zekâ bir sıçrama noktasıdır ve "insanlığa hayatı yeniden tasarlama ve şekillendirme gücü bahşetmiştir". Yenidünya düzeni insanlara, kurum ve kuruluşlara, makinelerin rutin ve sıradan işlerle ilgilenmesi, insanlarınsa kendilerine kalacak olan zaman diliminde daha yaratıcı çalışmalara odaklanması, yeni fırsatlar için kafa yorması gerektiğini söylemektedir. Rahatlığı, emek tasarrufunu vaat eden her teknoloji gibi… Elbette bu durum, "dar yapay zekâ" adı verilen, araba sürmek, satranç oynamak gibi bir ürün, hizmet vb. ile ilgili belli bir görevi gerçekleştirmeye odaklı (Frank vd., 2019) yapay zekâ türü için geçerlidir. Dar yapay zekâlı sistemler, "temel

[1] Dr., MEB, Bilişim Teknolojileri Öğretmeni
e-mail: seda.isguzar@gmail.com

fayda sağlamayı merkezde tutan, insanlara konfor alanı yaratan" sistemler olarak görülmektedir. Sağlıktan, finansa, eğitimden, otomotive neredeyse her sektör, bu yapay zekâ türü ile yollarını kesiştirmiştir.

Harari ise, daha çok, kendilerine pek de uzakta olmadığımız genel ve süper yapay zekâ türlerinin doğuracağı güçten bahsediyor olmalı. Çünkü genel yapay zekâda artık, insanlara ait bilişsel faaliyetleri taklit etme becerisi, insanınkinden ayırt edilemez durumdadır. Burada, karşılaştığı "her durumu" bir insan gibi algılayabilen, ona göre reaksiyon gösterebilen, bütün kararları "tek başına" alabilen, ilgili işi bir insan gibi ya da insandan daha iyi yapabilen sistemler söz konusudur. Şimdilik bütünüyle "genel yapay zekâ" denebilecek bir "akıllı makine" mevcut değildir (Ackermann 2019). Süper zeka ise kimilerinin üstinsan (Karabulut, 2020) olarak betimlediği, dünyadaki bütün kolektif akıllardan daha da zeki bir yaşam formunu ifade etmektedir (Dermedya 2016). Bilim kurgu filmlerinde işlenen, insanlığın sonunu getireceğinden korkulan yapay zekâ türü budur.

Ortada Varoluşsal Bir Risk mi Var?

Moravec (1988), Zihin Çocukları isimli kitabında "robot" denen bu makinelerin henüz tam anlamıyla "zeki" sıfatını hak etmediğini, bir bebek gibi ilgiye ve geliştirilmeye ihtiyaç duyduğunu ifade eder. Ancak şunu da ekler "bugünün emekleyen makineleri, 21. yüzyılda insanı da onun tahminlerini de aşan ve ataları insan olduğu için onurlanacak varlıklar olacaklardır."

Moravec'in öngörüsü henüz gerçekleşmiş değil; hatta bazı araştırmacılara göre bunun için "birkaç nobel ödüllük daha zaman" (Say, 2018; Bostrom, 2020) var. Çünkü insan beyninin çalışma prensiplerini bütünüyle öğrenme, robotlara his ya da sağ duyu kazandırabilme, işlem gücü ve kapasitesi yüksek ağların ve bilgisayarların gerekliliği gibi konularda daha fazla ilerleme gerekiyor (Say, 1998; Bostrom, 2020). Ancak birçok araştırmacının net şekilde emin olduğu bir şey varsa, o da "bağımsız, kendi kendine öğrenebilen işbirlikçi makineler" çağının başladığı (Eberl, 2019). Elbette bu görüşün sağlamasını, gelişmelere ufak bir göz atarak dahi yapmak mümkün. Sayısız alanda, sayısız yapay zekâlı sistem kullanılıyor. Üyesi olduğunuz sosyal ağlar, alış-veriş siteleri vb. birçok platform beğeni, gezinme, tıklama, izleme alışkanlıklarınızdan sizi tanıyarak sundukları içeriği sizin için kişiselleştiriyor. Cep telefonları birer asistan sunuyor, Google Çeviri gibi araçlar sayesinde ise herhangi bir dile karşı yabancılık çekilmiyor. Alibaba gibi e-ticaret devlerinin depolarındaki işlerin çoğu depo robotları tarafından idare ediliyor. Yıllardır, otomotiv başta olmak üzere birçok sektör, fabrikalarında endüstriyel robotlar kullanıyor. Sürücüsüz araçlar dahi, –ki bu konuda en iddialı isimlerden biri şüphesiz Google- rüştünü çoktan ispatladı.

Elbette akıllı makinelerin, iştahını yalnızca böylesine mavi yakalıların işleri kabartmıyor. Bilgi işçileri de kariyerlerini yan koltuklarındaki "yapay muadillerine" teslim etmeye hazırlanıyor.

Humantelligence gibi yazılımlar, iş başvurularını değerlendirip en az bir insan kaynakları yöneticisi kadar isabetli kararlar veriyor. IBM tarafından üretilen Ross isimli robot avukat, binlerce avukata hizmet ediyor (Rossintelligence,2021). Çin'de tıp sınavını geçen Asistan Doktor Al, göreve başlayalı üç yıl oldu. Humanoid robot BINA48'in bir konferensta konuk konuşmacı olarak sunum yapmasının üstünden ise sekiz sene geçti.

Buraya kadar akıllı makineler, hayatı kolaylaştıran, insanla uyumlu çalışabilen ve kötücül beklentiyle her yeni teknoloji gibi mevcut işleri otomatize eden ve insanları işsizlik tehlikesi ile karşı karşıya bırakan varlıklar olarak görülebilir. "Geçmişte de benzer durumları deneyimledik, endişeye gerek var mı?" diye düşünülebilir. Buhar makinesi, mekanikleşmeyi getirdiğinde, makinenin insan emeğinin yerini alıp, onları atıl hale getireceğinden kaygı duyulmuştu (Frank vd., 2019). Ancak bu olumsuz kehanet, korkulan düzeyde gerçekleşmemişti. Bazı meslekler kaybolsa da, açıkta kalan işgörenleri geri emecek yeni iş sahaları ortaya çıkmıştı. Sosyo-teknik uyarlamalar ile yenilik, normalleşmişti. Peki şimdi ne değişti?

Bu yüzyılda teknolojinin, insanlığı önceki asırlardakinden çok daha büyük ve eşi benzeri görülmemiş zorluklarla karşı karşıya bırakacağı düşünülüyor. Öyle ki gücün mikro düzeyde seçkin/ seçilmiş bir kitlenin eline geçip insanlığın geri kalanını gereksiz kılabilecek "dijital diktatörlüklerin" oluşması ihtimal dışı değil (Harari, 2018). Facebook'un iki yapay zekâlı robotunun buluşturulup "acaba hangimiz daha çok insana benziyoruz?" sorusundan, hangi takımı tuttuklarına kadar birçok konuda sohbet ettirilmelerinin üstünden uzun zaman geçmedi. Yani duyulan endişelerin abartı olarak görülmesini engelleyecek birçok gelişmeye tanık olundu. O nedenledir ki insanların ve onlardan daha zeki, becerikli, yorulmayan, acıkmayan androidlerin bir arada bulunduğu gelecek kurguları, oldukça distopik. Üstelik kapıdaki devrim yüzyıllar değil yıllar içerisinde, baş döndüren bir hızla gerçekleşiyor. Bu da demek oluyor ki, yeni çağda fiziki, sosyal ve psikolojik varlığının tehlikeye girmesini engelleyecek düzenlemeler için insanlığın sahip olduğu zaman oldukça az. Bahsedilen değişimin hızını ve boyutunu kestirmekse zor değil. Son altmış yıldır bilgisayarların işlem gücündeki gelişimini ve hızlanmayı öngörmemizi sağlayan Moore Yasasının, miadını doldurduğunu tartışan çalışmalar, tespitlerin doğruluğunu gösteriyor. Çünkü hem donanım hem yazılım hızları, hem belleklerin hem de ağların taşıyabileceği bilgi kapasitesi, Moore yasasının yapabileceği tahminin çok üzerinde (Ford,

2020). Quantum bilgisayarlar, plazmonik elektronik devrelerle tasarlanan mikroskobik bilgisayarlar… geleneksel bilgi teknolojilerine kafa tutmaya başladı. Bilgi teknolojileri de kendi ezberlerini bozmak üzere. Bütün bunlar, çok büyük verilerin çok yüksek hızlarda işlenebileceği ve yapay zekânın bu sayede beklenenden daha yüksek ve kısa sürede bir sıçrama yapabileceği anlamına geliyor. Eşiğine geldiğimiz bu devrimin nasıl sonuçlar doğuracağına dair hiçkimsenin net yargılar üretmesi ise mümkün değil.

Artık toplumların geleceğini, sosyologlar, eğitimciler, akademisyenler yöneticiler, seçmenler, politika yapıcılar değil mühendisler ve ekonomik olarak onları besleyen öğeler inşa ediyor. Sanırım henüz ne getireceği bilinmeyen bir mevzuda, bu kadar yazılıp çizilmesinin, temelinde de "bizim için olduğu söylenen; ama içinde bizim olmayacağımız bir gelecek yaratılıyor ve biz seyirci mi kalıyoruz?" sorusu yatıyor. Çünkü geçmişten günümüze gelen anlatılar, yeni toplumsal düzeni oluşturmada ve teknolojik yeniliğe bağlı her dönüşümü "insanlığın kadim varlığını" sarsmayacak şekilde sağlamada etkililerdi. Şimdi ise toplum bağışıklığı kazandıran kuramların, öğretiler ve modellerin, bu yeni durumda ekonomiden-kültüre her sahanın nasıl şekillenmesi gerektiğine dair veri tabanında taşıdığı bilgi oldukça az. Yani "toplumun dokusu", "parçalanma tehdidi" (Nilsson, 2019; Ford, 2020) ile karşı karşıya ve araştırmacılar, yeni ahlaki, ekonomik, kültürel, sosyal, hukuki normlara ihtiyaç olduğu konusunda hemfikirler (Harari, 2018; Ford, 2020).

Bir yanda Musk, Gates gibi teknoloji devlerinden Hawking gibi bilim insanlarına, felsefecilerden teologlara kadar birçok kesim çeşitli argümanlarla "yapay zekâyı belli sınırlar içinde tutma ve bunu sağlayacak regülasyonları yapma gerekliliğini" dile getiriyor. Öte yanda ise bütün bu tartışmalardan azade olan, akıllı makinelerin eğitimden-ev işlerine kadar geniş bir spektrumda sunduğu fırsat ve kolaylıktan istifade eden, bir kesim bulunuyor. Bugün tüketiciler, milyarlarca veri arasından istediği sonucu kısacık sürede listeleyen bir arama motorundan; kendisine tam da zevkine göre film, müzik önerileri sunan platformlardan; kendisi kahve içerken evi süpüren robottan değil, onların sağladığı konforu kaybetmekten korkuyor. Çoğu insan, kullandığı ürün ya da hizmetlerdeki yapay zekâ dokunuşunun farkında bile değil ki bu durum elbette eleştirilmeyi hak etmiyor. Zaten kıyamet senaryoları da böylesine belli hizmetlere odaklı "ilkel/ dar yapay zekâ" değil, "süper zeka" için yazılıyor. Elbette unutulmaması gereken şey, yaygarası koparılan insanüstü makinelerin gelişim gücünü şimdi sahip olduğumuz teknolojilerden alacak olduğudur. Demek ki, sınırlama ve düzenlemeler konusunda geç kalınırsa "yapay zekâ, kimseye ihtiyaç duymadan kendi gelişimini sürdürüp insanı aradan çıkarabilir" kaygısı hayat bulabilir.

Daha önceden de belirtildiği gibi kimsenin –şimdilik- bütün bunlara verebilecek net bir yanıtı yok; ancak teknolojinin kısa sürede alışılmadık bir dönüşüm yaratacağı konusunda neredeyse görüş birliği var: bireylerden şirketlere varlığını sürdürmek isteyen her sistemin bunu fark etmesi gerektiği hususunda olduğu gibi… Bu nedenle de yapay zekâ hata yapar mı, yaparsa ne olur? Ön yargılı mı? Peki sağ duyu gerektiren kararları alabilecek potansiyele ulaşabilir mi? Düşük vasıflılardan başlayarak yüksek vasıflılar dahil bütün işleri otomatize ettiğinde insanlar hem üretici hem de tüketici sıfatlarını bırakmak zorunda kalabilir mi ve bunun önüne geçmek mümkün olabilir mi? sorguları yüksek sesle yapılmakta ve şekillenmeye başlayan gelecekte, söz sahibi olabilmek için yapılabilecekler üzerine kafa yorulmaktadır.

Bu sorulardan birkaçına verilen cevaplara bakmak, yerinde olacaktır.

Yapay Zekâ, Karar ve Ön Yargı

Bilimde ve ticarette mühim araçlar haline gelen yapay zekâ uygulamalarından bazıları, bugün görünmez olacak kadar günlük hayatımıza işlemiştir (Nillson, 2020). Öyle ki bir konuma ulaşmak için harita uygulamalarının sunduğu rotaları kullanırken ya da sosyal ağlarda paylaşılan fotoğrafta yer alan kişiler otomatik olarak etiketlenirken yapay zekâlı bir sistemin varlığı hissedilmez. Bu ve buna benzer bir çok araç hayatın "normali" olarak görünür, üstüne düşünülmez. Diğer yanda ise bahsi geçen teknolojileri kullanan birine "seninle robot doktorumuz ilgilenecek" dendiğinde –şimdilik- bu durumu aynı kabullenmişlik ile karşılayabileceğinden şüphe duyulur. Prensipte aynı temel üzerine inşa edilmiş gibi görünen iki örnekten birine teslim olunabilirken ötekine neden güven duyulmuyor? Hatta daha genel bir soru soracak olursak: asıl mesele yapay zekânın yanlış yapma ihtimali midir?

Aslında bütün bunlara tatminkar bir yanıt verilip verilemeyeceği tartışılmalıdır. Çünkü teknolojik olsun olmasın her yeni şeyin aksayan yönleri, hataları, yan etkileri bulunur. Zaman içerisinde bunlar giderilmekle kalmaz, her seferinde eskisinden daha fazlasını yapan versiyonlar türer. O yüzdendir ki mevcut duruma yönelik sorgulamalara kısmi cevaplar bulunabilecek olsa da genel çıkarımlar yapmak, kurgusal ifadeler üretmekten öteye gidemeyecektir. Kısmen de olsa fikir sahibi olabilmek içinse karar almanın doğasına, yapay zekâlı sistemlerde kullanılan öğrenme yaklaşımlarına, algoritmalara ve bunların başarım oranlarına bakmak gereklidir.

Gündelik hayatta işe giderken giyinilecek kıyafeti seçmekten, hastanın hangi tedaviyi alması gerektiğine, personel alımından, yatırımlara kadar her konuda bir kararlar silsilesi ile karşılaşılır ki karar vericiler bağlamında bunlar bireysel ya da örgütsel nitelik taşıyabilir (Yaşar, 2016). Karar verme, mevcut problem ya da durumlara karşı, ilgili alternatifler doğrultusunda kişilerin takınmaları gereken tavırların ya da eylemlerin ne olması gerektiğine dair yapılan öngörülerdir (Daft, 1989). Yani, istenen sonucu alabilmek için ihtiyaç duyulan verinin elde edilmesini; daha

sonra sitematik ve bilimsel yollar izlenerek alternatifler oluşturulmasını ve bunların arasından en uygununun seçilmesini kapsayan bir süreçtir (Tekin, 2010). Elbette bu tanım, "rasyonel karar verme modellerine dayanan analitik karar verme stilleri" için geçerlidir. Rasyonel karar verme modeli, insanın akılcı ve ekonomik olduğunu kabul eder. Bu karar vericilerin mümkün olan bütün seçenekleri sonuçlarıyla beraber bildiği bilinciyle hareket ettiğini ve optimal kararlar verdiğini öne sürer (Tural, 1988).

Britanyalı Psikolog Stuart Sutherland'ın (2009) ise dikkat çektiği bir başka nokta vardır: sezgilere, dürtülere dayanan irrasyonel kararlar. Kişi, bilgisine dayanarak en doğru sonuca ulaşmayı amaçlar, bu rasyonel bir davranıştır. Bilgisinin yetersiz olduğu durumda karar vermek için daha fazla kanıt arayışına girmesi de aynı şekilde rasyoneldir. Ancak kişiler çoğunlukla sadece ellerindeki bilgileri destekleyecek kanıtlar aradıklarından büsbütün irrasyonel davranırlar. Bu nedenle Sutherland (2009), en rasyonel kararı vermenin her zaman en iyi sonucu getireceğini garanti etmeyeceğini, insani konularda "şans" faktörünün göz ardı edilemeyeceğini söyler. Hatta insanların kararlarının büyük çoğunluğunun irrasyonel olduğunu savunur. Elbette bu iddialar irrasyonel kararların, rasyonel kararlardan daha iyi olduğuna dair bir savı temellendirmek için yapılmaz. Aksine insanların düşünce sistemlerinde pek çok yanlışın var olduğunu ortaya koymaya çalışır. Bazen başka tür kararların daha iyi neticelere yol açtığını; fakat en iyi şekilde potansiyel kullanma isteği varsa mümkün olduğu kadar rasyonel kararlar verilmesini öğütler.

Bazıları için ise karar verme, sezgi ve rasyonellik olmak üzere, ikili bir sürecin çıktısıdır (Kahneman, 2003; Sadler- Smith &Shefy, 2004). Sezgi, tekrar eden yaşantılar yoluyla öğrenilen duygular ve otomatizmler yani alışkanlıklarla ilgilidir (Kahneman, 2003). Sezginin altında yatan mekanizmalardan biri "sinaptik modifikasyon" diğeri "patern tanımadır". İlki, öncesinde analitik olan eylemin "sabitlenerek alışkanlığa dönüşmesi" ve bilincin müdahalesi olmaksızın devreye girmesi ile oluşur. İkincisi ise önceki deneyim ve örneklerle o an ki durum arasındaki farklılıkları ve benzerlikleri keşfetmeyi içerir (Gür, 2021). Sağduyu, insiyatif, vicdan gibi etmenler ile burada karşılaşılır. "Böylesi daha doğru geldi, neden bu kararı aldım bilmiyorum; ama böyle hissettim." gibi kurulan cümleler çoğunlukla sezgisel kararların sonucu olarak görülür. Özet olarak, kararların hem analitik hem de sezgisel sürecin bir çıktısı olduğu söylenebilir; ancak kararın içeriğinin karar vericinin bilgisine dayandığını unutmamak gerekir (Benner vd., 1992).

İnsanların karar verme mekanizmaları ile günümüzün karar vericileri haline gelmeye başlayan yapay zekâlı sistemlerin karar mekanizmaları arasında bir mukayese yapılabilmesini sağlamak için buraya kadar insani karar verme süreçlerine değinilmiştir. Bugün yapay zekâlı sistemler daha karmaşık ve farklı konularda karar verebilir durumdadırlar. Bunu da insan beyninin bilgi işleme süreçlerini modelleyerek öğrenme ve karar verme prensibi üzerine kurulu öğrenme yaklaşımları ile sağlarlar. Daha önce de kısmen değinildiği gibi problemi tanımlama, kriterleri belirleme, alternatifleri oluşturma, çözüm arama ve değerlendirme, seçim yapma ve uygulama basamakları rasyonel karar verme sürecine aittir. Bugünkü yapay zekânın temel rolü ise sonuç çıkarım yeteneklerini kullanarak çoğunlukla "arama ve değerlendirme" basamaklarını yönetmektir. Yani yapay zekâ şu an sınırlı bir role sahiptir; ama çok geç olmayan bir gelecekte bu prosesin tüm adımlarında daha fazla etkin olması beklenmektedir (Yapay zekâ, 2021).

Bu noktada şu çıkarımı yapmak mümkündür.Yapay zekânın kararları, rasyonel verilere dayanırken, insani kararlar bilgi toplama ve işlemenin direkt bir sonucu olmayabilir. Sezgi dünyasından, bilinçaltından doğabilir, sağduyuya dayanabilir. Yapay zekâ alan yazınında, eğitim al(a)mamışlar dahil olmak üzere bütün kişilerin "dünya hakkındaki ortak bilgilerine" verilen teknik isim, sağduyudur. Örneğin bilindik bir meselede bir ansiklopedi maddesi yazıldığını varsayalım. İlgili konunun anlaşılır olması için gereken; fakat "bunu yazmaya gerek yok nasılsa herkes bilir" diye açıklanmayan her şey sağduyunun kapsamındadır. İnsanların birbirleriyle çok konuşmadan anlaşabilmesinin ya da aynı dili konuşuyor olsalar da birbirlerini yanlış anlamalarının ardındaki sebep herkesin ortak saydığı bu bilgi tabanındaki farklılıklar ve benzerliklerdir (Say, 1998). Yapay zekâ sistemlerinin ise sağduyu, önsezi gerektiren durumları anlama konusunda daha az becerikli olduğu düşünülmektedir. Yapay zekâ insan dili üzerinden öğrenerek dünyaya, kültürlere dair oldukça isabetli çıkarımlar yapsa da bir insanın zorluk çekmeksizin kavradıklarını anlamada başarısızlık göstermektedir. Örneğin Martin Luther King Jr'ın sivil haklar mücadelesi sırasında tutuklanmasını konu alan bir makaleden öğrenen yapay zekâ, Afro-Amerikalılara karşı olumsuz bir etiketleme yapabilmektedir. İnsanlar bunun Nobel Barış Ödülü alacak kadar haklı bir direniş hikayesi olduğunu bilse de yapay zekâ için Martin Luther King, siyahtır ve tutukludur (Altındiş, 2017). Bu nedenle yapay zekâ, insanlara kıyasla özellikle önceden tanımlanmış bir bilgi alanının dışındaki belirsiz veya öngörülemeyen ortamlarda daha az uygulanabilirdirler. Şu halleriyle çok yetenekli olsalar bile hayal gücü, yaratıcılık, üstün sezgiler gerektiren durumlarda nitelikli sonuçlar doğuramamaktadırlar (Jarrahi, 2018). Bütün bunlar yapay zekâya duyulacak güvenin boyutlarını belirleyen ilk unsurdur.

Bütün bunların aşılabilmesi, yapay zekaya sezgi, sağduyu gibi insana özgü kimi özellikler kazandırılabilmesi için yıllardır çalışılmaktadır. Her ne kadar Searle ve Penrose gibi kimi araştırmacılar aadece biyolojik sistemlerde ve insanlarda bulunan; duygular, benlik, nöronlardaki oynak elektronlar gibi durumlardan kaynaklanan taklit edilemez bir özün var olduğunu ve onun asla yapılamayacağını savunsa da (Say, 1998); bunun aksini düşünen çok kişi vardır. Onlara göre, insanlar aslında sihirli sezgilere sahip değidlir. Kararlar "gizemli bir özgür iradeye değil" çok yüksek

hızda ihtimalleri hesaplayan nöronlara bağlı olarak alınmaktadır. İnsani sezgiler ise, "örüntü tanıma" yetisinden başka bir şey değildir. Aynı bakış açısına göre duygular da arzular da biyokimyasal etkileşimlerden ibaret olabilir ve bu durum, bilgisayarların kutsal yaratmanın sonucu oluşan ruhu anlayabilmenin yolunu da açabilir (Harari, 2018; Say, 2018). Gerçekleşir mi ya da ne zaman ve nasıl gerçekleşir bilinmiyor; ancak o güne dek insanın, avantajlı durumda olmaya devam edeceği açıktır. Çünkü "öngörü, veriye dayanır" ve insanlar halen makinelerin bildiğinden çok daha fazla şey bilmektedir. Bunun yanı sıra duyu organlarının sağladığı bilgi, makinenin algılayıcılarının çok önündedir. Öğrenmek için insanların tercihlerini bilmeye muhtaç olan makineler ise insanlara ait verileri kullanabilmek için (reklamlar vb. aracılığı ile) bir nevi ücret ödemek zorundadır. İnsanların halen yeterince tutucu olduğu, mahremiyetlerini önemsedikleri ve zihin sağlıkları, cinsel hayatları, aykırı düşünceleri gibi verileri ulu orta paylaşmadıkları da göz önüne alınmalıdır. Bu durumda makineler insana ait bütüncül ve sağlam bilgilere sahip olamayacak dolayısıyla – en azından şimdilik- tam anlamıyla ilgili özelliklerle donanamayacaktır. Bütün bunlarla birlikte fazla veri yokken bile ne yapılması gerektiğine dair karar almada insanların, makinelerden daha iyi olduğu unutulmamalıdır (Agrawal, Gans, Goldfarb, 2019).

Yapay zekânın, sezgilere dayalı kararlar almada henüz yetkinliğe sahip olmamasının dışında bir de mevcut halleriyle yaptığı hatalar vardır ki bu hayat bulma sürecinde oldukça doğaldır. Bir insan yapay zekâyı kontrol eder ve yanlışları düzeltir. Yapay zekâ zamanla bir insanın denetimine gerek kalmayacak hale gelene kadar hatalarından ders çıkarır (Agrawal, Gans, Goldfarb, 2019). Yine de bunlara değinmek o hataların nedenini keşfetmek açısından önemlidir. Dünyanın ilk robot avukatı olarak 2017 yılında hizmete giren DoNotPay, İngiltere'de haksız yere yazıldığı düşünülen park cezalarına itiraz edilmesine olanak sunan bir bir chatbot. DoNotPay aracılığı ile 250.000 şikayet yapılmış 160.000'ini haklı bulunmuştur. Bu da %60'dan daha fazla başarım oranı demektir (Cloudsnames, 2021). Aynı DoNotPay, ise şu an işlediği suç ile başa çıkmakla meşgul. Üyeliği bulunmayan birine, ilgili kişinin izni olmadan hizmetlerini açıklayan ve kaydolmayı talep eden metin mesajları göndermekle suçlanıyor. Bununla birlikte bir araştırma şirketine yine kendisiyle ilgili %60 kötü, %40 mükemmel şeklinde bildirilmiş görüşler mevcuttur (Adhikari, 2020). Bu noktada tartışılması gereken bir diğer husus elbette sorumluluğun kime ait olduğudur; ama şimdilik buraya bir virgül koyalım. Yani yüzlerce katmana milyonlarca yapay sinir hücresine sahip yapay öğrenme modelleri, yanılmaz değiller (Kızrak, 2019). Bunlar veri setinin çeşitliliğe sahip olmaması, kullanılan öğrenme algoritmaları, modelde kural koyucuların bilerek ya da bilmeyerek gösterdiği zafiyetler gibi birçok nedenden kaynaklanabilmektedir. Son yıllarda adını çokça duyduğumuz birçok konuda insan seviyesine erişebilen kararlar alabilmesiyle popüler olan derin öğrenme ile oluşturulan modeller, çok sayıda doğrusal olmayan fonksiyon kullanır. Ancak bu kadar fazla fonksiyona ait bir yapı, gittikçe daha karmaşık sonuçlar üretmeye başlamıştır. Bu durum sonuçların yorumlanmasını ve anlaşılmasını zorlaştırmaktadır (Kızrak, 2019). Yani sistem neden bunu yaptı, neden başka bir şey yapmadı, ne zaman başarılı ne zaman değil, ne zaman ona güvenmeliyiz sorularının cevaplanması güçleşmektedir (Turek, 2021). Ayrıca derin öğrenme yaklaşımları ile görüntü işleme uygulamalarına tek-piksellik ataklar yapılabilmektedir. Bir görüntüde sadece bir piksel değiştiğinde bile insan gözü hala görüntüyü doğru biçimde algılayabilirken, oluşturulan model büyük hatalar yapabilmektedir. Örneğin, otonom bir aracın kamera girdisinde insan gözüyle fark edilmeyen küçük değişiklikler yapılarak aracın önündeki yayaları görmemesi ve yolun açık olduğunu düşünmesi sağlanmıştır (Say, 2018). Bu durum derin sinir ağlarında kırılganlık yaratmakta ve temelindeki öğrenilmiş karar verme süreçlerini şüpheli hale getirmektedir. O yüzden özellikle hastalık teşhisi, ceza uygulamaları gibi güvenlik bakımından şeffaflık gerektiren uygulamalarda regresyon ya da karar ağaçları gibi daha kolay açıklanabilir makine öğrenmesi modelleri tercih edilebilmektedir. Sinir ağı yaklaşımlarının yaygın kamu güvenini kazanması ve gerçekten adil olmaları için üretilen sonuçların açıklanabilir olması gereklidir (Kızrak, 2019). Nasıl? sorusuna ise XAI başlığında cevap verilecektir.

Yapay zekânın verdiği kararların güvenirliği ile ilgili olarak akıllara takılan bir diğer durum bu sistemler için optimum kararın ne olduğudur. Mesela bir kaza kaçınılmaz olduysa otonom araç yolcuları mı korumalıdır, yoksa toplamda ölecek kişi sayısını en aza indirmenin yollarını mı aramalıdır? Bir İHA sivil kayıplar yaşanacak olsa bile, bölgedeki teröristleri etkisiz hale getirmeli midir? (Everett, 2017). Alzheimer hastalığını azaltma amacıyla tasarlanmış bir ileri yapay zekâ için optimum kararın altmış beş yaşını geçen kişileri öldürmek olmayacağının garantisi var mıdır? Bu noktada, yapay zekâ istenmeyen durumlar yarattığında sorumluluğun kimde olacağı, canlılığı tehdit edebilecek karar alma potansiyelinin varlığı gibi etik meselelere değinmek gereklidir. Önce şunu kabul etmek gerekir ki bahsedilen konularda karar vermek insanlar için bile oldukça zordur ve elbette bunlar, dar yapay zekâ için geçerli olabilecek söylemler değildir. Çünkü dar yapay zekâ, bir insan tarafından kendisine kodlanan kurallar ile hareket eder ya da yalnızca bazı davranışlara karşı sorumlu olur. Yani irade ve bilinç sahibi değildirler, varlıklarının farkında olamamaktadırlar. Dolayısıyla şimdi hayatımızda olan bu tür yapay zekâlı sistemlerin etiksel bir statüye sahip olmadığı düşüncesi hakimdir (Çelebi, 2017). Etik kaygısı daha çok genel ya da süper yapay zekâ için konuşulması gereken bir durumdur. Çünkü onlar tam anlamıyla insanlardan müteşekkildir, bilinç sahibi olacakları düşünülmektedir ve insanla aynı etiksel seviyede davranmalıdır. Henüz hayat bulmamış türlerle ilgili bu bağlamda

çıkarımlar yapmak çok mümkün değildir. Ancak böyle yapay zekâlar üretilecekse onun kritik kararları –hele ki bu kararlar canlılık ile ilgiliyse- nasıl vereceği üzerine düşünülmesi, yapay ahlâk entegre edilmesi, iyi-kötü, doğru-yanlış kavramlarının öğretilmesi gereklidir (Uyar, 2017, Topakkaya & Eyibaş, 2019). Her şeye rağmen DoNotPay örneğinde olduğu gibi yapay zekâ istenmeyen durumlar yaşattığında sorumluluğun kime ait olacağına dair yorumlar ilgili yapay zekânın irade ve bilinç sahibi olup olmamasıyla ilişkilendirilir. Dar yapay zekâlı sistemlerin mantıksal çıkarımlar yapabiliyor olması onun bilinçle hareket ettiğini göstermez, sadece görünürde zekice davranmaktadırlar. Bu nedenle de yapay zekâlı sisteme sorumluluk yüklenmesi mümkün değildir. Peki bir başkasının hakkı gasp edilir ya da sistem suç işler hale gelirse üreticisine sorumluluk yüklenebilir mi? Bu durumda iki senaryodan söz edilir. İlkinde eğer hata makinenin beklenmeyen sonuç üretme potansiyelinden kaynaklıysa, üreticiden bağımsız bir durum gelişmiş demektir ve sorumluluk yüklenemez. İkincisinde ise üretici yapay zekânın yanlış yapma ihtimalini görmesine rağmen sistemi kullanıma sunuyorsa sorumluluk üreticidedir (Topakkaya & Eyibaş, 2019). Ancak ikinci durumda bu durumun ispat edilebilir olması önem taşır. Bu da hatanın kaynağının iyi keşfedilmesine bağlıdır. Tabi görünürde zeki davranmayan, gerçekten bilinç sahibi akıllı makineler türediğinde robot mahkemelerinin kurulması kaçınılmaz olabilir.

Yapay zekâya duyulan güveni etkileyen ve etik tartışmalarının merkezinde olan bir diğer unsur ise önyargıdır (BIAS). Bu konuya Princeton Üniversitesinde Bilgisayar Bilimleri alanında doktora çalışmasını yapan Aylin Çalışkan'ın (2017) geliştirdiği GloVe isimli sistemle başlayalım. GloVe, metinler arası ilişkileri inceleyen, Wikipedia'dan ya da tarafsız dile sahip olduğu düşünülen haber metinlerinden oluşan bir veri seti ile eğitilen yapay zekâlı bir yazılımdır. Geliştiricisi olan Çalışkan'ın elde ettiği sonuç ise net: "sistem, tamamen insanların temel önyargılarını yansıtıyor, dildeki örtük önyargıları cımbızlayıp kullanıyor." (Altındiş, 2017). Elbette yapay zekâlı sistemlerin önyargısına dair tek örnek bu değil. Şartlı tahliye için en uygun adayları tahmin etmek için tasarlanan COMPAS isimli sistemin, siyahi tutukluları "suç işlemeye daha meyilli" olarak etiketlemesi; AMAZON'un işe alım için geliştirdiği aracın, kadın adayları filtrelemesi, MICROSOFT'un geliştirdiği Tay isimli chatbotun Twitter'da insanlarla sohbet etmesinin üzerinden 24 saat bile geçmeden ırkçı, soykırım yandaşı ve küfürbaz bir araca dönüşmesi bunlardan sadece birkaçı (Jones, 2019). Bütün bunlar uzun yıllardır üzerinde durulan önyargı tartışmasına yeni bir boyut daha katmaktadır: insanlar farkında olmadan içselleştirmiş/ öğrenmiş oldukları önyargılarını dil üzerinden yapaya zekaya mı aktarıyor? Peki insanların makinelere öğretmek için kullandıkları iyi/kötü, doğru/yanlış etiketlemeleri önyargıdan arınabilir mi? Yapay zekânın yakıtı olan veriler –gizil olsa da- böyle yönlendirmeler ile karşı karşıya ise bu sistemlerin iddia ettikleri tarafsızlığa nasıl güvenilir? "Basit, eğitim verilerinin doğru tutarlı, ilgi düzeyi yüksek ve yansız olması bütün bu önyargı sorunlarını çözer" (Poslu, 2020) diyebilsek de maalesef gerçek dünya verileri bu ifadenin gerçeğe dönüşmesini engellemektedir. Bu noktada yapay zekâda önyargı türlerine ve bunun nedenlerine bakmak doğru olacaktır.

Herkes kararlarının mantığa ve adil, doğru algılara dayandığına inanmak istese de, maalesef tüm insanlar bilinçsiz önyargılara sahiptir. Bilinçsiz önyargı, belirli kişiler, şeyler veya gruplar lehine veya aleyhine sahip olunan görülerdir (McDade, Testman, 2019). Bugün insanlara ait, ırk, cinsiyet, din vb. öğeler içeren yüz seksen önyargının varlığından söz edilmektedir. Yapay zekâda görülen önyargı türleri ise bunların kaynağına göre belirlenmiştir.

- Örneklem Yanlılığı/Önyargısı (Sample Bias): Bu önyargı türünde sapma, veri kümesiyle başlar. Algoritmayı eğitmek için kullanılan veriler, modelin çalışacağı problem alanını doğru bir şekilde temsil etmediğinde oluşur. Örneğin otonom bir aracın hem gece hem gündüz çalışması bekleniyorsa; ama yalnızca gündüz verileriyle eğitilip gece sürüş verileriyle eğitilmiyorsa örneklem yanlılığı oluşur (Jones, 2019; Swooptalent, 2019; Rossi, 2020).
- Önyargılı Önyargı (Prejudicial Bias): Bir veri seti cinsiyet, ırk gibi dikkate alınması yanlılık doğurabilecek veriler içeriyorsa o zaman önyargılı önyargı oluşabilmektedir. Amazon örneğinde olduğu gibi sistemin erkek cvlerin daha fazla olduğu bir veri kümesi ile eğitilmesi kadınların aleyhine sonuçlar doğmasına sebep olabilir (Jones, 2019; Swooptalent, 2019).
- Ölçüm Sapması/ Önyargısı (Measurement Bias): Bu tür bir önyargı, gözlemlemek veya ölçmek için kullanılan cihazla ilgili bir sorun olduğunda meydana gelir. Eğitim için toplanan veriler üretim sırasında toplanan verilerden farklıysa ortaya çıkar. Bu önyargı, yalnızca daha fazla veri toplanarak önlenemez. Birden fazla ölçüm cihazına ve bu cihazların çıkışını karşılaştırmak için eğitilmiş insanlara sahip olmak bir çözüm olabilir (Jones, 2019; Swooptalent, 2019).
- Dışlama Önyargısı (Exclusion Bias): Veriler, makine öğrenme modelini eğitmeden veya test etmeden önce önişleme tabii tutulur ve genellikle gereksiz, ilgisiz görünen veriler silinir. Bu işlem bu tür önyargının oluşmasına sebep olabilir (Jones, 2019; Swooptalent, 2019).
- Algoritma Önyargısı (Algoritm Bias): Bu son önyargı türünün verilerle hiçbir ilgisi yoktur. Bir modelin eğitildiği verilerden değil, makine öğrenimi modelinin kendisinden kaynaklanır. Modelin aşırı öğrendiği durumlarda gerçekleşir (Jones, 2019; Swooptalent, 2019).

Özetle, yapay zekâdaki ön yargılar, evren, popülasyon ya da örneklemden; veri toplama araçlarından, toplumsal bakış açılarından ve bunlarla yetişmiş geliştiricilerinden ya da algoritmik hatalardan kaynaklanabilir. Bu önyargıların yapay zekâ sistemlerine müdahil olması karar verme şekillerini etkileyecektir. Yalnızca daha yüksek verimlilik, daha yüksek performans amacı ile hareket etme lüksünün olmadığı açıktır. Her önyargı türünün önüne geçmek için veri çeşitliliğini arttırmak, sonuçları yorumlayabilecek insan kaynağı yetiştirmek ya da istihdam etmek, veri toplama ve kullanma politikalar gibi çözüm önerileri mevcuttur. Ancak bunun daha büyük ölçekte mücadele edilmesi gereken ve ilerlemeyi engelleyen bir sorun olduğunun farkında olan IBM, Google gibi firmalar yapay zekâlı sistemlerdeki önyargıların keşfedilmesini ve azaltılmasını sağlayan araçlar (What-If, AI Fairness 360..) geliştirmişlerdir. Genel olarak, adalet, şeffaflık ve güven gibi değerlerin merkezde olduğu etik ve değer odaklı bir yaklaşım ile yapay zekâ önyargılarını gidermenin mümkün olduğuna inanılmaktadır. Böylece, yalnızca insan ön yargılarını kopyalayan veya güçlendiren sistemlerden kaçınılmış olunmaz; aynı zamanda insanların daha adil olmasına yardımcı sistemler üretilebilir (Rossi, 2020). Elbette tamamen şeffaf ve "açıklanabilir modeller" oluşturana kadar, makine öğrenimi modellerimizdeki önyargıyı ölçmek ve en aza indirmek için ilgili araçlara güvenmek ve gereken ihtiyatla hareket etmek gereklidir.

Karar alırken tarafsız olması ve manipülatif davranmaması beklentisi, yapay zekânın tercih edilirliğini arttıran önemli bir unsurdur. Ancak tartışıldığı gibi bu sistemler çeşitli sebeplerden ötürü objektif davranamayabilmektedir. Bu bağlamda ne kadar güvenilir olduklarını tespit edebilmek ve ilgili kararlarının neden alındığını bilmek için programların uygulanabilir olması kadar incelenebilir olması gereklidir (Topakkaya, Eyibaş, 2019). Bu noktada karşımıza açıklanabilir yapay zekâ modelleri çıkmaktadır.

XAI

MIT Technology Review'de uzun yıllardır yapay zekâ editörü olan Will Knight (2017) "Yapay zekânın Kalbindeki Karanlık Sır" başlıklı yazısında şöyle bir ifade kullanmaktadır: "Hiç kimse en gelişmiş algoritmaların, yaptıkları şeyleri nasıl yaptığını gerçekten bilmiyor: geliştiricileri bile. Bu bir sorun olabilir." Bahsedildiği üzere derin öğrenme bir çok konuda sorun çözme becerisinin oldukça yüksek olduğunu ispatlamış bir makine öğrenmesi modelidir. Kullanıldığı sistemlerde ise ilgili görevin nasıl öğrenileceğinden, gerçekleştirileceğine kadar neredeyse hiçbir aşamada insan müdahalesi bulunmamaktadır; fakat bu durum yapay zekâlı sistemlerin kamaşıklığını arttırmakta (Goodman, Flaxman, 2017), yorumlanabilirliğini ve şeffaflığını azaltmaktadır. Bu sistemler gittikçe "kara kutu" olarak nitelenen ilgili sonuca nasıl, neden varıldığı sorgularına cevap veremeyecek bir kapalılığa bürünmektedir. Yalnız unutmamak gerekir ki yasa koyucular, işletmeler, resmi kurum/ kuruluşlar ve genel kullanıcılar zamanla yapay zekâlı sistemlere daha da bağımlı hale gelecektir. Bu nedenle de açıklanabilirlik, güven ve şeffaflık ilkeleri önem kazanacak ve sağlanabilmesi içinse karar verme sürecinin hesap verebilir şekilde tasarlanması gerekecektir (Özel, 2021). Yapay zekâya güvenilip benimsenebilmesinin, yönetilebilmesinin önünde engel olan bu durumun, açıklanabilir yapay zekâ modelleri (Explainable Artficial Intelligence-XAI) ile giderilebileceği düşünülmektedir (Gunning, 2016).

"Şeffaf yapay zekâ" diye de tabir edilen XAI, yapay zekâ ile elde edilen sonuçları, insanların anlayabileceği basitlikte ortaya koyabilmek amacıyla izlenen yöntem ve tekniklerdir (Labservicetech, 2020; Özel, 2021). Yapay zekâlı sistemlerin, yüksek seviyede öğrenme performansını korurken daha açıklanabilir modeller üreten bir makine öğrenimi teknikleri paketi oluşturmayı vaat etmektedir (Gunning, 2016). XAI, yapay zekâ sistemlerinde güveni karakterize eden birkaç özellikten biridir (Phillips vd., 2020). Son kullanıcının iş, kredi gibi herhangi bir başvurusunda değerlendirme yapan yapay zekânın neye göre kişiyi elediğini bilme hakkı, geliştiricinin önyargı ya da istenmeyen sonuçların neden kaynaklandığını keşfetme gerekliliği gibi durumlar bugün bu kavramın konuşulmasının nedenidir. Son yıllarda daha fazla gündeme taşınsa da kökleri, yapay zekânın öncülerinden Minsky'ın aynı kaygılarla öne sürdüğü Hümanist Zeka modeline kadar dayanmaktadır (Özel, 2021). Elbette XAI'nin gerekliliği tartışmalı da bir konudur. İnsanların bile aldıkları kimi kararlar açıklanamazken bunu yapay zekâdan beklemenin anlamlı olmadığını düşünen bir kesim de mevcuttur (Kızrak, 2019).

Bir yapay zekânın güvenilir olabilmesi için, adil, hesap verebilir, değerlerle uyumlu, sağlam, tekrarlanabilir ve açıklanabilir olması gerektiğine dair bir görüş birliği vardır (Kızak, 2020). Bütün bunları sağlayacağı düşünülen XAI modellerinin temelde şu üç durumu karşılaması beklenmektedir (Kızrak, 2019):

- Sistemi geliştirenlerin ve kullananların nasıl etkilendiğini açıklamak
- Veri kaynaklarının neler olduğunu ve sonuçları nasıl etkilediğini açıklamak
- Girdilerin nasıl çıktılara sebep olduğunu açıklamak

XAI modellerinin bunları nasıl yapacağından neler içermesi gerektiğine kadar birçok konu, Google, IBM gibi yapay zekâ üreticilerinden psikologlara kadar geniş bir çevrede tartışılmaktadır. Kaliteli bir makine-insan simbiyozuna öncülük edeceği düşünülen XIA'nin sahip olması gereken ilkeleri Philips ve arkadaşları (2020) şu şekilde açıklamıştır:

- Açıklama: Yapay zekâlı sistemler, sundukları tüm çıktılara eşlik eden kanıt(lar) sunmalıdır.

- Anlamlılık: Yapay zekâlı sistemler, bireysel kullanıcılara anlaşılabilir açıklamalar sunmalıdır.
- Açıklamanın Doğruluğu: Sunulan açıklamalar, sistemin çıktıyı oluşturma sürecini doğru şekilde yansıtmalıdır.
- Bilgi Sınırları: Sistem, tasarlandığı koşullar dışında çalıştırıldığında sonuç üretmek için bilgi sınırlarının yeterli olmayabileceğini beyan edebilir nitelikte olmalıdır. Aksi takdirde yanlış yargılar üretilebilir.

Bütün bunlardan hareketle yapay zekâlı sistemlerin insan düzeyinde tasarlanabilmesi için derin öğrenme gibi karmaşık tekniklere ihtiyaç duyduğu açıktır. Bunun yanı sıra kullanıcılara karşı sorumlu hale getirilmeyen yapay zekâların, Hollywood filmlerinin yıllardır sunduğu kurgulara gerçeklik kazandıracak kadar büyük riskler barındığıracağı da aynı netliği taşımaktadır. Bu nedenle tarafsız ve denetlenebilir sistemler için XAI üzerinde daha fazla kafa yorulması gerekmektedir.

SONSÖZ

Yapay zekâ, yaşama, çalışma, savaşma, oynama, eş arama, gençlerimizi eğitme ve yaşlılarımıza bakma şeklimizi derinden etkileyecek, işgücü piyasalarını altüst edip sosyal düzenimizi yeniden şekillendirecek bir teknoloji olarak kapıda beklemektedir. Öyle ki günün sonunda, makineler yaratıcılarından azade hedefler peşinde koşup sadece insannın söz sahibi olduğu alanlarda bile onlardan daha iyisini yaparken insanlar evrendeki yerini tartışıyor olabilir (Kaplan, 2016). İnsanoğluna şimdiye kadar ki emek ve gayretleri teşekkür edilip gelecek cyborglara mı teslim edilecek, elbette buna dair net bir hüküm verebilmek şimdilik mümkün değil. Hatta bazıları böyle kötücül senaryoları anlamlı bile görmemeyip "yapay bir zekanın doğal bir aptallıktan daha iyi olduğu" (Say, 1998) gibi keskin söylemler kullanmakta; bazıları ise gelecekteki süper zekalarla baş edebilmenin yollarını arayarak insan beynine çip yerleştirmeye hazırlanmaktadır. Kimin haklı çıkacağı bilinmezlik taşıyor; ama geçmiş dönemlerden çok daha farklı bir teknolojik devrimin gerçekleşmeye başladığı ise muhakkak. İnsanların ve makinelerin birlikte adapte olması gereken bir dünya düzeni şekilleniyor. Bu nedenle bu sistemlerin nasıl çalıştığı, hangi sınırlar çerçevesinde üretilmeleri gerektiği, ne kadar güvenilir oldukları, insanları atıl hale getirmeyecek ve refahını sağlayacak yeni ekonomik, sosyolojik vb. poltikaların neler olması gerektiği, nasıl ahlaki ve yasal çerçevelerin oluşturulacağı gibi ciddi sorular üzerinde düşünülmektedir. Ancak gerekli görünen bu regülasyonlar için sahip olunan zaman ise oldukça az.

Bunun yanı sıra hali hazırda günümüzde hayatımıza sinmiş olan yapay zekâlı sistemlerin bugünden yarına taşınacak etiksel kusurları ve güvenirlik sorunu bulunmaktadır. Önyargılı karar alma potansiyelleri, sağduyu gerektiren kararlar verirken hata yapabiliyor olmaları, kullandığı öğrenme algoritmalarına göre doğruyu gerçekten doğru öğrenip öğrenemedikleri, çıktılarına ait gerekçelerinin şeffaf olmaması gibi durumlar şimdiden çözüm bulunması gereken meselelerdir.

Görüldüğü üzere dünya doğum sancısı çekmektedir ve yeni bir çağ hayat bulmak üzeredir. İster bireysel ister örgütsel düzeyde olsun değişim korkusu, umutsuzluk hissettirmekte; adapte olamamayı dolayısıyla varlığını sağlıklı şekilde sürdürememe riskini içeren değişememe korkusunun önüne geçebilmektedir (Ayde & Düşükcan, 2002). Bu değişim korkusu, şu an varlığı pek hissedilmeyen bu sistemler, birkaç yıl sonra iyiden iyiye baş gösterdiğinde Luddizm benzeri tepkilere neden olur mu, olsa da neyi engelleyebilir, bilinmiyor. Çünkü bilim-kurgu filmlerinin yerleştirdiği korku bir tarafa, bugün bile yapay zekâlı bir hayatın güvenilmez olarak algılanmasına sebep olan birçok gelişme yaşanıyor. Deepfake gibi sahte içerik üretebilip bilgi sağlığını yok eden, zararlı uygulamalar bunlardan sadece biri…

Elbette farklı boyutları ele alarak yapay zekânın güvenirliğinden bahsederken insanların yaptıkları hataları, adaletsizlikleri göz ardı etmek doğru olmayacaktır. Üstelik yapay zekânın bazı durumlada daha iyi işler çıkardığı bilinirken: tıpkı kompozisyonları, insanlardan daha yüksek doğruluk seviyesinde notlandırabilen makineler (Ford, 2020) örneğinde olduğu gibi… Buradan hareketle insanların makinelere göre daha irrasyonel davranabildiği, daha yavaş olduğu söylenebilir. Ayrıca insanlar tek tip de değillerdir (Epstein, 2015). Bu nedenle karar alma noktasında insan ne kadar aradan çıkmalı ya da yapay zekâ ne kadar irade sahibi olmalı sorguları önemlidir. Çünkü iki türün bir arada yaşayabilmesi için yapılacak düzenlemelerin çerçevesini belirlemektedir. Aynı bağlamda, Ian Brown'un yapay zekâ geliştiricilernin kendilerine sormalarını önerdiği şu sorular, bahsedilen simbiyotik yaşamın önünü açabilir. Tasarladığınız algoritmanın bütünüyle ne yaptığını biliyor musunuz; bunları kullanıcılara onların anlayacağı şekilde açıklayabilir misiniz ve son olarak açıkladığınız zaman bu sistem için yaptıklarınız onları mutlu edecek mi? İlgili sorulara verilecek yanıtlar, kaçışı olmayan bir dönüşüm için açıklanabilir yapay zekâ modellerinin temelini teşkil edecektir.

Özetle şimdilik yapay zekâlı sistemler vicdan, insiyatif gibi duyular gerektirmeyen, rutin ve rasyonel kararları –önyargılardan arındırmak kaydıyla- almada verim ve hız sağlayan, insan hayatını kolaylaştıran yararlı teknolojiler olarak hayatımızda yer almaktalar (Baştan, 2003). Bunu yapabilmek için de kişisel veriler toplamak ve işlemek zorundadırlar. Veri toplanmasından işlenmesine kadarki süreçten alınan sonuçlara kadar her aşamanın açıklanabilir

ve yorumlanabilir olması insan dostu yapay zekâlara yaşamımızda yer açmamıza imkân sunacaktır. Yapay zekânın kendi bildiği kurallarla, kendisine alan yaratılmasına fırsat vermeden hükümranlığını ilan etmesini istemeyen yasa koyuculardan bireysel kullanıcılara kadar herkes proaktif yaklaşımlar sergilemek ve bunları hayata geçirmek mecburiyetindedir. Kaçışı olmayan bu dönüşüm sürecinde, makinelerin objektif olduğundan ve iyi amaçlar için tasarlandığından emin olunduğunda gelişmeler korku ile değil heyecanla karşılanabilecektir.

Referanslar

Ackermann T., J., (2019) "The 3 types of AI: Narrow (ANI), General (AGI), and Super (ASI)", Erişim Adresi: https://www.bgp4.com/2019/04/01/the-3-types-of-ai-narrow-ani-general-agi-and-super-asi/, Erişim Tarihi: 15 Nisan 2019.

Adhikari, R. (2020), "Robot Lawyer Faces Legal Troubles of Its Own", Erişim Adresi: https://www.technewsworld.com/story/86956.html, Erişim Tarihi: 03.02.2021

Agrawal A., Gans J., Goldfarb, A. (2019). *Geleceği Gören Makineler.* (M. Ürgen, Çeviri).İstanbul: Babil Kitap, 1. Baskı

Altındiş, D. (2017). "Yapay Zekâ ama Önyargılı", Erişim Adresi: https://kurious.ku.edu.tr/haberler/yapay-zeka-zeki-ama-onyargili/, Erişim tarihi: 23.02.2021

Asimov, I., (1950). *I, Robot*, New York: Gnome Press

Ayden, C , Düşükcan, M . (2002). Örgütsel Öğrenme Kavramı Ve Öğrenme Engellerinin Giderilmesinde Örgüt Kültürü Ve Liderliğin Rolü . Sosyal Ekonomik Araştırmalar Dergisi , 2 (4) , 120-139

Aydın İ.H., Değirmenci, C. H., (2018). "*Yapay Zekâ*", Girdap Yayınları, 1. Baskı.

Bastan, S. (2003). "Yapay zekâ, Yeni İletisim Teknolojileri ve Orgutsel Degisim: AkilliOrguteDogru" Celal Bayar Üniversitesi IIBF Cilt:10 Say :1

Benner P, Tanner C, Chesla C. From beginner to expert: Gaining a differentiated clinical world in critical care nursing. Advances in Nursing Science 1992;14(3):13-28

Bostrom, N. (2020). *Süper Zeka.*(F. B. Aydar, Çev). İstanbul: Koç Üniversitesi Yayınları, 2.Baskı

Brown, I. (2020), "Yapay zekâda Sorunlar", Erişim Adresi: https://www.hairezmi.com/yapay-zekada-sorunlar/, Erişim Tarihi: 05.03.2021

Civelek Ömer, (2003). "*Yapay Zekâ*", Türkiye Mühendislik Haberleri, Sayı 423.

Cloudsnames, "Avukat Robot 160.000 Dava Kazandı", Erişim Adresi: https://cloudnames.com.tr/teknoloji/teknoloji-avukat-robot-160-000-dava-kazandi/, Erişim Tarihi: 22.02.2021

Çelebi, Ö., F. (2017). Yapay Zekâ ve Etik. İstanbul Medeniyet Üniversitesi Sosyal Bilimler Enstitüsü, s. 1-20.

Daft,F. R. L. (1989). OrganizationTheoryand Design. St. Paul: West Publishing Company.

DerMedya (2016). "Yapay zekâ nedir? (Yapay zekâ'ya Giriş)", Erişim Adresi: https://www.dermedya.com/post/yapay-zeka, Erişim tarihi: 14.04.2019.

Eberl, U. (2019). *Akıllı Makineler.(L. Tayla, Çev.)* Paloma Kitapları, 1. Baskı

Epstein, S.L. (2015). Wanted: Collaborative intelligence. Artificial Intelligence, 221, 36-45.

Everett, J., (2017). "İnsanlar, Yapay zekâya ve Robotlara Neden Güvenmiyor?", (Evrimağacı, Çeviri),Erişim Adresi: https://evrimagaci.org/insanlar-yapay-zekaya-ve-robotlara-neden-guvenmiyor-5279, Erişim Tarihi:19.02.2021

Ford, M. (2020). *Robotların Yükselişi.* (C. Duran, Çevirisi). Edtör: Murtaza Özeren. İstanbul:Kronik Kitap, 7. Baskı

Frank, M., Roehrid P., Pring, B., (2019). *Makineler Her Şeyi Yaptığında Biz Ne Yapacağız?.* (E, Yılmaz, Çev) Aganta Kitap

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine*, *38*(3), 50-57. https://doi.org/10.1609/aimag.v38i3.2741

Gunning, D., (2016), "Explainable Artificial Intelligence (XAI), DARPA/ I20", Erişim adresi: https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf, Erişim tarihi: 28.02.2021

Gür, M. H., "Sezgisel Karar Verme Ve Yönetimde Sezgi", Erişim Adresi: https://kararverme.wordpress.com/2019/05/17/sezgisel-karar-verme-ve-yonetimde-sezgi/, Erişim tarihi: 11.02.2021

Harari, Y. N. (2018). *21. Yüzyıl İçin 21 Ders* (S. Siral, Çev.). İstanbul: Kolektif Kitap, 7. Baskı.

İTÜ Vakfı Dergisi, "İnsanlaşan Makineler ve Yapay Zekâ", Sayı:75, Ocak-Mart 2017.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586.

Jones, M.T (2019). "Machine learning and bias", Erişim Adresi: https://developer.ibm.com/technologies/machine-learning/articles/machine-learning-and-bias/, Erişim tarihi: 01.02.2021.

Kahneman, D. (2003). A Perspective on Judgement and Choice.*American*Psychologist.Vol.58, No. 9, 697–720.

Kaplan, J. (2016), *Artificial Intelligence: What Everyone Needs to Know,* Birleşik Krallık: Oxford University Press.

Karabulut, A. (2020). "Niceliğin Egemenliği ya da Dataizm: Yeniçağın Nihilizmi", *Dijital Yaşam, Yeni İnsan ve Sonrası*, Editörler: Aydın Karabulut, Emrah Arğın (İstanbul: Literatürk), 135-153.

Kızak A., (2019), "Açıklanabilir Yapay zekâ Nedir ve İhtiyaç Mıdır?", Erişim Adresi: https://ayyucekizrak.medium.com/a%C3%A7%C4%B1klanabilir-yapay-zeka-nedir-ve-i%CC%87htiya%C3%A7-m%C4%B1d%C4%B1r-65adef9b086, Erişim tarihi: 24.02.2021

Kızak, A. (2020), "Açıklanabilir, Sorumlu ve Güvenilir Yapay Zekâ", Erişim Adresi: https://ayyucekizrak.medium.com/a%C3%A7%C4%B1klanabilir-sorumlu-ve-g%C3%BCvenilir-yapay-z, Erişim tarihi: 24.02.2021

Knight, W., (2017). "The Dark Secret at the Heart of AI", Erişim Adresi: https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/,Erişim Tarihi:17.02.2021

Labservicetech, "Açıklanabilir Yapay zekâ (XAI)", Erişim Adresi: https://tur.labservicetech.com/explainable-artificial-intelligence-367721, Erişim Tarihi: 01.03.2021

McCarthy J., Minsky, M. L., Rochester N., Shannon C. E., (1955), Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

McDade, M., Testman, A ., (2019). "Tackling bias in AI", Erişim Adresi: https://www.ibm.com/blogs/systems/tackling-bias-in-ai/, Erişim Tarihi:19.02.2021

Mehmet Karaca, Sayı 75, 18-21. İstanbul: İstanbul Teknik Üniversitesi Vakfı Yayınları.

Moravec, H. (1988). Mind children: The future of robot and human intelligence. Harvard University Press.

Nilsson, J. N. (2019). *Yapay Zekâ Geçmişi ve Geleceği,* Boğaziçi Üniversitesi Yayınevi, 2. Baskı.

Özel, U., (2020). "Açıklanabilir Yapay zekâ (Explainable AI)", Erişim Adresi: http://www.umutozel.com/, Erişim Tarihi:25.02.2021

Phillips P. J., Hahn C. A., Fontana P. C., Broniatowski, D. A., Przybocki M. A, (2020), "Four Principles of Explainable Artificial 14 Intelligence", Erişim Adresi: https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8312-draft.pdf, Erişim Tarihi: 01.03.2021

Poslu, M. (2020). "İstatistiksel Önyargı ve Yapay zekâ", Erişim Adresi: https://www.datasciencearth.com/istatistiksel-onyargi-ve-yapay-zeka/, Erişim Tarihi: 02.02.2021

Rossi, F., (2020). "How IBM Is Working Toward a Fairer AI", Erişim Adresi: https://hbr.org/2020/11/how-ibm-is-working-toward-a-fairer-ai, Erişim Tarihi:19.02.2021

Rostintelligence, Erişim Adresi: https://rossintelligence.com/about-us, Erişim tarihi: 10.02.2021

Sadler-Smith, E.; &Shefy, E. (2004). Understanding and Applying 'Gut Feel' in Decision- Making. The Academy of Management Executive (1993-2005), Vol. 18, No. 4, Decision- Making and Firm Success (Nov., 2004), pp. 76-91.

Say C., (2018). *50 Soruda Yapay Zekâ*. İstanbul: Bilim ve Gelecek Kitaplığı, 9. Baskı. ISBN:978-605-5888-58-9

Say, C. (1998). "Akla Doğru". *Cogito*, 13, ss:67-76

Sutherland, S. (2009). *İrrasyonel.* İstanbul: Domingo.

Swooptalent, "Machine Learning Bias", Erişim Adresi: https://www.swooptalent.com/hubfs/Machine%20Learning%20Bias%20Overview%20-%20Kevin%20at%20SwoopTalent.pdf?_ga=2.194658286.295599295.1557329951-129422730.1537454063, Erişim Tarihi:12.02.2021

Tekin, Ö. A. ve Ehtiyar, R. (2010), "Yönetimde Karar Verme: Batı Antalya Bölgesindeki Beş Yıldızlı Otellerde Çalışan Farklı Departman Yöneticilerinin Karar Verme Stilleri Üzerine Bir Araştırma", Journal of Yasar University, Cilt: 20, Sayı: 5, ss. 3394-3414.

Tektaş, M., Akbaş, A. & Topuz, V. (2002). Yapay Zekâ Tekniklerinin Trafik Kontrolünde Kullanılması Üzerine Bir İnceleme. I. Uluslararası Trafik ve Yol Güvenliği Kongresi.

Topakkaya, A., Eyibaş, Y. (2019). Yapay Zekâ ve Etik İlişkisi. *Felsefe Dünyası*, *70*(1), 81-99

Tural, N. (1988), "Rasyonel Karar Kuramı ve Eğitim Yönetiminde Karar Kuramı", Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, Cilt: 21, Sayı: 1, ss. 497-508.

Turek, M., "Explainable Artificial Intelligence (XAI)", Erişim Adresi: https://www.darpa.mil/program/explainable-artificial-intelligence, Erişim Tarihi:19.02.2021

Turing, A. (1950). Computing Machinery and Intelligence. *Mind, 59* (236), 433-460. Retrieved January 16, 2021, from http://www.jstor.org/stable/2251299

Uyar, T. (2017). Ya Yapay Ahlâk. *İnsanlaşan Makineler ve Yapay Zekâ*, Editör:

Wikipedia, "Jetgiller",Erişim URL: https://tr.wikipedia.org/wiki/Jetgiller, Erişim Tarihi, 02.01.2021

Yapay zekâ, Erişim Adresi: https://web.itu.edu.tr/~sonmez/lisans/ai/yapay_zeka_problem_cozme.pdf, Erişim Tarihi: 15.02.2021

Yaşar, O. (2016). *Davranışsal Karar Verme, Düşünme, Problem Çözme*. Ankara: Detay Yayıncılık.