

## Yapay Zekâ ve Özgür İrade: Yapay Özgür İradenin İmkânı<sup>1</sup>

Mehtap DOĞAN\*

### Öz

Yapay zekâ çalışmalarının nihai amacı insan zihnini bütünüyle yapay bir sistemde yeniden inşa etmektir. İnsan zihni ise zekânın yanı sıra bilinç, duygulanım, karar verme, özgür irade gibi fiziksel açıklamasını yapmakta zorlandığımız öznel nitelikler barındırmaktadır. Bu sebeple, insan zihnine eş değer bir yapay zekânın imkânını tartışmak için bu niteliklere sahip olup olamayacağına dair derin bir değerlendirme yapılması gereklidir. Özgür irade de insan zihnini karmaşık bir hesap makinesi olmaktan uzaklaştıran niteliklerin başında gelmektedir. Ancak özgür iradenin hem varlığı hem de tanımı teorik ve empirik düzeylerde ele alınan önemli bir sorunsaldır. Eylemlerimizin ve kararlarımızın özgür irademizle gerçekleştiğine o kadar eminizdir ki günlük hayatımızda irademizin varlığını sorgulamaya dahi gerek duymayız. Ancak hem klasik fizik hem de nörobilim, varlığından şüphe etmediğimiz iradenin büyük bir yanığı olabileceğine dair güçlü saptamalarda bulunmaktadır. Bu çalışmada, özgür iradenin varlığı ve tanımına dair fiziksel ve fiziksel-olmayan teorik yaklaşımların temel tezleri ortaya konulmuş, bu teorilerin yapay özgür iradenin imkânına dair savunabilecekleri muhtemel iddialar analitik bakış açısı ile öngörülmeye çalışılmıştır. Sonrasında ise bilişsel özgürlük ve ontolojik özgürlük ayrımı üzerinden yapay özgür irade için alternatif bir bakış açısı geliştirilmek istenmiştir. Buna göre, yapay zekâ sistemlerinin özgürlüğünün bilişsel seviyede mümkün olabileceği ancak ontolojik olarak özgür yapay zekâdan başka bir deyişle bir "özne" olarak yapay zekâdan bahsetmenin ilkece mümkün olamayacağı iddia edilmiştir.

**Anahtar Kelimeler:** Yapay Zekâ, Zihin Felsefesi, Özgür İrade, Determinizm, Bağdaşıcılık, Bağdaşmazcılık

\*Dr. Öğr. Üyesi, Ankara Yıldırım Beyazıt Üniversitesi, Felsefe Bölümü, mdoğan@ybu.edu.tr

Doğan, M. (2021). Yapay Zekâ ve Özgür İrade: Yapay Özgür İradenin İmkânı. TRT Akademi, 6 (13), 788-811. DOI: 10.37679/trta.969912

<sup>1</sup>Bu makale, 7-8 Ekim 2020 tarihlerinde düzenlenen Nörobilim, Hukuk, Psikoloji ve Ötesi Sempozyumunda sunulan bildirinin genişletilmiş ve geliştirilmiş hâlidir.

### Derleme Makale

Geliş Tarihi: 11.07.2021

Kabul Tarihi: 23.09.2021

## Artificial Intelligence and Free Will: The possibility of Artificial Free Will

Mehtap DOĞAN

### Abstract

The ultimate goal of researches on artificial intelligence is to reconstruct the human mind in an artificial system. The human mind involves many subjective qualities such as consciousness, decision-making, affection, free will together with intelligence. Therefore, in order to discuss the possibility of artificial intelligence equivalent to human mind, it is necessary to discuss the possibility of these subjective qualities in artificial systems. Free will is one most important properties of human mind differentiated from complicated calculator. Nevertheless, both the existence and the description of free will are huge problematics in theoretical and empirical status. In daily life, we are sure concerning freedom of our actions and decisions; hence, we are never need to be skeptical about the existence of our will. However, both classical physical theories and neuroscience have strong evidences about the idea that free will is just an illusion. In this study, by analyzing physicalist and anti-physicalist theories on the description and existence of free will, it is tried to predict their viewpoints on artificial free will. Then, it is tried to develop an alternative approach to artificial free will by the distinction between ontological and cognitive free will. Thereafter, it is claimed that the free will in artificial intelligence systems is possible only in cognitive level; ontologically free systems are not nomologically possible.

**Keywords:** Artificial Intelligence, Philosophy of Mind, Free Will, Determinism, Compatibilism, Incompatibilism

### Review Paper

---

Received: 11.07.2021

Accepted: 23.09.2021

---

*Akıl sahibi varlıkların bütün eylemleri, görünüşler olmaları bakımından, doğanın zorunluluğu altındadırlar ama bu aynı eylemler, akıl sahibi özneyle ve bu öznenin sırf akla göre eylemde bulunma yetisiyle ilgilerinde özgürdürler (Kant, 2015: 99).*

## 1. Giriş

Yapay zekânın ortaya çıkışındaki en temel motivasyon insan zekâsının taklit edilmesidir. Yapay zekâ, insanın tarih boyunca her zaman farklı biçimlerde de olsa var olan kendine benzeyen varlıklar üretme hayalinin ulaştığı en son noktadır. İnsan artık yapay zekâ ile birlikte yalnızca bedensel olarak kendisine benzeyen ya da kendi bedensel iş gücünü azaltmak için kullandığı makineler yapmak değil aynı zamanda zihinsel olarak kendisine benzeyen ve kendi zihinsel işgücünü de azaltacak makineler üretmek istemektedir. İnsanın hayal gücünün ve teknoloji üretme kapasitesinin en önemli ürünlerinden olan yapay zekâ, bir diğer yönüyle de insanın yaptığı “zekâ gerektiren” işleri yapabilen makineler ya da bilgisayarlar geliştirmek için araştırmalar yürüten çalışma alanına da verilen isimdir. Ancak “zekâ” kelimesinin bu alan içerisindeki kullanımının, insanın “hesap yapabilme, analiz etme, akıl yürütme yeteneği” olarak bilinen sözlük anlamlarından daha geniş bir içeriğe sahip olduğunu vurgulamak gereklidir. Hem yapay zekâ uygulamalarının hem de bir çalışma alanı olarak yapay zekânın başlangıçtaki ve de günümüzdeki hedefleri yalnızca insanın hesap yapabilme ya da akıl yürütme yeteneğini taklit etmekle sınırlı değildir. Bundan fazlası olarak yapay zekâ, insanın zihinsel yetisini tümüyle taklit etmeyi ve hatta yeniden üretmeyi amaçlamaktır. Yapay zekânın kavram olarak ilk kez kullanıldığı ve yapay zekânın bağımsız bir araştırma alanı olarak ilk kez görüldüğü 1955'te John McCarthy, Marvin Minsky, Nathaniel Rochester ve Claude Shannon tarafından “Yapay Zekâ Yaz Araştırma Projesi” başlığıyla sunulan metinde de açıkça görüldüğü üzere yapay zekâ alanı yalnızca insan zekâsının matematiksel bir modellemesini yapmayı amaçlayan bir alan olarak kurgulanmamıştır. Söz konusu proje metninin giriş bölümündeki şu cümleler yapay zekâ araştırmalarının amacının gelişmiş hesap makineleri üretmekten çok daha fazlası olduğunu açıkça ortaya koymaktadır:

Darhmouth Koleji Hanover, New Hampshire'da, 1956 yazında, yapay zekâ üzerine 2 aylık, 10 kişilik bir çalışma yapılmasını teklif ediyoruz. Çalışma, öğrenmenin tüm yönlerinin veya zekânın başka özelliklerinin prensipte, bir makinenin taklit edebileceği derecede ayrıntılı olarak tarif edilebileceği kestiriminden yola çıkacaktır. Makinelerin dili kullanmalarının, soyutlamalar yapmaları ve kavramlar oluşturmalarının, şu anda yalnızca insanlara özgü olduğu düşünülen problemleri çözmelerinin ve kendilerini geliştirmelerinin nasıl sağlanabileceğini bulmak üzere araştırmalar yapılacaktır. Dikkatle seçilmiş bir bilim insanı grubunun bir yaz boyunca bu problemler üzerine birlikte çalış-

masının önemli ilerlemeler sağlayacağını düşünmekteyiz. (McCarthy, Minsky, Rochester ve Shannon, 1955, s.2).

1955'ten bu yana yapay zekâ çalışmalarının geldiği nokta ve son yıllarda hızla gelişen yapay öğrenme teknolojisi birlikte düşünüldüğünde hedeflenenin aslında yapay bir insan zihni üretmek olduğu açıkça söylenebilir. Zihin ise zekânın yanı sıra bilinç, irade, duygulanım, karar verme, öz-farkındalık gibi berimsel ve sembolik olarak taklit edilebilmesi zor olan "öznel" niteliklere sahip pek çok unsuru kendisinde barındırmaktadır.

Bilgisayarlar, makineler, dijital aletler ve de robotik teknoloji geliştikçe, insana ait niteliklerin her birisinin bu yapay sistemlerde mümkün olup olamayacağına dair sorular giderek daha anlamlı bir hâl almaya başlamıştır. Klasik bilgisayarlar bizlere mühendisler tarafından programlanmış, kuralları onu programlayan insana ve doğa yasalarına ait berimsel makineler olarak görünseler de makine öğrenmesi, kuantum bilgisayarlar ve konuşan insansı robotlar gibi teknolojik gelişmeler insana dair her türlü niteliğin yapay sistemlerde de üretilebilmesi olasılığını sorgulamaya açmıştır. İnsan zihnine dair bilinmeyen unsurlar çözüldükçe ve makinelerin bilişsel kapasitesi insana yaklaştıkça, insan ve yapay sistemler arasında yaptığımız keskin ayrımların ve her ikisinin tanımlarına çizdiğimiz koyu çizgilerin bulanıklaşmaya başlayacağı açıktır. Giderek zekileşen ve karmaşıklaşan bilgisayarların kendi kararlarını almaları, duygu durumları, özgür seçimler yapmaları, bilinçli olarak adlandırabilmeleri ve dahası benlik sahibi olmaları artık yalnızca kurgusal düzeyde değil teorik olarak da tartışılan ihtimaller olarak göze çarpmaktadır. Kuşkusuz insan zihninin diğer öznel nitelikleriyle birlikte özgür irade de makinelerin evrende sahip olabileceği role yönelik ayırt edici unsurlardan olacaktır.

Özgür seçimler yapabilen ve özgür kararlar veren akıllı makinelerin varlığı insan kontrolünden bağımsız yapay zekâ sistemleri anlamına gelmektedir. Özgür yapay zekâ sistemleri yazılımları ile sınırlı kalmayan, hedeflerini güncelleyebilen makineler olacaklardır. Özgür yapay zekâ sistemleri ile ilgili kaygı duyulmasının en önemli sebebi; makineler gerçek anlamda motivasyona ve iradeye sahip olduklarında, onların motivasyonları anlayamayacak ve kararlarını tahmin edemeyecek olmamızdır (Yonck, 2019, s. 298). Buna karşılık, özgür iradenin insandaki varlığına dair dahi uzlaşmış bir görüş yokken, onun yapay sistemlerdeki varlığının imkânına dair öngörülerde bulunmak çok kolay olmayacaktır.

Kendimize içsel bakışımız bize doğrudan, seçim yapma ve karar verme yeteneğimizin olduğunu; dünyayı ve dünyadaki olayları şekillendirme ve etkileme gücümüzün olduğunu hissettirir. Eylemlerimiz, davranışlarımız ve dolayısıyla hayatımı-

zın kontrolünün bize ait olduğuna sorgusuz bir güvenimiz vardır. Eylemlerimizin ve seçimlerimizin bize bağlı olduğuna; zira başka türlü davranma, başka seçimler yapma ya da çok daha farklı kararlar alma şansımızın olduğuna dair örtük ve devamlı inancımız, eğer üzerine düşünmezsek, “Bu seçim gerçekten bana mı ait?” sorusunu sormamıza asla müsaade etmez. Bu inanç, “özgür irade” inancıdır. Bu hususta, özgür iradenin varlığını inanca bağlamak kimilerince doğru görünmeyecektir. Ancak özgür iradenin varlığına dair birbirini yanlışlayan ve bir sonuca ulaşmamış görünen tartışmalar, “Özgür irade vardır.” ya da “Özgür irade yoktur.” yargılarına neyi bildiğimizden çok neye inandığımızla ulaştığımızı doğrular niteliktedir. Çünkü tam tersi bir bakış açısı, doğrudan kendimizi davranış ve eylemlerimizin sahibi hissettiren bu inancın yanlış bir inanç olduğunu iddia etmektedir. Kendimize içeriden değil dışarıdan baktığımızda ya da öznel değil nesnel bir bakış açısı ile yaklaştığımızda tüm kontrolün bizde olduğunu yadsıyan bir portre ile karşılaşabiliriz. Düşünce tarihi, eylemlerimizin nedenini -tanrı, kader, doğa yasaları, psikolojik faktörler, toplumsal süreçler, genetik yatkinlar, beyin aktiviteleri, hormonlar, elektro-kimyasal süreçler, ekonomik faktörler vs.- birçok farklı değişkene dayandıran teorileri aynı anda barındırmaktadır. Özgür irademizin varlığına dair inancımızı ve buna benzer kendimize dair içsel çıkarımlarımızı bir kenara bıraktığımızda, kendimizi dünyadaki nedensellik ilişkilerinin içerisinde sıradan bir yere sahip, eylemleri fiziksel yasaların ya da fiziksel olayların zorunlu sonucu olan, davranışları kendi tarafından değil de başka bir yerden kontrol edilen bir varlık olarak görmemiz mümkündür. İnsan özgürlüğü üzerine düşünen birçok insana göre özgür irade problemi, dünyanın bizi bilmediğimiz bir biçimde şekillendirdiğini fark ettiğimizde başlar (Kane, 2002, s.5). Çünkü bu fark ediş, benlik algımız ya da ben hissimiz ve bununla birlikte özgür olduğumuz inancı ile çelişik bir görüntü sunmaktadır. Bu çelişkiyi çözmek için ise özgür iradenin varlığını inkâr etmek, benliği bir yanılsama olarak kabul etmek ya da determinizmi reddetmek gibi farklı yaklaşımlar benimsemektedir.

Bununla birlikte son yıllarda büyük gelişme kaydeden nörobilim de özgür irade sahibi olduğumuza dair inancımızın bir yanılsama olduğuna dair dikkat çekici veriler ortaya koymaktadır. Özellikle Libet ve ekibinin 1983 yılında yayınladığı verilerle özgür irade tartışmaları büyük oranda teorik düzlemde empirik düzleme kaymıştır. Bugün Libet deneyi olarak isimlendirdiğimiz çalışmada, deneklerden karar verdikleri ya da istedikleri anda kollarını hareket ettirmeleri istenmiş ve bu sırada kas hareketleri EMG, beyin hareketleri EEG cihazlarıyla ölçülmüştür; sonuç olarak da denegın karar verdiği ya da istediği sandığı andan ortalama 300m/s önce, en az da 150 m/s saniye önce beynin ilgili bölgesinin aktive olduğu gözlen-

miştir (Libet v.d., 1983, s. 623). Libet'in kendisi daha sonra bu deneyin tek başına özgür iradenin yanılısına olacağına dair kesin bir kanıt olamayacağını söylemişse de onun bu çalışması, karar verme olayının biz daha farkında olmadan bilinç dışı süreçlerle beyinde gerçekleşmiş olduğuna yönelik empirik araştırmaların sayısını artırmıştır.

Libet'in açtığı yolda Daniel Wegner ve Thalia Wheatley tarafından yürütülmüş başka bir deneyde de, "özgür iradenin yalnızca kırmızı rengi görme gibi bir deneyim olduğu" ve eylemi gerçekleştirenin "ben" olduğum hissinin bir yanılığın ibaret olduğu iddiasına dayalı olarak önce eylemi gerçekleştirdiğimiz, sonra bu eylemi yapmaya karar verdiğimizizi düşündüğümüz öne sürülmüştür (1999, 480-490). Buna göre aslında bizler yaptığımız şeyleri kararımız neticesinde yapmıyoruz; bir şeyi yaptıktan sonra kendimize buna karar verdiğimizizi söylüyoruz (Reese, 2020, s.249).

Aynı iddiayı devam ettiren çalışmalar içerisinde en etkileyici olanlardan bir tanesi ise John-Dylan Haynes'e aittir. Haynes'in çalışması Libet deneyinin belirsizliklerinden daha uzaktır ve beyin görüntüleme tekniği olarak EEG'den daha gelişmiş bir yöntem olan fMRI kullanır; Haynes deneyini "yeniden yapılandırılmış Libet Deneyi" olarak tanımlar (Haynes, 2011, s.11). Deneklerin karşısına 500 m/s'de bir rastgele değişen harflerin olduğu bir ekran konulmuş, biri sağ biri sol elle kullanılmak üzere iki düğmeden birisine basarak fMRI cihazına bağlıyken karar verdikleri anı belirlemeleri ve bu anda hangi harfi gördüklerini not etmeleri istenmiştir. Burada Libet deneyinden farklı olarak harflerin rastgele olması ile deneklerin sistematik tercih yapma ihtimalleri en aza inmektedir ve deneklerin basacakları buton beyin hareketlerine göre önceden tahmin edilebilmektedir (Haynes, 2011, s.11-12). Dolayısıyla Haynes'e göre, kişi karar vermeden daha önce nedensel etki süreci beyinde tamamlanmaktadır; kişi daha kararının bilincinde değilken her şey beyinde olup bitmektedir. Böyle bir beyin işleyişi içerisinde özgür iradeye yer yoktur.

Nörobilime ait bu ve benzeri bulgular kimi gruplarca özgür iradenin varlığının artık tartışmaya gerek bile olmayan bir yanılığ olduğunu doğrulasa da teorik düzeyde özgür irade tartışmaları hâlen devam etmektedir. Çünkü Byron Reese'in ifade ettiği gibi: "İçimize baktığımızda bir saatin mekanik hassasiyetiyle veya bir gezegenin yörüngesiyle aynı şekilde işlev gördüğümüzü hissetmeyiz. Bizim hissettiklerimiz canlılık, irade, niyet, güdü ve hırsır." (220, s. 251). Şu bir gerçektir ki özgür iradenin varlığını kabul eden ya da etmeyen hemen hiç kimse özgür iradesi varmış gibi hissettiğini inkâr etmez. Özgür iradeyi zor ve kadim bir felsefi problem hâline getiren de bu histir.

Özgür iradenin varlığına dair bakış açımızı, bilimsel veriler kadar dünya görüşümüz, inançlarımız ve çeşitli ön kabullerimizin belirlediği yadsınamazdır. Dolayısıyla yapay özgür iradenin imkânına dair sunulacak her görüşün arkasında da özgür irade teorilerinde olduğu gibi belli varsayımlar olacaktır. Bu varsayımlar da iki önemli soru etrafında şekillenmektedir: “Evrenin bileşenleri nelerdir?” ve “İnsan nedir?”. İlk soruyu en genel anlamıyla iki şekilde cevaplayabiliriz: “Evren yalnızca fiziksel şeylerden oluşur.” ve “Evrende fiziksel olmayan şeyler de vardır.” İkinci soruya da aynı şekilde iki genel cevap verilebilir: “İnsan makinedir.” ve “İnsan makineden fazlasıdır.” Bu bağlamda, bu çalışmanın ikinci bölümünde özgür irade problemi “Evren yalnızca fiziksel şeylerden oluşur.” ve “İnsan makinedir.” varsayımlarıyla şekillenmiş olan fizikalist yaklaşım ekseninde ele alınacaktır. Üçüncü bölüm ise bu varsayımlara karşıt olarak “Evrende fiziksel olmayan şeyler de vardır.” ve “İnsan makineden fazlasıdır.” aksiyomlarını kabul eden anti-fizikalist yaklaşımlara ayırmıştır. Bununla birlikte, ele alınan yaklaşımın yapay özgür iradeyi hangi ölçüde olanaklı ya da olanaksız kıldığına dair değerlendirmeler her başlıkta ayrı ayrı yapılacaktır. Son bölümde ise yapay özgür irade sentezci bir yaklaşımla analiz edilerek probleme alternatif bir bakış açısı önerilecektir.

## 2. Özgür İrade Problemine Fiziksel Yaklaşımlar

Özgür irade probleminin en önemli kaynaklarından bir tanesi determinist görüştür<sup>1</sup>. Determinizmin temel tezi; her olayın, belli koşullar ve nedenlere göre belirlendiğidir. Koşullar ve nedenler sabit kaldığı sürece, olayın değişme şansı yoktur. Başka bir deyişle belirlenmiş bir olayın meydana gelmesi, o belirlenim koşullarında kaçınılmazdır; belirlenim, şartlı zorunluluktur. Determinizmin en klasik tanımını Laplace’de (1749-1827) görmemiz mümkündür:

Evrenin şimdiki durumunu, onun geçmişinin etkisi ve geleceğinin sebebi olarak düşünebiliriz. Eğer bir akıl, belirli bir anda, doğayı kuran tüm güçleri ve doğayı oluşturan tüm öğelerinin konumlarını bilseydi ve bu akıl, bu bilgiyi analiz edebilecek kadar büyük olsaydı; evrenin en büyük parçalarının hareketinden en küçük atomunun hareketine kadar hepsini tek bir formülde birleştirirdi. Böyle bir akıl için hiçbir şey bilinmez olmazdı ve gelecek de tıpkı geçmiş gibi gözünün önünde belirirdi (1995, 2).

Determinizm, özgür iradeye karşı oluşturduğu tehditle onun aynı zamanda felsefi ve bilimsel bir problematik olarak ele alınmasına temel teşkil eder. Her olayın be-

<sup>1</sup>Determinizmin, kaderci, teolojik, psikolojik, fiziksel ya da bilimsel, mantıksal vs. birçok formu bulunmaktadır. Bu çalışmada yalnızca fiziksel determinizm ele alınmaktadır.

lirlenmiş olduğu ve insan eylemlerinin birer eylem olduğu öncüllerinden hareketle, insan eylemlerinin belirlenmiş olduğu sonucuna ulaşan akıl yürütme, insanda özgür iradenin varlığını olanaksız kılar. Özellikle 19. yüzyılda Newton fiziğinin yükselişiyle kendisine güçlü bir dayanak bulmuş olan determinist evren görüşü, ilk bakışta iradenin varlığı için bir alan sağlamamakta ve hatta irade ile çelişik bir görünüm sunmaktadır. Determinist anlayışa göre başlangıcı ve sonu önceden belirlenmiş, birbirine sebep-sonuç ilişkisi içerisinde bağlı olgular dünyasının içerisinde bulunmaktayız. Newton'un determinizm anlayışını çok daha ileri götürerek, çok katı bir şekilde evrensel nedenselliği savunan Laplace'ın da ifade ettiği gibi yukarıdan holistik bir gözle bakabilecek biri geçmiş, şimdi ve geleceği görebilir çünkü şu anda meydana gelen her bir olay zamanın ilk anından itibaren belirlenmiştir (1995, s. 2). Dolayısıyla bu belirlenmiş olaylar zinciri içerisinde insan eylemlerinin insanın özgür iradesinden kaynaklandığını söylemek mümkün değildir.

Katı determinizmin iradeyi yok sayan bakış açısının herkes tarafından görüş birliği ile kabul edilmesi mümkün olmamıştır. Bu sebeple, evrendeki fiziksel determinizmin varlığı inkâr edilmeden, özgür iradeye yer açmak adına birçok farklı yaklaşım öne çıkmıştır. Zira özgür iradenin varlığının inkârı, insanın kendisini evrende konumlandığı yere büyük zarar vermekte ve insanı makine ile eşdeğer bir varlık hâline getirmektedir. İnsanın yalnızca mekanik bir varoluşa sahip olduğunu savunan katı determinizmden kaçınmak için determinizm ile özgür iradenin çatışmadığını savunan görüşlerin yanı sıra determinist tezin yanlışlığını öne süren farklı bakış açıları geliştirilmiştir. Özgür irade probleminin literatürdeki yeri yalnızca determinizm-özgür irade çelişkisi üzerine kurulu değildir. Determinizmi kabul etmeyenler olduğu gibi, determinizm ile özgür iradenin çatışmadığını savunan görüşler de mevcuttur. Tartışmaları takip etmek adına özgür irade problemini oluşturan sorular tek tek ele alınarak genel bir çerçeve çizilmesi sağlanabilir.

1- Determinizm doğru mudur?

2- Determinizm doğruysa özgür irade ile bağdaşır mı?

### **2.1. Determinizm Doğrudur, Özgür İrade ile Bağdaşmaz**

“Determinizm doğrudur, özgür irade ile bağdaşmaz.” tezi klasik fiziğe dayalı bakış açısının bir uzantısıdır. Bu bakış açısına göre, “Her doğal olay meydana gelişindeki durumların veya fiziksel koşulların toplamı tarafından niteliksel olarak belirlenir.” (Loewenstein, 2018, s.293). Her sonuç olarak her olayın bir sebebi vardır. Bu da demektir ki bütünü bulmak istediğimizde yapmamız gereken şey yalnızca parçaları toplamaktır. Parçaların haricinde hiçbir şey evrendeki düzene etki edemez; tüm moleküllerin izlediği belirli bir yol vardır ve bu yolun dışına çıkmadan hare-



ketlerine devam ederler. Klasik fizik bakış açısının temeli olan Newton fiziği tamamen belirlenimcidir. Newton fiziğinin önemli bir sonucu şudur ki belirli bir anda evrendeki nesnelere her birine etki eden tüm kuvvetler ile bu kuvvetlerin konum ve hızları biliniyorsa nesnelere hareketleri kuramsal olarak tam bir şekilde sonsuz kadar hesaplanabilir (Halpern, 2017, s. 28).

Buna göre, insan ve insan beyni de moleküllerinin izlediği yolun dışına çıkacak fazladan bir niteliğe sahip değildir. Dolayısıyla, belirlenmiş bir yapı olan insan beyninde, özgür iradenin varlığını kanıtlayacak hiçbir özellik mevcut değildir.

Özgür irade üzerine tartışmaların birçoğunda determinizmin, özgür iradeye yer bırakmadığına dair güçlü iddialarda bulunmaktadır. William James, evrendeki her olayın belli bir nedene bağlı olduğu, bu nedenlerin doğa yasaları tarafından belirlenmiş olduğu bir düzen içerisinde özgür iradeye yer olmadığını savunan görüşü, kader, özgür iradenin esareti, zorunluluk gibi kavramları kullanmaktan kaçınmayan katı-determinizm (hard-determinism) olarak adlandırır (2014, s. 149). Bağdaşmazlık (incompatibilism) olarak bilinen bu yaklaşım, determinizm ile özgür iradenin tanımları gereği bir arada bulunamayacakları görüşünü benimsemektedir. Buna göre, bir bireyin özgür olduğunu söyleyebilmek için gerekli olan koşullar, determinizmin temel teziyle çelişir görünmektedir. Determinizm doğru kabul edildiğinde, özgür iradenin varlığını yadsıyan yaklaşımların başlıca iddialarını şöyle sıralayabiliriz:

1- Alternatif olasılıkların varlığı özgür irade için zorunlu koşuldur. Ancak evrende alternatif olasılıklara yer yoktur; her olayın bir nedeni ve şartı vardır, aynı şartlarda farklı bir eylemde bulunabilme olasılığı yoktur. Nedensellik kuralı, çoklu olasılıkların varlığına imkân tanımaz.

2- Eylemlerimiz doğa yasalarının ya da evrenin ilk var olduğu andan itibaren meydana gelen olayların bir sonucudur. Ne doğa yasalarının kontrolü ne de biz var olmadan önce yaşanan olayların kontrolü bize aittir. Bu demektir ki eylemlerimizin kontrolü bizde değildir.

3- Bir kişi davranışlarından, o davranışa neden olan arzular, değerler ya da inançlar gibi tüm etkenler tamamen ona aitse sorumlu olabilir. Ancak determinizme göre eylemlerimiz ya da davranışlarımız daha biz doğmadan önce meydana gelen olayların ürünüdür ve dolayısıyla bizim dışımızda olan, kontrol edemeyeceğimiz nedenleri vardır. Bu sebeple, determinizm gerçek bir fail için yer bırakmamaktadır (Tononi, 2013, s. 167-168).

“Determinizm doğrudur, özgür irade ile bağdaşmaz.” varsayımına sahip birisinin yapay sistemlerin özgür irade sahibi olup olamayacağı sorusuna vereceği cevap

çok açıktır. Determinist dünyada özgürlüğe hiçbir şekilde yer yoktur. İnsanın özgür olmadığı bir dünyada, robotların ya da makinelerin de özgür olamayacağı çok açıktır. Ancak yine de makinelerin özgürlüğü insan özgürlüğünden bağımsız olarak değerlendirilirse, onların programlandıkları andan itibaren determinist bir işleyişe sahip olduğunu söylemek mümkündür. Bilgisayarlar belirli komutları takip eden belirlenimli makinelerdir; mantığın fiziksel bir versiyonu olarak ele alınabilecek bilgisayarlar mantık kadar kesin ve net “0” veya “1” durumlarına sahiptir (Walsh, 2020, s. 84). O hâlde, “Bir bilgisayarın özgür irade sahibi olması mümkün müdür?” sorusu kolaylıkla “Hayır.” olarak cevaplanabilecek bir soru gibi görünmektedir.

Geleneksel yapay sistemler belli girdi-çıkıtı ilişkilerine dayalı olarak çalışmaktadır ve hangi girdiye hangi çıktıyı verecekleri programlandıkları andan itibaren bellidir. Buna göre, klasik olarak programlanmış bir sistemin özgür irade sahibi olacağını iddia etmek güçtür. Ancak bağdaşmazcı yaklaşımı savunanlar, bunun da ötesinde, öğrenebilen, seçimler yapabilen ve hatta bilinç sahibi olan bir makineye dahi özgürlük atfetmeyeceklerdir. Bu bakış açısına göre, böyle bir makinenin davranışları bilinçli bile olsa doğa yasalarına göre belirlenmiştir ve o makine daha yapılmadan önce tüm eylem ve seçimleri bellidir. Dolayısıyla, insanın hiçbir koşulda özgür irade sahibi olamayacağını savunan bağdaşmazcı deterministlere göre, bir makine de ne kadar karmaşık ve gelişmiş olursa olsun özgür olamaz; yapay zekâ sistemlerinde özgür irade mümkün değildir.

## 2.2. Determinizm Yanlıştır, Özgür İrade Vardır

Bağdaşmazcı teorilerin birçoğu determinizmi doğru kabul ettiklerinden dolayı, özgür iradeyi yok saymaktadır. Ancak öte yandan, özgür iradenin varlığının kesin olduğunu ve özgür iradenin varlığının determinizmi yanlışladığını savunan yaklaşımlar da mevcuttur. Özgürlükçü (libertarian) olarak adlandırılan bu yaklaşımların bazıları, özgür bir eylem ya da kararın bir nedeni olmadığını savunurken; bazı özgürlükçü bakış açıları da özgür bir eylemin ya bir nedeninin olmadığını ya da diğer olayların belirlenmemiş olarak o eyleme neden olduğunu iddia etmektedirler (Clark, 2002, s.356). Bu bakış açısının belirlenimci olmayan (indeterminist) bir evren görüşüne yaslandığı gözden kaçırılmamalıdır. İndeterminizm, bütün olayların doğa yasaları ve diğer olaylar tarafından belirlenmiş olduğu tezini kabul etmemektedir. Özgürlükçü bağdaşmazcılığı savunanlar, evrenin indeterminist olduğunu ve evren indeterminist olduğu için özgür olduğumuzu dile getirmektedir. Ancak savundukları indeterminist görüşün kaynağı, evrene düalist değil, determinizmde olduğu gibi fizikalist yaklaşımdır.

Determinizmi reddeden fizikalist özgürlükçülerin, fiziksel dünyada bulduğu en büyük dayanak Kuantum fiziğidir. Alışıldık tasviriyle Kuantum fiziği, indeterminizmi fiziksel dünyanın içerisine dâhil etmiştir. Buna göre, atom altı parçacıkların hareketleri kesin olarak tahmin edilemez, belirlenimci kurallarla açıklanamaz; yalnızca olasılıkla açıklanabilir. Kuantum fiziğinin evrensel belirlenimcilik için bir açmaz oluşturduğu açıktır. Ancak atom-altı seviyede geçerli olan olasılığa dayalı sistemin nasıl olup da makro seviyede insan davranışına etki ettiği hususu tartışmalıdır. Bu sebeple belirlenimci olmayan özgür irade teorileri, Kuantum fiziğinin özgür iradeyi açıklamak için tek başına yeterli bir temel oluşturmadığı düşünceyle eleştirilmektedir. Bu eleştiriler, kavramsal temelleri hâlâ tartışılan Kuantum fiziğinin belirlenmemiş olayların nasıl olup da özgürlüğe katkıda bulunduğu dair pozitif bir açıklama getirmeden, özgür iradenin varlığına ulaşamayacağı ileri sürmektedir (Strawson, 2010, s. 28). Kuantum fiziğinin bize gösterdiği indeterminist evren modeli, Newton fiziğine ek olarak evrene olasılıkları dâhil etmektedir. Ancak olasılıkların üzerinde özgür irademizin bir etkisi olduğunu gösterecek yeterli bilimsel veri bulunmamaktadır. Olasılıkların varlığını tek başına özgür iradenin bir kanıtı olarak ele almak eksik bir yaklaşımdır. Çünkü olasılıkların kendisi de yine doğa yasalarına benzer olarak bize bağlı ve bizim kontrolümüzde değildir. Bu açıdan bakıldığında, determinizm ile uyumlu olmayan özgür iradenin indeterminizm ile uyumlu kabul edilmesi zor görünmektedir. Zira indeterminizmin kabul ettiği belirlenmemiş olayların meydana gelmesi ya da gelmemesi tesadüfidir; rastlantısal olaylar da fail dâhil kimsenin kontrolü altında değildir (Kane, 2002, s. 23). Bu durumda, belirlenmemiş olayların varlığına dayanarak failin özgürlüğünü kabul etmek geçerli bir çıkarım olarak görünmemektedir. Ek olarak, insan davranışının makro seviyede belirlendiği ve makro seviyedeki (insan beyni ve bedeni) indeterminizmin ihmal edilebilir olduğu da kuantum fiziğinin özgür iradeyi açıklayabileceği iddiasına karşıt olarak iddia edilmektedir.

Özgürlükçü yaklaşımın, özgür iradeye bakış açısı, yapay özgür iradeye dair bakış açısını da kesin olarak göstermektedir. Bu yaklaşıma göre, özgürlüğün kaynağı olasılıklardır ya da başka bir deyişle belirlenmemiş olayların varlığıdır. Bir eylem, seçim ya da davranışın özgür olması için o eylem, seçim ya da davranışın nedeninin doğa yasaları ya da başka eylemler tarafından belirlenmemiş olması gereklidir. Özgürlükçü yaklaşıma göre, Kuantum fiziği bu türden belirlenmemiş olayların varlığına dair bir dayanak oluşturmaktadır. Dolayısıyla bu yaklaşımın özgür kabul edeceği yapay sistemler olasılıklara sahip olmalıdır. Günümüzde kullandığımız bilgisayarlar ya da yapay zekâya dayalı uygulamalar, belirlenmiş (deterministik) bir yapıya sahiptir. Seçim yaptığını düşündüğümüz programlar da bu seçimleri,

seçim yapabilmek üzere programlandığından yapmaktadırlar. Dolayısıyla, gelecekteki algoritmaya dayalı sistemlerde özgür iradeden bahsetmek “Özgürlük indeterminizmi gerektirir.” tezini savunanlar için mümkün değildir. Fakat bu yaklaşım, indeterminist bir dünyada indeterminist bir bilgisayarı mümkün görecektir. Kuantum bilgisayarların inşası da bu görüşe destek sağlamaktadır. Kuantum bilgisayarlar, bugün bizim kullanımımıza sunulmamış olsa da geliştirilmeleri yönünde ciddi çalışmalar söz konusudur. Örneğin, Google bir süre önce Sycamore isimli klasik bilgisayarın 10.000 yılda çözebileceği problemi 200 saniyede çözen kuantum işlemcisini ürettiğini duyurmuştur (Arute v.d., 2009, s. 505-510). Klasik bilgisayarda bir bit 0 veya 1 değerine sahip olabilirken kuantum bilgisayarlar, 0 ve 1’i aynı anda içeren kübitlerden oluşmaktadır. Bu da kuantum bilgisayarların, bir girdiye her durumda aynı çıktıyı vermeyeceğini göstermektedir. Prensipinde bir kuantum bilgisayar eş zamanlı olarak birden fazla durumda olabilir ve böylece eş zamanlı olarak birden fazla görevi yerine getirebilir (Loewenstein, 2018, s.309). Dolayısıyla özgürlükçü yaklaşıma göre, çoklu olasılıklara sahip ya da alternatif durumlara sahip olarak adlandırabileceğimiz kuantum bilgisayarların özgür iradeye sahip olabileceğini söylemek mümkündür.

Fakat indeterminizmin özgürlüğün kaynağı olduğu tezine getirilen itirazlar, yapay sistemlerin özgürlüğü meselesinde de geçerliliğini korumaktadır. Kuantum bilgisayarların, klasik bilgisayarlardan farklı olarak olasılıklara sahip bilişsel yapısının, nasıl olup da onları özgür olarak nitelendirmemizi gerektireceği net olarak ortaya konulmamaktadır. Katı deterministlerin eleştirisi, Kuantum bilgisayarların bilgi işleme süreçlerinde var olan olasılıkların, onların bir program tarafından belirlendikleri gerçeğini değiştirmeyeceği yönünde olacaktır. Ayrıca, indeterminizmin insan kontrolünün dışında olan rastlantılara dayanmasının, özgür irade için bir gerekçe sunamayacağı itirazı yapay sistemler için de geçerli bir itiraz olarak görünmektedir. İndeterminist bir yapay sistem, rastlantısal olaylar içerdiği için özgür sayılamaz çünkü rastlantısallık sistemin kendisi tarafından kontrol edilemez. Oysaki özgür irade, tanımı gereği, olayların ya da eylemlerin kontrolüne sahip olmayı gerektirir.

### **2.3. Determinizm Doğrudur, Özgür İrade ile Bağdaşır**

Bağdaşıcılık, farklı formlarda savunulabilen bir yaklaşımdır:

a- Determinizm doğrudur ve bizim özgür olmadığımızı göstermez ancak biz özgür değiliz.

b- Determinizm doğrudur ve bizim özgür olmadığımızı göstermez ancak özgür olup olmadığımızı ispatlanamaz.

c- Determinizm doğrudur ve biz özgürüz.

d- Determinizm doğrudur ve biz özgürüz; bizim özgürlüğümüz determinizmin doğruluğunu gerektirir.

e- Determinizm doğru olsun ya da olmasın, biz özgürüz.

f- Determinizm doğru değildir ve biz özgürüz; determinizm doğru olsaydı da özgür olurduk.

g- Determinizm doğru değildir ve biz özgür değiliz. Fakat yine de determinizm ve özgür irade bağdaşmaktadır (Strawson, 2010, s. 5).

Bu farklı formlar arasından, bağdaşırcılık denildiğinde ilk akla gelen determinizmin doğru olduğu ve buna rağmen özgür olduğumuzdur. Dolayısıyla fazladan bir açıklama yapılmadıkça, bağdaşırcılığın klasik tezi, "Determinizm, özgür iradenin varlığına bir engel teşkil etmez. Determinizm ile 'Özgür irade vardır.' önermesi aynı anda doğrudur." olarak yorumlanmalıdır. Bu tez kendisine farklı argümanlarla destek bulmaktadır. Fakat bu argümanların hemen hepsinde ortak olan husus, özgür iradenin yeniden tanımlanması ve yorumlanması üzerine kurgulanmış olmalarıdır. Thomas Hobbes, David Hume ve John Stuart Mill tarafından da savunulan klasik bağdaşırcılık, özgürlüğü istediğimizi ya da seçtiğimizi yapma gücü olarak ele alır. Örneğin, Hobbes'a göre insanlar, yalnızca kendi kararlarını verebildiklerinde özgürdür; kendi kararını verebilmek ise ancak irademizi, arzularımızı ya da yönelimlerimizi engelleyen şeylerin yokluğunda mümkündür (akt. Kane, 1996: 10). Buna göre, insan eylemini engelleyen bir engel var olmaması durumunda istediğini yapabiliyorsa özgürdür. Eylemlerimizin çeşitli nedenlere dayalı olarak belirlenmiş olması, o eylemi isteyerek gerçekleştirdiğimiz gerçeğine antitez oluşturmaz. Fakat özgür iradeyi, istediğini yapabilme gücüne indirgeyen böyle bir tanım; özgür iradeyi alternatif olasılıklara sahip olma ya da başka türlü de eyleyebilme gücü şeklinde ele alan sonuç argümanı (consequence argument) karşısında zayıf bir duruşa sahip görünmektedir. Sonuç argümanı şöyle ifade edilebilir:

Eğer determinizm doğruysa her olay, üzerinde kontrol sahibi olmadığımız doğa kanunlarının ve uzak geçmişteki olayların sonucudur. Kontrolünün bizde olmadığı bir nedene bağlı olarak meydana gelen hiçbir şeyin kendisinin kontrolü de bizde değildir. Bu sebeple, eğer determinizm doğruysa, eylemlerimiz ve düşüncelerimiz de dâhil olmak üzere meydana gelen hiçbir şeyin kontrolü bizde değildir. Bunun, düşündüğümüz ya da yaptığımız hiçbir şeyin, bize olan ya da bizimle olan hiçbir şeyin, tıpkı piyanistin her vuruşunda piyano telinin titremesi gibi, başka türlü olması mümkün değildir (Kapitan, 2002, s.127).

Bu argüman eylemlerimizin farklı şekillerde gerçekleşme olasılıklarını, bu eylemlerimizin nedenleri olan geçmişî ya da doğa yasalarını değiştirmenin mümkün olmadığını ileri sürerek yadsımaktadır. Buna göre, eylemlerimiz, kontrolü bizde olmayan ve dolayısıyla asla değiştiremeyeceğimiz geçmiş olayların ve doğa yasalarının kaçınılmaz sonucudur.

Bağdaşırıcı düşünürler sonuç argümanını farklı açılardan cevaplamaya çalışmaktadır. İlk cevap, özgür iradenin anlamını koşullu olarak yorumlamaktadır. Bir eylemi gerçekleştirmede özgür olmamız demek, “İstedığımız için o eylemi gerçekleştirdik.” demektir. Bu anlamda özgür iradenin, “Eğer isteseydik, başka türlü davranabilirdik.” şeklinde bir koşula yönelik olarak ele alındığında determinizm ile çelişkili olmayacağı iddia edilmektedir. Bu iddia, özgür iradenin koşullu bir anlama sahip olması durumunda, geçmişî ya da doğa yasalarını değiştirmeye gerek kalmadan özgür eylemden söz edilebileceğini ileri sürmektedir (Kane, 2002, s. 15). Ancak bağdaşmazcı bakış açısı, “Eğer isteseydik, başka türlü davranırdık.” şeklinde koşullu bir tanımın determinist evrende özgür iradenin varlığına yer açacağını kabul etmez. Zira determinist görüş başka türlü davranmayı istemeyeceğimizin de belirlenmiş olmasını gerektirmektedir.

Özgür iradenin koşullu analizinin sonuç argümanını bertaraf etmekte yetersiz kaldığını düşünen bağdaşırıcılardan Harry Frankfurt, özgür iradeyi determinizm içinde konumlandıramamamızın sebebinin alternatif olasılıklar prensibi olduğuna işaret etmektedir. Frankfurt’a göre, “determinist dünyada alternatif olasılıklara yer olmadığı, bu yüzden de özgür iradenin var olamayacağı” görüşü, determinizmin çoklu olasılıklara olanak tanınması açısından değil; özgür iradenin alternatif olasılıkları zorunlu olarak gerektirmemesinden ötürü yanlıştır (1969, s. 829). Bu hususta tartışılması gereken özgür iradenin varlığı değil, alternatif olasılıklar prensibinin kendisidir. Frankfurt’a göre, alternatif olasılıklar prensibi yanlıştır çünkü bir insan başka türlü yapmak için bir alternatifi olmasa dahi eylemlerinden ahlaki olarak sorumludur. Bu iddiasını sonradan Frankfurt-tipi örnekler olarak adlandırılan örneklerle desteklemeye çalışan Frankfurt’a göre, bir insanın farklı türden davranma şansı olmadığı için öyle davranması ya da başka bir şey yapma olasılığı bulunmadığı için yapmış olduğu şeyi yapması onun özgür olmadığını göstermez. Frankfurt-tipi örneklerden birisini şöyle özetleyebiliriz:

Jones, Amerikan seçimlerinde Demokrat Parti ya da Cumhuriyetçi Parti için oy kullanmaya karar verecektir. Bu esnada, Demokrat Parti lehine çalışan bir beyin cerrahı Jones’un farkına varmadan beynine bir cihaz yerleştirmiş ve Jones’un beynini görüntülemektedir. Eğer Jones Demokrat Partiye oy vermeye karar verirse bu cihaz yalnızca görüntülemeye devam edecek ve karar verme sürecine

müdahale etmeyecektir. Fakat eğer Jones oyunu Cumhuriyetçi Partiden yana kullanmaya karar verirse, bu cihaz beyne müdahale edecek ve Jones'un oy tercihini Demokrat Parti lehine kullanmasını sağlayacaktır. Çeşitli değerlendirmelerden sonra Jones oyunu Demokrat Partiye verme kararı alır ve oyunu kullanır; cihazın karara hiçbir müdahalesi olmamıştır (Fischer, 2002, s. 282).

Bu örnekte, Jones'un Demokrat Partiye oy vermek dışında yapabileceği hiçbir şey yoktur. Çünkü aksi yönde bir karar almak istediğinde dahi beyindeki cihazın müdahalesiyle yine Demokrat Partiye oy verecektir. Ama burada oyunu kendi isteği ile Demokrat Partiden yana kullanmıştır, cihazın hiçbir müdahalesi söz konusu değildir. Bu sebeple, alternatif olasılıklarını yok eden cihaza rağmen Demokrat Partiye oy verme eyleminde özgür ve ahlaki olarak sorumlu görünmektedir. O hâlde, özgür irade için alternatif olasılıkların varlığı zorunlu bir koşul değildir ve alternatif olasılıklar prensibi yanlıştır. Zira iddia şudur ki bir insan farklı türlü davranma şansına sahip olsa da yine istediği gibi davranacaktır, başka bir şey yapma olasılığı bulunduğu da istediği şeyi yapacaktır (Frankfurt, 1969, s. 837). Dolayısıyla, özgür iradeyi mümkün kılan alternatif olasılıkların varlığı değil istediğini yapabilme yetisidir.

Modern bağdaşıcılık burada ele alamadığımız pek çok farklı argümanla determinist bir dünyada özgür iradenin mümkün olduğunu savunmaya devam etmektedir. Ancak bu argümanların hepsinin bu çalışmada tartışılmaya açılması konumuzun sınırlamamız gerektiğinden mümkün değildir. Burada üzerinde durmamız gereken konu bağdaşıcı özgür irade yaklaşımının, yapay iradeye yönelik çizeceği çerçevedir. Bu sebeple, bağdaşıcılığın en genel hâliyle, özgürlüğü "istediğim gibi eylemde bulunabilmek" olarak tanımladığı ve istediğim gibi eylemde bulunduğum takdirde, eylemin nedenlerinin belirlenmiş olmasının benim özgür irade-me engel olmayacağını öne sürdüğünü göz önüne alarak yapay sistemlerde özgür iradenin imkânına dair bağdaşıcıların potansiyel görüşleri hakkında değerlendirme yapabiliriz.

Klasik bilgisayarların determinist yapıda olduklarını daha önce de vurgulamıştık. Bu sebeple, determinist bir yapıya sahip olan herhangi bir yapay sistemin özgürlüğünden bahsedilecekse kuşkusuz ki determinizm ve özgür iradenin bir arada bulunabileceğini savunan bağdaşıcılık bu iddia için en uygun zemini sağlamaktadır.<sup>2</sup> Zira bağdaşıcılık, yapay zekâ ile özgürlüğün tanımları gereği çelişmemesini sağlayan yegâne bakış açısıdır. Bağdaşıcılığın "İnsan istediğini yaptığı ölçüde özgürdür." iddiası yapay sistemler için düşünüldüğünde, "Makine istediğini yaptığı

<sup>2</sup> Özgür irade sahibi makineleri bağdaşıcı yaklaşım temelinde mümkün gören bir bakış açısı için bkz. Zambak, A. (2018). "Free Will and Artificial Intelligence". *MetaZihin: Yapay Zekâ ve Zihin Felsefesi Dergisi* (2): 167-181.

ölçüde özgürdür.” şeklinde bir iddiaya dönüşecektir. Burada istediğini yapmak ifadesi arzu ettiğini yapmak gibi bir duygu durumuna bağlanırsa yapay sistemlerde duyguların imkânına dair yeni bir sorunla karşılaşılır. Bu sorunun bir adım ötesinde de duyguların varlığını konuşmamızı mümkün kılan bilinçli deneyimin varlığı tartışmaya açılmış olur. Ancak burada gözden kaçırılmaması gereken husus şudur ki çalışmamızın başında da belirttiğimiz üzere bağdaşırıcı yaklaşımın arka planında diğer fizikalist bakış açıları da olduğu gibi Evrende var olan her şey fizikseldir ve İnsan gelişmiş bir makinedir. varsayımları yer almaktadır. Dolayısıyla, bağdaşırıcılık için duygular da bilinç de fiziksel niteliklerdir. Dolayısıyla, duygulara ya da bilinçli deneyime sahip bir makine bağdaşırıcılık için mümkündür çünkü insandaki varlıkları nasıl fiziksel gelişmiş bir makinede de aynı şekilde var olabilirler. O hâlde, bağdaşırıcılık; istediğini yapan makineler inşa ettiğimizde, bu istekleri programcıları tarafından belirlenmiş olsa bile, bu makinelerin özgür olduğunu iddia edecek ve bunun gerçekleşmesinin önünde fiziksel bir engel bulunmadığını ileri sürecektir.

### 3. Özgür İradeye Fiziksel Olmayan Yaklaşımlar

Özgür iradeye fiziksel yaklaşımlar, özgür iradenin doğasından ziyade varlığını tartışma konusu olarak ele almaktadır. Bunlardan katı-determinizm özgür iradenin varlığını tamamen reddederken bağdaşırıcı yaklaşımlar ahlaki sorumluluk probleminden kaçınmak için özgür iradeye determinist evren içerisinde bir yer açma çabası içerisindedir. Özgürlükçü yaklaşımın esas problemi ise indeterminizmi özgür iradenin varlığını gerekçe göstererek temellendirmektir. Her üç yaklaşımda da özgür iradeye verilen bir belirleme gücü söz konusu değildir.

Oysaki özgür irade sahibi olduğumuza dair inancımızın en temel kaynağı eylemlerimiz ve davranışlarımıza etki edebilme gücümüzün olduğunu düşünmemizdir. Özgürlük, nasıl eyleyeceğimizin kontrolünü sağlama kapasitesi olarak eylemlerimizi belirleyen bir güçtür (Pink, 2004: 105). Fakat elle tutulmayan bir kavram olan özgürlüğe belirleyici bir rol atfetmek, fizikalist evren anlayışı içerisinde çok mümkün görünmemektedir. Özgür iradenin eylem ve kararlarımız üzerinde belirleyici bir gücü olduğunu söylemek, daha geniş bir perspektiften, zihinsel durumların fiziksel durumları etkilediğini söylemek olacaktır. Bu ise zihinsel ve fiziksel olmak üzere iki ayrı ontolojik gerçekliği varsaymak anlamına gelmektedir.

Özgür irade problemine düalist yaklaşımda genel olarak fenomenal bilincin, başka bir deyişle öznel deneyimlerimizin, özgür iradenin kaynağı olduğunu iddia etmektedir. Bu hususta, fenomenal bilincin ontolojik olarak fiziksel olana indirgenip indirgenmeyeceği zihin felsefesinin temel tartışma konularından birisidir. Onto-



lojik olarak beyne indirgenemeyecek bir bilinç yaklaşımını savunmak şüphesiz ki özgür irademizin eylemlerimizin belirleyicisi olduğunu iddia etmek için güçlü bir temel sağlamaktadır (O'Connor, 2004, s. 116). Fenomenal bilincin ontolojik indirgenemezliğini savunmak, öznel deneyim ile nesnel beyin aktiviteleri arasında kapanmayacak bir açıklama boşluğunu<sup>3</sup> olduğunu varsaymaktır. Bu sebeple, özgür irade de fenomenal bilincimizin bize sunduğu bir içsel gerçeklik olarak nesnel terimlerle açıklanamaz görünümüdür. Burada kabul edilen düalist bakış açısı, zihin ve bedeni bağımsız gerçeklikler olarak ele alan töz düalizmi olabileceği gibi; fiziksel bir töz olan bedenin fiziksel ve zihinsel olmak üzere iki farklı niteliğe sahip olduğunu varsayan nitelik düalizmi de olabilir. Ancak her iki türden düalizmin kabulü de düalizmin klasik açmazları ile yüzleşmeyi gerektirmektedir. Zihin ve bedeni bağımsız varoluşlar olarak kabul etmek, bu iki tözün nasıl birbiriyle etkileşime girdiğini sorgulayan klasik zihin-beden problemine bir çözüm önerisi geliştirmeyi gerektirmektedir. Aynı şekilde, zihinsel ve fizikselin aynı tözün farklı nitelikleri olduğunu kabul etmek de fiziksel bedenin nasıl zihinsel niteliklere neden olduğu sorusuna cevap vermeyi gerekli kılmaktadır. Fakat hem töz düalizminin etkileşim problemi hem de nitelik düalizminin zihinsel nedensellik problemi ne felsefi ne de bilimsel açıdan çözümlenebilmiş değildir.

Bu çalışmadaki odak nokta, bu yaklaşımların yapay irade için çizecekleri çerçeve olacaktır. Öncelikle şunu belirtmeliyiz ki düalist yaklaşımların temelinde insanın makineden fazlası olduğu varsayımı yatmaktadır. Özgür irade de insanı makineden farklı kılan niteliklerden bir tanesidir. Dolayısıyla düalist bakış açısında, yapay özgür iradenin imkânına yer bulmak çok mümkün görünmemektedir. Özellikle töz düalizmi, zihne atfettiği bağımsız ontolojik statü ile tamamen fiziksel olan bir makinede zihinden ve zihinsel niteliklerden bahsetmenin önünü tamamen kapatmakta ve yapay özgürlüğü ne mantıksal ne de ilkece mümkün görmektedir. Ancak zihinsel niteliklerin, karmaşık beyin süreçlerinden belirlediğini savunan bir tür düalizm için, yapay özgür irade mantıksal olarak mümkündür.

Özgür irade sahibi bir makine aynı zamanda fenomenal bilinç sahibi de olacaktır. Dolayısıyla eğer fenomenal bilinç sahibi bir makine oluşturulabilirse bu makinenin özgür irade sahibi olması da mümkün kılınabilir. Fiziksel olmayan yaklaşımın birbirinden ayrı ele aldığı fenomenal bilinçli durumlar ile fiziksel nitelikler arasındaki ilişkinin doğasına dair kesin bir kavrayışa sahip değiliz. Bu yaklaşım, insan

<sup>3</sup>Joseph Levine fiziksel nitelikler ve bu fiziksel niteliklerin deneyimi arasında kapatılamayacak bir "açıklama gediği" olduğunu iddia eder. Bu iddiayı dayandırdığı temel, acının beyindeki fiziksel korelasyonları ile acının deneyiminin doğaları gereği birbirinden farklı olduğudur. Bu sebeple, bilincin tüm fiziksel korelasyonları bilinse dahi bu, deneyimin kendisini açıklamak için yeterli olmayacaktır. Bkz. Levine, J. (1983). "Materialism and Qualia the Explanatory Gap" Pacific Philosophical Quarterly (64): 354-361.

beyninin nöral haritası tamamlansa, bilişsel süreçler bütünüyle ortaya konulsa dahi fenomenal deneyimin doğasına bir açıklama getirilemeyeceğini iddia etmektedir. Ancak fenomenal durumların ortaya çıkışını fiziksel süreçlerle açıklayamayız demek; bizi, “Yapay sistemlerde fenomenal bilinç mantıksal olarak mümkün değildir.” iddiasına götürmemektedir. Fenomenal durumların ortaya çıkışını fiziksel süreçlerle açıklayamayız demek; bize yalnızca fiziksel süreçlerin fenomenal durumların doğasını açıklamak için yetersiz olduğunu söylemektedir. Fenomenal durumların doğasının nasıl açıklanacağına dair bir yöntem sunmamaktadır. Bilişsel ve nöral modellemesi tamamlanan bir yapay sistemde –ne şekilde ve hangi aşamada olduğunu bilemesek de– fenomenal bilinçli durumların oluşması hâlâ ihtimal dâhilindedir. Dolayısıyla yeterince gelişmiş bilişsel niteliklere sahip ya da beynin birebir modellemesiyle oluşturulmuş bir yapay sistemde, fenomenal bilincin ortaya çıkması ve buna bağlı olarak özgür iradenin ortaya çıkışı mantıksal olarak mümkündür. Ancak bu ortaya çıkış rastlantısal olacak ve fiziksel süreçlerin ne şekilde fenomenal durumlara yol açtığını ve özgür iradenin ne şekilde ortaya çıktığını açıklayamayacaktır.

#### 4. Sonuç Yerine: Alternatif Bir Yapay İrade Yaklaşımı Önerisi

Özgür irade, en genel ifadeyle, bağımsız seçim yapma ya da karar verme yetisi olarak tanımlanabilir. Yazılımları ya da algoritmaları ile sınırlı olan kapalı sistemlerde özgür seçim yapma yetisinin varlığı söz konusu değildir. Bağımsız karar alabilmek aynı zamanda bağımsız motivasyonlara ya da belirlenmemiş hedeflere sahip olmayı gerektirir. Hedefleri baştan belirlenmiş hiçbir yapı özgür iradesi ile eylemde bulunamaz. İnsanda özgür iradenin varlığına olan inancımızın en önemli gerekçelerinden bir tanesi de zaten insanın belirli olmayan hedefler geliştirme yetisi ile birlikte belirli olmayan motivasyonlara sahip olmasıdır. Karmaşık yapıya sahip özneler olarak biz insanlarda davranış, deneyim ve hedeflere dayanır; hedefler belli değildir.

Buna göre, yapay akıllı sistemlerin hedef ya da motivasyon bakımından esnekliklerinin değerlendirilmesi bu sistemlerde iradenin varlığını denetlemek için yararlı bir yöntem olarak ele alınabilir. Manzotti, yapay sistemleri teleolojik/ereksel esnekliklerine göre Sabit Kontrol Yapıları, Öğrenen Yapılar ve Hedef Üreten Yapılar -Teleolojik Olarak Açık Sistemler olarak sınıflandırır (2007, s. 186).

Bugünkü teknolojik gelişmişlik düzeyinde, kullanılmakta olan yapay zekâ sistemlerinin büyük çoğunluğu önceden belirlenmiş bir hedefe ulaşmak üzere oluşturulmuş algoritmalara dayalı olarak çalışmaktadır. Sabit Kontrol Yapıları olarak adlandırdığımız bu sistemlerin hedefleri ve bu hedeflere nasıl ulaşacakları önceden

belirlenmiştir ve yapılarındaki hiçbir şey deneyimlerinden etkilenmez. Dolayısıyla bu sabit yapıların özgür irade sahibi olduklarını iddia etmek mümkün değildir. Hesap makinesi, termostat, içecek otomatları gibi girdi-çıkıtı ilişkisiyle çalışan makinelerin tamamı sabit kontrol yapılarının basit örnekleri olarak ele alınabilir. 1986 yılında Honda mühendislerinin yürüten bir robot yaratma projesiyle ortaya çıkan ASIMO<sup>4</sup> ise sabit kontrol yapılarının en gelişmiş örneklerindedir. Kendisine yüklenen sabit algoritma ile çevresiyle etkileşime giren ASIMO, çevresini sınırlı ölçüde sahip olduğu kodlar ile tanımakta; farklı bir fiziksel çevrede farklı nesnelere ilişki kurması gerektiğinde yazılımı güncellenmektedir. ASIMO ayrıca öğrenme yeteneği olan bir robot değildir; veri tabanından yararlanarak kodlanmamış çıktılar üretmez<sup>5</sup>. Yeni hedefler üretmediği gibi, hedefe giden yolları değiştirme ve güncelleme yetisine de sahip değildir.

Sabit Kontrol Yapılarından daha gelişmiş yapay zekâ uygulamalarına sahip öğrenme yapıları ise davranış temelli olup görevleri nasıl yerine getireceklerini öğrenebilirler. Yapay sinir ağlarına dayanan sistemler, bu tür yapıların en çok bilinen örnekleridir. Öğrenme yapıları, özel bir hedefe ulaşmaları istendiğinde sonuca ulaşma yollarına kendileri karar vermekte ve yollar arasında seçim yapabilmektedirler. Ancak hedefe ulaşmak için yeni yollar öğrenebilseler de hedef, görev ile birlikte sisteme önsel olarak tanımlandığından hedef üretmeleri söz konusu değildir. Sadece tanımlanan hedefin farklı yollarını deneyebilirler. MIT'nin üretmiş olduğu Cog<sup>6</sup> bir tür öğrenme yapısıdır. Cog, teleolojik olarak kapalı bir yapı olarak, verili olan hedef doğrultusunda öğrenme ve hedefe giden farklı yolları deneyebilme yetisine sahiptir. Bu kapalılık, onun kendisi için yeni hedefler üretebilmesine engeldir. Belli hedefler doğrultusunda çevresiyle ilişki kuran Cog, davranışlarını çevresel etkilere göre seçebilmekte ve ilk defa karşılaştığı durumlara öğrenme kabiliyeti ile adapte olabilmektedir ancak otonom olarak davranışta bulunma ve karar verme yetisine sahip değildir. Bu bağlamda, Cog oldukça gelişmiş bir bilişsel kapasiteye sahip olmakla birlikte, insanın çevresi ile kurduğu sosyal etkileşimi yerine getirebilecek bir robot değildir. Dolayısıyla Cog için tamamıyla özgür irade sahibi bir robot tanımlaması yapmak yanlış olacak olsa da en azından bilişsel olarak özgür olma potansiyeline sahip bir robot olduğu söylenebilir.

<sup>4</sup>Advanced Step in Innovative MObility.

<sup>5</sup>ASIMO ile ilgili bilgilere <https://asimo.honda.com/asimo-history/> adresinden erişilmiştir.

<sup>6</sup>Cog'a ait fiziksel özelliklere ve Cog'un yeteneklerine dair bilgilere <http://groups.csail.mit.edu/lbr/humanoid-robotics-group/cog/methodology.html> ve <http://www.ai.mit.edu/projects/humanoid-robotics-group/cog/overview.html> adreslerinden ulaşılmıştır.

Teleolojik olarak açık olup hedef üretebilen yapılar ise ilk iki sistemden farklı olarak verilen görevi nasıl yerine getireceğini ve hedefin ne olduğunu öğrenebilirler. İlk iki yapıdan çok daha fazla çevreyle etkileşim içerisinde bulunabilirler. Çevrelerinden kopuk davranışlar sergilemezler; çevresel süreçlerin oluşumu ve karmaşıklığının arkasında dururlar (Manzotti, 2007, s. 188). Bu türden yapılar için bugün için üzerinden analiz yapılacak bir örnek bulunmamaktadır. Zira var olan tüm sistemler ne kadar gelişmiş öğrenme yapılarına sahip olurlarsa olsunlar, belli hedefler doğrultusunda eylemde bulunmaktan öteye geçememektedirler. Ancak yeni hedefler üretebilen yapıların oluşturulması için çok uzun bir zamana ihtiyaç olmadığı ve bu teknolojik aşamanın çok yakında aşılabilecek olduğu yapay zekâ araştırmacıları tarafından sıklıkla dile getirilmektedir. Teleolojik olarak açık yapıların inşası ya Cog gibi bilişsel olarak çok gelişmiş bir robotun öğrenme teknikleri ilerletilerek mümkün olabilir ya da en başından açık bir yapı inşa edilebilecek bir algoritma geliştirilebilir. Teleolojik olarak açık yapıların inşası, bilişsel kapasitesi gelişmiş, otonom kararlar alabilen, çevresiyle sosyal etkileşim kurabilen yapıların oluşturulması anlamına gelecek; bu ise dar anlamda özgür irade sahibi akıllı yapay sistemlerin ortaya çıkışı yönünde büyük bir adım olacaktır.

İnsanın davranış ve eylemleri şüphesiz ki çevresel, biyolojik, genetik, hormonal, duygusal, kimyasal, nöral, bilişsel vs. birçok etkene bağlıdır. Bu etkenler içerisinde insanın sahip olduğu karar verme, seçim yapma ve problem çözme yeteneğinde bilgi-işleme kapasitesinin büyük bir role sahip olduğu tartışmasızdır. Karar verme ve seçim yapma yeteneği, insanda özgür iradenin varlığına dair gösterilen en önemli gerekçelerdendir. Bu yetenek, insan zihninin bilişsel yönüne aittir. Bugün her ne kadar insan zihnini bilişsel olarak modellemekten hâlen çok uzak olsak da insan zihninin bilişsel yapısını tam olarak çözümlediğimiz takdirde bilişsel yapımızın yapay sistemlerde yeniden inşasının önünde ilkesel ve mantıksal hiçbir engel bulunmamaktadır. İşlevselci görüş, çok kuvvetli bir şekilde insanın bilgi-işleme süreçlerinin karmaşık yapısına rağmen yapay olarak modellenilebilecek nitelikte olduğunu savunmaktadır (Baars, 1998; Dennett, 1991; Sun, 1997; Chrisley, 2003). Henüz pratikte uygulanamamış bu iddia, mantıksal ve teorik olarak çürütülebilmiş değildir.

Bilgisayarların sahip olabileceği/olduğu tahmin edilemez/yaratıcı seçimler yapma yetisi “Go” oyununa dayalı bir örnekle kolaylıkla anlaşılabilir. Go oyunu, 10397 değişik hamle olasılığına sahip, karar verme yetisine dayalı olan dünyanın en zor oyunlarından birisidir. Satranç oyununun (farklı hesaplamalara ait farklı hamle sayıları bulursa da) 10120 olası hamle içerdiği düşünüldüğünde oyunun karmaşıklık derecesi daha belirgin olacaktır. Alphago isimli program, “Go” oyununda,

belli kurallar çerçevesinde makine öğrenmesine dayalı olarak dünya şampiyonunu yenmiştir. Bunun üzerine Alphago'nun geliştiricisi olan DeepMind şirketi, Alphago'ya yüklenen en baştaki kurallar olmadan sıfırdan, pekiştirmeli öğrenme yöntemine dayalı bir algoritma yazarak AlphagoZero'yu<sup>7</sup> oluşturmuştur (Yazılan algoritma go algoritması değil, pekiştirmeli öğrenme algoritmasıdır.). AlphagoZero ise sadece öğrenme algoritmasını kullanarak, AlphaGo'yu defalarca kez yenmeyi başarmıştır. AlphagoZero'nun kodunda hangi durumda hangi hamleyi yapacağına dair hiçbir girdi bulunmadığı göz önüne alındığında, bu programın tüm kararlarının kontrolünün sadece kendisinde olduğu, hamlelerinin programcısı tarafından tahmin edilemez olduğu ve farklı durumlara farklı hamlelerle karşılık verdiği açıkça görülür. Fizikalist yaklaşımda özgür iradenin varlığından bahsedilmesi için gerekli koşullar olarak öne sürülen, alternatif olasılıklara sahip olmak, eylemlerin kontrolüne sahip olmak, eylemlerin belirlenmiş bir nedene dayanmıyor olması gibi özelliklerin hepsinin AlphagoZero'da var olduğunu söylemek yanlış olmaz. Fakat AlphagoZero'nun özgürlüğünün statüsünü belirlemek önemlidir. AlphagoZero'nun özgürlüğü bu durumda yalnızca bilişsel özgürlüktür. Bilişsel özgürlük ile kastedilen, bilgi-işleme süreci ve sürecin çıktılarının belirlenmemiş olduğudur. Bilişsel yapısı ise kapalı ve sınırlıdır. Çünkü AlphagoZero her ne kadar yaratıcı kararlar alabiliyor olsa da AlphagoZero yalnızca Go oynamak ile sınırlı, dar bir yapay zekâ uygulamasıdır. Dar yapay zekâ, belirli bir veya birkaç görev için programlanmış zekâ anlamına gelmektedir; böyle bir zekâ, belli bir görev için amaca yönelik yapılmıştır ve iş odaklıdır (Frank, Roehrig ve Pring, 2019, s. 67). AlphagoZero özelinde savunulan bu iddia, benzer algoritmalara sahip tüm programlar için genelleştirilebileceği gibi, bu programların yürütüldüğü robotik yapılar için de geçerlidir. Bir bilgisayar programı, bir makine ya da bir robot öğrenerek yazılımcısının ona verdiği kodların dışında, bağımsız olarak tahmin edilemez ve yaratıcı kararlar alabiliyorsa onun bilişsel olarak özgür olduğu söylenebilir. Yine de bu bilişsel özgürlük bugün için sınırlıdır çünkü var olan yapay zekâ uygulamalarının hepsi yalnızca belirli görevler için programlanmıştır. Go oynayan bir program aynı zamanda borsa tahminleri yapamamaktadır. Dolayısıyla günümüzde derin öğrenme yöntemi ile belirli görevleri yerine getiren yapay zekâ uygulamaları dar anlamda bilişsel özgürlüğe sahiptirler diyebiliriz. Ancak tam bir bilişsel özgürlük için yapay genel zekâyâ ulaşılması zorunlu bir koşuldur. Yapay genel zekâ ise "insan benzeri beceri ile geniş aralıkta görevler yürütebilen yapay zekâ" anlamına

<sup>7</sup>AlphagoZero'nun çalışma şekli ve algoritmasının ayrıntılı anlatımı için bkz. Silver, D. Hubert, T., Schrittwieser, v.d., (2018). "A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-play". Science (362): 1140-1144.

gelmektedir (New Scientist, 2021, s. 291). Bu ise birçok bilim kurgu filminde karşımıza çıkan; Bill Gates, Stephen Hawking, Elon Musk gibi birçok düşünürün endişe ile yaklaştığı süper zekâların -insanüstü düzeyde bilişsel performansa sahip genel zekâyâ sahip sistemlerin- doğuşu anlamına gelmektedir (Bostrom, 2020, s. 39).

Bilişsel özgürlük, özgür iradenin bir katmanını oluştursa da ontolojik olarak özgürlükten bahsedebilmek için tek başına yeterli değildir. Ontolojik özgürlük, fenomenal bilinci ve buna bağlı olarak bir özne tasavvurunu zorunlu olarak gerektirmektedir. Fenomenal bilinç, öznel deneyimdir ve özgür irade de ancak fenomenal bilinçli bir deneyimin sonucu olarak farkına varılabilecek bir gerçekliktir. Dolayısıyla, ontolojik olarak özgür olacağını iddia edeceğimiz bir makinenin aynı anda fenomenal bilinç sahibi olması da gereklidir. Fenomenal bilincin yapay sistemlerdeki varlığı rastlantıya dayalı olarak mantıksal bir olasılığa sahip olsa da fenomenal bilinç, deneyimleyen özneyi gerekli kıldığından ilkece mümkün görünmemektedir. Fenomenal bilinç, öznel deneyimlerin kaynağıdır. Öznel deneyimin var olması için deneyimleyen öznenin de var olması gerekir. Dolayısıyla bir makinenin ontolojik özgürlüğü, onun aynı zamanda özne olması anlamına gelmektedir. O hâlde soru şudur: Bir makinenin deneyimleyen bir özne olarak ele alınması mümkün müdür? Bu sorunun cevaplanması ise özne üzerine derin bir metafizik sorgulama gerektirdiğinden başka bir çalışmanın konusu olarak tartışmayı burada bırakmalıyız.

Sonuç olarak, eğer makinelerde özgür iradeye dair bir iz aranacaksa, bu iz ancak teleolojik olarak açık yapıdaki akıllı makinelerin bilişsel kapasitesinde bulunabilir. Özgür iradenin varlığı tartışmaları felsefi ve bilimsel literatürde sıkça karşılaşıldığı üzere, özgür irade kavramına farklı anlamlar yüklenmesi üzerinden ilerlemektedir. Özgür iradenin yapay zekâyâ dayalı sistemlerde varlığını tartışılır kılmak ancak böyle bir yöntemle özgür irade bağımsız karar verme ve seçim yapma yetisi olarak tanımlandığında mümkündür. Fakat bu bilişsel nitelikli tanıma dayanarak, ontolojik olarak özgür makineleri olanaklı görmek eksik bir bakış açısı olacaktır. Ontolojik özgürlük tartışması bilişsel özgürlüğün epistemolojik sınırları içerisinde değildir. Ontolojik özgürlük metafizik bir bakış açısı gerektirmektedir. Öznenin varlığı kabul edilmeden yapılan her özgür irade tanımı eksik kalacaktır. Bu sebeple, “Yapay zekâ özgür olabilir mi?” sorusu ancak deneyimleyen bir mekanik öznenin imkânına dair yapılacak metafizik bir sorgulamayla cevaplanabilir.

**Çıkar Çatışması Beyanı**

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan etmiştir.

**Kaynakça**

- Arute F., Arya Kunal, Babbush, R. v.d., (2009). "Quantum Supremacy Using a Programmable Superconducting Processor", *Nature* (574): 505–510.
- Baars, B. (1998). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Bostrom, N. (2020). *Süper Zekâ Yapay Zekâ Uygulamaları, Tehlikeler ve Stratejiler*. İstanbul: Koç Üniversitesi Yayınları.
- Chrisley, R. (2003). "Embodied Artificial Intelligence". *Artificial Intelligence* 149 (1): 131-150.
- Clarke, R. (2002). "Libertarian Views: Critical Survey of Noncausal and Eventcausal Accounts of Free Agency". *The Oxford Handbook of Free Will*. ed. Robert Kane. New York: Oxford University Press. 356-385.
- Dennett, D. C. (1991). *Consciousness Explained*. New York: Back Bay Books/Little, Brown and Company.
- Fischer, J.M. (2002). "Frankfurt-Type Examples and Semi-Compatibilism". *The Oxford Handbook of Free Will*, ed. Robert Kane. New York: Oxford University Press. 281-308.
- Frank, M., Roehrig, P. & Pring, B. (2019). *Makineler Her Şeyi Yaptığında Biz Ne Yapacağız?*. İstanbul: Aganta Kitap.
- Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility". *The Journal of Philosophy* 23 (66): 829-839.
- Halpern, P. (2017). *Einstein'in Zarı ve Schrödinger'in Kedisi*. İstanbul: Kırmızı Kedi Yayınevi.
- Haynes, J. D. (2011). "Decoding and Predicting Intentions." *Annals of the New York Academy of Sciences* (1224): 9-21.
- James, W. (2014). "The Dilemma of Determinism", *The Will to Believe and Other Essays in Popular Philosophy*. Cambridge: Cambridge University Press. 145-183.
- Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- (2002). "Introduction: The Contours of Contemporary Free Will Debates". *The Oxford Handbook of Free Will*. ed. Robert Kane. New York: Oxford University Press. 3-41.
- Kant, I. (2015). *Gelecekte Bilim Olarak Ortaya Çıkabilecek Her Metafizığe Prolegomena*. Ankara: Türkiye Felsefe Kurumu.
- Kapitan, T. (2002). "A Master Argument for Incompatibilism?". *The Oxford Handbook of Free Will*. ed. Robert Kane. 127-157.
- Laplace, P.S. (1995). *Philosophical Essay on Probabilities*, New York: Springer.
- Levine, J. (1983). "Materialism and Qualia the Explanatory Gap" *Pacific Philosophical Quarterly* (64): 354-361.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act". *Brain : A Journal of Neurology* (106): 623–642.
- Loewenstein, W. L. (2018). *Zihnini Fiziği*. İstanbul: Papersense Yayınları.

- Manzotti, R. (2007). "From Artificial Intelligence to Artificial Consciousness". *Artificial Consciousness* (ed. Antonio Chella ve Riccardo Manzotti). Exeter: Imprint Academic, 174-190.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>
- New Scientist (2021). *Düşünen Makineler Yaklaşan Yapay Zekâ Çağı ve İnsanlığın Geleceği*. İstanbul: Say Yayınları.
- O'Connor, T. (2004). *Persons and Causes The Metaphysics of Free Will*. New York: Oxford University Press.
- Pink, T. (2004) *Free Will: A Very Short Introduction*. New York: Oxford University Press.
- Reese, B. (2020). *Yapay Zekâ Çağı*. İstanbul: Say Yayınları.
- Silver, D. Hubert, T., Schrittwieser, v.d., (2018). "A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-play". *Science* (362): 1140-1144.
- Strawson, G. (2010). *Freedom and Belief*. New York: Oxford University Press.
- Sun, R. (1997). "Learning, Action and Consciousness: A Hybrid Approach toward Modeling Consciousness". *Neural Networks* 10 (7): 1317-1331.
- Tononi, G. (2013). "On the Irreducibility of Consciousness and Its Relevance to Free Will". *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*. ed. Antoine Suarez, Peter Adams. New York: Springer. 147-176.
- Walsh, T. (2020). *2062 Yapay Zekâ Dünyası*. İstanbul: Say Yayınları.
- Wegner, D. & Wheatley T. (1999). "Apparent Mental Causation Sources of the Experience of Will". *American Psychologist* 54 (7): 480-492.
- Yonck, R. (2019). *Makinenin Kalbi Yapay Duygusal Zekâ Dünyasında Geleceğimiz*. İstanbul: Paloma Yayınevi.
- Zambak, A. (2018). "Free Will and Artificial Intelligence". *MetaZihin: Yapay Zekâ ve Zihin Felsefesi Dergisi* (2): 167-181.