



AKDENİZ ÜNİVERSİTESİ  
SOSYAL  
BİLİMLER ENSTİTÜSÜ



Ferhat BAYIK

YAPAY ZEKAYA UYGULAMALI ETİK BAĞLAMINDA BİR YAKLAŞIM

Felsefe Ana Bilim Dalı  
Doktora Tezi

Antalya, 2022



AKDENİZ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ



Ferhat BAYIK

YAPAY ZEKAYA UYGULAMALI ETİK BAĞLAMINDA BİR YAKLAŞIM

Danışman

Prof. Dr. Hasan ASLAN

Felsefe Ana Bilim Dalı

Doktora Tezi

Antalya, 2022

**T.C.**  
**Akdeniz Üniversitesi**  
**Sosyal Bilimler Enstitüsü Müdürlüğüne,**

Ferhat BAYIK'ın bu çalışması, jürimiz tarafından Felsefe Ana Bilim Dalı Doktora Programı tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Hüseyin Gazi TOPDEMİR

Üye (Danışmanı) : Prof. Dr. Hasan ASLAN

Üye : Prof. Dr. Süleyman DÖNMEZ

Üye : Doç. Dr. İsmail SERİN

Üye : Dr. Öğr. Üyesi Ali Bilge ÖZTÜRK

Tez Başlığı: Yapay Zekaya Uygulamalı Etik Bağlamında Bir Yaklaşım

Onay : Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 24 / 5 /2022

Mezuniyet Tarihi : 16/06/2022

(İmza)  
Prof. Dr. Ebru İÇİGEN  
Müdür

## AKADEMİK BEYAN

Doktora Tezi olarak sunduğum “Yapay Zekaya Uygulamalı Etik Bağlamında Bir Yaklaşım” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımda yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

İmza

**Ferhat BAYIK**





**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**SOSYAL BİLİMLER ENSTİTÜSÜ**  
**TEZ ÇALIŞMASI ORJİNALLİK RAPORU**  
**BEYAN BELGESİ**



**SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE**

| <b>ÖĞRENCİ BİLGİLERİ</b>             |                                                                   |
|--------------------------------------|-------------------------------------------------------------------|
| <b>Adı-Soyadı</b>                    | <b>Ferhat BAYIK</b>                                               |
| <b>Öğrenci Numarası</b>              | <b>20155232004</b>                                                |
| <b>Enstitü Ana Bilim Dalı</b>        | <b>FELSEFE</b>                                                    |
| <b>Programı</b>                      | <b>Doktora</b>                                                    |
| <b>Programın Türü</b>                | <b>() Tezli Yüksek Lisans (X) Doktora () Tezsiz Yüksek Lisans</b> |
| <b>Danışmanın Unvanı, Adı-Soyadı</b> | <b>Prof. Dr. Hasan ASLAN</b>                                      |
| <b>Tez Başlığı</b>                   | <b>Yapay Zekaya Uygulamalı Etik Bağlamında Bir Yaklaşım</b>       |
| <b>Turnitin Ödev Numarası</b>        | <b>1853830663</b>                                                 |

Yukarıda başlığı belirtilen tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 136 sayfalık kısmına ilişkin olarak 10.06.2022 tarihinde tarafımdan Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nda belirlenen filtrelemeler uygulanarak alınmış olan ve ekte sunulan rapora göre, tezin benzerlik oranı;

alıntılar hariç % 3

alıntılar dahil % 11 'dir.

Danışman tarafından uygun olan seçenek işaretlenmelidir:

(X) Benzerlik oranları belirlenen limitleri aşmıyor ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylarım.

( ) Benzerlik oranları belirlenen limitleri aşıyor, ancak tez/dönem projesi danışmanı intihal yapılmadığı kanısında ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylar ve Uygulama Esasları'nda öngörülen yüzdelik sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.

**Gerekçe:**

Benzerlik taraması yukarıda verilen ölçütlerin ışığı altında tarafımda yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim.

...../...../.....

(imzası)

Prof. Dr. Hasan ASLAN

## İÇİNDEKİLER

|                           |      |
|---------------------------|------|
| KISALTMALAR LİSTESİ ..... | v    |
| ÖZET .....                | vi   |
| SUMMARY .....             | viii |
| TEŞEKKÜR .....            | x    |
| ÖNSÖZ .....               | xi   |
| GİRİŞ .....               | 1    |

## BİRİNCİ BÖLÜM

### İNSAN ZİHNİ ve YAPAY ZEKA İLİŞKİSİ

|                                                   |    |
|---------------------------------------------------|----|
| 1.1 Akıl Nedir? .....                             | 10 |
| 1.1.1 Sezgisel Akıl ( <i>Nous</i> ) .....         | 12 |
| 1.1.2 Pratik Bilgelik ( <i>Phronesis</i> ) .....  | 13 |
| 1.1.3 Sanat ( <i>Techne</i> ) .....               | 15 |
| 1.1.4 Bilimsel Bilgi ( <i>Episteme</i> ) .....    | 17 |
| 1.1.5 Felsefi Bilgelik ( <i>Sophia</i> ) .....    | 17 |
| 1.1.6 Akıl ( <i>Logos</i> ) .....                 | 18 |
| 1.2 Zeka Nedir? .....                             | 20 |
| 1.2.1 Zekanın Felsefi Temelleri .....             | 21 |
| 1.2.2 Psikolog Tanımları .....                    | 25 |
| 1.2.3 Yapay Zeka Araştırmacısı Tanımları .....    | 26 |
| 1.3 Bilinç Nedir? .....                           | 28 |
| 1.3.1 Felsefi Söylemde Bilinç .....               | 29 |
| 1.3.2 Yapay Zeka ve Bilinç İlişkisi .....         | 33 |
| 1.4 Algı Nedir? .....                             | 36 |
| 1.4.1 Teknolojide ve Felsefede Algı Kavramı ..... | 37 |
| 1.4.2 Özne-Nesne Bağlamında Algı .....            | 40 |
| 1.5 Sonuç .....                                   | 42 |

## İKİNCİ BÖLÜM

### TARİHTE YAPAY ZEKA İLE İLGİLİ TARTIŞMALAR

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| 2.1 Yapay Zekanın İşleyişi ve İnsanın Kavrayışı .....                   | 45 |
| 2.2 Zihin-Beden Problemi Temelinde Yapay Zeka .....                     | 48 |
| 2.3 Turing Makinesi, Taklitçilik Oyunu, Turing Testi ve İtirazlar ..... | 56 |
| 2.3.1 Teolojik İtiraz .....                                             | 59 |
| 2.3.2 Kumun Altındaki Kafalar İtirazı .....                             | 60 |
| 2.3.3 Matematiksel İtiraz .....                                         | 61 |
| 2.3.4 Bilinçlilik Argümanı .....                                        | 62 |
| 2.3.5 Çeşitli Engellere İlişkin Tartışmalar .....                       | 63 |
| 2.3.6 Leydi Lovelace'ın İtirazı .....                                   | 65 |
| 2.3.7 Sinir Sistemindeki Süreklilik Argümanı .....                      | 66 |
| 2.3.8 Davranışın Belirsizliğine İlişkin Argüman .....                   | 67 |
| 2.3.9 Duyuüstü Algıya İlişkin Argüman .....                             | 68 |
| 2.4 Güçlü ve Zayıf Yapay Zeka .....                                     | 69 |
| 2.5 Çince Odası Argümanı ve İtirazlar .....                             | 72 |

|                                                         |    |
|---------------------------------------------------------|----|
| 2.5.1 Sistemler Yanıtı (Berkeley) .....                 | 74 |
| 2.5.2 Robot Yanıtı (Yale) .....                         | 74 |
| 2.5.3 Beyin Simülatörü Cevabı (Berkeley ve M.I.T.)..... | 75 |
| 2.5.4 Birleşim Cevabı (Berkeley ve Stanford).....       | 76 |
| 2.5.5 Diğer Zihinler Cevabı (Yale).....                 | 76 |
| 2.5.6 Çoklu Konumlar Cevabı (Berkeley) .....            | 77 |
| 2.6 Sağduyu Tartışması .....                            | 82 |
| 2.7 Sonuç .....                                         | 87 |

## ÜÇÜNCÜ BÖLÜM

### ETİK KURAMLARIN SINIFLANDIRILMASI

|                                                           |     |
|-----------------------------------------------------------|-----|
| 3.1 Etik Kavramına Genel Bakış .....                      | 91  |
| 3.2 Yapay Zeka Sistemlerine Yönelik Etik Yaklaşımlar..... | 98  |
| 3.2.1 Ahlaki Bir Araç olarak Robot Düşüncesi .....        | 98  |
| 3.2.2 Felsefi Yaklaşımlar .....                           | 103 |
| 3.2.3 Etiğe Yönelik Tepeden ve Temelden Yaklaşımlar ..... | 109 |
| 3.2.4 Sonuççu Kuramlar .....                              | 111 |
| 3.2.5 Deontolojik Kuramlar .....                          | 112 |
| 3.2.6 Psikolojik / Sosyolojik Yaklaşımlar .....           | 114 |
| 3.2.7 Sonuç .....                                         | 116 |

## DÖRDÜNCÜ BÖLÜM

### TARTIŞMA

|                                         |     |
|-----------------------------------------|-----|
| 4.1 Otomasyon ve İşsizlik Sorunu .....  | 120 |
| 4.2 Özgünlüğün Kaybolması Sorunu.....   | 121 |
| 4.3 İstenmeyen Sonuçlar Sorunu .....    | 123 |
| 4.4 Sorumluluk ve Gizlilik Sorunu ..... | 124 |
| 4.5 İnsan Türünün Sona Erme Sorunu..... | 126 |
| 4.6 Sonuç .....                         | 128 |

|                    |            |
|--------------------|------------|
| <b>SONUÇ .....</b> | <b>130</b> |
|--------------------|------------|

|                      |            |
|----------------------|------------|
| <b>KAYNAKÇA.....</b> | <b>136</b> |
|----------------------|------------|

|                       |            |
|-----------------------|------------|
| <b>ÖZGEÇMİŞ .....</b> | <b>145</b> |
|-----------------------|------------|

## KISALTMALAR LİSTESİ

|      |                         |
|------|-------------------------|
| Alm. | Almanca                 |
| Fr.  | Fransızca               |
| akt. | Aktaran                 |
| bk.  | Bakınız                 |
| vb.  | Ve benzerleri           |
| vd.  | Ve devamı, Ve diğerleri |
| yy.  | Yüzyıl                  |
| Çev. | Çeviren                 |
| İng. | İngilizce               |
| İsp. | İspanyolca              |
| İt.  | İtalyanca               |

## ÖZET

İnsanoğlunun teknolojiyle ilişkisi mitolojik anlamda imgesel, tarihsel anlamda uygulamalı örnekleriyle günümüze dek sürmüştür. İnsanın yaratma eylemi için kullandığı techne, öznenin nesneyi belirlediği bir serüveni temsil ederken bu anlayış çağdaş dünyada nesnenin özneyi belirlemesine dönüşmüştür. Bu doğrultuda insanoğlu yarattıklarıyla yüzleşmek, hesaplaşmak hatta kendi doğasına geri dönebilmek için adeta bir yarışa mecbur bırakılmıştır. Eskiçağdan bu yana gelişen sürecin çıktısı teknolojik anlamda yapay zeka olarak ifade edilebilir. Yapay zeka teriminin bileşenlerinden zeka, akıl kavramının daraltılmış bir formu olarak sayısal sistemlerdeki süreçlere karşılık gelmektedir. Bu darlık zekanın salt hesaplamalı ya da rasyonel süreçleri kapsamaması, aklın ise estetik, sanat, duygular, psikolojik gibi pek çok sürece ev sahipliği yapmasından ileri gelmektedir. Zekanın akıl karşısındaki bu zayıflığı tarihin akışıyla yerini tam tersi bir anlayışa bırakmıştır. Bu anlayış çerçevesinde zeki olan insanın yüceltildiği insan modeli öne çıkarken, akıllı insandan daha az söz edilmeye başlanmıştır. Bu yönüyle yapay zeka ifadesinin pek çok çevrede ilgi çekici bir şöhrete sahip olduğundan söz edilebilir. Zeka teriminin bu kullanımı yapay zeka açısından kullanıldığında uzun vadede bazı etik problemlere yol açabileceği fark edilmektedir. Söz gelimi aklın enginliği karşısında hesaplamaya dayalı zekanın insanı tanımlaması hem akli hem de zekayı yeniden sorgulamayı gündeme getirmektedir. Yapay zeka meselesinde konu insan akli ve makine zekasının yarışına dönüşmektedir. Bu bağlamın bütün dinamiklerinin eleştirel bir yaklaşımla değerlendirilmesi etik hususlara dikkat çekmek ve olası çözüm önerilerini ortaya koymak açısından literatüre katkı sağlayacaktır.

Başlangıç olarak felsefi bir izlekte zihin süreçleri incelenmiş olup algoritma temelli sayısal sistemlerin işlevleriyle karşılaştırılmıştır. Bu kurgudan yola çıkarak hem öznenin hem de nesnenin etraflıca anlaşıldığı bir yaklaşım neticesinde insan doğası ve algoritmanın doğurduğu bazı kaygılar ve risk faktörleri geleceğe yönelik potansiyeli incelenmiştir. Bu noktada kavramları düşünme, araştırma ve üretme gibi felsefinin amaçları düşünüldüğünde yapay zekanın doğuracağı etik kaygılar ve risk faktörleri belirlenmiştir. Tespit edilen potansiyel riskler ve etik kaygılar ışığında birtakım olası felsefi sonuçlara yer verilmiştir.

“Yapay Zekaya Uygulamalı Etik Bağlamında Bir Yaklaşım” özü itibarıyla bir özne – nesne problemini konu almaktadır. Bu bağlamda tarih boyunca öznenin nesneyi belirlediği bir sürece tanıklık etmenin ötesinde nesnenin özneyi belirlemeye kalkıştığı yeni bir anlayışın sorgulaması yapılmıştır. Bu sorgulama kavramsal temeller barındırdığından felsefi söylemden bağımsız düşünülemez. Kaldı ki; akıl, zeka, bilinç, algı gibi felsefi kavramları indirgemeci bir

anlayışın mekanik sembollerine kurban etmek mümkün değildir. Bu nedenle yapay zekanın arka planını oluşturan felsefi düşüncüyü anlamak, algoritmanın gelecekteki olası sonuçlarını anlamamızı sağlayacak önemli bir noktadır. Tarihsel süreçte özne – nesne etkileşiminde filizlenen etik kuramları sınıflandırmak bu yapay zekanın olası sonuçlarını ön görebilmek için etkili bir yöntemdir. Sonuç olarak bu yöntemler ve bilgiler ışığında insan ve yapay zeka ilişkisinin doğuracağı muhtemel olgulara ve potansiyel felsefi sonuçlara yer verilmiştir.

**Anahtar Kelimeler:** Yapay Zeka, Akıl, Zeka, Bilinç, Etik, Gelecek



## **SUMMARY**

### **AN APPROACH TO ARTIFICIAL INTELLIGENCE IN THE CONTEXT OF APPLIED ETHICS**

Humanity's relationship with technology has continued to the present day with its imaginary examples in the mythological sense and applied examples in the historical sense. *Techne* used by man for the act of creation represents the formation of object by the subject. This understanding has turned into the formation of the subject by the object in the modern world. In this direction, humanity has been forced to face its creations, reckon with them, and pushed to return to its original nature. This process has been developing since ancient times has shown itself as artificial intelligence in a technological sense. Intelligence, one of the components of the term artificial intelligence, corresponds to processes in numerical systems as a narrowed form of the concept of reason. This narrowness is because intelligence covers purely computational or rational processes, and the mind hosts many processes such as aesthetics, art, emotions, and psychology. The flow of history has replaced this weakness of the intellect versus the mind with an opposite understanding. This understanding, where intelligent person is glorified has come to the fore, and the mindful person has become less mentioned creates ethical problems. In this aspect, the expression of artificial intelligence has an intriguing reputation in many circles. When this use of the term intelligence is used in terms of artificial intelligence, it may lead to some ethical problems in the long run. For example, the definition of calculative intelligence raises the question of both reason and intelligence again. In the matter of artificial intelligence, the subject turns into a human mind and machine intelligence race. A critical assessment of this context's dynamics will contribute to the literature in terms of drawing attention to ethical issues and putting forward possible solution proposals.

First, mental processes were examined philosophically and compared with the functions of algorithm-based numerical systems. As a result of this philosophical approach, the subject and the object are understood in detail. Some concerns and risk factors caused by human nature and algorithms and their potential for the future has been discussed. At this point, considering the aims of philosophy, such as thinking, researching, and producing concepts, ethical concerns and risk factors that artificial intelligence will cause have been determined. Finally, some possible philosophical implications in the light of the potential risks and ethical concerns are examined.

The philosophical implications of artificial intelligence for the future are essentially a subject-object problem. In this context, the process of the subject forming the object throughout

history changed to a new understanding in which the object attempts to form the subject. Since this questioning has conceptual foundations, it cannot be considered independently of philosophical discourse. It is impossible to sacrifice philosophical concepts such as mind, intelligence, consciousness, and perception to the mechanical symbols of a reductionist understanding. Therefore, understanding the philosophy behind artificial intelligence is an important point that will enable us to understand the possible future consequences of the algorithm. Furthermore, classifying the ethical theories that emerged in the subject-object interaction in the historical process is an effective method of predicting the possible consequences of this artificial intelligence. As a result, in the light of these methods and information, possible facts and potential philosophical consequences are included.

**Keywords:** Artificial Intelligence, Mind, Intelligence, Consciousness, Ethics, Future



## TEŞEKKÜR

Yapay zeka felsefesiyle tanışmama olanak tanıyan danışman hocam Prof. Dr. Hasan ASLAN'a desteği ve katkıları için öncelikle teşekkür ederim. Ayrıca kıymetli görüşleriyle desteğini hissettiğim Prof. Dr. Hüseyin Gazi TOPDEMİR hocama saygılarımı sunarım. Tezimi okuyarak önemli katkılar sağlayan Prof. Dr. Süleyman DÖNMEZ hocama özellikle teşekkür ederim. Bunun yanında terminoloji ve çeviri hatalarıma ilişkin önemli tespitleriyle tezime yön veren Doç. Dr. İsmail SERİN hocamın katkısı benim için çok kıymetlidir. Saygılarımı ve teşekkürlerimi sunarım. Her daim kaynak ve bilgi paylaşımını esirgemeyen Dr. Öğr. Üyesi Ali Bilge ÖZTÜRK hocama desteği için teşekkür ederim.

Tez yazım sürecinin gizli kahramanları; eşim Nazan, oğullarım Toprak ve Uygur her daim mutluluk kaynağım olmuştur. Kattıkları motivasyona, mutluluğa ve güce ne kadar teşekkür etsem azdır.

Ferhat BAYIK

Antalya, 2022

## ÖNSÖZ

Bu tez çalışması, yapay zekayı uygulamalı etik bağlamında özne ve nesne açısından değerlendirmeyi amaçlamaktadır. Bu yaklaşım felsefi kavramlar olan akıl, zeka, bilinç ve algı kavramlarının anlaşılmasıyla teknolojik bir gelişme olan yapay zekanın insana etkilerini konu almaktadır. Bugüne dek öznenin özneye ilişkilerini konu alan etik yaklaşımların artı yerini öznenin nesneye ilişkisini inceleyen bir alana bıraktığı bir alan oluşmaktadır. Bu nedenle sadece öznenin nesneye ilişkisi olduğu kadar nesnenin özneye ilişkisini belirleyen bir takım etik kuralların felsefi kavramların anlaşılmasıyla daha kapsamlı bir şekilde belirlenmesini mümkün kılacaktır. Bu yönüyle insanlığın geleceğini ilgilendiren bazı etik kaygılar bu türden bir değerlendirmeyi felsefi düzlemde gerekli kılmıştır. Söz konusu etik problemlerin tanımlanması, sınıflandırılması ve olası önlemlerin alınması felsefi bir söylemin ya da tartışmanın konusu olduğundan uygulamalı etik bağlamında değerlendirmeler yapılmıştır. Yapay zekanın teknolojik bir gelişme olarak çok yeni olmadığı tarihsel gelişmelerden anlaşılmaktadır. Öte yandan yapay zeka etiği özne ve nesne ilişkisini tikel ve tümel anlamda kapsamaktadır. Bu nedenle yapay zekanın gözlemlenebilir, ölçülebilir mekanik yönlerinden ziyade geleceğe ilişkin olası, öngörülmesi zor olan problemleri konu almaktadır.

Yapay zeka başlıca mekanik, sinirbilim ve etik alanlarında çalışılan bir alan olması sebebiyle bu alanların en günceli etiktir. Etik sorgulama zorunlu olarak yapay zekanın hem mekanik hem de sinirbilime ilişkin yönlerini incelemeyi gerektirmektedir. Bu yönüyle etik yani felsefi yönlerini anlamak mümkün olacaktır. Kaldı ki mühendislik faaliyeti olmasının ötesinde yapay zeka, felsefi tartışmalara konu olan iddialara ve tartışmalara da gebedir. Bu iddiaların başında bilinç problemi gelmektedir. Dolayısıyla tezin izleği öncelikle insan aklının felsefi temellerini değerlendirmekle başlamakta, yapay zekanın işleyişi ile insan anlayışının farklılıklarını incelemekle devam etmektedir. Akabinde yapay zekaya ilişkin etik yaklaşımlar sınıflandırılarak tartışma bölümünde başlıca kaygılara yer verilmiştir. Sonuç olarak tez boyunca elde edilen bulgular ışığında yapay zekanın insanlığı etkileyebilecek olan geleceğe yönelik yansımaları değerlendirilmiştir.

Tez esas olarak insanın referans alındığı bir yaklaşımla yapay zekayı anlamayı, sınırlılıklarını ve kapasitesini ortaya koyabilmeyi öncelemektedir. Yapay zekanın gelecekte hangi türden sonuçlar doğurabileceğine ilişkin öngörülerin belirlenmesi bu bilgiler ışığında amaçlanmıştır. Bu hususta en önemli referans noktası insan “aklı” ve yapay zeka algoritması arasında bir ayrım fark edilmiştir. Söz konusu algoritma temelli “zeka” ile akıl yani *logos*’un değişkenlerini detaylıca irdelenmiştir. Kaldı ki akıl teriminin felsefi anlamı ile yapay zeka

ifadesinde geçen zeka teriminin mühendislikteki anlamı farklı şeylerdir. Bu nedenle insan zihnine ilişkin detaylara felsefi yönleriyle yer verilmiş, daha sonra da algoritmanın sistematığına değinilmiştir. Bu türden bir ayrımın ya da benzerlik kurma çabasının ne türden etkiler yaratacağı konusu ayrıntılarıyla tartışılmıştır. Sözü edilen amaç ve hedefler doğrultusunda tez dört bölümden meydana gelmiş olup her bölüm içerisinde değerlendirmelere yer verilmiştir.

Bu tezde zeka terimi algoritmanın işleyişine benzer bir anlamda tartışılmaktadır. Bu yönüyle zekanın mekanik bir anlamda kullanılmasına bağlı olarak akıl terimi karşısında zekanın yükseldiği bir anlayışa dikkat çekilmektedir. Bu tespit yapay zekanın etik sonuçlarıyla ilgilidir. Öz bir ifadeyle kavramların mekanik düzleme indirgenmesinin aklın olanaklarını körelttiği iddia edilmektedir. Bu yönüyle zekanın algoritma gibi anlaşıldığı çağa eleştirel tutum söz konusudur. Bu görüşler özellikle tartışma bölümünde; otomasyon ve işsizlik riski özgünlüğün kaybolması riski istenmeyen sonuçlar riski sorumluluk ve gizlilik riski insan türünün sonunu getirme riski gibi başlıklar altında incelenmiştir. Halihazırda etik tartışmalardan sonuncusu olan “yapay zekanın insan türünün sonunu getirip getirmeyeceği” tartışmasında bilincin sayısal sistemlerde mümkün olup olmadığına ilişkin görüş ve eleştirilere yer verilmiştir. Bu yönüyle bilinç probleminin önemine dikkat çekilmiştir. Kaldı ki yapay zeka ve insan karşılaştırmasında düşünülen konunun bilinç olduğu fark edilmektedir. Yapay zekanın insan bilincine erişen ya da onu aşan bir konuma gelmesinin neden önemli olduğu, bu tartışmanın hangi çevrelere, hangi amaçlara katkı sağlayacağı sorgulanmıştır. Öte yandan insan aklının uçsuz bucaksız olanakları, algoritmanın hızına ve hassaslığına kurban edilemeyecek kadar büyüktür. Bu bağlamda Searle “Kısacası sentaks, sistemin fiziğine içkin değildir; yalnızca onu kullanan kişide anlamını bulur.”<sup>1</sup> ifadesiyle algoritmanın sınırlarına dikkat çekmiştir.

---

<sup>1</sup> Searle, 1992: 77,78.

## GİRİŞ

Yapay zeka, mantıksal işlemler yapabilen, öğrenebilen ve özerk hareket edebilen sistemler olarak zeki makineler yaratmaya çalışan bir bilgisayar bilimi dalıdır. Senelerdir üzerinde çalışılan bir alan olmasına rağmen yapay zeka, büyük veri ve makine öğreniminin ortaya çıkmasıyla nispeten son zamanlarda daha yaygın bir şekilde kullanılmaya başlanmıştır. Günümüzde finans, sağlık, üretim ve ulaşım gibi çeşitli sektörlerde kullanılan yapay zekanın gelişimi devam ettikçe potansiyel kullanım alanlarının artacağı olasıdır. Öyle ki yapay zeka başlıca uzman sistem, bulanık mantık, yapay sinir ağları ve makine öğrenmesi gibi tekniklerin kullanıldığı genel bir terimdir. Yapay zekanın nasıl çalıştığını, hangi ortamlarda kullanıldığını bilmek terim yapısını anlamak açısından katkı sağlayacaktır. Bu hususta Bose (1994) tarafından kaleme alınan uzman sistem, bulanık mantık ve yapay sinir ağlarına ilişkin bilgilere yer vermek gereği doğmaktadır.

İlk olarak “Uzman Sistem” adı verilen, insanın düşünme süreciyle ilişkilendirilen tamamen yeni bir yazılım yapılandırma yöntemi ortaya çıkmıştır. Yazılım mühendisliğinin yeni dalına “Bilgi Mühendisliği” denmekteydi. “Bilgi Mühendisleri”, belirli bir alandaki insan uzmanlardan bilgi edinerek onu yazılıma dönüştürmekle görevliydi. 1980'lerde, endüstriyel süreç kontrolü, tıp, jeoloji, tarım, bilgi yönetimi, askeri bilim ve uzay teknolojisinde uzman sistem uygulamaları çoğaldı. İkinci olarak 1960'ların ortalarında "Bulanık Mantık" veya bulanık küme teorisi adı verilen ve uzman sistemi bir yapay zeka aracı olarak tamamlamaya yardımcı olan yeni bir teori ortaya atılmıştır. Bu teori, insan düşüncesinin çoğunun doğası gereği bulanık veya belirsiz olduğundan Boole mantığının ("0" ve "1" ile temsil edilen) düşünme sürecini yeterince taklit edemediğini savunmuştur. Son yıllarda, bulanık mantık, önemli bir yapay zeka aracı olarak ortaya çıkmıştır. Süreç kontrolü, tahmin, tanımlama, teşhis, borsa tahmini, tarım, askeri bilim vb. alanlarda yaygın olarak uygulanmaktadır. Üçüncü olarak geleneksel dijital bilgisayar, uzman sistem problemlerini çözmede çok verimliyken bulanık mantık problemlerini çözmede sınırlıydı. Bu nedenle araştırmacılar son zamanlarda “yapay sinir ağı” (*ANN - artificial neural networks*) veya “sinir ağı” olarak adlandırılan yeni bir yapay zeka dalı üzerinde odaklanmıştır. Temel olarak nöron adı verilen milyarlarca sinir hücresinden oluşan insan beynindeki bu nöronlar biyolojik sinir ağını oluşturmak için birbirine bağlıdır. Düşünme sürecimiz bu sinir ağının eylemiyle oluşturulur. Yapay sinir ağı, elektronik hesaplama devreleri ile sinir ağını taklit etme eğilimindedir. Yapay sinir ağı teknolojisi, insan düşüncesine denk gelen en genel yöntem olarak bilinmektedir. Süreç kontrolü, teşhis, tanımlama, karakter

tanıma, robot görüşü, uçuş planlaması, finansal tahmin vb. için uygulanmıştır.<sup>2</sup> Buraya kadar yapay zekanın tekniklerine yer verdik. Dolayısıyla yapay zekanın bu tekniklerin uygulandığı bir üst kavramdır. Öte yandan güçlü ve zayıf yapay zekaya ikinci bölümde ayrıca yer verilmiştir.

Yapay zeka teknikleri, işleyişi, yaygın olarak mühendislik faaliyetinin bir parçası olarak bilinmektedir. Ancak bu tür sistemlerin özne – nesne ilişkisiyle neden olacağı etkiler yapay zeka tartışmalarının farklı ve önemli bir alanıdır. Bu tartışma mühendislik ya da sinirbilimin çalışma alanını aşmaktadır. O nedenle yapay zeka felsefesi, bu sistemlerin potansiyel faydaları ve zararları hakkında etik açıdan ilgilidir. Meselenin iki yönüne doğrudan dikkat çekmek ve tezin izleğini somutlaştırmak adına mühendislik faaliyeti olan yapay zekanın etik sonuçlarına ilişkin bir örnek faydalı olacaktır.

İsviçre Federal NBC (nükleer, biyolojik ve kimyasal) Koruma Enstitüsü, Kimyasal ve Biyolojik Bilimler tarafından düzenlenen bir konferansta ilaç araştırma ve geliştirme sürecinde yapay zekanın faydaları tartışılmıştır. Konferansta amaçlanan faydaların yanında beklenmedik bir detay vahim olabilecek sonuçlardan birinin habercisi olmuştur. Tahmin edilen zehri etkisiz kılmak ve istenileni pekiştirmek amacıyla “MegaSyn” yanında “De Novo” adında bir model tasarlanmıştır. Yirminci yüzyılın en zehirli kimyasal savaş araçlarından biri olan sinir ajanı VX<sup>3</sup> gibi bileşiklere yönlendirilmiştir. Bu model 6 saatten daha kısa bir sürede istenilen eşikte puan alan 40.000 molekül üretmiştir. Bu süreçte yapay zeka sadece VX'i değil, aynı zamanda kamu kimya veri tabanlarındaki yapılarla görsel doğrulama yoluyla tanımlanan diğer birçok bilinen kimyasal savaş aracını da tasarlamıştır.<sup>4</sup> Bu tespitin bir bilim kurgu olmadığı vurgulanan makalede, ilaç keşfi için yapay zeka yazılımı kullanan yüzlerce şirketin bulunduğu dikkat çekilmektedir. Dolayısıyla kötüye kullanım olasılıklarının yapay zeka tarafından tasarlanmış molekülleri üretmek için küresel kimya ağını kimlerin kullanabileceğine ilişkin kaygılara ve olası felaketlere dikkat çekilmektedir.

Yapay zekanın hızına ve hassaslığı kapsamında geliştirilebilecek şeyleri hayal etmek güçtür. Öyle ki yapay zekanın yanlış veya kötüye kullanımı insanlığı tehdit edebilecek sonuçlara neden olabilecektir. Bu noktada çalışmamız, mühendislik faaliyetinin ötesinde bu faaliyetlerin doğurabileceği sonuçlara felsefi söylemle dikkat çekmektir. Bu husus hem felsefi hem de toplumsal çerçevede etiği öne çıkarmaktadır. Bu doğrultuda yapay zekanın felsefi sonuçlarına ulaşma yolunda zihin felsefesi bağlamında insan zihni ve yapay zeka, tarihte yapay

<sup>2</sup> Bose, 1994: 1303,1304.

<sup>3</sup> Birkaç tuz tanesi büyüklüğünde VX taneciği (6-10 mg) bir insanı öldürmek için yeterlidir. Detaylı bilgi için bk. <https://doi.org/10.1038/s42256-022-00465-9>

<sup>4</sup> Urbina vd., 2022: 189–191.

zeka tartışmaları ve etik kuramların sınıflandırılması bu hedeflere ulaşma maksadıyla çalışılmıştır. Yapay zeka felsefesini ele alırken yapay zeka açısından felsefeyle ilgili ortak bazı kavramları da analiz eder. Bu izlekte zihin felsefesinin başat kavramlarını tanıtarak algoritmanın kısıtlılığına tarihteki yapay zeka tartışmalarıyla dikkat çekilmektedir. Bu türden bir analiz hem yapay zekayı anlamak hem de etik tartışmalara dayanak oluşturmak amacıyla destekleyici niteliktedir. Özünde kavramsal bir çalışma olan bu tez, etik düzlemde yapay zekanın özne ve nesne bağlamında olası sonuçlarını belirlemeyi amaçlamaktadır. Kaldı ki yapay zekanın tasarım süreci bilinmekte ve gözlemlenebilmektedir. Oysa bu sistemlerin gelecekte doğurabileceği sonuçlar belirsizdir. Bu belirsizlik insanlığın gidişatını nelerin beklediği konusunda kaygı ve endişe doğurmaktadır. Bir tarafta yapay zekanın verimlilik açısından faydalar sunduğunu iddia edenler, diğer tarafta ise insan işlerinin kaybı, veri gizliliği sorunları ve hatalı algoritmaların neden olduğu hata riski gibi olasılıklardan endişe duyanlar bulunmaktadır. Bu kaygıların ötesinde insanlığın geleceğine ilişkin yapay zekanın vahim sonuçları da bulunmaktadır. Gerek yapay zeka savunucuları gerekse yapay zeka karşıtları gelecekteki olasılıkları ön görebilmek açısından güçlü yöntemlere ihtiyaç duymaktadırlar. Öyle ki, yapay zeka sistemlerini geliştirmek kadar bu sistemlerin geleceğe yansımalarını ortaya koymak felsefe aracılığıyla mümkündür. Kaldı ki yapay zeka meselesini genel hatlarıyla mekanik, sinirbilim ve felsefi çerçevede tartışmak mümkündür. Söz gelimi, Bose'a (1994) göre psikologlar ve psikiyatristler gibi davranış bilimciler, insan beynini referans alarak çalışmalarını sürdürmektedirler. Ancak beyin hakkındaki bilgilerin yetersizliğinden bu sürecin uzun süreceği tahmin edilmektedir. İnsan beynin nihai akıllı makine olduğu kabulüyle şu soru sorulmaktadır: "Böyle bir zekayı veya en azından bir kısmını, bir bilgisayar yardımıyla yapay olarak üretmek ve karmaşık problemlerimizi çözebilmesi için mümkün mü?"<sup>5</sup> Sinirbilim açısından yapay zeka araştırmaları yoğunlukla sayısal sistemleri insan beyninin kapasitesine yükseltme uğrunda ilerlemektedir. Peki, bunun mümkün olduğu düşünülürse olası sonuçları inceleyen alan nedir? İşte bunun cevabını uygulamalı etikte görmekteyiz.

Tezin hedefi yapay zekanın neticelerine dikkat çekmek olduğundan etik konular son bölümde sınıflandırılmış, felsefi sonuçlar da sonuç bölümünde temellendirilmiştir. Ayrıca uygulamalı etiğin meseleleri temellendirmek amacıyla yapay zekanın mekanik ve sinirbilim yansımalarından faydalanılmıştır. Bu hususta felsefi yaklaşımın ana yol olduğu unutulmamalıdır. Öyle ki teknoloji felsefesinin bir branşı olarak yapay zeka, felsefede ortak kavramlara sahiptir. "Zeka, etik, bilinç, epistemoloji ve özgür irade birçok ortak kavramdan

---

<sup>5</sup> Bose, 1994: 1303.

sadece birkaçıdır.”<sup>6</sup> Sözü geçen kavramların çalışılması, yapay zekanın etik yönlerini ve felsefi sonuçlarını belirlemek açısından önemlidir. Bu doğrultuda tez özünde etik ve kavramsal bir çalışmadır.

Buraya kadar gerek yapay zekanın tekniklerini gerekse uygulamalı etikle neyi hedeflediğimizi özetledik. Şimdi felsefi söylemde “yapay zeka” terimini ve bileşenlerini tarihsel arka planıyla özetlemeye çalışalım. XXI. yüzyılda adından sıkça söz ettiren yapay zeka özellikle bilişim alanında yaygınlaşmıştır. Bu yönüyle yapay zeka gücünü sembollerin kusursuz dizilimi olan algoritmadan almaktadır. Algoritmanın tarihine kısaca bakıldığında; Muhammed Bin Musa el Harezmi’nin (780-850) “Kitab Ü'l Muhtasar Fi Hesab El-Cebr Ve'l Mukabele” adlı eserinde geçen El-Cebr ifadesinden doğduğu bilinmektedir. Ancak bu ifade Latin dünyasına Algorismus olarak yansıtılmış ve algoritma olarak günümüze dek ulaşmıştır.<sup>7</sup> Bu noktada tezin bütününe şekillendirecek olan algoritma teriminin tanımına yer vermek önemlidir. Algoritma terimini Smith (1990), her zaman doğru veya en iyi sonucu üreten biçimsel bir yöntem olarak ifade etmektedir. Bir algoritma belirli bir sonucu garanti eden veya belirli bir sorunu çözen adım adım bir yöntem olarak tanımlanmaktadır. Algoritma yöntemi kısıtlı bir sürede hesaplama yapma esasına dayanmaktadır. Bu bağlamda programcılarının geleneksel bir programı geliştirirken programın izleyeceği algoritmayı belirledikleri ifade edilmektedir.<sup>8</sup>

Anlaşılabacağı üzere mekanik anlamda uygulamalı, düşünsel anlamda imgesel örnekleriyle algoritmanın gelişimi tarihsel örneklerle müjdelenmiştir. Diğer bir deyişle mekanik anlamda çok eskilere dayanan icatlar algoritmanın gücüyle yapay zekaya imkan tanımıştır. Nihayet XX. yüzyıl; sibernetik<sup>9</sup>, zeki makineler, yapay zeka ve robot<sup>10</sup> gibi kavramların ortaya çıktığı ve Turing Testinin önerildiği dönem bir olmuştur. Yapay zekanın gelişimine yönelik ilk söyleme Bush’un (1945) “Mantık yasalarına uyarak basit tekrarlayıcı düşünceler makineler tarafından yapılabilir.”<sup>11</sup> ifadesinde görmekteyiz. Yapay zekaya giden yolda çeşitli teknolojileri (bir dereceye kadar) öngörmüştür. Bilimsel bir çalışma olarak literatüre giren yapay zeka<sup>12</sup> kavramı John McCarthy (1956) tarafından ortaya atılmıştır. Sonuç

<sup>6</sup> McCarthy, 2006: 2.

<sup>7</sup> Aslan, 2003: 244.

<sup>8</sup> Smith, 1990: 12.

<sup>9</sup> Norman Weiner tarafından oluşturulmuş bir terim makinelerdeki ve biyolojik organizmalardaki kontrol mekanizmalarının çalışmasını belirtmek için kullanılır. Yunanca bir kelime olan 'dümenci' sözcüğünden türetilmiştir. Latince eşdeğeri, vali ve hükümet yöneticisi gibi terimlere yol açmıştır. Detaylı bilgi için bk. Raynor, 1999:70.

<sup>10</sup> Bir robot, önceden bilinmeyen koşullar altında birçok görevi gerçekleştirebilen genel amaçlı bir sistem olarak belirtilmektedir. Robotlar, sensörler, bilgisayar tabanlı bir kontrol cihazı, diğer giriş cihazları ve ortam üzerinde kullanabileceği mekanik el gibi bir tür tetikleyici gibi şeylere sahip olan bir mekanizma olarak değerlendirilmektedir. Detaylı bilgi için bk. Raynor, 1999:259.

<sup>11</sup> Bush, 1945: 116.

<sup>12</sup> Russell ve Norvig, 2010: 18.

olarak Eski Çağ'dan günümüze değin yaşanan gelişmelere bakıldığında insan zihninin işleyişine benzer süreçleri mekanizmalar üzerinde gerçekleştirme çabası birçok araştırmacı tarafından gerçekleştirilmiştir. Bu gelişmelere yapay zekanın kısa tarihçesi niteliğinde bir bakışla değinmek yerinde olacaktır.

Söz gelimi Homeros'un *İlyada* eserinde ateş tanrısı total Hephaistos'un yarattığı kendiliğinden harekete geçebilen bir sandalyeye değinilmiştir.<sup>13</sup> Öte yandan eserin devamında Hephaistos'un misafiri olan deniz tanrıçası Thetis'e hizmet etmesi için görevlendirilen zeka sahibi 'uşaklar' şu ifadelerle anlatılmaktadır: "Efendisine alttan destek olsun diye görevlileri harekete geçirdi. Bunlar altın rengindeler ve canlı genç kadınlara benzerler, kalplerinde zeka var, konuşma ve güç var, ölümsüz tanrılardan da bir şeyler yapmayı öğrendiler."<sup>14</sup> Ayrıca Hephaistos'un yarattığı muhafız Talos insan gücünü aşan, görevi güvenliği sağlamak olan bir otomat olarak anlatılmaktadır. Hephaistos, Zeus'tan Avrupa'ya bir hediye olarak görevi Girit sahillerini günde üç kez devriye gezmek olan bir bronz adam olan Talos'u yapmıştır. İşgalcilere büyük kayalar fırlatarak ya da izinsiz girenleri alev dolu bedeniyle yalancı bir kucaklamayla yakarak engeller."<sup>15</sup>

Öte yandan hava, boşluk ve denge üzerine çalışmalarla tarihte önemli bir yeri olan İskenderiye Mekanik Okulu'ndan Ctesibios (M.Ö. 3.yy), Philon (M.Ö. 2. yy) ve Heron (M.Ö. 1. yy) mekanizmalar ve otomatlar üzerine çeşitli çalışmalarda bulunmuşlardır. Bu çalışmaların tercümeleri sayesinde İslam Dünyası'na geçişi mümkün olmuş, özellikle Benu Musa (9. yy), Farabi (875-950), Hazini(1100'ler) ve Cezeri'nin (13. yy) çalışmalarıyla adeta doruk noktasına çıkmıştır.<sup>16</sup>

Diğer bir yandan McCorduck'un çalışmasında özellikle XV. yüzyıl sonrasındaki gelişmeler sayısal sistemlerde algoritmanın kullanım amacının değiştiğine işaret etmektedir. Çağdaş dünyada bunun en önemli yansıması algoritmanın gücüyle bilinci anlayabilme, taklit edebilme hatta bilincin sınırlarını aşabilme gayesi fark edilmektedir. Bu süreç, Avrupa'da ilk modern ölçüm makineleri olan mekanik saatler, insanı temsil eden küçük zeki yaratık *homunculus* modelleriyle başlamıştır. XVII. yüzyılda otomatlara duyulan hayranlık gerek eğlence gerekse estetik gereksinimlerle şekillenmiştir. Ayrıca bu dönemde Blaise Pascal'ın *Pascaline* adında mekanik bir hesap makinesi, Descartes'ın zihin-beden problemini sistemleştirdiği *İnsan Üzerine İnceleme* eseri ve Leibniz'in gelişmiş bir mekanik hesap makinesi olan *Step Reckoner*'ı icadı gibi gelişmeler yaşanmıştır. XIX. yüzyıl ise Hoffman'ın

<sup>13</sup> Homer, 1951: 385.

<sup>14</sup> Homer, 1951: 386.

<sup>15</sup> McCorduck, 2004: 5.

<sup>16</sup> Unat, 2006: 225, 226.

*The Sandman*'ı, Goethe'nin *Faust*'u ve Mary Shelley'nin *Frankenstein*'ı gibi edebi yapay zekaların çoğaldığı bir dönem olmuştur.<sup>17</sup>

Yapay zekanın gelişimine ilişkin kısa tarihçenin ardından terim yapısı ve tanımlamaya ilişkin bilgi vermek yerinde olacaktır. İlk olarak yapay zeka teriminin yapısına bakıldığında yapay<sup>18</sup> *artificiālis*<sup>19</sup> ve zeka<sup>20</sup> *intellect* sözcükleriyle meydana geldiği fark edilmektedir. Bu birleşimden ortaya çıkan yapay zeka ifadesinden doğal veya kendiliğinden olmayan bir tür zeka türü düşünülebilir. İfadenin mahiyetini belirleme çalışmalarında pek çok fikir ortaya atılmış ve tartışmalar gerçekleşmiştir. Yapay zeka ifadesinin tanımında yaşanan belirsizliklere ve tartışmalara dikkat çeken düşünürlerden Murphy (2000), “Makinelerin akıllıca hareket etmesine ilişkin bilime genellikle yapay zeka ya da kısaca YZ denir. Yapay zekanın yaygın olarak kabul edilmiş bir tanımı yoktur.”<sup>21</sup> ifadesini kullanmıştır. Yapay zekanın mühendislik faaliyetleri kapsamında değerlendirildiği çevrelerde “YZ” (*AI*) kısaltması oldukça popüler bir kullanımdır. Ne var ki çalışmamız bu konuya teknik, mekanik, davranışçı ya da salt teknolojik bir yaklaşımı benimsemediğinden kısaltma kullanılmamış, doğrudan “yapay zeka” terimine yer verilmiştir.

Tanımaya ilişkin diğer bir yaklaşım da bilişsel bilimler ve bilgisayar bilimleri üzerine çalışmalarda bulunan John McCarthy tarafından 1956 senesinde öne sürülen yapay zeka, “Zeki makineler, özellikle zeki bilgisayar programları yapma bilimi ve mühendisliği.”<sup>22</sup> şeklinde tanımlamıştır. Bu tanımın McCarthy nezdinde sadece genel bir çerçeveden ibaret olup üzerinde uzlaşılmış bir tanımın olmadığına bilgisi de verilmektedir. McCarthy, bu sorunun kaynağını genel olarak hangi hesaplamalı süreçlerin zekayla ilgili olabileceği ya da olamayacağının ayırt edilememesinden kaynaklandığını not etmektedir. Öte yandan Murphy (2000) “Tartışmalı bir ifade olan yapay zeka, bir makinenin zeki olup olamayacağı hususunda devam eden felsefi tartışmaları beraberinde getirmektedir.”<sup>23</sup> yorumuyla yapay zekanın felsefi söylemden bağımsız düşünülemeyeceğinin altını çizmiştir. Üstelik, yapay zekanın işlevlerini “zihin” ile bağdaştırmanın oluşturduğu belirsizliği deyim yerindeyse bir tür çılgınlık görmektedir. Bu hususta mühendislerin yapay zekayı bir spekülasyon olduğu gerekçesiyle görmezden geldiğini, yapay zekayı çevreleyen tartışmaları önlemek için çalışmalarını sıklıkla “zeki sistemler” veya

<sup>17</sup> McCorduck, 2004: XXIII – XXX.

<sup>18</sup> Alm. *künstlich*, İng. *artificial*, İsp. *artificial*, Fr. *artificiel*, İt. *artificiale*.

<sup>19</sup> *ars* kökünden "sanat" anlamında, *artifex*'e türeyerek "zanaatkar" anlamında, *artificium*'a türeyerek "bir sanat eseri; beceri, teori, sistem" anlamında, son olarak *artificial*'a dönüşerek "doğal veya kendiliğinden olmayan" anlamında kullanılmıştır. Bk. <https://www.etymonline.com/word/artificial>

<sup>20</sup> Alm. *intelligenz*, İng. *intelligence*, İsp. *inteligencia*, Fr. *intelligence*, İt. *intelligenza*.

<sup>21</sup> Murphy, 2000: 15.

<sup>22</sup> McCarthy, 2007: 2.

<sup>23</sup> Murphy, 2000: 16.

“bilgi tabanlı sistemler” olarak etiketlediği bildirilmektedir.”<sup>24</sup> Araştırmacıların görüşlerinden anlaşılacağı üzere yapay zekanın tanımı kadar kapsamı da tartışmalıdır. Yani kimi görüşe göre insana özgü niteliklerin sayısal sistemlerde gerçekleştirilmesi iddia edilirken, kimi görüşe göre bilgi tabanlı sistemlerin yapay zeka olarak kabul edilebileceği savunulmaktadır.

Buraya kadar yapay zekanın gerek kavramsal gerekse kapsamına ilişkin giriş niteliğinde açıklamalarda bulunduk. Yer verdiğimiz açıklamalar yapay zeka meselesinde öne çıkan konulardır. Bu konuların ötesinde yapay zekanın gelecekte özne nesne ilişkisine etkilerinin ne olacağı problemi muğlak bir konudur. Yapay zekanın mekanik boyuttan sembolik boyuta taşınmasıyla öznenin nesneyle ilişkisi yön değiştirmiş olup bu değişim nesnenin özne üzerinde söz sahibi olabileceği bir anlayışı doğurmuştur. Bu anlayış sayısal sistemlerin insan ve toplum üzerindeki sonuçlarına bağlı olarak uygulamalı etik konuları ilgilendirmektedir. Yapay zekanın doğurabileceği etik kaygılar felsefenin tartışma alanıdır. Bu yüzden geleceğe ilişkin sonuçlar yani etik sorunlar yapay zekayı zihin felsefesi konu etmektedir. O nedenle yapay zekanın etik boyutları bu çalışmanın merkezini oluşturmaktadır. Örnek bir ifadeyle başlamak gerekirse M.I.T. yapay zeka laboratuvar yöneticisi Edward Fredkin’in görüşlerine bakmak yeterli olacaktır. Fredkin (1983) yapay zeka sistemlerinin gelecekte alt birimlerini tasarlamasıyla ortaya çıkabilecek senaryoya bir belgesel programında verdiği demeçle dikkat çekmektedir.

“Bu makineler geliştikçe ve bazı zeki makineler diğerlerini tasarladıkça hatta daha zeki hale geldikçe, en akıllı insandan milyonlarca kat daha akıllı bir makineyi sahiplenme düşüncesini hayal etmek oldukça zorlaşıyor. Geleceğin yapay zekaları bizim anlayamayacağımız ciddi problemler için endişelenecek ve bizimle konuşmaya tenezzül etmeyebilecektir. Ara sıra bizi eğlendirebilecek veya sevdiğimiz oyunları oynayabilecek, bir anlamda bizi evcil hayvan olarak tutabilecekler.”<sup>25</sup>

Fredkin’in şaşırtıcı hatta kaygılandırıcı sözlerine yer vermek tesadüf olmamakla beraber amaç, tezin esaslı düşüncesine doğrudan dikkat çekmektedir. Fredkin’in sözleri etik anlamda doğrudan belirleyici olmamakla birlikte yapay zekanın gelecekteki sonuçlarından olumsuz olanlarına göndermede bulunmaktadır. Öte yandan böyle bir iddianın felsefi dayanaklarla yapılması isabetli olacaktır. Kaldı ki yapay zeka gibi bir olgu özellikle zihin felsefesiyle anlaşılmadıkça etik değerlendirmeler eksik kalacaktır. Bu bağlamda tarihte yapay zeka tartışmalarını tanımak, çağdaş dünyaya yansımaları incelemek etraflı bir bakış açısı sağlayacaktır. Bu izlekte söz konusu tartışmaların sınıflandırılması yapılarak tarihte yer almış argümanlar, itirazlar ve uygulamalı örneklere değerlendirilecektir. Son olarak tezin esaslı konularından etik yaklaşımlar sınıflandırılarak yapay zekanın sosyolojik, hukuki ve özne-nesne

<sup>24</sup> Murphy, 2000: 16.

<sup>25</sup> Laurie, J.(1983) Better Mind the Computer [Horizon]. BBC. İngiltere.

ilişkisi tartışma bölümünde değerlendirilecektir. Girişte verdiğimiz tanıtıcı bilgilerin hemen ardından yapay zekanın amacını, konumunu, işlevlerini ve gelecekteki rolünü belirlemeye yönelik girişim destekleyici bilgileri incelemektir. Bu çerçevede insan zihni ve yapay zeka ilişkisi bağlamında zihin felsefesi sorgulaması yapılacaktır. Burada amaç yapay zeka ifadesinin bileşenlerinden “zeka” teriminin zihin felsefesinin diğer unsurlarıyla karşılaştırmaktır. Bu nedenle tanımlama girişimleri felsefi anlayıştan uzaklaşmakta ve problemler ortaya çıkmaktadır. O halde bu sorunun önüne geçmenin olanağı nedir? Bu nedenle çalışmamızın ilk bölümü her şeyden önce zihnin başat unsurları olan akıl, zeka, bilinç ve algı kavramlarının incelenmesine ayrılmıştır. Kaldı ki yapay zeka hususunda çeşitli yaklaşımlar öne çıkmış olup fizikalist<sup>26</sup> yaklaşımın özellikle mühendislik çevrelerinde yaygın olduğu göze çarpmaktadır. Bu bağlamda çalışmanın gayesi fizikalist ya da indirgemeci yaklaşımın ötesinde felsefi söylem aracılığıyla konuyu etraflica ele almaktır. Bu çerçevede zihin kavramının felsefi söylemden uzak, salt indirgemeci bir anlayışa sıkıştırılmasına karşı duruş sergilenmektedir. Zihnin konumunu beyin organına indirgeyen anlayış çerçevesinde bir yapay zeka sisteminin mümkün olabileceğine ilişkin görüşler gerek sinirbilim çevrelerinde gerekse uygulama geliştiricileri arasında oldukça yaygındır. Buna rağmen yapay zekanın düşünsel temellerini çözümlemede felsefenin rolünü göz ardı etmeyen düşünürler de bulunmaktadır. Dennet (1998) bu konunun önemine şu ifadelerle dikkat çekmiştir.

“Fiziksel bir şey-bir insan, bir hayvan, bir robot-dünyanın bilgisini algıdan çıkarması ve daha sonra bu bilgiyi başarılı eylemin rehberliğinde kullanması nasıl mümkün olabilir? Bu, filozofların nesiller boyunca boğuştukları bir sorudur, ancak yapay zekanın tanımlayıcı sorularından biri olarak da kabul edilebilir. Yapay zeka, büyük ölçüde felsefedir. Genellikle ilk bakışta tanınabilen felsefi sorularla doğrudan ilgilidir: Akıl nedir? Anlam nedir? Akıl yürütme ve rasyonellik nedir? Algıdaki nesnelerin tanınması için gerekli koşullar nelerdir? Kararlar nasıl alınır ve gerekçelendirilir?”<sup>27</sup>

Dennet’in ifadelerinden anlaşılacağı üzere özellikle zihin felsefesinin başat unsurlarından akıl, zeka, bilinç, algı gibi kavramların felsefi bağlamda değerlendirilmesi ve yapay zekanın bu çerçevede anlaşılması son derece önemlidir. O nedenle birinci bölümde özellikle zihin felsefesi bağlamında akıl, zeka, bilinç ve algı kavramları çözümlenerek bir yapay zeka sisteminin bu fakültelerin tam anlamıyla gerçekleştirmesinin ne kadar mümkün olabileceğine ya da olamayacağına çeşitli görüşlerle değinilmiştir. İkinci bölümde ise gücünü

<sup>26</sup> Fizikalizm terimi ve materyalizm, genellikle birbirinin yerine kullanılır. İnsan doğası anlayışını oluşturan bazı merkezi fikirleri fizikalizmde şu şekildedir: (1) İnsanlar fiziğin içinde konumlandırılan türden varlıklardır veya tamamen onlar tarafından oluşturulmuştur. Hiçbir maddi olmayan Kartezyen ruh, yaşamsal ruhlar veya yetkinlik yoktur. (2) İnsan vücudu nedensel olarak eksiksiz bir fizikokimyasal sistemdir: vücut dış nedensel etkiye oldukça duyarlı olmasına rağmen, vücuttaki tüm fiziksel olaylar ve tüm vücut hareketleri prensipte fizikokimyasal terimlerle tam olarak açıklanabilir. (3) Bir insan tarafından herhangi bir özelliğin somutlaştırılması, eninde sonunda fizikokimyasal terimlerle açıklanabilir. Detaylı bilgi için bk. Guttenplan, 1995: 471.

<sup>27</sup> Dennet, 1998: 265, 266.

algoritmadan alan yapay zekaya ilişkin iddialara, tartışmalara ve itirazlara yer verilmiştir. Üçüncü bölümde de genel anlamda makine etiğine yönelik yaklaşımların sınıflandırmasına yer verilmiştir. Son olarak tartışma bölümünde günümüzde tartışılan etik kaygılara ve risk faktörlerine yer verilerek sonuç bölümünde tezin esaslı iddiaları sıralanmıştır. Kavramsal bir değerlendirme amacıyla olan bu tezin kapsamında; yapay zekanın tarihi, felsefi yönleri de değerlendirilmiştir.



*“Yapay zeka büyük ölçüde felsefedir. Genellikle ilk bakışta tanınabilen felsefi sorularla doğrudan ilgilidir: Akıl nedir? Anlam nedir? Akıl yürütme ve rasyonellik nedir? Algıdaki nesnelerin tanınması için gerekli koşullar nelerdir? Kararlar nasıl alınır ve gerekçelendirilir?”*

D.C. Dennet

## BİRİNCİ BÖLÜM

### İNSAN ZİHNİ ve YAPAY ZEKA İLİŞKİSİ

#### 1.1 Akıl Nedir?

Zihin felsefesinin önemli konularından akıl<sup>28</sup> (us) yapay zeka meselesinde önemli bir kavram olarak karşımıza çıkmaktadır. Ayrıca akıl terimiyle sıkça benzer anlamlarda kullanılan zeka da bu bağlam için oldukça önemlidir. Yapay zeka ifadesinde geçen zekanın akıl ile benzerliği düşünüldüğünde iki terimin muğlaklığı dikkat çekicidir. Akıl teriminin incelendiği bu bölümde özellikle zekaya kıyasla aklın farkını ortaya koymak amaçlanmıştır. Böylece aklın ne olduğu ve yapay zeka sistemlerine uyarlanma olanağı tartışılacaktır. Yapay zeka sistemlerinin insan zihniyle ilişkilendirilmesi Alan Turing’in (1950) “insanlar gibi düşünen makineler”<sup>29</sup> iddiasına dayanmaktadır. Bu iddia öylesine ilgi çekicidir ki yapay zekaya ilişkin tanımların çoğu bu düzlemde şekillenmiştir. O halde insanlar gibi düşünebilen sistemlerin ne derece mümkün olduğunu anlamak için düşünceye ev sahipliği yapan zihnin bileşenlerini tanımak gerekmektedir. Öyle ki akıl kavramını anlama hedefi iddialı bir teşebbüs olacağından tarihteki tartışmalara ve görüşlere yer vermek doğru bir başlangıç noktası olacaktır. Tarihsel akışa bakıldığında akıl kavramının Eski Çağ’lardan bu yana pek çok anlamsal değişime uğradığını özellikle Aristoteles’in sınıflandırmasında görmekteyiz. Bu değişim akıl teriminin maruz kaldığı bir daralma olarak ifade edilebilir. Yani Eski Çağ’daki kullanımıyla akıl günümüzde zeka yaygınlaşmıştır. Bu hususta akıl ve zeka terimlerinin benzer kullanımlarına ilişkin olarak yapay zekanın kapasitesine ya da potansiyeline yönelik olası öngörüler ortaya atılabilir. Bu yönüyle zihinsel süreçler bağlamında hem kavramların mahiyetini belirlemek hem de yapay zekanın işlevini tespit etmek önemli hale gelmektedir. Kaldı ki yapay zekanın bileşenlerden zeka ve insan aklı en yalın haliyle çeşitli dillerin sözlüklerinden akademik

<sup>28</sup> Alm. *vernunft*, İng. *reason*, İsp. *ment*, Fr. *esprit*

<sup>29</sup> Turing, 1950, 433 – 460.

çevrelere değin önemli benzerlikler barındırmaktadır.<sup>30</sup> Bu benzerlikler kavramsal karmaşaya yol açtığından felsefi temelleri açısından çeşitli görüşlere başvurulacaktır.

Bu yönelimde ilk durağımız Aristoteles'in insanın potansiyeline ilişkin rasyonel hayvan<sup>31</sup> tespitidir. İnsan doğasının belirleyici unsuru düşünme olarak işaret edilerek bedensel olanın yanında düşünselin önemine değinilmektedir. Aristoteles'e göre beynin esaslı görevlerinden birisini vücudun en sıcak organı olan kalbin hararetini soğutmaktır. Bu bağlamda zihinsel süreçlerin konumunu kalp olarak işaret etmektedir. "İşte bunun nedeni, insanın kalbindeki sıcaklığın en saf olmasıdır. Akıllı ne kadar iyi kıvama getirebildiğini gösterir çünkü insan en bilgedir."<sup>32</sup> Öte yandan zihinsel süreçlerin konumuna ilişkin Platon'un görüşüne bakıldığında zihnin beyinde yer aldığı düşüncesi hakimdir. "Beyin dediğimiz kafanın iliği aklın yerleşim yeridir; ruhun alt formları ise omuriliğe iştirilmiştir."<sup>33</sup> Burada önemli bir nokta da o dönemlerde "beyin" kavramı kullanılmadığı için içi sıvı dolu kafatası tabiriyle beyin ima edilmektedir. O halde Platon'a göre aklın beyinde konumlandırıldığı bilgisini kabul edilebilir. Anlaşılabacağı üzere zihinsel süreçler Eski Çağ düşünürlerinde farklı konumlandırılmaktadır. Öte yandan zihinsel süreçler zihin-beden problemi kapsamında anlaşılmaya çalışıldığı gelenek öne çıkmıştır. Bu yaklaşımlar: töz ve nitelik düalizmi, materyalizm (fizikselcilik), idealizm, panpsişizm, nötr monizm, epifenomenalizm, belirimcilik, davranışçılık, işlevseleilik ve dışlayıcı materyalizm olarak sıralanmaktadır.<sup>34</sup> Bu yaklaşımların her biri hayli detaylı olduğundan yapay zeka temelinde akıllı ele almak bağlamında monist ve düalist yaklaşımlara değineceğiz. Monizm, zihinlerin ve bedenlerinin farklı maddeler olduğunu reddederken, düalizm zihinlerin ve bedenlerinin birbirinden ayrı şeyler olduğunu kabul eden görüşlerdir.<sup>35</sup> Bu bağlamda özellikle yapay zeka hususunda Kartezyen düalizmi takip eden bir gelenek göze çarpmaktadır. Aristoteles'in akıl anlayışını ise monist bakışla özdeşleştirmek mümkündür.

Aristoteles'in akla ilişkin açıklamasını bir örnek üzerinden ifade etmek gerekirse hem zihin hem de bedenin birliğini taşıyan bir çember düşünülebilir. Bu çember uçsuz bucaksız engin bir kavram olan akıllı simgelerken, çember içindeki konumlandırılmış bir nokta işareti de zeka olarak düşünülebilir. Çemberin kendisi ve kapsadığı nokta aklın yani *logos*'un kendisidir. Bu yönüyle bilgi ve bilme eylemini<sup>36</sup> zihin-beden bütünlüğünde kapsayan bir olgudur. Diğer

<sup>30</sup> Bk. Türkçe'de akıl, us (TDK Sözlüğü), İngilizce'de akıl *mind* (Websters 1986:1174, 1436), Almanca'da akıl *geist* (Ülkü 2006: 317, 415), Fransızca'da akıl (*esprit*) (Girard 1965: 316, 427), İspanyolca'da akıl (*mente*) (Uriarte vd. 2010: 788, 843)

<sup>31</sup> Langer, 1942: 58.

<sup>32</sup> Aristotle, 1908: 2088.

<sup>33</sup> Plato, 1888:268.

<sup>34</sup> Rakova, 2006: 118.

<sup>35</sup> Crane, 2000:73-85.

<sup>36</sup> Bilgi statik bir olgu olarak kabul edildiğinde (*know-what*), bilme eylemi deneyimsel (*know-how*) bir tür beceridir. Dreyfus açısından bu konu ilerleyen bölümlerde etraflıca değerlendirilecektir.

bir deyişle tarihselliği olan, deneyimsel olguları barındıran akıl düşünsel olana da ev sahipliği yapmaktadır. Bu yönüyle Aristoteles'te ruh ve akıl sözcükleri yakın anlamlarda hatta birbirlerinin yerine kullanılmaktadır. Ruhun (*psyche*) akla ait fakültelerin tümünü barındıran bir çember ya da taşıyıcı olduğu düşüncesini de ifade etmek mümkündür. Ruha ait olan akıl da kendi içerisinde çeşitli kısımlara ayrılmaktadır. Her bir kısım bilgiye ulaşma uğrunda farklı görevleri yerine getirir. Aristoteles akli oluşturan unsurları yani aklın doğru bilgiye ulaşma yöntemlerini beş yolla sıralamıştır; sanat (*techne*), bilimsel bilgi (*episteme*), pratik bilgelik (*phronesis*), felsefi bilgelik (*sophia*) ve sezgisel akıl (*nous*).<sup>37</sup> Söz konusu fakülteler yani akli meydana getiren alanların her biri bilgiye ulaşma uğrunda düşünsel faaliyet için farklı görevlere sahiptir. Düşünceyi (*logos*) gerçekleştirmede Aristoteles'in zihin-beden bütünlüğüne olan inancı önemli yer teşkil etmektedir. Bu nedenle fakültelerin her birini görevlerini, diğer fakültelerle ilişkisi bağlamında irdelemek gerekmektedir.

### 1.1.1 Sezgisel Akıl (*Nous*)

Sezgisel akıl (*nous*) Aristoteles'in etkin ve edilgin akıl başlığı altında açıklanmıştır.<sup>38</sup> Sezgisel akli etkin (faal) ve edilgin (pasif) olarak ikiye ayıran Aristoteles için salt akılsal nitelikler sadece etkin olan bölüme özgüdür. Edilgin akıl yalnızca bedenin duyular vasıtasıyla elde ettiklerini işlemektedir. Diğer bir deyişle edilgin bölüm kendine sağlanan bir veri olmaksızın herhangi bir işlem yapamaz. Bu yönüyle düşünüldüğünde her insanın farklı bedeni olduğu gibi farklı edilgin akıllar da bulunmaktadır. Çeşitli edilgin akıllar bedene bağlı olarak ortaya çıkmakta ve yok olmaktadır. Bazı yorumlarda etkin akli ruh dışında ilahî bir akıl olarak düşünme eğilimi görülse de Ross'un (1995) yorumuna göre asıl ve ilk var olan edilgin akıldır. Etkin akıl, yalnızca edilgin akli ortaya çıkaran bir kıvılcım olarak ifade edilmektedir. Aristoteles'in *Metafiziki*'nde de belirttiği gibi etkin akıl insan ruhuna içkin ilahî bir akıldır. "Bu, faal akli bireysel insani varlığın tümüyle dışında bulunan ilahî bir akla özdeş kılmaya çalışan her türlü yorum imkanını ortadan kaldırır ama faal akılda insan ruhuna içkin bir tanrısal akli gören yorumu ortadan kaldırmaz."<sup>39</sup> Etkin akli ruhun dışında Tanrısal bir akılla özdeşleştirme düşüncesi bu ifadeyle boşa çıkmaktadır. Ancak insan ruhuna ait Tanrısal bir aklın varlığını etkin akıl bağlamında savunmak Aristoteles'e göre doğru bir yaklaşımdır. Çünkü faal akıl, Tanrı'nın mahiyetini belirlemede araç olarak savunulsa da tek başına bir Tanrı düşüncesiyle de eş değer tutulamaz. Konunun detaylarını aşağıdaki açıklamadan okumak meseleyi belirginleştirecektir.

<sup>37</sup> Aristotle, 1893: 184.

<sup>38</sup> Ross, 1995: 93.

<sup>39</sup> Ross 2011: 235.

“Bilginin olgusal gerçekliğini anlamak istiyorsak pasif akıl ve nesne arasındaki ilişkinin yanı sıra, etkin akıl bir aracı olmamasına rağmen yine de üçüncü bir şeydir. Görme yetisinin olgusal gerçeğini anlayabilmek istiyorsak, göz ve nesne arasındaki ilişkide olduğu gibi ışık da üçüncü bir şey olarak ele alınmalıdır. Işık, bir aydınlatıcı olarak aslında şeffaf hale getirilmiş bir ortamın aracı koşuludur ve aslında görebileceğiniz gözün ve görünür nesnenin görülmesini mümkün kılan şey ışığın gerçekliğidir ya da aktüalitesidir.”<sup>40</sup>

Buradan anlaşılacağı üzere duyuların ve nesnelerin etkileşiminde önemli rolü olan üçüncü bir şey de ışık olarak ifade edilmektedir. *Nous* etkin ve edilgin olmak üzere iki bölümden oluştuğu belirtilmektedir. Bu açıklamalardan anlaşılacağı üzere edilgin bölüme herhangi bir veri sağlanmaksızın işlevsiz kalacağı sonucu çıkarılmaktadır. O halde yapay zekanın büyük veri yığınları olmaksızın yani herhangi bir veriyle beslenmeksizin işlevsiz kalacağı bağdaşımı kurulabilir. Bu anlamda yapay zekayı insan aklının bölümleriyle eşleştirmek gerekirse *nous*’un edilgin bölümü uygun düşecektir. Kaldı ki her ikisi için de veri ön koşuldur. Öte yandan insan aklının yapay zeka ile karşılaştırılmasında kurulabilecek bir denklik var ise bu yalnızca *nous*’un edilgin bölümüyle mümkündür. Aksi halde diğer bölümler ve geniş anlamda akıl ile yapay zekayı bağdaştırmak ya da aklın enginliğini algoritmanın sentaks düzlemine indirgemek kasıtlı ve yanlış bir tutum olacaktır. En azından felsefi söylem açısından kabul edilemez.

### 1.1.2 Pratik Bilgelik (*Phronesis*)

Bilişsellüğün yanında deneyimselliğin önemine yapılan vurgu sayısal sistemler açısından değerlendirildiğinde hesaplama kapasitesinin ötesinde algoritmanın çeşitli yetilere gereksinim duyacağı sonucu çıkarılabilir. Bu hususta algoritmanın zihinsel süreçlerden *phronesis*’e ilişkin nitelikleri yapıp yapamayacağını anlayabilmek amacıyla öncelikle pratik bilgeliği anlamak gerekmektedir. Pratik bilgelik konusunda Aristoteles için makul olan az ya da çok değil ikisinin ortasıdır. Ona göre orta olan şey de doğru akıldır. Aklın esaslı görevi iyiyi, güzeli ve doğruyu yakalama uğrunda aşırı ve yetersiz olanın arasında bulunan ortayı tutturmaktır. İnsan eylemlerinde daima ölçülü olanı yakalamayı amaçlayan *phronesis*, *nous*’tan farklı olarak deneyimi ve tecrübeyi yönetir. Yani yaşantılar ve deneyimler bir insan için akıl sahibi olmayı, akliselim olmayı mümkün kılar. Aristoteles’in şu ifadesinden yola çıkarak bu kanaate varmak mümkün olabilir.

“Dahası bir insanın kendi sorunlarıyla başa çıkabilmesi, inişli çıkışlı bir yoldur ve dikkat gerektirir. Bu durum, bir gencin geometride ya da matematikte ustalaşabileceği ve bu konularda bilge olabileceği ama onun ihtiyatlı ya da sağduyulu olamayacağı gerçeğinden anlaşılmaktadır. Bunun sebebi, bazı gerçekleri

<sup>40</sup> Ross, 1995: 156.

bize sadece deneyim tanıtabilir ve nihayetinde insan ihtiyatlı veya sağduyulu olabilir; ama genç birisi yılların meyvelerini tatmadığı için deneyimsiz olmalı.”<sup>41</sup>

Aristoteles’in deneyime dayalı ihtiyatlı, sağduyulu bilgi anlayışı makine ve insan arasındaki ayrımın belirleyicisidir. Bu bağlamda Turing başta olmak üzere çoğu araştırmacının algoritmaya olan inancı söz konusu niteliklerin makinelere yüklenebileceği yönündedir. Bu inançtan hareketle tarihte pek çok çalışma yapıldığı gibi günümüzde de özellikle sinirbilim üzerine çalışma yapılmaktadır. Tarihteki örneklerine bakıldığında kusursuz algoritma kümelerinin ortaya çıkaracağı gücün insana özgü pek çok şeyi insandan ayırt edilemez şekilde ortaya koyacağı iddiası özünde Turing’in “Taklit Oyunu” adını verdiği argümandır. Bu argüman zihnin işleyişini algoritmanın işleyişiyle bağdaştırdığı için felsefi dayanaktan yoksun olup, pek çok ayrıntıyı gözden kaçırmaktadır. O nedenle özellikle *phronesis* ve *nous* fakülteleri incelenmiştir. Öte yandan sanat (*techne*), bilimsel bilgi (*episteme*) ve felsefi bilgelik (*sophia*) gibi fakültelerin de tartışılması gerekmektedir.

Deneyimin gücüne Aristoteles’in yaptığı atıf aklın önemli bir yönünü işaret etmektedir. Aristoteles bunu genç birisinin rasyonel eylemler sonucunda zeki olarak kabul edilebileceğiyle örneklendirmektedir. Oysa akliselim olmayı yılların meyvesini tatmış, deneyim sahibi kişilere özgü bir olgu olarak öne sürmektedir. Akliselim ifadesiyle yorumlanan *phronesis* kavramı, akıl olgusunun deneyimden bağımsız düşünülemeyeceğini ifade etmektedir. Aristoteles en öz ifadeyle aklıbaşındalık (*phronesis*) terimini şöyle tanımlamaktadır, “Aklıbaşındalık, iyi ya da kötü şeyler için akılla beraber var olan ve eylemlerde doğru davranmamızı sağlayan bir huydur.”<sup>42</sup> Bu türden bir ifade insan aklının fakültesi olan *phronesis* terimini açıkladığı düşünülmürse yapay zekanın salt hesaplama ibaret olması ne *logos* ne de *phronesis* ile karşılaştırılamayacağını ima eder.

Öte yandan benzer bir çıkarım Dreyfus’un (1999) sağduyu terimiyle bağdaştırılabilir. Doğrudan yapay zekanın sınırlılıklarını işaret eden bu yaklaşım Aristotelesçi bakış açısının bilim dünyasındaki karşılığı olarak düşünülebilir. Özü itibarıyla sağduyu ile Dreyfus aklın enginliği, sınırsızlığı ve gücü için tarihselliğin yani deneyimin önemine dikkat çekmektedir. Hatırlanacağı üzere *phronesis*’in görevinden söz ederken bu niteliklere göndermede bulunmuştuk. Buraya kadar zihnin bileşenlerine değinmişken zihne ait süreçlerin yapay zekaya yansımalarından söz etmek gerekir. Makine öğrenmesinin sayısal sistemlerde halihazırda kullanıldığı ve bunun da öğrenmeyi gerçekleştirdiği iddiası öne sürülebilir. Ne var ki Dreyfus’un ifade ettiği sağduyu bilgisinin “öğrenme” ve “edinme” kavramlarının bağlamında

<sup>41</sup> Aristotle, 1893: 194.

<sup>42</sup> Aristotle, 1893: 194.

açıklanması gerekmektedir. Bu konu ikinci bölümde sağduyu tartışması başlığı altında etraflıca incelendiği için şimdilik kısaca değinmek yeterlidir.

### 1.1.3 Sanat (*Techne*)

Sanat fakültesi (*techne*) birbirinden farklı olan yapma ve yaratma (*to do / to make*) sözcüklerinden ileri gelmektedir. Bir şeyin yapılması ya da yaratılması ilk halinden farklı bir durumda gelmesiyle ilgilidir. Bu sebeple ortaya çıkan her şey bir sanat olamaz. Yaratmanın, mimarlık gibi akılla ilgili unsurları vardır. Yaratmayla ilgili olan şeyler zorunlu olarak sanattır. Bunun koşulu da yaratan insandır. Sanat yalnızca doğru akılla eş zamanlı bir yaratma eylemidir. Bu yönüyle bir yapma değildir.<sup>43</sup> Ross'un ifadesiyle sanat, doğru bir kural aracılığıyla şeyleri meydana getirme ya da yaratma eğilimidir.<sup>44</sup> Burada sözü edilen doğru kural aklın fakültelerinden *techne* aracılığıyla akla (*logos*) uygun bir yaratma biçimidir. Bu düşünce aklın bir uzantısı olarak sanatı doğrudan vurgular. Bu bağlamda *techne* yani teknik derin anlamlar taşımaktadır. Bu hususta *logos*'un *techne* ile olan ilişkisine değinmek önemlidir.

Grekçe *techne* terimi bir şeyin gizini açma ya da bir şeyi öne çıkarma anlamına gelmektedir. Yunanca *techne* ve *logos* terimlerinin birleşiminden günümüzde kullanılan "teknoloji" kavramı meydana gelmiştir. *Techne*; sanat, beceri, zanaat veya bir şeyin kazanıldığı yol, tarz veya araçlar anlamına gelir. *Logos* ise; kelime, içsel düşüncenin ifade edildiği söz veya ifade anlamına gelir. Yani kelimenin tam anlamıyla teknoloji, şeylerin kazanılma şekliyle ilgili sözler veya söylemler anlamına gelir. Ayrı terimler olan *techne* ve *logos* Ksenofanes, Platon, Aristoteles ve Stoacıların klasik retorik metinlerinde görülmektedir. Aristoteles *techne* ve *logos* arasında zayıf bir ilişki kurmasına rağmen iki kelime bir istisna dışında hiçbir klasik metinde birleşik bir terim olarak bir arada görülmemektedir. Bu istisna da Cicero'nun Atticus'a Mektupları teknolojici (*technologist*) terimini kullanmasıdır. Cicero'nun, Yunanca *techne* ve *logos* terimlerini birleştirmekle ne demek istediğinin tam olarak açık olmakla birlikte bugünkü "teknik" anlayışına yakın görünmektedir.<sup>45</sup>

Aristoteles için *techne* bir yaratma yani bir şeyin ilk halinden farklı bir duruma gelmesi, doğru kural aracılığıyla şeyleri meydana getirme ya da yaratma eğilimi olarak ifade edilmişti. Yakın tarihin varoluşçu düşünürlerinden Merleau-Ponty'nin (2005) "perspektif"<sup>46</sup> kavramı bu görüşe katkı sağlayabilir. Merleau-Ponty, çizdiği kavramsal çerçeveye *techne* ve *logos* terimlerinin gerçek anlamları dışında kullanımını perspektif kavramıyla açıklamıştır. Merleau-

<sup>43</sup> Aristotle, 1893: 186, 187.

<sup>44</sup> Aristotle, 1893: 225.

<sup>45</sup> Tulley, 2008: 93-94.

<sup>46</sup> Ponty, 2005: 23.

Ponty’ e göre şeylerin hakikati ile onların perspektifleri yani insan algısının süzgecinden geçmiş hali, şeylerin özünden koparılması anlamına gelmektedir. Deyim yerindeyse algı filtresinden geçirilen şeylerin gerçekliğini kaybettiği düşüncesi vurgulanmaktadır. Bu nedenle sanat en yalın haliyle geometrik sembollere ya da hesaplamalara indirgenecek bir olgu değildir. Yani sanatın aklın diğer fakülteleriyle iç içe geçmiş bir boyutu vurgulanmaktadır.

Öte yandan insanın makineye makinenin de insana benzetilmesinin tarihteki örneklerine ek olarak Hobbes (2013), Tanrı’nın sanatını doğayla temsil ettiğinin, otomatların da insan vücudunu temsil ettiğinin bilgisini vermektedir. Bu bağlamda pek çok organla sistemin parçası arasında ilişki kurulurken bu sistemi faaliyete geçiren olgunun göz ardı edilmesi dikkat çekicidir.

“Tanrı’nın dünyayı onunla yaratmış olduğu ve yönettiği sanat olan doğa, başka pek çok şeyde olduğu gibi bunda da yapay bir canlı yaratacak şekilde, insanın sanatı ile taklit edilir. Çünkü hayat, organların, başlangıcı içerdeki bir temel parçada bulunan, hareketinden başka bir şey değildir; bütün otomatların (yaylar ve çarklar yardımıyla kendi kendine hareket makinaların, mesela bir saat) yapay bir hayata sahip olduklarını söyleyemez miyiz? *Kalp* nedir ki *yaydan* başka; *sinirler* nedir ki çok sayıda *yaylardan* başka; ya eklemler, yapıcının planlandığı şekilde bütün gövdeyi harekete geçiren çok sayıda *çarklardan* başka?”<sup>47</sup>

Farklı bakış açılarından anlaşılabileceği üzere *techne* doğası gereği çok boyutlu bir kavram olup ontolojisinden ve semantiğinden bağımsız düşünülmemelidir. Semantiğin önemini sanatla ilişkilendiren bir örnek sunmak gerekirse son zamanlarda hayli popüler olan insansız hava araçlarının sürü zekası sayesinde topluca iletişim kurmalarına bağlı olarak sergiledikleri görsel şölenler düşünülebilir. Söz gelimi binlerce insansız hava aracının (*drone*) eş zamanlı hareketiyle gökyüzünde sergilenen büyüleyici bir görsel gösteri karşısında yaşanan hayranlık, beklenen bir durumdur. Ancak binlerce hava aracının koordine bir şekilde oluşturdukları görselin ontolojisine sahip olduğunu söylemek mümkün değildir. Yani hava araçları insanlar için çok anlamlı olabilecek bir gösteri ortaya koysalar dahi bu hava araçlardan hiçbir gösterinin semantiğine ilişkin bilgi sahibi olamayacaktır. Aristoteles’in sanata ilişkin ifadeleri düşünüldüğünde bu türden bir eylemin sanat kabul edilmeyeceği öne sürülebilir. Bu bağlamda hava araçlarının eylemleri bir yaratmadan ziyade, insan yordamıyla yüklenen verileri işleme ya da yapma olarak yorumlanabilir.

<sup>47</sup> Hobbes, 2013:17.

#### 1.1.4 Bilimsel Bilgi (*Episteme*)

Bilimsel bilginin (*episteme*) nesnesi zorunlu ve sonsuz şeylerle ilgilidir. Yani bilim için bilgi zorunlu koşuldur. Çünkü doğası gereği zorunlu olan şeyler aynı zamanda ebedidir. Bu nedenle ebedi olanın başlangıcı ve sonu olmaz. Ayrıca tüm bilimlerin öğretilebileceği ve bilim yoluyla gelecek nesillere aktarılması *episteme* ile mümkündür. Genelin bilgisine ulaşmanın yöntemi tümevarım, özelin bilgisine ulaşmanın yöntemi de tümdengelimdir. Bu yönüyle bütün öğretimler bilinen bir şeyden başlar çünkü ya tümevarım ya da kıyas yoluyla ilerlemektedir. Öğrenciyi evrensel ilkelere götüren tümevarımdır, kıyas bunlardan yola çıkar. Bu durumda kıyasın başladığı, kıyas yoluyla ulaşılmayan ve dolayısıyla tümevarım yoluyla ulaşılması gereken ilkeler vardır. O halde bilim, Kategoriler’de sıralanan diğer tüm niteliklerle birlikte bir alışkanlık ya da oluşturulmuş bir kanıtlama yetisi olarak tanımlanabilir. Özü itibarıyla bilimsel bilgi ispatlara dayanmaktadır. Çünkü bilginin ilkeleri kabul edildiğinde ve bizim tarafımızdan belirli bir şekilde bilindiğinde bilimsel bilgiye sahip olduğumuz söylenebilir çünkü bu ilkeler bizim için onlara dayanan sonuçlardan daha iyi bilinmedikçe, bilginiz yalnızca tesadüfi olacaktır. O halde bu bizim bilim anlayışımız olarak düşünülebilir. Bilimsel bilgi üzerine Aristoteles’in ifadeleri ana hatlarıyla bu cümlelerden ibarettir.<sup>48</sup> Kısaca özetlemek gerekirse bilimsel bilgi, deneysel bir etkinlik olarak bir şeyi ispatlayabilme yeteneğini olarak tanımlanabilir. Bu bağlamda bilimin bilgisini amaçlamaktadır. Bilimsel bilgiye ulaşmanın yolları birtakım ilkeler ve nedenler ağına bağlıdır. Bunlar bilindiği zaman bilimsel bilgiye ulaşmak da mümkündür.

#### 1.1.5 Felsefi Bilgelik (*Sophia*)

Felsefi bilgelik (*sophia*), sezgi ve bilim bileşkesinin en yüksek nesnelerde yoğunlaşması olarak anlaşılabilir. En yüksek gök cisimlerinin konumu ne ise felsefi bilgeliğin konumu da pratik bilgeliğe kıyasla odur. Felsefi bilgelik matematik, metafizik ve doğa bilimlerini kapsamaktadır. Aristoteles’e göre bu konular özellikle insanın amaçladığı en iyi yaşamın bir göstergesi niteliğindedir.<sup>49</sup> Felsefi bilgeliğin yapay zeka ile ilişkisi bağlamında Charniak ve McDermott (1985) tarafından sunulan tanım *sophia* ile ilişkilendirilebilir. “Bilişim modellerinin kullanılmasıyla zihinsel fakültelerin incelenmesidir. Yapay zeka, bilişim modellerin kullanımıyla zihinsel yetkinliklerin araştırılmasıdır.”<sup>50</sup> Bu tanım ileriki bölümlerde de tartışacağımız üzere yapay zekanın önemli kullanım amaçlarından birini müjdelemektedir. Özellikle çağımızda yapay zeka teknolojisinin salt mekanik bir süreç olduğu ve beşeri

<sup>48</sup> Aristotle, 1893: 184-185.

<sup>49</sup> Ross, 2011: 339.

<sup>50</sup> Charniak ve McDermott, 1985: 6.

bilimlerin bu alanla ilgilenmesini gerektirecek bir şeyler olmadığı düşüncesi yaygındır. Bu hususta Charniak ve McDermott'un ifade etmeye çalıştığı şey; felsefi etkinliğin insana özgü bir eylem olması yanında yapay zekanın bu türden bir düşünme yetisini gerçekleştirmesiyle daha güçlü önermelerde ve argümanlarda bulunabileceği inancıdır. Yani felsefi problemleri çözmek amacıyla sayısal sistemlerin kullanılabileceği düşüncesi vurgulanmaktadır. Fark edileceği üzere zihnin üst düzey fakültelerinden *sophia*'nın robotik bir zeka tarafından gerçekleştirilebileceğine dair görüşler bildirilmektedir. Ancak buraya kadar sayısal sistemlerin *sophia* ya da benzer bir yetiyi gerçekleştirmesi olasılığı en azından Aristotelesçi anlayışa göre mümkün görünmemektedir. Zira yapay zekanın hız ve güvenilirlik kapsamında gücü tartışılmazken, semantik bir anlamlandırma ya da felsefi öz düşünümü gerçekleştirmesi Searle açısından reddedilmektedir. İlerleyen bölümlerde bu konuya detaylarıyla değinilmiştir.

### 1.1.6 Akıl (*Logos*)

Bilgiye ulaşma yollarından *nous*, *phronesis techne*, *episteme* ve *sophia* fakültelerinin akıl olgusunu meydana getirme sürecindeki görevlerini açıkladık. Yukarıda açıklamalarına yer verdiğimiz fakültelerin bileşkesi nihayetinde *logos*'u meydana getiren bir olgudur. Bu çerçevede her bir kavram yapay zekanın kapasitesini ya da sınırlarını anlamada önemlidir. İkinci bölümde sayısal sistemlerin zihnin işleyişini gerçekleştirebileceği hatta aşabileceğine ilişkin iddialara değinilmiştir. Bu nedenle tüm fakültelerin bileşkesi olan biricik ve masif akıl (*logos*) tartışmak önemlidir. Yunanca isim kökünden gelen *logos*, *lego* (*söylüyorum*) fiilinden türemiştir. Başlıca; “sözcük, konuşma, argüman, açıklama, öğreti, itibar, sayısal hesaplama, ölçü, oran, ifade, prensip ve akıl” gibi sözcükler *logos*'un sözlük anlamlarıdır. Felsefi açıdan *logos* Herakleitos'a göre; evren hakkındaki düşüncelerimiz, evrenin kendi rasyonel yapısı ve bu rasyonel yapının kaynağıdır. Onun için *logos* bir akıl türüyken, tek başına bir düşünme etkinliği değildir. Anaksagoras ise evrenin akıl prensibinin *logos*'tan ziyade *nous* olduğunu ve öteki şeylerden özerk bir halde bulunduğunu kabul etmektedir. Platonik evren anlayışında aklın rasyonel ilkelerle organize edildiği fark edilmektedir. Bu organizasyon ya da düzen, Sezgisel akıl (*nous*) ile şekillenmektedir. Ancak insan ruhu söz konusu olduğunda Platon özellikle *Devlet* eserinde *nous* ve *logos*'u çok yakın anlamlarda kullanmıştır. Stoacı düşünce sisteminde *logos* doktrini en geniş uzantısına erişmiştir. Bu yönüyle *logos* Stoacılar için evrendeki bütün aklın ilkesidir. Öyle ki *logos* bütün etkinliklerin kaynağı olan Tanrı'yla tanımlanmıştır. Bu düşünceye göre *logos*, diğer maddelerin içine işleyen bir başka maddedir. Öz bir ifadeyle *logos*; bir şarap damlasının, bütün denizin her yerine yayılarak nüfuz etmesidir.<sup>51</sup> Anlaşılacağı üzere

<sup>51</sup> Kerferd, 1967: 83.

*logos* diğer tüm fakültelerin bileşkesiyle meydana gelmektedir. Bu fakülteleri diyalektik anlamda hem zihinsel hem bedensel temelde karşılaştırmalı olarak işleyebilmek, yapay zekanın akıl karşısında zafiyetini ortaya koyan bir detay olarak değerlendirilebilir. Yapay zekanın sınırlılığını anlamak açısından aklın keşfi bu nedenle çok önemlidir. Bu önem yapay zeka ve aklın yarışını gereksiz kılabilen bir unsur olarak düşünülebilir.

Öte yandan akıl terimine ilişkin Platon'un düşüncelerine bakıldığında *logos* sözcüğü akıl ve akla ait olan anlamını sürdürür. Öyle ki *logos*, Platon için akıl manasının yanı sıra "sözcük", "dil" ve "tanım" anlamlarına da gelmektedir.

"Platon, Theaetetos'da bilginin 'hesabı verilmiş' (Yunancada *logos*) doğru inanç olduğu görüşünü tartışmıştı; Menon'da ise bilginin açıklayıcı bir akıl yürütmeyi (*logismos*) içerdiğinin ifade edildiğini görüyoruz. (Yukarıdaki her iki metinde de geçen "*logos*" kökü Eski Yunancada çok geniş bir anlama sahiptir; bir yandan "sözcük", "dil" ve "tanım" gibi kavramları ifade ederken, öte yandan da "düşünce", "akıl" ve "akılsallık" kavramlarını çağrıştırmaktadır.)"<sup>52</sup>

*Logos*'un tümelin bilgisini barındırdığına ilişkin bir farklı bir yaklaşım da Stoa felsefesinde fark edilmektedir. "Stoacı evren, bir yasa tarafından yani içkin bir *logos* tarafından belirlenen bir dünyadır. Bu, onların felsefesinin her üç yönünden geçer. Sonuçta, bunlar sadece yönlerden ibarettir, yani; başka çıkar yol olmadığında birliği – doğa, evren ya da Tanrı – olan bir şeyi sunmanın yollarıdır."<sup>53</sup> *Logos*'un ilahî bir düzlemde değerlendirilmesi o dönemin Tanrı'yı açıklama belirsizliğinde gerek Herakleitos gerekse Stoacılar tarafından başvuru bir yol olarak düşünülebilir. Ancak *logos*'un Tanrı'nın kendisini temsil etmesinin ötesinde masif bir akıl olgusunu sağlama hususunda kapsayıcı bir kavram olduğu düşüncesi ağır basmaktadır.

*Logos* Aristoteles'te değerlendirildiğinde insanın amaçsal ya da araçsal işlevleriyle salt usamlamanın ötesinde tarihsellik, deneyim, ruh gibi alanlara nüfuz ettiğini Aygün (2017) şöyle yorumlamaktadır.

"Şu an için hem dünya hem de insan ruhu bir şekilde *logos* ile donatılmış gibi görünür ama hiçbir tepeden tırnağa donatılmış değildir. Yine de insan ruhunun analizi, *alogos* kısmı ile *logos* olan kısım arasındaki ayrımın ötesinde bir ayrım daha gerektirir. Örneğin beslenme ve büyüme sadece *alogos*'a özgüdür. Ancak bu *alogos* bölümünün yanı sıra *logos*'a sahip olan insan ruhunun başka bir parçası daha var. Bu üçüncü bölüm, ikisi arasında bir şekilde aracılık etmektedir: "*alogos* iken, bir şekilde *logos*'a katılır"<sup>54</sup>

Başka bir yorum da Wedin'in *logos*'u dil olarak ifade etmesidir: Bu hususta *Corpus*'taki çoğu pasajın *logos*'u dil ile bağdaştırdığını hatırlatarak *De Partibus Animalium II. 17, 660a2-3* eserinde *logos*'un ses yoluyla harflerin birleşiminden oluştuğunu belirtmektedir. *De*

<sup>52</sup> Cottingham, 2003:22.

<sup>53</sup> Long, 1986: 144.

<sup>54</sup> Aygün, 2017: 116,117.

*Generazione Animalium* V.6, 786b19-22 eserinde doğanın, insanı en çok ses kapasitesiyle donattığını çünkü hayvanlar içinde yalnızca onun *logos*'u kullandığını bildirmektedir. Yine, *Politica* I.2, 1253a9-15 eserinde bu bağlantıyı yeniden doğrulamaktadır. Doğanın insana zevk ve acıdan daha fazlasını bahsettiğini ifade ederek sese sahip olduğunu ima eder çünkü imkanlarını ve imkanı olmadıklarını ortaya koyabilen *logos*'a sahip varlıklar yalnızca insan olarak işaret edilmektedir.<sup>55</sup> Aklın yani *logos*'un mantıksal, kuşatıcı, sonsuz, masif ve ilahî yönlerine dikkat çekilmesi kavramın önemini ve derinliğini gözler önüne sermektedir. Giriş bölümünde ön bilgisini verdiğimiz üzere zihin felsefesinin tartışma alanları yapay zekanın mekanik boyutunu aşmaktadır. Bu detay aklın derinliğini ortaya koyarken zeka kavramının sınırlarını da belirlemektedir. Yani bir anlam çıkarılması gerektiğinde köklü bir terim olan aklın enginliği, yeni bir terim olan zekanın dar açısını ima etmektedir.

Sonuç olarak akıl bir üst terim olarak Aristotelesçi *logos* altında yer alan sanat (*techne*), bilimsel bilgi (*episteme*), pratik bilgelik (*phronesis*), felsefi bilgelik (*sophia*) ve sezgisel akıl (*nous*) fakültelerine ev sahipliği yapmaktadır. Bu yönüyle alt fakültelerin bileşkesi *logos*'u meydana getirirken *logos* da varlığını bu fakültelere borçludur. Tersinden bakıldığında *logos*'un yapay zekayı anlamaya katkısı zekanın sınırlarını belirlemeyle ilgilidir. Bu yönüyle çeşitli sözlüklerde zeka teriminin en çok ilişkilendirildiği ve karıştırıldığı terimin akıl olduğu bilinmektedir.<sup>56</sup> Söz konusu muğlaklığı “zeka” terimi özelinde tartışmak amacıyla zekanın felsefi, psikolojik, sosyolojik ve diğer anlamlarda neyi ima ettiğinin incelenmesi gerekmektedir. Dolayısıyla amaç aklın enginliği karşısında zekanın kısıtlılığına dayanak oluşturabilecek farklı bakışlara dikkat çekmektir.

## 1.2 Zeka Nedir?

Aklın alt bileşenlerine yönelik incelemede bilgiye ulaşma yöntemlerine değinilmiştir. Bu yöntemlerden öne çıkan sezgisel akıl (*nous*) felsefi anlamda zekaya denk düşen bir ifadedir. Yapay zeka ifadesinde yer alan zeka sözcüğünün dolaylı karşılığı *nous* ise yapay zekanın geleceğine ilişkin pek çok şeyi zekanın tarihçesine bakarak ifade etmek mümkün olabilecektir. Kaldı ki zekanın kökeni ve tarihsel süreçleri üzerine yapılacak değerlendirmeye zekanın ne olduğu ya da ne olmadığı sorusu isabetli bir hareket noktası olacaktır. Yapay ve zeka sözcüklerini tarihte ilk defa birleştiren John McCarthy'nin “Zeki makineler, özellikle zeki bilgisayar programları yapma bilimi ve mühendisliği.”<sup>57</sup> tanımına bakıldığında yapay zekanın

<sup>55</sup> Wedin 1988: 146

<sup>56</sup> Kaynak için Türk Dil Kurumu sözlüğü başta olmak üzere; Almanca, İngilizce, Fransızca, İspanyolca ve İtalyanca sözlüklerde her iki kavramın karşılaştırılması yapılabilir. Her iki kavramı açıklayan ifadelerin benzer olduğu fark edilecektir.

<sup>57</sup> McCarthy, 2007: 2.

akıldan öte dar bir zeka türünden ibaret olduğu fikri belirlemektedir. Öte yandan *logos*'a dair bir yetiye sahip olmaması, yapay zekayı zorunlu olarak sınırlandırmaktadır. Mühendislik çalışmaları bu bağlamda sayısal sistemleri zeki kılabilen niteliklerin araştırılma ve geliştirilme alanı olarak ön plana alınmıştır. Diğer bir deyişle kavramsal analiz gölgelendiği salt mekanik bir anlayışın egemen olduğu girişimlerin olduğu sonucu çıkarılabilir. Kaldı ki zekanın niteliği zihin felsefesinin yanı sıra pek çok disiplin tarafından da ele alınan bir konudur. Bu bağlamda Copin, “Yüzyıllardır biyologları, psikologları ve filozofları şaşırtan, iyi tanımlanmış bir cevabı olmayan karmaşık bir sorudur bu.”<sup>58</sup> yorumuyla kavramsal muğlaklığa doğrudan parmak basmıştır. Kavramın analizine giden yolda sözlük anlamına bakıldığında “insanın düşünme, akıl yürütme, objektif gerçekleri algılama, yargılama ve sonuç çıkarma yeteneklerinin tamamı, anlık, dirayet, zeyreklik, feraset”<sup>59</sup> sözcükleri, zeka<sup>60</sup> kavramını tanımlayan ifadelerdir. Bir önceki bölümde akıl ve zekanın sözlük anlamlarına bakıldığında pek çok ortak ifadeler barındırdığına dikkat çekmiştik. Kaldı ki yapay zeka ifadesinden de çıkarılacağı üzere sayısal bir sisteme yazılım yoluyla yüklenen algoritmanın kapasitesi ya da sınırları zeka kavramının sınırlarından ibaret olacaktır. Bu doğrultuda terimi tarihsel ve kavramsal gelişimleriyle anlamak çok disiplinli bir yaklaşımı gerektirecektir. Bu çerçevede felsefe, etik, sosyoloji ve psikoloji gibi alanlarda zekanın nasıl anlaşıldığı değerlendirilmelidir. Bu yaklaşım yapay zekanın gerek kavramsal temelleri gerekse geleceğe ilişkin sonuçları hakkında öngörülerde bulunmaya katkı sağlayacaktır.

### 1.2.1 Zekanın Felsefi Temelleri

Zekanın burada geçen tanımına bakıldığında ortak bir anlayış her iki terimde de varlığını sürdürmektedir. Ne var ki zihin felsefesinde doğrudan zeka anlamına gelen bir ifadeye ancak Descartes döneminde rastlanmaktadır. Descartes döneminde yükselen zeka sözcüğünün Eski Çağ'daki karşılığının *nous* olabileceğine dikkat çekmiştik. Yine de bu denkleğin bir eşitlik olmadığı unutulmamalıdır. Kaldı ki Descartes öncesi dönemlerde doğrudan zeka sözcüğüne benzeyen ya da o anlamda kullanılan herhangi bir sözcük görülmemektedir. Eski Yunan'dan bu yana insan zihnini temsil etmek için akıl teriminin kullanıldığına dikkat çekmiştik. Ayrıca

<sup>58</sup> Copin, 2004: 4.

<sup>59</sup> Erişim Tarihi: 10.02.2018 <http://www.tdk.gov.tr>

<sup>60</sup> İngilizce'de zeka (*intelligence*); anlama fakültesi, bilme veya kavrama kapasitesi, bir kişinin yeni ortama uyum sağlama, problem çözebilme, öğrenme, sembolleri ya da ilişkileri kullanabilme gibi niteliklerin zeka testleriyle ya da diğer sosyal kriterlerle ölçülmüş mevcut becerileri. (Websters 1986:1174) Fransızca'da zeka (*intelligence*); zeka, akıl; açıkça anlama veya anlama, beceri, yetenek; karşılıklı anlayış ilişkisi, yazışma, bilgi aracı. (Girard 1965: 427) Almanca'da zeka (*intelli'genz*); zeka, akıl. İnsanın, duyguları bilme ve akılla kontrol etme yeteneği. (Ülkü 2006: 415) İspanyolca'da zeka (*inteligencia*); anlaşılma, kavranma, anlama kapasitesi, kavrama yeteneği zeka. (Uriarte vd. 2010: 843)

*logos* tarihsel olan her ne varsa doğrudan onlara nüfuz etmiştir. Bu da demektir ki tarihteki düşünsel ve uygulamalı eylemler *logos*'a çok şey borçludur. *Logos* ve *nous* ilişkisi düşünüldüğünde *logos*'un *nous*'a ev sahipliği gerçeği karşısında çağımızda zekanın akıl karşısında yüceltilmesi örneği aklın zekaya indirgendiğinin resmidir. Kaldı ki yer verdiğimiz tanımlara bakıldığında yapay zekanın 1970'lerde bilişin ve bilgisayarların yaptıklarının bir şekilde aynı olduğu ve zihniyet ile yazılımın aynı alanda yer aldığı fikrine atıfta bulunulmuştur. İnsan benzeri bir bilgisayar yapmanın mümkün olacağı ima edilmiştir. Çağımızda ise bu tartışmaların 'istatistik' bağlamında sadece daha iyi veri analizine dönüştüğü fark edilmektedir. Çünkü pazarlamanın başat amaç olarak benimsendiği bir döneme girilmektedir.

Bu konuya ilişkin detaylara önceki başlıkta yer verildiği gibi zekanın Descartes'çı karşılığı en yalın haliyle akıl kavramının kısıtlı bir formu olarak gelişmiştir. Anlaşılacağı üzere Descartes felsefesinin doğurduğu zeka sözcüğü büyük ölçüde yapay zekanın meydana gelmesinden gerek kavramsal gerekse uygulamalı anlamda sorumlu tutulabilir. Bunun açığa kavuşturulması Descartes okumasıyla mümkünken, karşılaştırılmalı olarak Aristoteles'in görüşleri de önemlidir. Bu bağlamda önce Aristoteles'in bilgiye ulaşma yöntemlerinden *nous*'a, daha sonra Descartes'ta zekanın nasıl ortaya çıktığına yer vermek yerinde bir izlek olacaktır.

Eski Yunan'da zekaya eş değer ya da yerine geçen bir ifadenin ne olabileceği düşünüldüğünde Aristoteles'in edilgin akıl (*passive intellect*) kavramı zihinde canlanmaktadır. Edilgin aklın günümüzdeki zeka kavramına işlevsel anlamda denklik oluşturduğu düşünülebilir. Bu denkliği ortaya çıkaran şey, Aristoteles'in edilgin akli ve Descartes'çı zekanın benzerlikleridir. Bu iki kavramı karşılaştırmalı olarak ele alacak olursak, *nous*'un yalnızca etkin akla ait olan bölüme özgü olduğunu ve edilgin aklın veri olmaksızın görevini yerine getiremeyeceğini hatırlamamız gerekecektir. Diğer bir deyişle edilgin bölümde kendine sağlanan bir veri olmaksızın herhangi bir işlem gerçekleştirilmemektedir. Bu yönüyle zeka *nous*'un bileşeni olan edilgin akıl (*passive intellect*) ile ilişkilendirilebilir. Nihayetinde akla ait olan bir fakültenin yalnızca alt bir unsuruyla ilişkili olduğu düşünülebilir. Zeka ve edilgin akıl birlikte düşünüldüğünde aklın sadece bir fakültesine ilişkin nitelikler taşıdıklarından her iki kavramın da denk görevlerde olduğu öne sürülebilir. Bu bağlamda edilgin aklın *nous* içerisindeki görevi her ne kadar sınırlı ise zeka da akıl karşısında o ölçüde sınırlıdır. Bu husus yapay zekanın düzlemini ya da işlev kapasitesini belirlemektedir. Diğer bir deyişle yapay zeka her ne ölçüde geliştirilirse geliştirilsin algoritma düzeyinin dışına çıkamadığı müddetçe bu dar görevi temsil etmekten öteye gidemeyecektir.

Diğer bir konu da Descartes düalizmi ekseninde gelişen zeka kavramının kullanımı ve tarihsel evrimidir. Descartes'ın XVII. yüzyılda ortaya koyduğu bilinç anlayışı bu bağlamda

değerlendirildiğinde “düşünen şey” ile “yer kaplayan uzamlı şey” ifadeleri zihin-beden problemini yepyeni bir zemine taşımıştır. Bu anlayış Aristotelesçi zihin-beden birliğine karşı çıkarak zihni bir yerde bedeni ise başka bir yerde konumlandırmaktadır. Bu görüş salt dönemsel tartışmalarla sınırlı kalmamış olup Descartes’ın kendisinden sonraki beş yüz yıllık bir döneme damgasını vurmuştur. Öyle ki bu etkinin yansımaları yapay zekaya ilişkin önemli tartışmalar doğurmuştur. Öncelikle düalizme ilişkin Descartes’ın (1998) ifadesine bakıldığında ruh ve bedenin kesin ayrımına dikkat çekildiği fark edilecektir.

“Sonra, ne olduğumu dikkatle inceleyerek hiçbir bedenim olmadığını, bulunabileceğim hiçbir dünya, hiçbir yer olmadığını varsayabileceğimi ama buna göre varolmadığını varsayamayacağımı, tersine başka şeylerin doğruluğundan kuşkulandırmayı düşünüyorum oluşumdan varolduğum sonucunun apaçık ve kesin bir biçimde ortaya çıktığını, oysa düşünmeyi bıraksaydım tasarladığım tüm başka şeyler doğru olsalar bile varolduğuma inanmam için elimde hiçbir neden bulunmadığını görerek tüm özü ya da doğası düşünmekten başka bir şey olmayan ve varolmak için herhangi bir yere gereksinimi olmayan, herhangi maddesel bir şeye bağımlı olamayan bir töz olduğumu anladım. Öyle ki bu ben yani kendisiyle neysem o olduğum ruh, bedenden tümüyle ayrıdır, hatta bedenden daha kolay tanınır ve beden olmadığında bile o kendisi olmaktan çıkmaz.”<sup>61</sup>

Descartes’ın ifadelerinden anlaşılacağı üzere düşünsel olanın bedensel olandan kökten ayrıldığı düşünce ortadadır. Felsefe tarihinin önemli gelişmelerinden olan bu düşünce daha sonraları uygulamaya dökülecek yapay zekanın düşünsel arka planını şekillendirecektir. Burada yapay zeka sistemleri bedeni, algoritmalar ise zihni temsil etmektedir. Bu anlayış çerçevesinde düalist bakış yapay zekanın bilinç tartışmalarının önünü açmıştır. Kaldı ki Descartes’çı anlamda zihnin bedenden ayrı bir şey olduğu düşünüldüğünde yazılım ve donanımın birbirinden ayrı işlemesi insana özgü bilinci gerçekleştirebileceğinin işareti olarak görülmektedir. Bedensel olanın sayısal sistemlerin donanımıyla, algoritmanın yani yazılımın da düşünsel olana benzetilmesiyle doğan bir düşünceye ilişkin benzetmenin detayları *sanal makineler* olarak Boden (2016) tarafından açıklanmıştır.<sup>62</sup>

Descartes’ın yöntemine kısaca değinmek, bu yöntemin bilim dünyasına etkilerini anlamamızı sağlayacaktır. Zihin-beden ayrımına olan inancıyla Descartes, yaşadığı dönemin karmaşık ve hantal zihniyetini oluşturan kurallardan yakınlıkla kendi yöntemini ortaya koyarak tam anlamıyla uygulama kararı almıştır. Bu yöntemlerden ilki; ön yargılardan ve aceleden kaçınarak bir şeyin apaçık doğruluğundan emin olmak, ikincisi; problemleri mümkün olan asgari bölümlere ayırarak çözümlemek, üçüncüsü; basitten karmaşığa doğru ilerleyen bir sıra izlemek ve son olarak adımların eksiksiz bir şekilde ilerlediğinden emin olmak amacıyla

<sup>61</sup> Descartes, 1998: 39.

<sup>62</sup> Boden, 2016: 5.

kontroller ve sayımlar yapmaktır.<sup>63</sup> Descartes'ın yönteminden anlaşılacağı üzere bölümlere ayırarak çözümlemek bu düşüncenin en temel unsurudur. Bu düşünce kılavuzluğunda Descartes, Aristoteles'in aksine zihin ve bedeni kökten ayırarak bilinci tek taraflı olarak düşünmüştür. Modern döneme kılavuzluk edecek böylesine radikal bir felsefi yöntemin ortaya çıkmasında Descartes'ın yaşadığı dönemin etkisi azımsanmamalıdır. O döneme bakıldığında Orta Çağ, din tahakkümünde geçirdiği belirsizlik ve bunalımdan ciddi tahribat görmüştür. Bu sayede Descartes'ın bilimsellik, nesnellik ve kesinlik ilkeleri için uygun şartlar Avrupa'da gelişmiştir. Bu hususlarda açık, seçik ve kesin olgularla gelişen bilim anlayışı Hristiyanlık inancını da etkileyerek gelecekte dinin bilim karşısında kendisini güncellemesi gerektiğini fısıldamıştır.

“Duyguların yanılabilirliğini kullanarak deneysel her şeyi kuşku alanına çeker sonra da açık ve seçikliğin kuşku götürmez apaçıklığına başvurur. Bunun sonucunda keşfettiği şey, ‘sorgulanamaz / kuşku götürmez biricik kesin şey’dir. Descartes’a göre şüphesiz her bilim ampirik (deneye dayalı) incelemeyi gerektirir. Ama bu ampirik faaliyetin amacı Francis Bacon’un dediği gibi ‘olgusal veriler’ yığını değildir; aksine Tanrı’nın yaratıcı eyleminin her bir bilimsel alanda vücut bulduğu “açık ve seçik” ideallerdir.”<sup>64</sup>

Descartes'ta bir felsefeye dönüşen bölümleme ya da sınıflandırma yaklaşımı zekanın tanımlarından indirgemeci materyalist düşünceye değin pek çok alanda karşılık bulmuştur. Bu anlayış gerçeğin ya da tümelin bilgisini belirlemede zihin-beden birliğini gölgelemiş, düşünsel olanın bedensel olandan koparılmasına neden olmuştur. Bu açığı zeka terimi bağlamında değerlendirdiğimizde, zihnin uçsuz bucaksızlığı zekanın dar açısına sığdırılmaya çalışılmıştır. Bu ülkü yapay zekanın imkanını sağlamak için girişilen kavramsal bir mücadele olarak düşünülebilir. Diğer bir deyişle akılsal unsurlar zeka tarafından devralınmıştır. Akılsal niteliklerin enginliğini ve yapay zekanın kısıtlılıkları Searle'ün “semantik”, Penrose'un “bilinç” ve Dreyfus'un “sağduyu” argümanlarıyla ikinci bölümde değerlendirilmiştir.

Felsefi temelde zeka teriminin incelenmesi ilerleyen bölümlerde detaylandıracağımız yapay zeka ifadesine zemin oluşturmaktır. Bu doğrultuda farklı disiplinlerin zeka anlayışına yer vermek önemlidir. Legg ve Hutter (2006) tarafından zekanın farklı alanlardaki karşılığına yönelik yapılan çalışmada<sup>65</sup> yetmiş civarında tanım bulunmaktadır. Bu tanımlar “Genel Tanımlar”, “Psikolog Tanımları” ve “Yapay Zeka Araştırmacısı Tanımları” olmak üzere üç grupta sınıflandırılmıştır. *Genel tanımları kısaca burada özetlemek yerinde olacağından, zekaya ilişkin başlıca sekiz tanım verilmiştir.* Bu tanımlarda geçen anlamlar yoğunlukla; hafıza, bilgi, tecrübe, anlayış, muhakeme, hayal gücü ve genel zihinsel yetenekler sözcüklerini

<sup>63</sup> Descartes, 1994: 21-22.

<sup>64</sup> Toulmin, 1990: 47.

<sup>65</sup> <https://arxiv.org/abs/0706.3639> (erişim tarihi: 13.01.2019)

içermektedir. “Psikolog Tanımları” ve “Yapay Zeka Araştırmacısı Tanımları” başlıklarını ayrı ayrı incelemek zekanın tarihsel serüvenine dikkat çekmek açısından önemlidir. Bu nedenle aşağıda her iki başlıkta tanımlara yer verilmiştir.

### 1.2.2 Psikolog Tanımları

Zekanın konumu, kültürle ilişkisi gibi çeşitli ayrıntılara ilişkin tanımlar psikolojinin tartışma alanlarındandır. Bu başlıkta zekanın psikolojik yönlerden sınıflandırılmasına ve zekanın anlamlarına yer verilecektir. Psikolojide zeka terimi yoğunlukla belirli alan, beceri ya da rasyonel işlev odağında pek çok araştırmacı tarafından tanımlanmıştır. Örneğin “Spearman (1904) genel zeka, Thorndike (1920) sosyal zeka, Catell (1963) akışkan ve kristalize zeka, Sternberg & Wagner (1986) akademik ve pratik zeka, Levinson (1995) etkileşimsel ve analitik zeka, Perkins (1995) sinirsel, deneyimsel ve yansıtıcı zeka, Sternberg, (1999) yaratıcı zeka, Mayer (1999) duygusal zeka, Kaufman (2000) sözel ve algısal zeka, Gardner (2006) görsel-mekansal, bedensel-kinestetik, müzikal, kişilerarası, kişisel, dilsel ve mantıksal-matematiksel zeka.”<sup>66</sup> Yer verilen tanımlar felsefi bakışla karşılaştırıldığında Descartes’ın yöntemlerinde kullanılan ilkeler zihinlerde canlanabilir. Yani sözü geçen zeka türlerinin zihin-beden birliğinden uzak, ayrık bölümlerden oluşan bir anlayışla sınıflandırıldığı söylenebilir. Ayrıca tanımların geneline bakıldığında zekanın yüceltilecek bir terim olarak öne çıkarıldığı hissedilmektedir. Bu yönüyle tarihsel sorgulamadan yoksun olduğu gibi aklın *logos* niteliklerinin zekaya yüklendiği gözlemlenmektedir. Kaldı ki özellikle Gardner’in (2006) zekayı farklı kategorilere ya da yeteneklere indirgemesi Descartes’çı anlayışın tanımlardaki etkisini ortaya koymaktadır. Bu tespitin çalışmamıza sunacağı en önemli kanıt uçsuz bucaksız akıl kavramının gölgelenmesi ve Descartes’tan beslenen zekanın yeni bir kavram olarak yüceltilmesidir. Bu türden bir girişim hem akıl hem de zeka kavramını ilişkisel olarak muğlaklaştıracak ve aralarındaki içkinlik problemini karmaşık hale getirecektir. Dolayısıyla zeka terimi kullanıldığında ya da sayısal sistemler tarafından zeki davranış olarak nitelendirilen bir eylem gerçekleştirildiğinde sayısal sistemlerin bilince ya da semantiğe eriştiğine ilişkin iddiaların zaferi ilan edilecektir. Öte yandan psikologların zekaya getirdiği tanımlar Legg ve Hutter’in (2006) çalışmasında<sup>67</sup> otuz dört adet olarak belirlenmiştir. Bu tanımların ortak anahtar kelimelerini kısaca belirlemek gerekirse duyum, algı, ilişkilendirme, hafıza, hayal gücü, ayırt etme, yargı ve muhakeme gibi bilişsel süreçleri kapsadığını özetleyebiliriz. Aynı çalışmada incelenen diğer bir zeka sınıflandırması da yapay zeka araştırmacılarına aittir.

<sup>66</sup> Sternberg, R.J ve Kaufman S.B., 2011: 108.

<sup>67</sup> Legg ve Hutter, 2006: 17-24.

### 1.2.3 Yapay Zeka Araştırmacısı Tanımları

Son olarak zeka kavramının hesaplamaya dayalı yönlerden değerlendirildiği alan yapay zekanın büyük ölçüde tasarımıyla ilişkindir. Bu alanda zeka kavramı başlıca; Fogel, Gudwin, McCarthy, Minsky, Nakashima, Newell, Simon, Poole ve Kurzweil gibi düşünürler tarafından yorumlanmıştır. Kabaca zeka, çeşitli ortamlarda başarıya ulaşma yeteneğini bilgiyi işleme kapasitesi olarak ölçen bir anlayış olarak kabul edilmektedir. Söz konusu tanımlar, McCarthy'nin ardından pek çok araştırmacı tarafından kişisel bakış açısına indirgendiği görülmektedir. Kronolojik temelde ilerlendiğinde Bellman (1978) insan düşüncesini karar verme, problem çözme, öğrenme, yaratma, oyun oynama vb. gibi şeyler olarak nitelendirmektedir. Bu bağlamda insan düşüncesiyle ilişkilendiren faaliyetlerin otomasyonunu karar verme, problem çözme, öğrenme gibi etkinlikler olarak bildirmektedir.<sup>68</sup> Yani insan düşüncesini meydana getiren parametrelerin otomasyondaki karşılığını öznel bakışıyla belirtmiş ve yapay zekayı tanımlamaya çalışmıştır. Bu tanımın yedi sene ardından Haugeland (1985), "Bilgisayarların düşünmesini sağlama uğrunda heyecan verici yeni çaba... Tam ve gerçek anlamda zihinleri olan makineler."<sup>69</sup> ifadesiyle yapay zekanın nihai hedeflerinden olan bilincin makinelerde gerçekleştirilmesine yönelik algoritmanın zihinle bağdaştırıldığı anlayışı vurgulanmaktadır. Kurzweil (1990) ise "İnsanlar tarafından yapıldığında zeka gerektiren işlevleri yerine getiren makineler yaratma sanatı."<sup>70</sup> ifadesiyle zeka gerektiren eylemin neliğine ve bu bağlamda makinenin zekaya sahip olmasının gerekçesine dikkat çekmiştir. Ayrıca bu türden bir girişimin sanatsal bir uğraş olduğunu belirterek geliştirme çabalarını yüceltmıştır. Kurzweil'in ardından Rich ve Knight (1991) "Günümüzde insanların daha iyisini yapabildiği şeyleri bilgisayarların nasıl yapabileceği üzerine bir çalışma."<sup>71</sup> tanımıyla insan eylemlerinin bilgisayar sistemlerince karşılanması ve daha iyisinin gerçekleştirilmesi hedefinde bir ilerleyiş olduğu belirtilmiştir. Winston (1992)'ye göre "Algılama, akıl yürütme ve hareket etmeyi mümkün kılan işlemlerin incelenmesi."<sup>72</sup> ifadesi ile yapay zekayı hem düşünsel hem eylemsel niteliklerin bütünü tıpkı zihinsel süreçlerde olduğu şekliyle gerçekleştirmeye çalışan bir kavram olarak tanımlamıştır. Öte yandan Poole, Mackworth ve Goebel (1998) öz bir ifadeyle "Bilişimsel zeka, zeki araçların tasarımının incelenmesidir."<sup>73</sup> ifadesiyle öznel jargonları olan bilişimsel zeka türünden söz ederek bu türden bir zekanın akıllı olan şeyleri irdelemeyi hedeflediğini ifade etmişlerdir. Yani zekaya ilişkin olanın akılsal olana nüfuzunu doğrudan ima

<sup>68</sup> Bellman, 1978: 3.

<sup>69</sup> Haugeland, 1985: 2.

<sup>70</sup> Kurzweil, 1990: 14.

<sup>71</sup> Rich ve Knight, 1991: 3.

<sup>72</sup> Winston, 1992: 5.

<sup>73</sup> Poole vd., 1998: 6.

edilmektedir. Yapay zekanın tanımına ilişkin diğer bir önemli katkı da Nilsson (1998) tarafından betimlenmiştir. Bu ifadede yekpare bir tanımın oluşturulmasının güçlüğüne dikkat çekerek “Kabaca tanımlanan yapay zeka kavramı insan eliyle yapılmış şeylerdeki zeka ürünü davranışlarla ilgilidir.”<sup>74</sup> demiştir. Yapay zekaya yönelik ifadeleri irili ufaklı tanımlar olarak nitelendiren Nilsson, bir bakıma kendi ifadesinin de yapay zekayı tüm boyutlarıyla kapsayamayacağını varsaymaktadır.

Yukarıda değindiğimiz tanımların yanı sıra Legg ve Hutter’in (2006) değerlendirmelerine bakıldığında zekanın mahiyeti üzerine bir uzlaşının olmadığını göstermektedir. Öte yandan bu görüşlerin genelinde çoklu zeka kuramında olduğu türden zekayı sınıflandırma anlayışının benimsendiği fark edilmektedir. Burada esaslı konu insan zekasının bölümlerden ibaret olduğunun imasıyla mekanik boyuta indirgenmesi, bu sayede yapay zeka ile bağ kurulması düşüncesinedir. Diğer bir deyişle yapay zekanın zihinsel yetiler yönünden sınırlılıklarını kavramsal mücadele ile yok etme çabasıdır. Bu düşünce doğrudan Descartes’ın çözümleme yöntemiyle özdeşleştirilebilir. Dolayısıyla tanımların geneli incelendiğinde iki yaklaşım dikkat çekmektedir. Birincisi zihin-beden ilişkisini birbirinden ayıran, ikincisi ise tarihselliği olan akıl kavramının niteliklerini zekaya yükleyen anlayıştır. Zeka terimini akılla eş değer tutmak hatta zeka sahibi olmayı, zeki olmayı aklı başında olmanın ya da us sahibi “uslu” olmanın karşısında daha gözde bir konuma yerleştirme ülküsü gözlemlenmektedir. Bu anlayışın ortaya koyduğu kavramsal değişim akıl ve zeka arasındaki farkların görmezden gelinmesine neden olmaktadır. Haliyle yapay zekanın gelecekte insana özgü pek çok niteliği gerçekleştirebileceğine yönelik inancı güçlendirmektedir. Bu inanç sağlam temellere dayansa da ciddi etik problemlerin habercisi olduğu gerçeğini ortadan kaldıramaz. Bu minvalde kavramsal analizin önemini tekrar vurgulayarak kavramların felsefi boyutlarıyla açık ve seçik bir şekilde temellendirilmesi yanlış iddiaların çürütülmesine, gölgelenen etik kaygıların açığa çıkarılmasına imkan tanıyacaktır. Diğer yandan, kavramsal analizin diğer bir durağı “bilinç” kavramı olup bu konu zihin felsefesi ve yapay zeka arasındaki ana tartışmadır. Bu nedenle aklın neliğini anlama uğrunda incelediğimiz zekanın ardından “bilinç” teriminin kavramsal dinamiklerine, tartışmalarına ve genel haliyle mahiyetine yapay zekayla ilişkisi bağlamında değinmek gereği doğmuştur.

---

<sup>74</sup> Nilsson, 1998: 1.

### 1.3 Bilinç Nedir?

İnsan zihninin bileşenlerini irdelediğimiz ilk bölümün diğer bir tartışma alanı da bilinç terimidir. Bilinç<sup>75</sup>, gerek sinirbilim gerekse zihin felsefesi perspektifinden tartışılan önemli konulardandır. Tarihsel gelişimiyle XVII. yüzyılda farkında olmak anlamında ortaya çıkan “bilinç”, Cudworth (1837) tarafından “kendi eylemlerinde özenli, ya da onları eleştirebilen, keyiflenme ya da nimetlenme ve kendini algılamak için eyleme geçmek ya da katlanmak uğrunda kendisiyle var olan bir varlık.”<sup>76</sup> şeklinde yorumlanmıştır. Tanımın içerdiği muğlaklıktan olsa gerek on yedinci yüzyıl ve sonrasında bilinç, en az dört ayrı temada tartışmanın merkezi olmuştur: “kişisel kimlik, ruhun ölümsüzlüğü, epistemik kesinlik ve aşkın deneyim koşulları.”<sup>77</sup>

Bilincin vicdan, şuur, farkındalık, uyanıklık gibi anlamlarda kullanımına ilişkin bazı görüşler de bulunmaktadır. Örneğin John Searle “vicdan”, “kendi olma bilinci” ve “biliş” sözcükleriyle bilincin karıştırıldığı iddiasıyla bilincin, bilgi ve bilise ev sahipliği yaptığını ifade etmiş, bilinci sistemin bir şalteri olarak kabul etmiştir.<sup>78</sup> Buna ek olarak Shaffer (1991), “İnsanı çevresindeki birçok şeyden ayıran şeydir. Bu da derin bir uykudan koparak uyanıklık haline geçmemiz yani sahip olduğumuz şuurdur.”<sup>79</sup> ifadesiyle farkındalığı bilinçle ilişkilendirmektedir. Öte yandan Freud’a göre “zihinsel süreçler kendi içinde bilinçsizdir”. Ayrıca “doğası ne olursa olsun zihinsel olan şey kendi başına bilinçsizdir”. Bilinç, zihinsel oluşumların “değişken bir niteliğidir”.<sup>80</sup> Rosenthal’e (2005) göre Freud’un bilinçli durumlara ilişkin ifadesi bilinci kapsamlı bir şekilde açıklamamaktadır. Tanımdan anlaşılacağı üzere bilinç kimi zaman bir sistemin açma kapama devresine, kimi zaman uyanık olma durumuna kimi zaman da bilincin ön koşulunun değişkenlik niteliğine bağlı olduğu öne sürülmüştür. Ekollerin ve akımların bilinci anlamada diğerlerini yadsıyan ve öznel bakışlarını yücelten yaklaşımları

<sup>75</sup> Türkçe sözlük anlamı “İnsanın kendisini ve çevresini tanıma yeteneği, şuur.” olan bilinç terimi, İngilizcede *conscience* “vicdan” sözcüğünün, *consciousness*’ın “bilinç” atasıdır. *Conscience* ise “biliyorum” *scio* ile “birlikte” *cum* sözcüklerinin birleşiminden meydana gelmektedir. *Conscious*, ise *conscio* fiilinden türetilen bir sıfatken; *conscientia* isim biçimidir. *Conscio* Latince bilginin birden fazla kişiyle bölüşüldüğü bir anlamı bildirmektedir. Burada bahsedilen bilgiden söz edilirken utanılacak ya da endişe duyulacak bir bilgi türü ifade edilmektedir. Yani aleni bir bilgiden ziyade örtük, gizli bir bilgiyle birtakım hileler yapma kaygısı vurgulanmaktadır. Dolayısıyla gizli işler içinde olan kişiler için *conscious* sözcüğünün kullanıldığına dikkat çekilmektedir. Hatta Thomas Hobbes’in bu sözcük üzerine yorumunu, “İki veya daha fazla kişi aynı olayı [yani fiili] biliyorsa, o kişilerin birbirleri karşısında o olayın bilincinde oldukları söylenir.” sözüyle özetlemektedir. Bu durumda pek çok kişi tarafından bir olayın ya da eylemin bilinmesi o şeyin bilincinde bulunduğu anlamına gelmektedir. *Conscious*’un ikinci anlamı da kişinin kendisinin ne yaptığının farkında olması olarak belirtilmektedir. Bu iki anlam sözcüğün “somut” anlamı olarak nitelendirilmektedir. Latince bu ifadenin sadece “biliyorum” veya “çok iyi biliyorum” anlamında kullanıldığı belirtilmektedir. Sözü geçen üç anlam da İngilizceye *conscience* olarak *conscientia*’nın muadili olarak girmiştir. (Zeman 2006: 35, 36.)

<sup>76</sup> Cudworth, R. 1837: 227.

<sup>77</sup> Heinämaa vd., 2007: 6-7.

<sup>78</sup> Searle J. 2002: 83-84.

<sup>79</sup> Shaffer J. 1991: 24-26.

<sup>80</sup> Rosenthal, 2005: 309.

fark edilmektedir. Bu yönüyle bilincin felsefi söylemi ve yapay zeka ile ilişkisi değerlendirilmelidir.

### 1.3.1 Felsefi Söylemde Bilinç

Yapay zeka ve bilinç ilişkisinin tarihsel arka planında şüphesiz Alan Turing bulunmaktadır. Öte yandan bu tartışmaların sayısı tarihsel akışla birlikte artmaktadır. Turing'in ifadesine görüleceği üzere bilincin sayısal sistemlerde gerçekleştirilmesi çabası özellikle Turing'in geliştirdiği mekanizmaya duyduğu güvenden hareketle ortaya attığı iddiadan ileri gelmektedir.

“Gelecekte yapay zeka sistemlerinin insan performansına eşit olacağı varsayıldığında; gerçek zekaya, gerçek anlayışa, gerçek yaratıcılığa sahip olurlar mıydı? Benlikleri, ahlaki duruşları ve özgür seçimleri olur muydu? Bilinçli olurlar mı? Ve bilinç olmadan, bu diğer özelliklerden herhangi birine sahip olabilirler mi?”<sup>81</sup>

Felsefi söylemde bilince girişi Turing ile yapmanın arkasında yatan gerçek, öncelikle bilince ilişkin tartışmanın yapay zekayla ilişkisini belirlemektir. Bu ilişkinin temeli de özellikle Turing'in iddialarıyla şekillenmiştir. Bilincin felsefi söylemi düşünüldüğünde akla gelen en önemli isim Chalmers'tir (1995). Chalmers'in argümanı bilincin neliğine ilişkin doğrudan bir cevap sunma niteliğinde olmayıp, olası bir cevaba ulaşabilme yöntemi olarak bilinç problemine yol göstermektedir. Bunlar bilincin kolay ve zor problemleri olmak üzere ikiye ayrılır. Chalmers'e (1995) göre; “Bilincin kolay sorunları, bilişsel bilimin standart yöntemlerine doğrudan duyarlı görünen ve fenomenin hesaplamalı ya da sinirsel mekanizmalar açısından açıklandığı problemlerdir.”<sup>82</sup> Başlıca ayırt etme, sınıflandırma, çevresel uyaranlara tepki verebilme yeteneği, bilginin bilişsel bir sistem ile entegrasyonu, zihinsel durumların bildirilebilmesi, bir sistemin kendi içsel durumuna erişme yeteneği, dikkatin odağı, niyetli davranış kontrolü, uyanıklık ve uyku arasındaki fark gibi nitelikleri kapsayan kolay problemler bağlamında bilinci açıklamaya çalışmaktadır. Chalmers bilinç sorunun yalnızca kolay problemten ibaret olması durumunda onu açıklamanın çok basit olacağını belirtmektedir. Farkındalık, ihtiyatlılık, uyanıklık, dikkatlilik, ulaşılabilirlik, özdüşünümsellik gibi kavramlar fonksiyonalist bakış ile ilgili olup bilgi işleme ve sinirbilimin tartışmaları olarak kabul edilmektedir. Fenomenal bilinç ya da duyular ya da diğer adıyla "qualia" materyal bir evrende metafizik bir sorgulamayı gerektirmesi nedeniyle genellikle fizikalist bakışın karşısındadır.<sup>83</sup>

<sup>81</sup> Boden, 2016: 119, 120.

<sup>82</sup> Chalmers, 1995: 2.

<sup>83</sup> Boden, 2016: 121, 122.

Maddeci yaklaşımı benimseyenler bilinci özellikle kolay problemlerden ibaret görmekte ve fenomenal boyutunu reddetmektedirler. Kolay problemler felsefi söylemden uzak maddeci bakışa yakın olduğundan Descartes'ın mekanik felsefesinin izlerini taşıyan bu geleneğe ait çoğu düşünür bilinci beyinde konumlandırmaktadır. Bu görüş için “Hem bilinç hem de yönelimsellik beyinde alt düzey sinirsel hücrelerin işlemleri ile belirlenen biyolojik süreçlerdir ve de başka bir şeye indirgenemezler.”<sup>84</sup> ifadesi maddecilerin bakışını anlamada önemli bir belirteçtir. Bu bakışa yer yer eleştirel tutumlar getirilmiştir. Örneğin Searle, beynin en küçük yapı taşının incelenmesiyle sayısal sistemlerin bilince ne ölçüde erişebileceğini tahmin etmeye yardımcı olabileceğini düşünmektedir. Bu düşüncesini de dipnotta belirttiği şu ifadeye dayandırmaktadır: “Bernard Williams'ın (1987) bu (yapay zeka) araştırmasında kısmen tartışılmakta olan şey, tam olarak zeki sistemlerin zeki olmayan maddeden oluşabilir olup olmayacağıdır.”<sup>85</sup> Diğer bir deyişle beynin en küçük yapıtaşının keşfiyle bilincin konumu belirlenirse o konumun çoğaltılmasıyla sayısal sistemlere bilincin nakledilebileceği düşüncesi ima edilmektedir. Maddeci bakışta bilincin konumu, işlevi ve bütünsel olarak mahiyeti beyin organıyla özdeşleştirilmektedir. “Zihinsel fenomenler, bilinçli veya bilinçsiz tüm zihinsel fenomenler, görsel veya işitsel, acılar, gıdıklamalar, kaşıntılar, düşünceler, hatta tüm zihinsel yaşamımız beyinde meydana gelen süreçlerden kaynaklanır.”<sup>86</sup> Bernard Williams'ın (1987) ifadeleri bağlamında eleştiride bulunan Searle, maddeci bakış açısıyla yapay zekanın gelecekte bilinç sahibi olup olamayacağına cevap veremeyeceğini ifade etmektedir. Kaldı ki bilinç, fenomenal bir olgu olarak bütünsel bir yaklaşımı gerektirmektedir. Bu konuya ilişkin detaylara Chalmers'in bilincin zor probleminde yer verilecektir.

Bu noktada esaslı tartışma bilincin zor problemiyle ilerlemektedir. Chalmers bir düşünce deneyi geliştirerek bilincin mahiyetini ortaya koyma yolunda yepyeni bir soru ve kavramsal tartışma ortaya atmış, bilincin mahiyetine bir adım daha yaklaşmıştır. Bilincin zor problemine ilişkin düşünce deneyinin detaylarına geçmeden önce Chalmers'in bilincin aydınlatılmasına ilişkin inancı şu ifadelerinde görülebilir.

“Mevcut bilinç teorilerinin çoğu ya fenomeni reddeder, başka bir şeyi açıklar ya da sorunu ebedi bir gizeme yükseltir. Sorunu ciddiye alırken bile ilerleme kaydetmenin mümkün olduğunu göstermeyi umuyorum. Daha fazla ilerleme kaydetmek için daha fazla araştırmaya, daha rafine teorilere ve daha dikkatli analize ihtiyacımız olacak. Zor olan problem zor bir problemdir, ancak bu hiçbir zaman kalıcı olarak çözülemeyeceği anlamına gelmez.”<sup>87</sup>

<sup>84</sup> Searle J. 2004:8.

<sup>85</sup> Searle J. 2004:51.

<sup>86</sup> Searle, 1984: 13.

<sup>87</sup> Chalmers, 1995: 25.

Chalmers, bilincin zor problemini kalıcı çözüm için bir önemli adım olarak nitelendirerek daha çok çalışmanın yapılması gerekliliğini vurgulamaktadır. Söz konusu düşünce deneyinde Chalmers, bilincin zor problemini bir tür “deneyim” problemi olarak bildirmektedir. Chalmers’in temel sorusu; görsel bilişsel sistemlerimizin ve işitsel bilgiyi işlerken, neden görsel veya işitsel deneyime sahip olduğumuzdur? Zihinsel bir imgenin tadını çıkarmak ya da duyguyu deneyimlemek gibi şeylerin nedenini nasıl açıklayabileceğimizdir? Tecrübelerimizin fiziksel temelde gerçekleşmesinin genelgeçer bir görüş olduğuna dikkat çeken Chalmers, fiziksel işlemlerin zengin içsel bir yaşantıya nasıl yol açabildiğini sorgulamaktadır.<sup>88</sup> Akıl ve zekaya yönelik incelemeden hatırlanacağı üzere düalizm karşısında zihin-beden birliğini benimseyen görüşler öne çıkmıştı. Kaldı ki Aristotelesçi zihin-beden anlayışında olduğu gibi Chalmers, bilinci salt maddeye içkin olmaktan öte gerek deneyimsel gerekse zihinsel süreçlerin ilişkisel bütünlüğünde anlamaktadır. Bilinci meydana getiren iki yön Chalmers’in sınıflandırmasıyla düşünüldüğünde kolay problemlerin sinirbilime ve pek çok alana önemli katkılar sağlayacaktır. Odaklanılması gereken şey bilincin zor problemidir. Zor problemin bilince en büyük katkısı da deneyimin, zihinsel süreçleri etkilemesi gerçeğidir. Bu açıdan Aristotelesçi pratik akıl özünde benzer bir düşünceyi barındırmaktadır.

Ayrıca Aristoteles’in *phronesis* kavramına bir örnek de deneyimin öznelliğinde bilinci açıklayan Thomas Nagel (1974) gösterilebilir. Nagel, yarasaların algılarıyla ortaya çıkan deneyim ile insanların algılarıyla ortaya çıkan algı farkına dikkat çekerek bilincin doğasının zorunlu olarak öznel deneyime bağlı olduğunu belirtmiştir. Yani yarasa olmanın nasıl bir şey olduğunu yarasa olmadan anlamanın pek mümkün gözükmediğini söylemektedir. Bu bağlamda doğuştan sağır ve kör bir kişinin deneyiminin öznel karakteri de diğer insanlar için erişilebilir değildir. Nagel’in öznel alan hakkındaki gerçekçiliği, insana özgü kavramların erişemeyeceği gerçeklerin varlığına olan inancıdır.<sup>89</sup> Bu ifade Nagel için o derece önemlidir ki bir şeyin kesin öyle olduğunu söylemekten kaçınılmalı ve gibi görünüyor ifadesi kullanılmalıdır. Nagel’in düşüncesine bakıldığında deneyimin zihin süreçleri açısından önemi fark edilmektedir. Nagel’in ifade ettiği deneyimsel öznelliğin ve Aristoteles’te *nous* ile *phronesis* iş birliğinde ortaya çıkan türden bir bilinç anlayışının bulunduğu ifade edilebilir. Yani *nous* düşünsel olan ile bağdaştırıldığında *phronesis* deneyime karşılık gelmektedir. Bu noktada Nagel, öz bir ifadeyle öteki olmanın olanaksızlığına göndermede bulunmaktadır. Öznenin özne olarak kalacağı, öteki olmaksızın öteki gibi duyumsanamayacağını ifade etmektedir.

<sup>88</sup> Chalmers, 1995: 200-19

<sup>89</sup> Nagel, 1974: 435-50.

Bilincin tanımına yönelik kapsamlı bir tartışma da Pepperell (2000) tarafından öne sürülmüştür. Pepperell, bilinç terimini başlıca on farklı yönden ele almaktadır. İlk olarak bir organdan ziyade, organizmanın işlevi olarak bilinci kabul etmektedir. İkincisi; insanın tek boyuta indirgenemeyeceği gerçeğine vurgu yapılarak zihin ve çevreyi farklı kutuplara yerleştirmeye karşı durmaktadır. Üçüncüsü; bilinç, beden ve çevrenin bütünsel ve sürekli olduğunu belirterek çevresi olmayan bir beden, bedeni olmayan bir beyin gibi işlevselliğin de sona ereceği düşüncesiyle bilincin duracağını iddia etmektedir. Dördüncüsü; bilincin tarifini su ısıtıcısı ve suyun kaynatılması örneğiyle açıklamaktadır. Öncelikle kaynama olgusunun ısıtıcısının parçaları arasında su, enerji gibi elemanların neresinde olabileceğini sorgulamaktadır. Daha sonra kaynamanın bu sistemin hiçbir parçasıyla açıklanamayacağını, bilincin kaynamanın kendisiyle özdeşleştirilmesi gerektiğini düşünmektedir. Yani kaynama bütün elemanların ortaya çıkardığı bir özellik olarak kabul edilmektedir. Bilinç olarak bildiğimiz özellik zihinsel ve bedensel unsurların bütünlüğünde ortaya çıkmaktadır. Beşincisi; her şeyin atomik seviyedeki parçacıklardan oluştuğu gerekçesiyle her şey nihayetinde enerji olacağından, her türlü sınıflandırmanın ve bölümlenmenin algılardan kaynaklandığını düşünmektedir. Yani her şeyin özünde aynı olduğu ancak bizim şeyleri farklıymış gibi algıladığımızı belirtmektedir. Altıncısı; hiçbir şeyin kendi içinde ayrı şeyler olarak bulunamayacağını belirterek aslında her şeyin bütünsel olduğunu ama bizim bölümlere ayırdığımızı vurgulamaktadır. Yedincisi; bütünlüğün karşısında bölünmüşlüğümüzü Pepperell, dil kavramına bağlayarak çocuklukta edinilen dilin bir parçalanma süreci ile çalıştığını, nesilden nesle geçerek her biri ayrı sözcüklerden oluşan farklı şeylere böldüğünü öne sürmektedir. Sekizincisi; dilin, gerçekliğin tümü olmadığı sadece bir parçası olduğunun altını çizerek kendimiz ve çevremiz arasında yaptığımız ayrımın dilden ortaya çıktığı için yanıltıcı olacağını iddia etmektedir. Dokuzuncusu; dil aracılığıyla yaptığımız tüm ayrımların toplamı olarak bilinci kabul etmektedir. Son olarak; yapısı itibarıyla dil dünyaya çarpık bir bakış açısı getirdiği için, parçalanmış bilincin ötesine geçmenin ve onu birleştirmenin yolunu kültür olarak işaret etmektedir. Bu minvalde kültürü, bilinci birleştirmenin bir yolu olarak işaret etmektedir.<sup>90</sup>

Felsefi söylemde bilinç tartışması Chalmers'ın kolay problemlerinden daha fazlasıdır. O nedenle bilincin zor problemi ortaya atılmıştır. Öte yandan Aristotelesçi *phronesis*, Nagelci deneyim, Chalmersçi öznel bilinci anlama yolundaki başlıca kavramlar olarak düşünülebilir. Chalmers, Nagel ve Aristoteles'ten öğrendiğimiz ve ortak kavram olarak değerlendirebileceğimiz “deneyimsellik” yapay zeka ve bilinç ilişkisinde bizlere önemli bir fikir verebilir. Bu konuya detaylarıyla tezin sonunda yer alan sonuç bölümünde değinilmiştir.

<sup>90</sup> Ascott, 2000: 13-16.

Chalmers ve Nagel'in bilinci açıklamada öne sürdüğü deneyimsellik ile Aristoteles'in *phronesis*'e yaptığı vurgu düşünüldüğünde bilincin makine öğrenmesine sığmayacak kadar engin bir olgu olduğu ortadadır. Burada asıl problem yapay zeka ve bilinç meselesinde bir yaklaşım ya da bakış açısı sorunudur. O nedenle değindiğimiz görüşlerin bütünü değerlendirildiğinde Greklerden bu yana felsefe çerçevesinde tanımlanan insanın, artık nesnenin özellikleri çerçevesinde belirlendiği görülmektedir. Çoklu zeka kuramı, bilgi işleme kuramı, insan belleğinin işleyişini sayısal sistemlerin işleyişine indirgendiği anlayış örneklerden bazılarıdır. Yani nesnenin özneyi belirlediği bir ekolün baş gösterdiği, teknolojinin psikolojik durumlara nüfuzu ve insan doğasının manipülasyonu ile ilerleyişini sürdürdüğü bir ekol şekillenmiştir. Bu türden bir ilişki bilinç ve algoritmayı deyim yerindeyse birbirinin yerine kullanırcasına ilişkilendirmektedir. Ne var ki Searle'un Çince Odası argümanında sözünü ettiği "tek başına söz dizimi, anlambilim için yeterli değildir"<sup>91</sup> ifadesi algoritmanın maddenin doğasına içkin olmadığı noktasında önemli bir ifadedir. Searle "Kısacası sentaks, sistemin fizikine içkin değildir; yalnızca onu kullanan kişide anlamını bulur."<sup>92</sup> ifadesiyle anlamı algoritma karşısında incelemektedir. Yani teknolojinin kazanımları karşısında felsefi düşünümü yüceltmekte ve felsefeyi belirleyici bir konuma taşımaktadır.

### 1.3.2 Yapay Zeka ve Bilinç İlişkisi

Kusursuz algoritmalar sayesinde sayısal sistemlerde bilincin olasılığı tartışılan konulardandır. Sayısal sistemlerin bu seviyeye ulaşma dönemi 'insan ötesi' (*posthuman*) çağ olarak işaret edilmektedir. Bu yönüyle insan ötesi durum doğrudan hümanizm ve transhümanizm terimleriyle ilişkilidir. Hayles'in (1999) ele aldığı posthümanizm 'doğal' bir benlikten uzak olan zeki makinelerin geliştirilmesiyle doğacak olan bir terimdir. Oysa hümanizm; insan, benlik ve bireyin özgürlüğü konularını kapsamaktadır.<sup>93</sup> Hümanizm ve posthümanizm kavramlarını karşılaştırarak değerlendiren Hayles bilinci, Descartes'çı anlayışının aksine başlıca zihinsel süreçlerden ileri gelmeyen, daha çok epifenomen olarak düşünmektedir. Bu düşüncede insan vücudu sonradan değiştirilebilen bir protez olarak görülmektedir. En önemlisi de zekaya haiz olma yolunda insanı yapılandırmaktadır. Posthüman durumda bedensel varlık ve bilgisayar simülasyonu, sibernetik mekanizma ve biyolojik organizma, robot teleolojisi ve insan hedefleri arasında zorunlu bir fark veya mutlak sınırlamalar bulunmamaktadır.<sup>94</sup> Hayles'in posthümanizm yorumuna bakıldığında insan

<sup>91</sup> Searle, 1984: 32.

<sup>92</sup> Searle, 1992: 77,78.

<sup>93</sup> Hayles, 1999: 22.

<sup>94</sup> Hayles, 1999: 3.

bilincinin makineden bağımsız var olamayacağı anlaşılmaktadır. Bilinci bu bağlamda değerlendiren Hayles, hümanizmin öte yakası olan posthüman durumu insanlığın sonunu getirebilecek bir gelişme olarak görmektedir. Posthüman durumda bilincinin nasıl anlaşılabileceği sorusuna Hayles'in cevabı şu ifadelerden anlaşılabilir.

“Bilinç, farklı dağıtılmış sistemler arasındaki çatışmaları doğrulamak için devreye giren kontrol sistemine benzer bir gelişmedir. Bilinç, Brooks'un dediği gibi, "ucuz bir hile", yani sistemin işlevselliğini artıran, ancak sistemin temel mimarisinin bir parçası olmayan ama ortaya çıkabilen bir özelliktir. Bilinç, aslında temsilden ibaret değildir, olması da gerekmez. Robotun kontrol sistemi gibi bilinç, dünyanın doğru bir resmini gerektirmez; sadece güvenilir bir ara yüze ihtiyaç duyar. İnsan bilincinin bu şekilde çalıştığına dair kanıt olarak Brooks, yetişkinlerin çoğunun görme alanlarının ortasında büyük bir boş nokta ile hayattan geçtiklerinden habersiz olduklarını söyler.”<sup>95</sup>

Sayısal sistemler ve insan bağlamında bilincin bir tür “özellğe” indirgendiği anlaşılmaktadır. Bu çıkarım insan bilincinin sayısal sistemlere benzetilmesi olarak düşünülmemelidir. Kaldı ki Hayles bunu net bir ifadeyle reddetmektedir. “Defalarca savunduğum gibi, insan her şeyden önce somutlaşmış bir varlıktır ve bu somutlaşmanın karmaşıklığı, insan farkındalığının (bilinç) sibernetik makinelerde somutlaşan zekadan çok daha farklı şekillerde ortaya çıktığı anlamına gelir.”<sup>96</sup> Hayles, insan bilincinin somutlaşmış bir olgu olduğundan söz ederken onun hem bedensel hem de zihinsel bir bütünlüğü taşıdığını ifade etmektedir. Sibernetik makineler tarafından gerçekleştirileceği öngörülen yapay zekanın çok daha sınırlı bir alanda işlev göreceğini vurgulamaktadır.

Öte yandan çevresel dinamiklerin bilinçle ilişkisini uygulamalı bir deney örneğiyle temellendirmeye çalışan Joseph Wakeling ve Per Bak, ‘mini beyinler’ sinir ağı sayesinde birbirleriyle rekabet ederek çalışan bir sistemi, temel özellikleri seçmekle görevlendirilmiştir. Mini beyinler sadece kendi davranışlarıyla ilgili bilgi sahibi olduklarında kötü performans gösterirken, çevredeki diğer mini beyinler ile iletişime geçtiklerinde çok daha iyi performans gösterdikleri ortaya çıkmıştır. Nihayetinde zeki davranışın beynin işleme gücünden daha fazlasını gerektirdiği sonucuna varmışlardır. Bedensel araç ve çevredeki faaliyet bilinci, iyi bir genel performans için çok önemlidir. Bu, aktif dünya ile aktif/işleyen zihin arasında derin bir birlik anlamına geliyor gibi görünmektedir.”<sup>97</sup> Özetle deneyimlerin zihinsel süreçleri etkilediği gerçeği tikel bir örnekle gözlemlenmiştir. Aristotelesçi *phronesis* kavramının akıl için önemine değindiğimiz gibi, bilinç için de çevrenin ve deneyimin önemini bu deney sayesinde görmekteyiz. Öte yandan bedensel olanın çözülmesinin ötesinde bilincin öznelliğine ilişkin problemlerin devam edebileceği Charles T. Rubin (2003) tarafından da vurgulanmaktadır.

<sup>95</sup> Hayles, 1999: 238-239.

<sup>96</sup> Hayles, 1999: 283, 284.

<sup>97</sup> Pepperell R. 2003:18.

“Kısacası eğer insanlar basitçe geliştirilebilen mekanizmalarsa parçalarımız başkaları tarafından değiştirilebilirse biyolojik olarak veya başka bir şekilde inşa edilmiş olup olmadıklarının çok az önemi vardır. Bu, beden varoluşu için gayet geçerlidir. Peki ya zihin? Bu hayat sadece beynin biyolojik mekanizmasından kaynaklanmıyor, aynı zamanda bu mekanizma aracılığıyla deneyimlediğimiz şeyin zaten sanal gerçeklik olduğunu savunuluyor. Gerçek dünya hakkında hiçbir bilgimiz yok; sadece beynimizin vücudumuzun duyuşal girdilerini işlemesine sahibiz. Bilinç kökten olarak öznel ve esasen tekindir.”<sup>98</sup>

Fark edileceği üzere akıl ve zeka terimlerinde olduğu gibi bilince ilişkin problem benimsenecek yönetime ilişkindir. Tezin girişinden buraya kadar akıl ve zeka terimlerini felsefi çerçevede inceledik. Bilinç meselesinde ise yukarıdaki alıntıdan anlaşılacağı üzere bilince ilişkin muğlaklıkların önündeki problem konunun fizikalist bakışla sınırlandırılmasıdır. Oysa yapay zeka konusunda yolların çoğu “Sayısal sistemlerin bilinci olabilir mi?” sorusunda düğümlenmektedir. Bu konuya ilişkin görüş bildiren Arf’in (1959) sistemlerin geleceğine ve insan ile kıyasına ilişkin yorumu önemlidir. Turing ile aynı dönemde çalışmalarda bulunan Arf’in o tarihlerde çoğu zihin düşünürlerinden öncesinde böyle bir tespitte bulunması ilginçtir.

“Fakat kanaatimce insan beyni ile makine arasındaki asıl fark, insan beyninin estetik mahiyette müessirleri alıp onlar üzerinde işleyebilmesi ve yine estetik mahiyette olan kararlar verebilmesine, verilen bir işi yapıp yapmamak hususunda kendisini serbest hissetmesine mukabil makinede bu vasıfların benzerlerinin yok oluşudur. Bu vasıfları karakterize eden husus hepsinin de bir belirsizlik unsuru ihtiva etmesi, bunların şaşmaz bir şekilde uydukları kaidelerin mevcut olmayışıdır. Belirsizlik karakterini haiz olan insan dışı tabiat hadiseleri mevcuttur. Bunlar atom içinde ceryan eden olaylardır. Bu itibarla nisbeten küçük sayıda atom içinde ceryan eden olaylar böyle makinelerin işleyişinde müessir hale getirilebilirse makinelerin estetik bakımdan da insan beynine benzetileceği ümit edilebilecektir. Böyle bir makine, mesela filan müzik parçasını güzel bulmadığını söyleyebilecektir. Fakat bu işin uzun yüzyıllar sonra bile ve belki de hiçbir zaman yapılamıyacağını zannediyorum.”<sup>99</sup>

Arf’in ifadelerinden “...verilen bir işi yapıp yapmamak hususunda kendisini serbest hissetmesine mukabil makinede bu vasıfların benzerlerinin yok oluşudur. Bu vasıfları karakterize eden husus hepsinin de bir belirsizlik unsuru ihtiva etmesi, bunların şaşmaz bir şekilde uydukları kaidelerin mevcut olmayışıdır.” cümlesi Irrgang’ın öznellik kavramıyla bağdaşmaktadır. Bu yönüyle sayısal sistemler öznellik kavramı çerçevesinde değerlendirilirse Arf’in da ileri sürdüğü gibi makineler doğası gereğince kusursuz bir mekanik işlevi yerine getirmek üzere tasarlandığından, etik değerlerin bu tür sistemler üzerinde uygulanması gerçekçi görünmemektedir. Konunun dikkat çekici diğer bir yönü de Arf’in “*Makineler Düşünebilir mi ve Nasıl düşünebilir?*” başlıklı makalesinde çıkardığı sonuçtur. Burada Arf, makinelerin

<sup>98</sup> Rubin, 2003: 88 – 100.

<sup>99</sup> Arf, 1959: 103.

düşünüp düşünemeyeceği sorusuna “Fakat bu işin uzun yüzyıllar sonra bile ve belki de hiçbir zaman yapılamayacağını zannediyorum.” cevabını vermiştir. Makinelerin düşünebilmesinin pek mümkün olmayacağını dile getirmesi etik kuramların kaderini belirleyici niteliktedir. Öyle ki Sullins’in ifade ettiği “yönelimsellik, özerklik ve sorumluluk” kavramları doğrudan insanda karşılık bulmaktadır. Arf’a (1959) göre düşünme yeteneğinin makineler için mümkün olmaması, etik kuramların yapay zeka sistemlerinde en azından insanlarda olduğu şekliyle mümkün olmayacağını işaret etmektedir.

Diğer taraftan felsefi çerçevede Chalmers, Pepperell, Nagel, Hayles ve Searle gibi düşünürlerin bilinç tartışmaları bu meselenin anlaşılmasına ve çözümüne ilişkin önemli katkılar sağlamışlardır. Yine de felsefi dayanakların gücüne rağmen bilinç hala gizemli bir alan olarak karşımızdadır. “Bu durumda, güçlü yapay zeka lehinde ve aleyhinde argümanlar sonuçsuzdur. Bilinç bir sır olarak kalır.”<sup>100</sup> Bu ifadeden anlaşılacağı üzere bilincin ne olduğu çözülmeksizin güçlü yapay zekanın imkanından bahsetmek mümkün değildir. Kaldı ki bu tür sistemler gerçek zihni olan teknolojiler olarak iddia edilmektedir. Bu konunun detaylarını Turing bölümünde inceleyeceğiz. Diğer taraftan bilincin insanın algısıyla ilişkisi değerlendirilmesi gereken önemli bir konudur. Bu noktada yapay zekanın makine öğrenmesi ve gelişmiş algılama sistemleri sayesinde dış dünya ile iletişim kurabileceği düşünülebilir. Bu düşünce de sürecin algı kavramıyla özdeşleştirilmesine ve sistemlerin sensörler aracılığıyla algı sahibi olabileceği varsayılabilir. Bu bağlamda zihin felsefesinin ortak kavramlarından birisi olan algıdan kastın ne olduğuna dikkat çekmek gerekmektedir. Öyle ki sayısal sistemler sensörler aracılığıyla veriyi işleyerek sembollere dönüştürürler. Ne var ki alıcıların elde ettiği verilerin sembole dönüştürülmesinin insan algısına veya belleğine indirgenmesi felsefi açıdan sorunludur. O nedenle konunun yeni bir başlık altında detaylarıyla değerlendirilmesi gerekmektedir.

#### 1.4 Algı Nedir?

Yapay zeka sistemlerinin öne çıkan özelliklerinden birisi de gelişmiş sensörler aracılığıyla nesne tespitini gerçekleştirme ve sembollere dönüştürerek algoritmanın kullanıma sunmasıdır. Bu türden bir işlev özellikle yapay zeka araştırmacıları tarafından insan duyusunun ve algısının temsili bir karşılığı olarak düşünülmektedir. Haliyle sayısal sistemlerin “algılayabileceği” söylemiyle sıklıkla karşılaşmaktadır. Algının ne anlama geldiğini, insanda algının nasıl gerçekleştiği, sayısal sistemler için algının mümkün olup olmayacağı problemleri tartışılması gereken noktalardır. Dolayısıyla önceki başlıklarda izlediğimiz yol doğrultusunda

<sup>100</sup> Russell ve Norvig, 2010: 1040.

algı terimine ilişkin tespitleri öncelikle kavramsal çerçevede değerlendireceğiz. Böylece algı terimi ve sayısal sistemlerin ilişkine dair birtakım yorumlarda bulunmak mümkün olacaktır.

#### 1.4.1 Teknolojide ve Felsefede Algı Kavramı

Sayısal sistemlerin işlevlerinden nesnelerin tanınması ya da sembolleştirilmesi özellikle mühendislik faaliyetlerinde algıyla özdeşleştirilmektedir. Bir algı sisteminin temel bileşenleri duyuşsal veri işleme, veri temsili (ortam modelleme) ve makine öğrenimi tabanlı algoritmalarıdır. Robotik algının alt alanlarına bazı örnekler: engel algılama, nesne tanıma, anlamsal yer sınıflandırması, üç boyutlu ortam gösterimi, jest ve ses tanıma, aktivite sınıflandırması, arazi sınıflandırması, yol tespiti, araç tespiti, yaya tespiti, nesne takibi, insan tespiti ve ortam değişikliği tespiti şeklinde sıralanmaktadır. Mevcut algı ve yapay zekanın insan performansına ne kadar yaklaştığı sorusuna Szeliski, iki yaşındaki bir çocuğun performansına ulaşmakta zorlandığı cevabını vermektedir. Ancak günümüz Evrişimli Sinir Ağları<sup>101</sup> sayesinde sınırlı alanlarda (Örneğin Büyük Ölçekli Görsel Tanıma'da) süper insan sınıflandırma performansına ulaştığını savunmaktadır.<sup>102</sup> Premebida ve arkadaşlarının çalışmalarında algı teriminden kastın ne olduğuna ilişkin bazı açıklamalar görülmektedir. Bu ifadelerden anlama ve muhakeme yeteneği algıyı tanımlamaktadır. Sayısal sistemlerde algılama zihin felsefesinde tartışılan “algı” kavramından çok farklıdır. Bu yönüyle felsefi bir terim olan algının kavramsal kökenlerinden koparılarak mekanik düzeye sıkıştırıldığını görmekteyiz. Bu tür girişimler terimin yalnızca doğasını tahrip etmeyip, insana dair algının kavramsal derinliğini de olumsuz etkilemektedir. O halde algı felsefi temelde değerlendirilmelidir. Önceki başlıklarda incelediğimiz terimlerde olduğu gibi kavramsal çözümlemeyi öncelikle anlamsal boyutlarıyla daha sonra düşünürlerin tespitleriyle değerlendirmek faydalı olacaktır.

Felsefe sözlüğünde algı için “Kişinin kendi çevresi ve bedeni hakkındaki bilgilerin çıkarılması ve kullanılması.”<sup>103</sup> dilimiz sözlüğünde ise “Bir şeye dikkati yönelterek o şeyin bilincine varma, idrak.”<sup>104</sup> tanımları verilmiştir. Ayrıca daha kapsayıcı bir ifadeyle algı Goldstein (2010) tarafından şu şekilde tanımlanmıştır: “Algılama alanı, duyuşların işleyişini ve

<sup>101</sup> Evrişimli sinir ağı bir görüntünün alınarak yorumlanıncaya kadar onun işlenmesidir. Örneğin, güneşi bir kağıt üzerinde çizmek gerektiğinde çıplak gözle bakmak güç olacağından, parlaklığı azaltmak için güneş gözlüğü takmak gerekecektir. Bu süreç evrişimli bir ağı ilk işlemlerinden olan renk azaltma filtresi olarak düşünülebilir. Bu sayede güneşin kaba taslak görünümünü çizmek mümkün hale gelecektir. İlerleyen süreçlerde ortaya çıkan şeklin üzerinden birkaç kez geçerek, görüntüyü aşamalı olarak onun bir temsiline indirgemek mümkün hale gelecektir. Evrişim kelimesi baktığımız güneşin görüntüsünü bir çizime dönüştürme anlamına gelir. Güneşi yakalamak için birçok göz hareketi yapmanın sonucunda her alanı yakalayarak çizime ulaşılmıştır. Detaylı bilgi için bk. Rothman, 2018: 187,188.

<sup>102</sup> Premebida vd., 2018:112-123.

<sup>103</sup> Audi, 1999: 654.

<sup>104</sup> <http://sozluk.gov.tr/> (erişim tarihi: 10.02.2019).

duyuların uyarılmasından kaynaklanan deneyimler ve davranışları açıklamakla ilgilidir.”<sup>105</sup> Bu çerçevede algının yalnızca beş duyu organına indirgenen bir olgu olduğu fark edilmektedir. Dolayısıyla beş duyuyu taklit edebilen bir sistemin algısal nitelikleri taşıyabileceği fikri düşünülebilir. Ancak algı ve duyu terimlerine felsefi bakış bu düşünceyi boşa çıkarmaktadır. En yalın haliyle algı kavramı için felsefi bir ifade şöyledir: “Algı yalnızca beyne sinyal gönderen reseptörler tarafından belirlenmez. Algının gerçekleştiği ortam, algılayan kişilerin kültürel bilgisi, beklentileri gibi pek çok faktörler de algılanan şeyi etkileyebilir.”<sup>106</sup> Goldstein’in yorumundan anlaşılacağı üzere algı mekanik bir zemine indirgenemez. Çünkü Audi (1999) algıyı; “algılayıcı”, “algılanan nesne”, “nesnenin algılayıcıya görüldüğü duyusal deneyim” ve “nesne ile algılayıcı arasındaki nedensel ilişki” olmak üzere dört kritik unsurla açıklar. Bu yönlerden bazılarının psikoloji ve algı nörofizyolojisine aitken, diğerlerinin felsefi soruşturmaların konusu olduğunu belirtmektedir.<sup>107</sup> Bu hususta Ayer ve Cohen (1977) algı kavramının beyin merkezli sinirlerde gerçekleşen değişim sürecine sıkıştırılmaksızın zihinsel süreçler bağlamında anlaşılması gerektiğini ifade etmiştir.

“Ayer’in nedensel algı kuramı bağlamında değerlendirdiği algı; en basit haliyle kağıt parçası olarak tanımlanan baktığımız bu şeyin nedensel açıklaması, uygun türdeki fiziksel bir nesneden gelen ışığın gözleri ısınlaması ve bazı süreçlerin beyinde gerçekleşmesidir. Ancak algı bu süreçlerle sınırlı olmadığı gibi, ortaya çıkması için zihinsel faktörlerin de dahil edilmesi hususu tartışmaya açıktır. Ancak bu koşulların gerekli olduğu yönünde fikir birliği bulunmaktadır. Bu hususu ortaya çıkarabilecek şey felsefi sonuçlardır.”<sup>108</sup>

Ayer’in ifadesinden açıkça anlaşılmaktadır ki algı terimi mekanik bir sistemde gerçekleşen algoritmaya dayalı “tarama” işlevinden çok daha fazlasıdır. Öyle ki her ne kadar tartışmalı da olsa algıyı anlamak için felsefenin önemli görevleri olduğuna dikkat çekmektedir. Bu noktada Hume (2009) şu yorumuyla algının kapsamına dikkat çekmiştir: “Buna ek olarak her izlenim, dış ve iç, tutkular, sevgiler, duyular, ağrılar ve zevkler, başlangıçta aynı temeldedir ve aralarında gözlemleyebileceğimiz diğer farklılıklar ne olursa olsun, hepsi, gerçek renklerinde izlenimler veya algılamalar olarak görünürler.”<sup>109</sup> Öte yandan Reid (1941), algıyı duyum ile karşılaştırarak; duyumun yoğunlukla his alanına çekildiğini, algının da büyük ölçüde inançlar ve fikirler kapsamına alındığını ifade etmektedir.

“Duyum başlı başına, dışsal herhangi bir nesnede ne inancı ne de kavramı ima eder. O, hisseden bir varlığı ve bu varlığın etkilendiği belirli bir usulü gerektirir ama başka bir şey gerektirmez. 'Algı', dışsal olan şeyin

<sup>105</sup> Goldstein, 2010: 27.

<sup>106</sup> Goldstein, 2010: 28.

<sup>107</sup> Audi, 2003: 32.

<sup>108</sup> Ayer ve Cohen, 1977:105.

<sup>109</sup> Hume, 2009: 304.

doğrudan bir inancı ve idesi anlamına gelir- hem algılayan akıldan hem de algı eyleminden farklı bir şey.”<sup>110</sup>

Duyum ve algının ayırımından farklı olarak Russell (1912) tespitini görünüm ve gerçeklik ayırımı üzerine yapmıştır. “Bu yüzden doğrudan gördüğümüz ve hissettiğimiz şey, sadece arkasındaki bazı ‘gerçekliğin’ bir işareti olduğuna inandığımız ‘görünüm’dür.”<sup>111</sup> ifadesiyle Russell, duyumun bize sağlayabileceği verinin görünümlerden öteye gidemeyeceğini işaret ederek algının derinliğini ima etmektedir. Ayrıca Fichte’ye göndermede bulunan Feuerbach’ın (1866) duyum ve algıya ilişkin ifadesi Reid ve Russell’ın görüşlerini destekler niteliktedir:

"Sen, diyor klasik idealist Fichte, sırf şeyleri gördüğün, duyduğun, hissettiğin için onları gerçek olarak kendi dışında şeyler olarak düşünüyorsun/varsayıyorsun. Oysaki görme, hissetme, duyma [dediğimiz şeyler] yalnızca duyumlardır, sadece senin duyumsamaların, kendine dair bilincinin belirlenimleridirler çünkü senin gördüğün, hissettiğin ve kendinin bilincinde [olduğun durumlarda] görüyor ve hissediyorsun. Öyleyse sen [kendi dışındaki] nesneleri duyumsuyor değilsin, aksine duyumsamaları duyumsuyorsun. Her algıda sen her zaman senin kendine özgü durumunu algılıyorsun.”<sup>112</sup>

Fichte yukarıdaki ifadesinde duyumu beş duyu organlarıyla sınırlarken; bunun ötesinde duyularımızı duymamıza olanak tanıyan algıya göndermede bulunmaktadır. Algı, bu yönüyle kendinin bilincinde olan bir varlığın eylemini nitelemektedir. Fichte her ne kadar algı ve duyum arasında bir tür katman farkını ima etse de Kant’ın (1998) “İçeriksiz düşünceler boş, kavramsız sezgiler kördür.” ünlü özdeyişi; nesneler için duyuların, düşünce için de anlayışın gerekliliğini vurgulamaktadır. Öyle ki duyum ve anlayışın birbirine üstünlüğü olmadığı gibi; anlama, yani duyum olmaksızın algının hiçbir şeyi sezemeyeceğini, algı olmaksızın da duyuların hiçbir şeyi düşünemeyeceğini kastetmektedir.<sup>113</sup> Ayrıca Berkeley’e (2002) bakıldığında var olabilmenin ön koşulu zihinsel süreçlerle yani algılamayla mümkündür. Bu da algılayan öznenin var olduğu, dolayısıyla varoluşun -yani gerçekliğin- ancak algılayan özneye mahsus olduğu anlamına gelmektedir. Berkeley bu ifadesini, “Var olmak algılanmaktır. (*Esse est Percipi.*)”<sup>114</sup> özdeyişiyle dile getirmiştir.

Algı teriminin bilişimsel ve felsefi kullanımlarına bakıldığında ilk paragraflarda sözünü ettiğimiz türden bir anlam daralmasının örneğini teknolojiye de görmekteyiz. Oysa algının felsefi yönleri çok daha kapsamlıdır. Bu bilgiler ışığında algı teriminin insan doğasını belirlemede önemli görevler taşıdığı gibi tarihsel öneme sahip olduğu öne sürülebilir. İnsanın

<sup>110</sup> Reid, 1941: 155.

<sup>111</sup> Russel, 1912: 16

<sup>112</sup> Feuerbach, 1866: 185.

<sup>113</sup> Kant, 1998: 193-194.

<sup>114</sup> Berkeley, 2002: 12-13.

zihin-beden bütünlüğünü belirleyen kavramlardan algıyı, mekanik sistemlerin basit eylemlerine indirgemek hem insanı mekanize edecek hem de kavramın tarihsel arka planını gölgeleyecektir. Öte yandan felsefi bir terim olan algının sanata kattığı değer ve yansımalar algının sanat içinde değerlendirilmesi zorunluluğunu doğurmaktadır. Bu bağlamda varoluşçu düşünürler algının varlığını sanat çevresinde tartışmışlardır. Bu yönüyle algı diğer bir yönüyle özne ile nesnenin etkileşiminde anlaşılması gerekmektedir. Öyle ki bu türden bir sorgulama yeni bir başlık altında incelenecektir.

#### 1.4.2 Özne-Nesne Bağlamında Algı

Görme eyleminde görenin göz olup olmadığı sorunsalı önemli bir tartışma alanıdır. Maurice Merleau-Ponty bu konuyu tartışarak görme eyleminin göze mi yoksa görünen nesneye mi içkin olduğunu yanıtlamaya çalışmıştır. Bu bağlamda görme eylemi ve görünen nesnenin ilişkisine dair bir açıklamayı algıyla açıklamaya girişmiştir.

“İki gözüm olduğu için, nesneyi çift görmeliyim ve eğer sadece birini görüyorsam bunun nedeni, iki görüntü aracılığıyla uzaktaki bir nesnenin fikrini inşa etmemdir. Algı, duyularımızın bedensel uyarılarına sağladığı işaretlerin bir "yorumu" haline gelir. Bu da zihnin "izlenimlerini kendisine açıklamak" için evrimleştiği bir "hipotez"e dönüşür. Retinal izlenimler, gerçek yansıma tarafından içeriden kavranan algı eyleminin kendisi olmak yerine, bir kez daha bedenini sağlamadığını sağlamaktan sorumlu olan salt bir algı "faktörü" haline gelir. Bu aşkın bir etkinlik olmak yerine, basit bir sonuca varmanın mantıklı bir etkinliğidir.”<sup>115</sup>

Merleau-Ponty, algıyı açıklamanın yolunu zihinde oluşan izlenimlerin yoruma dönüşme sürecinde bulmuştur. Bu bağlamda salt görünen nesnenin izlenimleri anlamlı hale getirmesi için yeterli değildir. Zihnin izlenimleri işleyerek bir yoruma vardığını bu yorumun da algıdan beslendiği düşünülmektedir. Bu türden bir tartışma algının sanatla ilişkisini özne ve nesne bağlamında sorgulatmaktadır. Paul Klee'nin (1920) “Sanat, görünenin kopyasını çıkarmaz; daha doğrusu onu görünür kılar.”<sup>116</sup> ifadesinde belirttiği üzere sanat, teknolojiye olduğu gibi gerçekliğin sembollerle çoğaltılmasından ibaret değildir. O halde özne-nesne bağlamında algının sanata yansımaları terimin kapsamını genişletmekle kalmadığı gibi, tanımını belirleyen nitelikleri de artırmaktadır. Varoluşçu düşünürlere bakıldığında sayısal sistemlerdeki temsili gerçekliklerin algıyla bağdaştırılması kavramı sıkışmış bir forma bürünmektedir. Oysa algı; öznenin nesneye, nesnenin de özneye işlediği, birbirinin içinde dolaştığı süreç olarak yorumlanabilir. Algının neliğini sanata ilişkin örneklerle ifade eden Cezanne'nin “İnsan renk sürdüğü ölçüde desen çizer.” sözüne atfen Merleau-Ponty, algılanan dünyada renkler ve

<sup>115</sup> Ponty, 1958: 39.

<sup>116</sup> Klee, 1920: 1.

desenler bağlamında zihin-beden bütünlüğünü işaret etmiştir. Sürekliliğin ve bütünlüğün, tablo üzerindeki renk cümbüşüyle var olduğu ve birbirinden kesin çizgilerle ayıramayacağını, asıl olanın klasik resim sanatının perspektif anlayışına karşı durmak olduğunu iddia etmiştir. Çünkü perspektif bir yanılsama sunar ve algılanan şeyin uzlaştırılarak resmedilmesi anlamına gelir. Bu nedenle Merleau-Ponty, algısal deneyimin oluşturduğu duygusal dinamiklere şu ifadelerle dikkat çekmektedir: “Cezanne’dan beri birçok ressamın geometrik perspektif yasasına başkaldırmasının nedeni, manzaranın gözlerimizin dibinde nasıl doğduğunu yakalayıp yansıtmak istemeleri, çözümleyici bir yaklaşımın ötesinde görsel algı deneyiminin verdiği duyguya ayak uydurmak istemeleridir.”<sup>117</sup>

Bu ifadeden algı teriminin duyuların ötesinde duygusal çeşitliliklerle ilişki kurulması gereken bir kavram olduğu anlaşılmaktadır. Bir insan olan ressamın amacı duygularını sanatına yansıtabilmektir. Bu hususta algı-duygu ilişkisi, Merleau-Ponty’e göre gözden kaçırılmaması gereken bir noktadır. Bu nedenle “İnsan bir vücut *ve* ruh değil, bir vücut *ile* bir ruhtur.”<sup>118</sup> demiştir. Burada Merleau-Ponty, “*ve*” bağlacını birbirinden ayıran farklı konumlara yerleştiren, “*ile*” bağlacını ise; bir araya getiren, birlikte olduğu düşünülen anlamında kullanmıştır. Bu ilişki düalist bakışın ayrıklığını “*ve*” ifade ederken, zihin-beden birliğini “*ile*” ifade ettiği sonucu çıkarılabilir. Bu bağlamda yapay zeka bir tür “*ve*” ayrımı gölgesinde var olurken, insan her yönüyle “*ile*” birliğinde var oluşunu sürdürmektedir.

Merleau-Ponty’nin algıya ilişkin görüşlerini ve sayısal sistemlerde kullanıldığı kadarıyla algı düşünüldüğünde robotik sistemlerin eylemlerini algı terimiyle açıklamanın uygunsuz olacağı sonucu çıkmaktadır. Kaldı ki beden ve zihin ayırık olgular olmayıp, Aristoteles’ten okuduğumuz kadarıyla bütünsel ve iç içe geçmiştir. Ayrıca Merleau-Ponty’nin ifade ettiği üzere perspektif algısının duygusal anlamda ortaya koyabileceği yanılsamalar bulunmaktadır. Yani duyuların sağladığı verilerin sembollere dönüşmesi perspektif yanılgısında olduğu türden yanılgıları beraberinde getirecektir. Bu düşünceye temel oluşturabilecek açıklamayı Satre’nin şu ifadesinde görmek mümkündür: sarı limonun ekşiliği, kavanozdaki reçelin yapışkan soğukluğu, tadının parmaklarda hissedilmesi, havuzdaki suyun akışkanlığı, ılıkılığı, mavimsi rengi, dalganın oynaklığı gibi nitelikleri, ruhun süzgecinden geçirebilen duyguyu hissettirebilecek bir sanat eseri<sup>119</sup> insan yeteneğinin meyveleri olarak belirtilmektedir. Bu nitelikleri yapay zeka sistemlerinin ortaya koyması olasılığı özne-nesne ilişkisinde ortaya çıkan “yorum” bağlamında muhtemel değildir. Hatırlanacağı üzere algoritma temelli sistemlerin sensörler aracılığıyla belirli nesneleri tarayarak sembollere dönüştürmesi,

<sup>117</sup> Ponty, 2005: 23.

<sup>118</sup> Ponty, 2005: 26.

<sup>119</sup> Satre, 2011: 266.

nesnelerin bizatihi kendisini değil özünden uzak olan temsili izdüşümünü tanımasına imkan verdiğini belirtmiştik. Bu da yapay zeka sistemlerin salt tasarımlara ve sembollere sıkıştığının diğer bir göstergesidir. Merleau-Ponty, “algı dünyasında şeyleri ve şeylerin görünme biçimlerini birbirinden ayırmak olanaksızdır.”<sup>120</sup> ifadesiyle algının varlıkların özüne (*essentia*) ve varoluşuna (*existence*) işlediğini belirtmektedir.

## 1.5 Sonuç

Giriş kısmından buraya kadar üzerinde durduğumuz düşünceleri toplamak gerekirse akıl, zeka, bilinç ve algı gibi zihin felsefesinin ortak kavramlarını inceledik. Tezin başlığı düşünüldüğünde zihne ilişkin kavramların ilk bölümde yer almasının nedenleri sorgulanabilir. Zihne ve alt bileşenlerine değindiğimiz ilk bölümde özellikle insan zekası ile akıllı arasındaki farka felsefî açıdan bir bakış getirdik. Bu farkın anlaşılması hem akıl hem de zekanın mahiyetinin belirlenmesiyle insanla karşılaştırılan yapay zekanın gelecekteki etkilerini anlamada bir yöntem sunabilir. Bu nedenle zihinsel süreçlere yer vererek yapay zekada geçen “zeka” terimini anlamaya çalıştık. Söz konusu kavramsal hususlar, yapay zekanın geleceğine ilişkin sonuçlarını öngörebileceğimiz alanlar olup özne – nesne ve nesne – özne ilişkisine bağlı olarak insanın teknolojiyle bağını anlamamıza yardımcı olacaktır. Yani insan doğasını belirleyen kavramların, nesne olan yapay zeka sistemlerinde gerçekleşmesinin olanağını zihne ait kavramlarla anlamaya çalıştık. Bu nedenle insan doğasını belirleyen kavramlara ilk bölümde yer verilmiştir. Kaldı ki tartıştığımız terimlerin göz ardı edilmesi, yapay zekanın anlaşılmasını kısıtlı hale getirecektir. Nitekim yapay zeka tartışmalarının yoğunlukla insan referansında yapıldığı unutulmamalıdır. Bu yönüyle zihin felsefesi bağlamında bir değerlendirme gereği görülmüştür. Tartışılan kavramlara yer verilmesi, tarihsel, bilimsel, kavramsal köken sayesinde yapay zekanın dün, bugün ve özellikle geleceğine ilişkin sağlayacağı bilgiler açısından için önemlidir. Bu çalışmanın amacı, yapay zekanın gelecekte ortaya koyacağı etik ve risk faktörleri bağlamında sonuçlar olduğu düşünüldüğünde geleceği okumanın koşulunun kavramsal çözümleme olduğunu hatırlanmalıdır. Bu çabanın sağlayacağı en önemli katkı zihnin bileşenlerinin etraflıca değerlendirmesiyle dar bir açıda sıkışmış, kısıtlı, mekanik bir yapay zeka düşüncesi ortaya çıkmaktadır. Yani giriş bölümünde yer verdiğimiz tanımlarda insana ilişkin pek çok niteliğin yapay zekayı tanımlamada kullanıldığına dikkat çekilmişti. Oysa her bir kavram tek tek incelendiğinde yapay zekanın görüldüğü kadar güçlü olmadığı fark edilmektedir. Bu sonuç akıl, zeka, bilinç ve algı gibi yetilerin mekanik düzeneğe indirgenemeyeceğinin felsefî çerçevede imasından ibarettir.

<sup>120</sup> Ponty, 2005: 60.

Bir düşünce deneyine başvuracak olursak insan beklenmedik her bir karmaşık durumun üstesinden gelebilmektedir. Oysa sayısal sistemler, her bir karmaşık durum için küçük yazılımlara ya da modüllere ihtiyaç duymaktadır. Bu bağlamda fenomenolojik çeşitlilik düşünüldüğünde bütün fenomenlerin bilgisini taşıyacak modülü sağlamanın imkansızlığıyla karşı karşıya kalınacaktır. Ancak insan söz konusu olduğunda hiçbir eklentiye bağlı kalımsızın beklenmedik durumlar herhangi bir sorun teşkil etmemektedir. Bu da insanın sağduyusuna bağlı bir durumdur. O halde sayısal sistemlerin büyük veri yığınları olmaksızın işlemde bulunamayacağını ifade etmek mümkündür. Oysa insan zihni kendisine bir veri sağlanmaksızın, deneyimlemediği bir şey üzerinde düşünce ya da fikir sahibi olabilmektedir. İnsan izlenimleriyle, yorumlarıyla ve belirlenimleriyle olgulara yeni anlamlar yüklemektedir. Oysa algoritma bu türden bir algıyı gerçekleştirememektedir. Bu husus, duysal veri çekmenin ötesinde içsel bir kavrama esasına dayanmaktadır. Daha açık bir örnek üzerinden ilerlemek gerekirse aşağılayıcı bir ifade karşısında utanma duygusuna bağlı olarak insanın yüzü kızarabilir. Bu kızarma yalnızca yüzdeki damarlara dolan kandan ibaret olmayıp, duygulanımların eşgüdümünde semantik bir süreci de tetiklemektedir. Öyle ki insan kimi zaman böyle bir ifadeye tepkisiz kalabilir, yüzü kızarabilir, utanabilir ya da bambaşka tepkiler verebilir. Bunun ölçütü insanın içinde bulunduğu ruh hali, coğrafya, iklim, eğitim düzeyi gibi pek çok kültürel ve ahlaki dinamikle ilgilidir. O halde insana özgü eylemler zihin ile bedenin iç içe geçmişliğiyle ortaya çıkan eşgüdümlü süreçlerken, algoritmaların etkileri ve tepkileri salt elektriksel süreçlerden ibarettir. Haliyle yapay zekanın insana özgü zihinsel süreçleri gerçek anlamda gerçekleştirmesinin ihtimalini olumlamak güçtür. Bunun en açık nedeni, özne ve nesnenin ilişkisi ontolojik, semantik ve duysal vb. nitelikler barındırmasıdır. Nesnenin nesneyle ilişkisinde bu türden bir etkileşimi ifade etmenin kavramsal altyapısı felsefi olarak oldukça zayıftır.

Buraya kadar insan aklının enginliğine yüceltici bir tutum söz konusudur. Şimdi sayısal sistemlerin insan kavrayışı karşısındaki konumu, kapasitesi, gelişmiş yönleri ve kısıtlılıkları gibi konuların tartışılması ihtiyacı hissedilmektedir. Bu konuya ikinci bölümde yer vereceğimizden kısaca konunun kapsamını özetleme akış açısından faydalı olacaktır. İkinci bölümde yapay zekaya ilişkin tartışmaların ve teknik detayların incelenmesi özellikle Kartezyen düalizminin yapay zekaya temel oluşturmasıyla ilgilidir. XVII. yüzyılda Descartes'ın benimsediği bu ekolün, 1950 senesinde Alan Turing'e referans olması yapay zekanın uygulama alanının arka planını oluşturan bir gelişme olarak değerlendirilebilir. Bu nedenle tarihteki yapay zeka tartışmaları ve yapay zeka tekniklerinin akıl referansında değerlendirilmesi bu sistemlerin olanaklarını anlamamızı sağlayacaktır. Bu yönüyle geleceğe

ilişkin etkileri ve alınması gereken inisiyatifleri belirlememize yardımcı olacaktır. Bu yaklaşım felsefi açıdan kavramsal, uygulamalı açıdan toplumsal, hukuki, psikolojik vb. incelemeyi gerektirmektedir. O nedenle çalışmamız yapay zekanın uygulanmasıyla doğacak sonuçları incelerken, esasında felsefi sonuçlara erişme noktasında kavramsal bir çizgide ilerlemektedir.



“Sentaks, semantik için tek başına yeterli değildir.”

John Rogers Searle

## İKİNCİ BÖLÜM

### TARİHTE YAPAY ZEKA İLE İLGİLİ TARTIŞMALAR

#### 2.1 Yapay Zekanın İşleyişi ve İnsanın Kavrayışı

Başta akıl olmak üzere zihnin fakültelerine ilk bölümde yer verdik. Bu bölümde ise yapay zeka sistemlerinin zihinsel süreçlerle karşılaştırılması yapılacaktır. Burada amaç yapay zekanın kavramsal ve uygulamalı anlamda gücüne ve sınırlılıklarına değinmektir. Öyle ki algoritmanın sayısal sistemlerdeki karşılığı insan zihninin sınırlarını aşabilecek güce erişme olasılığı tartışılmaktadır. İlk olarak insanın felsefe yapma yeteneği düşünüldüğünde yapay zekanın felsefe yapma olasılığı yapay zeka ve insan arasındaki fark belirginleştirebilecek bir yaklaşım olabilir. Bu bağlamda beynin öğrenme kabiliyetinin keşfine dayanan zeka patlaması Bostrom’un (2014) “süper zeka” ifadesiyle bağdaştırılabilir. Yüklemler (*uploads*) olarak da bilinen tam beyin taklidi, biyolojik bir beyin hesaplama yapısını tarayarak ve yakından modelleyerek üretme anlamına gelmektedir. Öyle ki bu türden bir keşif felsefi ilerlemeyi de etkileyecek bir gelişme olarak görülmektedir. Felsefi ilerlemenin felsefe yapmaktan ziyade dolaylı bir yoldan doruğa çıkarılabileceği ifade edilmektedir. Süper zekanın mevcut düşünürlerden çok daha iyi performans göstereceği bir alan, bilim ve felsefedeki temel soruları yanıtlamak olarak gösterilmektedir. Yine de bu türden bir süper zeka, şu anda hazır olmadığı ve uzun süre hazır olamayacağımız bir meydan okuma olarak kabul edilmektedir. Böylesi bir zeka patlamasının ne zaman olacağı konusunda çok az fikrimiz olduğunu düşünen Bostrom, ancak cihazı kulağımıza tutarsak hafif bir tıkırtı sesi duyabileceğimiz alayına başvurmaktadır.<sup>121</sup> Bu tür hipotetik reflekslerin ortaya çıkmasına neden olan sinirbilimci bakışa Crick’in (1994) argümanı örnek gösterilebilir. “Şaşırtıcı hipotez şudur: ‘Siz’, sevinçleriniz ve üzüntüleriniz, anılarınız ve hırslarınız, kişisel kimlik duygunuz ve özgür iradeniz, aslında sinir hücrelerinin ve bunlarla ilişkili moleküllerin geniş bir topluluğunun davranışından başka bir şey değilsiniz. Siz aslında bir paket nörondan başka bir şey değilsiniz.”<sup>122</sup> Crick hipotezinin esaslı düşüncesi; beynin yeterince doğru bir simülasyonu dijital bir bilgisayarda tam olarak örneklendirilebilirse dijital bir zihinde semantik süreçleri ortaya koyabilme mümkün olabilecektir. O halde Crick’in de dediği gibi beynin keşfi zihnin keşfi için yeterli midir? Bu

<sup>121</sup> Bostrom, 2014: 47, 48, 297.

<sup>122</sup> Crick, 1994: 3.

türden bir sorgulama için öncelikle Turing'e bakmamız gerekecektir. Sayısal sistemlerin işlem kapasitesinin kritik eşiği aşmasıyla zihin gücüne erişebileceği hatta aşabileceği düşüncesi Turing tarafından öne sürülmüştür.<sup>123</sup> Söz konusu iddianın geçerlilik tespiti iddiaya konu olan insan zihni ve algoritmanın karşılaştırmalı olarak değerlendirilmesini gerektirir. Bu bağlamda yapay zekanın esaslı amacı ve işlevine ilişkin görüşlerini Boden (2016) teknolojik ve bilimsel olmak üzere iki başlık altında sınıflandırmıştır. Teknolojik amacı, faydalı şeyler ortaya koymak amacıyla bilgisayarın birer araç olarak kullanılması olarak düşünülmektedir. Bilimsel amaç, insanlar ve diğer canlılar hakkında yanıtlanması güç soruların yapay zeka sistemlerince gerçekleştirilmesidir. Yapay zeka araştırmacılarından pek çoğu bu alanlardan sadece birine odaklanırken, bazıları ise her iki amacı birden değerlendirmektedir.<sup>124</sup> Boden'in ifadesi düşünüldüğünde bilimsel amaç *episteme* kavramıyla ilişkilendirilebilir. Bu amaç doğrultusunda hız ve güvenilirliğin en üst düzeyine ulaştığı kusursuz sistemlerce bilimsel etkinliğin geliştirilmesi hedeflenmektedir.

Öte yandan yapay zekanın düşünmeyi gerçekleştirmesine ilişkin tartışma, düşüncenin beyinde konumlandırıldığı anlayıştan hareketle bazı biyolojik söylemler doğurmuştur. Schade ve Smith'in (1970) beyin-bilgisayar ilişkisi bağlamında gerçekleştirdiği araştırmanın biyolojik ve mekanik yönlerinde düşüncenin konumu olarak iddia edilen insan beyninin bilgisayar donanımına benzetildiği anlaşılmaktadır. Schade ve Smith'e göre girdi, işleme, depolama ve çıktı cihazları içeren birimlerin birleşiminden oluşan bilgi işleme sistemleri, kendi kendini kontrol ederek verileri yüksek hızlarda işlemek amacıyla tasarlanmıştır. Hem tasarım açısından hem de bilgi işleme kapasitesi açısından en dikkat çekici sistemin insan beyni olarak gösterildiği çalışmada; elektronik makinelerin hızından yoksun olmasına rağmen insan bilgisayarının (beyninin) kapasitesinin bilinen herhangi bir elektronik hesaplama sistemini fazlasıyla aştığı da kabul edilmektedir. Temelleri ikili aritmetik olan (0 ve 1) dijital bilgisayarların ayrı veri bitleriyle çalıştığı ve bu sayede son derece hassas hale getirilmesiyle hız, doğruluk ve güvenilirliğin amaçlandığı belirtilmektedir. Sınır sisteminin üstün özelliklerini sıradan bilgisayar sistemleriyle kıyaslamak için beynin dijital, analog ve hibrit özelliklerini bilmek gerektiği kadar girdi, işleme, depolama ve çıktı cihazlarının da bilinmesi önemli görülmektedir.<sup>125</sup> Bu noktada bilgisayarın programı işlemesi ve insan beynin de zekayı işlemesi arasında kurulan özdeşliğe Searle (1996) cephesinden önemli bir eleştiri gelmiştir. Bu bağlamda "Böylece, örneğin bira kutularından yel gücüyle çalışan bir bilgisayar yaparsanız ve

<sup>123</sup> Turing, 1950: 433-460.

<sup>124</sup> Boden, 2016: 5.

<sup>125</sup> Schade ve Smith, 1970: 9.

onun doğru bir programı varsa bu bilgisayar zekaya sahip olacaktır.”<sup>126</sup> ifadesiyle bu özdeşliğe eleştirisini yöneltmiştir.

Öte yandan insanın bilgiyi hatırlaması ya da çağırmasıyla ilgili olarak Dreyfus (1986) tarafından yapılan bir değerlendirmede, bilgisayarın veriyi çekmesi ile insanın bilgiyi anımsaması farkına değinmiştir. Dreyfus’a göre bilgisayarlar bellekte bulunan bilgiyi tespit edebilmek amacıyla veri bankasında bir tarama gerçekleştirmektedir. Tarama sonucunda bilginin varlığı tespit edilebilirse bilginin seçilebileceğini ifade etmektedir. Öte yandan bazı bilgilerin veri bankasında bulunmadığının belirlenmesi için bilgisayarın bu eksikliği tespit etmek amacıyla daha fazla tarama yapması gerekecektir. Yani iki durumda da aradaki farkı çözmek için sorgulama ya da tarama gereksinimi bulunmaktadır. Ancak aynı eserde görüşüne değinilen Richard Shaffer, zihnin bu şekilde çalışmadığı; bilgisayarların aksine içerik ve anlamları doğrudan fark edebildiğini savunmaktadır. İnsanın bilgiyi hatırlaması ya da soruları cevaplaması bağlamında listelere gerek olmadığı gibi, hazır bir cevabımız olmasa da yeterince uzun düşünmeyle bilgiyi hatırlayabileceğimiz ifade edilmektedir. Birisinin ne zaman doğduğunu bilmiyorsa, bilmediğimizin farkında olacağımızdan gereksiz sorgulamaya ihtiyaç duyulmayacaktır. İnsanın bilgiyi anımsama yetisini olağanüstü kabul eden Dreyfus, yapay zekada kullanılan listelerin veya diğer veri yapılarının kısıtlılığına ilişkin bir örnek olduğunu ifade etmiştir.<sup>127</sup>

Diğer bir yönden bakıldığında sayısal sistemler ve insan karşılaştırmasında öne çıkan kavramın “bilinç” olduğu gözlemlenmektedir. Bu hususta bilinç problemini sayısal temelde çözmek için beyini bilgisayar donanımıyla, zihni de yazılımla bağdaştırmak yapay zeka araştırmacılarının yaklaşımlarındandır. Bu doğrultuda bilincin konumuna ilişkin ayrıntılı tespitleri olan Penrose (1999), bilincin biyolojik temellerle açıklandığı ve beyin merkezli alanlarda (korteks, talamus, hipokampus gibi) konumlandırıldığına ilişkin yaygın görüşlerin varlığına dikkat çekmiştir. Bu görüşe karşı tavır alan Penrose gerekçesini farklı canlılarda bilinci değerlendirerek açıklamaya çalışmıştır. Bu bağlamda sürüngenlerden, deniz canlılarına değin farklı seviyelerde de olsa bilincin var olmadığına ilişkin bir delilden yoksun olduğumuzu belirtmektedir. Dolayısıyla merkeze kendini koyan insanın, bilincin yalnızca kendi türüne bahsedildiği iddiasıyla pek çok hakkı kendine reva gördüğünü iddia etmektedir. Dünyada “bilinç sahibi olmanın” sadece insan türüne özgü bir şey olduğundan kuşku duyan Penrose bilince ilişkin öncelikle yöntem probleminin bulunduğunu ima etmektedir.<sup>128</sup>

<sup>126</sup> Searle, 1996:39.

<sup>127</sup> Dreyfus, 1986: 69.

<sup>128</sup> Penrose, 1999: 99.

Davranışçı savunmalara eleştirel tutumunu sürdüren Penrose bilincin derinliğine dikkat çekerek güçlü yapay zekayı eleştirmektedir. Penrose'a göre eğer beyin gerçekten bilince ev sahipliği yapıyorsa aynı zamanda zeka da beyinde gerçekleşen bir olguysa ruhun da beyinde yer alması gerekmektedir. Bu düşüncenin güçlü yapay zekaya dayanak oluşturabileceğinden endişelenmektedir. Bu bağlamda bilinci, algoritmanın bir karmaşıklığı, derinliği ya da zor kavranan uçsuz bucaksızlığı olarak değerlendiren yaklaşımı tehlikeli bulan Penrose bu bakış açısının güçlü yapay zeka görüşünü güçlendirebileceğini iddia etmektedir. Kaldı ki bu değerlendirmeye göre bilincin yönetimi beyin korteksindeki algoritmalarından geçmekte olup bu durumda güçlü yapay zekanın yapamayacağı bir şey kalmayacaktır.<sup>129</sup>

Yapay zekanın işlevi ve insan anlayışına ilişkin görüşlere bakıldığında bilinç tartışmasında Descartes'ın zihni bedenden tamamen ayrı konumlandırmasıyla, zihnin başka yapılarda ya da donanımlarda var olabileceği düşüncesi öne çıkmaktadır. Kaldı ki bedensel olanın donanıma, zihinsel olanın da yazılıma benzetilebileceğinin düşünsel arka planını Descartes'ta okumak mümkündür. İnsan bedeninin sistemini, hidrolik mekanizmaların işleyişine veya bir saat çarkının dönmesindeki devinime benzeten Descartes, "...tıpkı bir saat veya otomat yani kendiliğinden işleyen başka bir makine..."<sup>130</sup> ifadesiyle bedeni mekanizmaya indirgemıştır. Bu yorum felsefenin önemli bir alanını meşgul eden düalizm olarak da bilinen zihin-beden problemi olarak bilinmektedir. Dolayısıyla bu kavram temelinde yapay zekayı anlamak son derece önemli bir husus olduğundan, yeni bir başlıkta değinmek gerekmektedir.

## 2.2 Zihin-Beden Problemi Temelinde Yapay Zeka

Yaygın inanca göre yapay zekanın dünyaya yansımaları teknolojinin çıktılarıyla bağdaştırılmaktadır. Oysa Alan Turing ve John McCarthy'nin öncülük ettiği yapay zeka, arka planında Descartes'ın izlerini taşımaktadır. Düalizm bu ilerlemelerin temelini oluşturan ekol olarak düşünülebilir. Bu düşüncüyü destekleyen yapay zeka araştırmacılarına göre zihin tamamen biçimsel olarak tanımlanırken, diğerlerine göre zihin doğal biyolojik dünyaya indirgenemeyecek kadar uçsuz bucaksız bir olgudur. Bu hususu açmak gerekirse farklı düşünürlerin görüşlerine başvurma lüzumu doğmaktadır.

Searle (1996), zihin-beden probleminin çözümünü usandırıcı olarak nitelendirdiği bircilik, ikicilik, düşüncencilik, davranışçılık ve fizikselcilik gibi ekollere dayandırmaktadır.<sup>131</sup> Bu ekolleri 1998 yılında yayınladığı *Zihin Dil Toplum* adlı eserinde yeniden değerlendirerek öne çıkan ekolleri tanımlamıştır. Zihin-beden problemini başlıca düalist ve materyalist olarak

<sup>129</sup> Penrose, 1999: 97-101.

<sup>130</sup> Descartes, 2013: 9, 10.

<sup>131</sup> Searle, 1996: 17.

değerlendiren Searle, materyalistlerin bilinç sözcüğüne yer vermelerine rağmen bilinçten anladıkları şeyin farklı olduğuna değinmektedir. Her insanda var olduğu düşünülen bilinç olgusunun materyalist görüşte yok hükmünde bulunduğu değinmektedir. Düalizm ve materyalizm ekollerinin yaygınlığına dikkat çeken Searle, düalizmin “cevherci düalizm” ve “nitelikçi düalizm” kavramlarıyla tanımlamaya çalışmaktadır. Bu hususta Eski Çağ’a kadar uzanan cevherci düalizmin en ünlü savunucusu Descartes olarak gösterilmektedir. Bu görüşte zihin ve bedenin birbirinden tamamen ayrı konumlandırılması esasken, nitelikçi düalizmin metafiziksel olarak farklı nesnelerin iki çeşit özelliğini taşıdığı ifade edilmektedir. Söz gelimi bir şeyin ağırlığa ilişkin fiziksel niteliği ya da tam tersi olan üzüntü duymak gibi nitelikleri kastedilmektedir. Ayrıca tüm düalizm türlerinde ya fiziksel ya da zihinsel nitelikler olarak ayrı ayrı değerlendirmenin ya da sınıflandırmanın esas olduğu belirtilmektedir. Diğer bir deyişle bir tür tezatlık, zıtlık zorundallığı vurgulanmaktadır. Bu iki düalizm türlerinden öne çıkanını cevherci düalizm olarak işaret eden Searle (1998), felsefe çevresinde çoğunlukla materyalist görüşün benimsendiğini bildirmektedir. Materyalist felsefecilere göre maddenin üstünde ya da onu aşan herhangi bir bilincin olmadığını benimsendiğine dikkat çekmektedir. Materyalizmin öne çıkan türlerini örneklendiren Searle, zihni davranışlara ve davranışın alt unsurlarına bağlayarak açıklayan görüşü davranışçılık olarak tanımlamaktadır. Zihni, insan beyninin dinamikleriyle açıklayan görüşü de fizikselcilik olarak nitelendirmektedir. Fiziksel bir sistemin yapısındaki nedensellik ilişkisi sonucu ortaya çıkan davranışlar ve işlevsel durumları açıklayan görüşü de işlevselcilik olarak ifade etmektedir. Bu tanımlamaların ardından Searle gerek cevherci gerekse nitelikçi ekollerden düalizmin ya da materyalizmin doğruluğu konusunda şüphe duymaktadır. Yapılan yanlış kavramların geleneksel anlamında kullanımına bağlayarak diğer bir deyişle basmakalıp ifadelerden uzaklaşarak yepyeni bakış açılarıyla konuya değinmemiz gerektiğini önermektedir.<sup>132</sup>

Russell ve Norvig (2010) *Artificial Intelligence: A Modern Approach* eserinde söz konusu ekollere ilişkin kapsamlı değerlendirmelerde bulunmuşlardır. Descartes’ın mekanik anlayışla yönettiği zihin-beden probleminde düalizmi güçlü bir şekilde savunduğuna önemle dikkat çekilmiştir. Öyle ki bu mesele insan aklının (veya ruhunun) doğanın tamamen dışında veya fiziksel yasalardan muaf bir alanda konumlandırıldığı anlamına gelmektedir. Öte yandan makine olduğu varsayılan hayvanların bu ikili niteliği taşımadıkları belirtilmektedir. Düalizmin alternatiflerinden birisi olarak gösterilen materyalist görüşe göre zihin, fizik yasaları temelinde beynin işlevleriyle oluşturulmuştur. Öte yandan Aristoteles’te zihin kavramı ele alındığında bilgi ve eylem arasındaki bağlantının önemine dikkat çekilmektedir. Bilgi ve eylem arasındaki

<sup>132</sup> Searle, 1998: 55-57.

bağın yapay zeka için hayati öneme sahip olduğu, zekanın hem akıl yürütmeyi hem de eylemi gerektiren bir olgu olduğu vurgulanmaktadır. Dahası sadece eylemlerin nasıl gerekçelendirildiğini anlayarak davranışları gerekçelendirilebilen bir özne tasarlayabileceğimiz tartışmasını ortaya çıkarmaktadır. Aristoteles, eylemin sonucuna ilişkin amaçlar ve bilgi arasındaki mantıksal bağlantılarla söz konusu eylemlerin gerekçelendirildiğini iddia etmiştir. Bu iddiasına örnek niteliğinde bir gelişme, 2300 yıl sonra Newell ve Simon tarafından GPS programlarında algoritmanın uygulanması olarak gösterilmektedir.<sup>133</sup>

Bu doğrultuda benzer bir sorgulama da Searle'den (1996) gelmektedir. Searle maddi olan ile düşünsel olanın kökten ayrımına bağlı olarak anlam içermeyen maddi dünyanın nasıl olur da anlama ilişkin şeyleri kapsayabileceğini tartışmaktadır.

“Örneğin bilinçsiz fiziksel parçacıklardan başka hiçbir şey içermeyen dünyanın, bilinci de içerdiği doğru olabilir mi? Mekanik bir evren, nasıl olur da bilinçle devinen, yani bu dünyayı kendi kendilerine temsil eden insanları da içerebilir? Kısacası aslında anlamsız olan dünya, nasıl anlamı da içerebilir?”<sup>134</sup>

Bu sorulara cevaben miadı dolmuş terimlerin kullanıldığı maddeci ve düşünceli grupları eleştiren Searle (1996) akılsal fakültelerin başlıca dört özelliği olan; bilinç, yönelimsellik, öznellik ve akılsal nedensellik kavramlarıyla açıklamaktadır. Bilinci, insanın varoluşunun temel gerçekliği olarak kabul ederken, bilincin olmaması durumunda insana ait bütün niteliklerin boşa çıkacağını öne sürmektedir. Yönelimselliğin bilinçli ya da bilinçsiz olsun; aşk, nefret, tiksinti gibi aklın dışındaki duygusal dinamikleri ve dünyayı anlattığını vurgular. Öznelliği de bir insanının kendi hislerini ancak kendisinin hissedebileceği, başkalarının bunu hissedemeyeceği fikrine dayandırmaktadır. Düşüncelerimiz ve duygularımız gerçekten zihinsel şeyler ise akılsal nedenselliğin fiziksel şeyleri nasıl etkileyebileceği sorusuna dikkat çekmektedir. Bu noktada Searle, kolunu kaldırmaya karar verince kolunun kalktığını iddia eder. Yani duygu ve düşüncelerimizin fiziksel dünya üzerinde doğrudan nedensel bir etki oluşturduğunun altını çizmektedir. Netice itibarıyla Searle, zihinsel yaşamlarımızın bütün gerçek yönlerinin sözü konusu dört özelliğin bir arada değerlendirilmesiyle anlaşılabilirliğini önermektedir.<sup>135</sup> Bu düşüncenin düalist bakışa ters bir yaklaşım olarak düşünülmesi yanlış olmayacaktır.

Tartışmayı özellikle kavramsal eksende değerlendiren Searle'ün amacı, matematiksel bakış açısını hedef almaktır. Zira Searle, bu konuda çok dikkatli olmak gerektiğini aksi halde matematiksel bakış açısına göre herhangi bir şeyin (masada duran kalem dahi) bir bilgisayar olarak kabul edilebileceğini öne sürmektedir. Esaslı düşüncesini "Bir sayısal bilgisayar, açıklandığı gibi, düşünebilir mi?" sorusuyla açıklayan Searle, bu sorunun daha kapsamlı hali

<sup>133</sup> Russell ve Norvig, 2010: 5-7.

<sup>134</sup> Searle, 1996: 15,16.

<sup>135</sup> Searle, 1996: 17- 22.

olan "Doğru bilgisayar programının doğru girdi ve çıktılarıyla yürütülmesi, bu bilgisayarın yapıcı bir biçimde düşünebileceği anlamına gelir mi?" sorusuna "Hayır" cevabını vermektedir. Ayrıca sayısal bilgisayar programlarının söz dizimsel alana sıkışıp kalacağını ileri sürmektedir. Düşünme eyleminin ise sembollerin ötesinde anlamlı içeriklerle mümkün olduğunu vurgulamaktadır. Searle için bu 'anlam' kastedilen şeyin ta kendisi olarak ifade edilmektedir. Ayrıca bilgisayar teknolojisinin belirli bir aşamasını hedef almadığını vurgulayan Searle, argümanının bilgisayar biliminin gelecekteki ilerlemesiyle hiçbir ilgisi olmadığını altını çizmektedir. Bu konunun seri ve paralel süreçler arasındaki farklarla, programların boyutuyla, bilgisayar işlemlerinin hızıyla veya çevreleriyle ve hatta robotların icatlarıyla nedensel etkileşime girebilecek bilgisayarlarla ilgisi olmadığını özellikle vurgulamaktadır. Kuşkusuz, gelecekte bilgisayarların insan davranışını daha iyi taklit edebileceğini kabul eden Searle, zihinsel durumlara sahip olmaktan, akıl sahibi olmaktan söz edilirse tüm bu taklit etme girişimlerinin katiyen yersiz olduğuna inanmaktadır. Bu bağlamda teknolojinin ne kadar hızlı olduğunun öneminin bulunmadığını düşünmektedir. Zihnin söz konusu olduğu durumlarda davranışçılığın benimsendiği düşünen Searle, davranışçılığa mesafeli durmaktadır. Bu bağlamda zihnin, herhangi bir biyolojik ürün gibi tamamen biçimsel kabul edildiğinden hareketle düalizm için tortulaşmış ifadesini kullanmaktadır. Searle, yapay zeka taraftarlarının zihnin biyolojik dünyanın bir parçası olduğuna inandıklarını; aklın tamamen biçimsel olarak tanımlanabileceğine inandıklarını da ifade etmektedir. Yapay zeka çevresinde “düalizm” olarak adlandırılan bazı görüşlere ciddi tepkiler olmasına rağmen Searle; güçlü yapay zekanın düalizm üzerine kurulduğunu savunmaktadır.<sup>136</sup>

Searle (1984), güçlü yapay zekanın düalist bakıştan beslendiği düşüncesiyle hiçbir algoritmanın sisteme zihin sağlamak için yeterli olmayacağı fikrini benimsemektedir. Özetle yazılımların bir zihin oluşturmasının mümkün olmadığını ve kendi başlarına akıl sahibi olamayacağını ifade etmektedir. Searle, sadece programlar tasarlayarak zihni yaratmaya çalışan projenin başından beri başarısız olduğunu iddia etmektedir. Yani deyim yerindeyse güçlü yapay zekayı ölü doğmuş bir proje olarak görmektedir.<sup>137</sup>

Zihin-beden probleminde esaslı tartışma zihni oluşturan şeylerin nasıl belirleneceğinin yöntemidir. Bu bağlamda insan ve yapay zeka arasında kurulmaya çalışılan benzerlik dikkat çekicidir. Özellikle Descartes'ın ifadeleriyle insan bedeninin bir otomat olduğu, zihnin ise bedenden bağımsız bir alana sahip olduğu ileri sürülmektedir. Bu ifadeler ışığında yapay zekanın insan zihnini taklit edebilmesi imkanı savunulmaktadır. Sistemlerin bağımsız hareket

<sup>136</sup> Searle, 1984: 34-36.

<sup>137</sup> Searle, 1984: 34, 36.

edebilme özelliğinin zeka, yaratıcılık, duygu, bilinç gibi fakültelerin sayısal sistemlerde varlığına ilişkin bir tartışma başlatan Özaktaş (1998), zihin-beden sorunu bağlamında şu soruları sorar: “zeka, yaratıcılık, duygu, bilinç dediğimiz şeyler nedir? Bu soruların altında yatan daha temel soru şu: Bu insan özelliklerinin mekanistik veya algoritmik bir ifadesi var mı? Nedensel bir zincire veya bir bilgisayar programına indirgenebilirler mi?”<sup>138</sup>

Bu soruların odağında zihinsel ve bedensel olana ait şeylerin karşılığını sembolik ya da niceliksel yöntemlerle temsil etmenin olasılığı bulunmaktadır. Yani Descartes’ın mekanik felsefesi bağlamında değerlendirilirse; güçlü yapay zeka düşüncesinin mümkün olabileceği öne çıkarken, Aristoteles açısından bunun mümkün olmadığı sonucu çıkarılabilir. Searle, kendi önermesiyle şekillendirdiği zihnin dört ana özelliğini; bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik olarak sıralarken felsefi söyleme uygun bir yaklaşımı benimsemektedir. Bütün bunların bileşkesinde görüşünü şu ifadelerle belirtmektedir: “Özetlemek gerekirse: Bana göre akıl ve beden karşılıklı birbirlerini etkilerler; ama akılsal olgular beynin özellikleri olduğundan, bunlar iki farklı şey değildirler.”<sup>139</sup> Bu ifade özünde Searle’ün “Çince Odası” adlı düşünce deneyinde temellendirdiği argümandır.

Diğer taraftan giriş bölümünde değindiğimiz üzere yapay zekanın uygulamalı örneklerinin imgesel boyutlarına Eski Çağ’dan örneklerle yer verdiğimiz gibi, zihin-beden problemi üzerine Eski Çağ tartışmaları da yapay zekanın gelişimi ve sonuçları bağlamında önemlidir. Bu noktada ilk durağımız düalist fikirleri savunan iki düşünür Platon ve Descartes bu tartışmanın merkezindedir. Zihin-beden felsefesinde zihnin bedene gömülü olduğunu savunan materyalist bakışa karşılık düalist görüş, insan zihnini bedenden ayrı konumlandırmaktadır. Gerek Platon gerekse Descartes’ın görüşleri, düalizmin yapay zekaya sağladığı arka plana tanıklık etmek açısından dikkat çekicidir. Bu arka plana Platon’un, *Phaidon* eserinde yer verdiği Sokrates ve Kebes’in diyalogunda tanıklık etmekteyiz. “Biri ruh diğeri de beden olmak üzere ayrı şeye sahip değil miyiz?” ifadesiyle ruhun görülemeyen, bedenine ise görülebilen iki ayrı unsur olduğunu, ruhun zaman zaman görme gibi farklı duyu organları aracılığıyla bedenden istifade ettiği belirtilmektedir. Duyusal şeylere ruhun sadece beden aracılığıyla erişebileceğini, ancak ruhun bir şeyi kendi başına irdelemesi halinde hiç değişmeyen şeylere yöneldiğini ve bunun da ‘düşünme’ dediğimiz şey olduğunu öne sürer. Ruh; ilahî, ebedî, düşünen, sade, masif olana benzetilirken, beden ise ölümlüye, düşünmeyene, sadeliğin dışında olana benzetilmektedir.<sup>140</sup>

<sup>138</sup> Özaktaş, 1998: 77.

<sup>139</sup> Searle, 1996: 36.

<sup>140</sup> Platon, 2015:75-77.

Bunun yanında Descartes'ın kurguladığı sistematik şüphe yöntemi dikkat çekmektedir. Tüm duyu deneyimlerimizin sanrı olabileceğinden yola çıkan Descartes, modern bir örnekle sistemini açıklamaktadır. Descartes, çılgın bir bilim insanının doğum esnasında beynimizi açarak bizi sanrı görmek üzere programlayabileceği ve aklımızı çelerek maddenin varlığına olan inançlarımızdan kuşku duymamızı sağlayabileceğini hayal etmektedir. Artık bu noktada bizlerin, maddenin varlığını kesin olarak bilen kişiler olmadığımız ortaya çıkmaktadır. Özellikle bedenlerimiz maddi olduğundan önce onların şüpheli olduğu düşünülmektedir. Ancak kendisinden kuşku duyulan bir şüphe kendi kendini çürüteceğinden, düşündüğümüzü fark ederiz ve dolayısıyla varlığımızdan emin olabiliriz; bu argümana Descartes “cogito” demektedir. Yani en azından başlangıçta beden kuşku duyulabilen bir şeyken, zihin kuşku duyulmayan bir şeydir ve bu nedenle zihnin ve bedenin aynı şey olmadığı iddia edilmektedir. Descartes için zihin-beden düalizminin anlamı yalnızca farklılık değildir. Zihnin, bedendeki varlığına bağlı olmadığı anlamına gelen bu düşünce; bedenin, zihinden daha esaslı bir şey olmadığını da iddia etmektedir.<sup>141</sup> Bu düşünce yapay zeka sistemlerinin algoritma ve donanım temelinde değerlendirildiğinde Descartes'ın iddialarının önemli bir yeri olduğunu ifade edebilir. Bu konuya ilişkin Descartes'ın aşağıdaki ifadeleri değerlendirildiğinde insan bedeninin bir makine ile bağdaştırıldığı doğrudan fark edilmektedir.

“Bu durum her hayvan bedeninde bulunan kemiklerin, kasların, sinirlerin, atardamarların, toplardamarların ve tüm öbür parçaların büyük çokluğuyla karşılaştırıldığında insan ustalığının çok çeşitli otomatları ya da devingen makineleri pek az parça kullanarak oluşturduğunu bilip, bedeni bir makine gibi görecektir; oysa Tanrı'nın elleriyle yapılmış olduğundan o öbürleriyle karşılaştırılamayacak kadar iyi düzenlenmiştir ve insanların tasarladığı makinelerin hiçbirinde bulunmayacak biçimde kendi kendine eşsiz devinimler yapmaktadır.”<sup>142</sup>

Descartes, ifadesinde diğer otomatlara nazaran insan bedeninin kusursuzluğunu ilahî kaynak nedeniyle olağanüstü görmektedir. Bir tür makine olarak görülen insan bedeninin eşsiz davranışlarının otomatlarla kıyaslanamayacağı da vurgulanmaktadır. Öte yandan zihin ve bedene ait alanı birbirinden tamamen ayrı kabul eden Descartes, bedeni makineye benzeterek beyin boşluğunda yer alan akıl, hayal gücü, hafıza gibi zihinsel fakülteleri, maddi veya mekansal olmayan “düşünen madde” (*res cogitans*) olarak sınıflandırmaktadır.<sup>143</sup> Yani bedenden bağımsız olduğu düşünülen bu fakülteler bir anlamda kendi başınadırlar. Bu da zihnin insan bedeninde bulunduğu gibi hayvanda ya da makinede de bulunabileceği anlamına gelmektedir.<sup>144</sup> Kartezyen bakış özünde algoritmanın işlevini zihnin işleviyle bağdaştırmıştır.

<sup>141</sup> Guttenplan, 1995: 265-266.

<sup>142</sup> Descartes, 2013: 55.

<sup>143</sup> Descartes, 2006: 56.

<sup>144</sup> Floridi, 1999: 132.

Düalizmin yanı sıra diğer bir örnek de tikel örnek üzerinden açıklanabilir. “Gage” adında bir kişinin geçirdiği kaza sonrası benliğinde ciddi farklılıklar deneyimlemesiyle bilim çevrelerinde zihin-beden ayrımı tekrar sorgulanmaya başlanmıştır. Rivayete göre sevimli ve ılımlı birisi olan Gage, geçirdiği kazanın ardından kötü ifadelere eğilimli, kavgacı ve hatta sapkın bir kişiliğe bürünmüştür. Bu olay karşısında araştırmacılar beynin belirli alanlarında gerçekleşen hasarların insan benliğinde doğrudan etki yaratabileceği gerekçesiyle zihin-beden birliğini gündeme getirmişlerdir.<sup>145</sup> Yani “Bedensel yaralanma sonucu karakter, mizaç gibi zihinsel süreçler değişime nasıl uğrayabilir?” Descartes’ın ayırımına göre bunun mümkün olmaması gerekirdi. Zira bu görüşe göre beden bir tür makineyken, zihin kendi başına düşüncül bir faaliyetti. Bu argümanın sağladığı dayanaktan yola çıkarak yapay zeka sistemlerinin gücünü iç içe geçmiş zihinsel ve bedensel süreçlerden değil, iyi düzenlenmiş algoritmalarından aldığı ortaya çıkmaktadır. Bu yönüyle deneyimsel olguların sayısal sistemlerde içsel bir yaşantıya yol açması beklenemez.

Yapay zekanın böyle bir kapasiteye erişmesinin zorluğu geliştiriciler tarafından bilinen bir gerçektir. Bu yönüyle bilimsel, sosyal, sanatsal birçok alanda olguları doğrudan veya dolaylı olarak mekanikleştirme eğilimi fark edilmektedir. Söz gelimi Descartes “mekanik felsefe”<sup>146</sup> kavramının savunucusu olduğu gibi, makinenin işleyişiyle insan arasında özdeşlik kurmaya çalışmaktadır. Bu hususta Searle (1996), “Felsefede, psikolojide ve yapay zeka konusunda baskın olan görüş, insan beyninin işlevleriyle sayısal bilgisayarların işlevleri arasındaki benzerlikleri vurgulayan görüştür.”<sup>147</sup> ifadesiyle çeşitli disiplinlerin kasten bu hedefi gözettiklerini vurgulamaktadır. Söz gelimi insan beyni düşünüldüğünde Broca alanının “ön konuşma korteksi” konuşmaya ilişkin, Wernicke “arka konuşma korteksi” alanının ise anlamaya ilişkin alanlardan sorumlu olarak sınıflandırılmıştır. Beynin sol lobunun dilsel, sağ lobunun ise matematiksel yetileri yönettiği fikri vurgulanmıştır.<sup>148</sup> İfadelerden fark edileceği üzere dilsel ve zihinsel süreçleri biyolojik bir yaklaşımla sınıflandırma yönelimi bulunmaktadır. Bu konuya eleştirel yorumlar getiren Penrose (1989), insan bilincinin pek çok düşünüre göre konuşma diliyle ilişkilendirilmesine karşı durmaktadır. Ayrıca bilinci diğer şeylerden ayıracılabilecek yegane şeyin konuşma dili olduğunu kabul eden görüşü de etik kaygılardan ötürü reddetmektedir. Çünkü bu görüşe göre konuşmadığı gerekçesiyle bilinçsiz sayılan hayvanlara zulmetme hakkı doğmaktadır. Aslında Penrose’un tepkisi konuşma dilini açıklayan görüşedir. Bu görüşe göre konuşma dilini kontrol eden alanlar (Broca ve Wernicke alanları) beynin sadece

<sup>145</sup> Kaku, 2011: 114.

<sup>146</sup> Churchland, 2013: 13.

<sup>147</sup> Searle, 1996:38.

<sup>148</sup> Yule, 2006:138,139.

sol lobunda olduğundan, bilinci de bu alana indirgemektedir.<sup>149</sup> Penrose bu ifadesiyle özünde düalist bakış açısını hedef almaktadır.

XVII. yüzyıldan bu yana bilimsel, sosyal, teknolojik pek çok alanda bu yaklaşımın yansımalarını gözlemlemek mümkündür. Zihin-beden ayrımında olduğu gibi olguları sınıflandırma, kategorize etme ya da birbirinden ayırma, Kartezyen bakışın fikridir. Bu amaç insan zihnini “zeka” kavramıyla bağdaştırdığından, yapay zekanın bir zihne sahip olabileceği iddiası sezilmektedir. Yapay zekanın bilgiyi işleme kuramı temelinde gelişmesi düşünüldüğünde; insan beyninin çalışma sistemi bilgisayarın dosyalama ve geri getirme sistemiyle özdeşleştirilmektedir. Oysa John Searle’ün ifade ettiği üzere; eğer gerçekten beyne veya diğer organlara ait sinirlerin bütün detaylarına erişmiş olsaydık, bu sinirlerden hangilerinin zeka fakültesini harekete geçirdiğini fark ederek yapay sinir ağlarıyla yapay zekayı oluşturabilirdik.<sup>150</sup> Ayrıca tek boyutlu bir zihin anlayışı nihai olarak algoritma temelli bir zihin fikrine imkan tanıyacaktır. Kısacası sahada gerçekleştirilmesi mümkün olmayan pek çok şey masada gerçekleştirilebilecektir. Dolayısıyla burada esas problem, “zihinden” ne anladığımızdır. Bu bağlamda yapay zekanın insanı taklit edebilmesi değil, insan aklının bilgisayarın çalışma sistemine benzetilme çabasıdır. Descartes’ın “Benim yöntemim aritmetiğe kesinliğini veren şeylerin tümünü kapsamaktadır.”<sup>151</sup> sözüyle bütün insan zekasını “evrensel bir matematik” yığınınna benzetme çabası düşünülebilir. Descartes felsefesinin kendisinden sonraki yıllara damga vurması ve yüzyıllar sonrasında böyle bir teknolojinin gelişebileceğini fark etmesi yapay zekanın günümüzdeki yeri bağlamında önemlidir. Ancak “...bu makine en şaşkın insanların bile yapabildiği gibi karşısındakini yanıtlamak için söyleyeceği şeyleri anlamına göre çeşitli biçimlerde düzenleyemez.”<sup>152</sup> ifadesi düşündürücüdür. Bu ifade sayısal sistemlerin “şaşkınlık” olarak ifade edilen irrasyonel eyleme sahip olamayacağı imasıdır. Burada ortaya çıkan düşünce, irrasyonel davranıştan yoksun sistemlerin bu yönünü zenginleştirmenin imkansızlığı karşısında zihin ve bedenin mekanik zemine indirgenmesi olarak ifade edebileceğimiz, insanın robotlaştırılmasına yol açmaktadır. O halde sayısal sistemlerin temeli olan algoritma doğaya içkin bir olgu olmadığı, doğaya içkin olmayan bir olgunun semantik anlamda doğayı algılamasının mümkün olmadığı sonucu ortaya çıkmaktadır. Bu düşünceye karşıt duruşu savunan Turing’in “Turing Makinesi”, “Taklitçilik Oyunu”, “Turing Testi” argümanları sözünü ettiğimiz türden bir mekanikleşmenin yolunu açmak üzere kurgulanmıştır. Bu bağlamda argümanların yeni bir başlık altında tartışılması gerekmektedir.

<sup>149</sup> Penrose, 1999: 100-101.

<sup>150</sup> Searle, 2004:51.

<sup>151</sup> Stumpf, 1988:389.

<sup>152</sup> Descartes, 2013: 55.

### 2.3 Turing Makinesi, Taklitçilik Oyunu, Turing Testi ve İtirazlar

Tarihte otomatların ya da mekanizmaların gelişimine bakıldığında günümüz bilgisayarının atası olarak gösterilebilecek pek çok örneğe rastlamak mümkündür. Ancak onlardan en önemlisi iki bin yıl öncesine kadar uzanmaktadır. Tarihte bilinen ilk bilgisayar olan Antikythera mekanizması, Girit ile Mora Yarımadası arasında Ege Denizi'nde yer alan Yunan Antikythera adası yakınlarında M.Ö. 70 civarında bir Roma gemi enkazında batmıştır. 1900 yılında bir grup sünger dalgıç tarafından keşfedilen Antikythera mekanizmasının parçaları yaklaşık sekiz ay sonra fark edilebilmiştir. Bilim insanları Antikythera mekanizmasının teknik olarak bin yıl sonrasının bilinen cihazlarından bile çok daha karmaşık olduğunu belirtmektedir.<sup>153</sup>

Otomatların ve mekanizmaların gelişimine bakıldığında Turing'in dönemine dek mühendisliğin ön planda olduğu görülmektedir. Oysa zihin-beden probleminin yapay zekanın düşünsel zeminini anlamamıza imkan tanınması nedeniyle mekanizmaya yaklaşımın değiştiğini söylemek mümkündür. Bu değişimin belki de en önemli sıçrayışını Turing'in argümanlarına ve bu argümanlara yönelik itirazlarda görmekteyiz. Bu önemin bir parçası olarak Turing'in yaşam öyküsüne kısa bir bakış tarihsel arka plana ışık tutacaktır. Copeland'ın (2004) anlatımından faydalanacağımız tarihsel arka plan, İngiliz asıllı Alan Mathison Turing'in (1912-1954) hakkında önemli bilgiler sunacaktır. Mantık, matematik, biyoloji, felsefe, kripto analiz ve sonraları bilgisayar bilimi, bilişsel bilim, yapay zeka ve yapay yaşam olarak bilinen alanlara Turing biçimsel anlamda katkıda bulunmuştur. Turing, diğer bir yönüyle ikinci dünya savaşında Nazilere ait Deniz Enigma'sının şifrelerini kıran 'bombe' adlı şifre çözücü makineyi tasarlaması nedeniyle ulusal kahraman unvanına da layık görülmüştür. O dönemlerde Turing, bugün 'yapay zeka' adı verilen alanı kurmuş ve insan beyninin aslında dijital bir bilgisayar olduğu teorisini öne sürmüştür. Eğitim hayatının merkezine bilişimi alan Turing, bilgisayar zekası hakkındaki halka açık ilk dersini vererek kısa süre sonra "Bilişim Makineleri ve Zeka" başlıklı makalesini yayınlamıştır. Her şey bu makalenin yayınlanmasıyla hareketlenmiş, günümüz bilgisayarının temelini oluşturan evrensel Turing makinesi yani soyut dijital hesaplama makinesi tanımlanmıştır.<sup>154</sup> Yaşam öyküsünden fark edileceği üzere Turing, mühendislik temelinde geliştirdiği teknolojiyi özellikle Descartes felsefesinden beslenerek yepyeni bir teoriye dönüştürme girişiminde bulunmuştur. Bu teori sezgilerimizin ve içgüdüsel duygularımızın dışında sadece matematik kurallarının özenle düzenlenmesiyle neler yapabileceğimiz fikrine dayanmaktadır. Turing bu düşüncesini argümanlarıyla açıklamıştır. Bu argümanlar pek çok

<sup>153</sup> Steiglitz, 2019: 108, 109.

<sup>154</sup> Copeland, 2004: 1-3.

çevreden destek aldığı gibi, pek çok çevreden de eleştiri almıştır. Bu yönüyle Turing'in argümanları yapay zeka tarihinde tartışılan önemli konulardan birisidir. Bu argümanlar başlıca Turing Makinesi, Taklitçilik Oyunu, Turing Testi'dir. Şimdilerde en yalın haliyle Turing testi olarak bilinen, 'taklit oyunu'nun ilk tanımlarına “Zeki Makineler”<sup>155</sup> çalışmasında yer verilmiştir. Taklit oyununa benzeyen Turing testi, dijital bilgisayarların test edilmesi amacıyla uygulanan bir sürümdür. Hatta bu makalede Turing testi adında bir ifadeye rastlanılmamaktadır. Özetle Turing'in teorilerini temellendirdiği ve testine tabii tuttuğu gerçek anlamda mekanik bir bilgisayar olan Turing makinesi, prensip olarak günümüz dijital bilgisayarının temelini oluşturmaktadır. Turing makinesinin teknik açıklaması şöyledir:

“Turing makinesi ileri geri hareket eden bir tarayıcıdan ve sınırsız bir hafıza bandından oluşur. Bant karelere bölünmüştür. Her kare boş olabilir veya tek bir sembol ("0" veya "1", örneğin bir White alfabesinden alınmış başka bir sembol) içerebilir. Tarayıcı bir kerede yalnızca bir kare ("taranmış kare") inceleyebilir.”<sup>156</sup>

Sözü geçen çalışma prensibi insanın işlem kabiliyetinden daha hızlı olan bir bilgisayarın kusursuzca programlanması hayal edilmiştir. İlerleyen zamanlarda da bir adım daha ileri gidilerek bu sistemin düşünebildiği, belki de gerçek anlamda akıllı olabileceği hatta acı, mutluluk, tutku, gurur gibi duyguları da hissedebileceği, eylemlerinin farkında olabileceği varsayımı ortaya çıkmıştır. Penrose'a (1991) göre bir sistemin insana ait özellikleri sergileyip sergilemediği işlevselci bir bakış açısıyla yani gözlem yoluyla değerlendirilebilir.<sup>157</sup> Söz konusu varsayım Alan Turing'in (1950) savunduğu *Computing Machinery and Intelligence* başlıklı makalesinde yer almaktadır. Bu makalede Turing testinin amacı mantıksal açıdan bir makinenin düşünebildiği ya da düşünemediği söyleminin ne ölçüde gerçek olduğunu ispatlama çabası olarak değerlendirilmektir. Ne var ki makalenin yayınlanmasının ardından Turing pek çok eleştirinin odağı olmuş, yapay zeka alanında “güçlü yapay zeka” ve “zayıf yapay zeka” olan önemli iki görüşün ortaya çıkmasına neden olmuştur; Her iki görüş de bir sonraki başlıkta ele alınmıştır.

"Makineler düşünebilir mi?" sorusunun hem cevabından hem de nihai sonuçlarından kaçmanın zorluğuna değinen Turing (1950), problemi 'taklit oyunu' adlı bir oyunla tanımlamıştır. Amacı insan gibi düşünen bir makinenin varlığını ispatlamak olan bu oyunda bir erkek, bir kadın ve bir de sorgulayıcı bulunmaktadır. Sorgulayıcının cinsiyetinin bir önemi bulunmamaktadır. Sorgulayıcı, kadın ve erkeğin bir arada bulundukları odanın dışında farklı bir odadadır. Bu kişileri sadece X ve Y etiketleriyle tanıyan sorgulayıcının amacı kimin erkek

<sup>155</sup> Copeland, 2004:401.

<sup>156</sup> Copeland, 2004:6.

<sup>157</sup> Penrose, 1991: 5, 6.

kimin kadın olduğunu tespit edebilmektir. Sorgulayıcıya herhangi bir ipucu vermemek adına bütün cevaplar yazılı bir şekilde, hatta telgraf yöntemiyle iletilmektedir. Ayrıca sorgulamada bir kadın ve bir erkek bulunmasının nedeni herhangi bir kişinin sorgulayıcıyı yanıltmak amacıyla “Ona inanma! Erkek olan benim.” şeklinde numara yapmasını sağlamaktır. Bu oyunda “Bir makine erkeğin yerine geçerse ne olacak?” sorusunu soran Turing'e göre; ancak sorgulayıcının bunu tespit edebilmesi ya da edememesi sonucunda "Makineler düşünebilir mi?" sorusu yanıt bulabilecektir.<sup>158</sup>

Alan Turing'in insan zihni ve sayısal sistemlerin bağıını açıklamak üzere geliştirdiği Turing testinde yer alan en önemli unsur dil ve dilsel öğelerdir. Kaldı ki testte ifadelerin sadece yazılar ya da bir bilgisayar aracılığıyla aktarılması esastır. Bu bağlamda dilin insan zekasını açıklama rolünün ne olduğu sorulabilir. Şöyle ki Searle (2006) insanın konuşma yetisini sadece çenelerin hareketi ve birtakım titreşim sayesinde ortaya çıkan şok dalgalara indirgememiştir. Tersine, ilgili organların ortaya çıkardığı konuşma yetisinin bir söylem, soru, açıklama, komut, telkin, emir, vaat gibi zihinsel süreçleri de kapsayan birçok başka alanla da ilişkili olabileceğini vurgular.<sup>159</sup> O halde düşüncelerin belirleyicisi olarak dil çok önemlidir.

Dildeki yapısal çeşitliliğin kültürle ilişkisi Sapir-Whorf hipotezinde açıklanmıştır. Kısaca değinmek gerekirse yerli bir Amerikan kabilesi olan Hopi'ler kullandıkları dilin yapısından ötürü dünyayı farklı yorumlamaktadırlar. Whorf'un bakış açısıyla bu farklı algılayışın sorumlusu kullanılan dildir. Zira Hopi kabilesi kendi dillerinde şeyleri sınıflandırırken "canlı" veya "cansız" etiketlerini kullanmıyorlardı. Örneğin Hopi'ler bulutları ve taşları canlı olarak algılıyordu. Ancak İngilizce dilbilgisinde dilinde ya da kültüründe bulutların ve taşların "canlı" olduğuna dair bir işaret yoktur. Buradan çıkan sonuca göre, dilin bir iletişim aracı olmasının ötesinde o dili konuşan birey veya topluluğun dünya görüşünü etkileyen bir olgu olduğu inancına varılmıştır. Diğer bir deyişle dil, düşüncelerimizi şekillendiren bir olgu olarak yorumlanabilir.<sup>160</sup> Öyle ki her dildeki altta yatan dilbilimsel sistemin yalnızca fikirleri ifade etmeye yarayan bir mekanizma olmadığına dikkat çeken Whorf'a (1940) göre dilbilimsel yapı bireyin zihinsel faaliyetine, izlenimlerinin çözümlenmesine, işleyen zihinsel belleğinin sentezine ve zihinsel faaliyetine yönelik bir tür kılavuzdur.<sup>161</sup> İnsan doğasında yürümeyi biyolojik olarak organik, içgüdüsel bir işlev kabul eden Sapir (2009), dilin içgüdüsel olmayan, sonradan toplumsal olarak edinilmiş, “kültürel” bir işlev olduğunu da öne sürmektedir. Bu düşünce doğrultusunda "Dil ve düşünce oluklarımız

<sup>158</sup> Turing, 1950: 433-434.

<sup>159</sup> Searle 2006: 155,156.

<sup>160</sup> Yule, 2006: 218, 219.

<sup>161</sup> Whorf, 1959: 212, 213.

ayrılmaz bir şekilde iç içe geçmiş durumdadır, bir anlamda aynıdır."<sup>162</sup> ifadesiyle dilin düşünceyle yani akılla olan ilişkisine doğrudan göndermede bulunmaktadır.

Dil ve kültür bağından anlaşılacağı üzere Turing'in sayısal sistemleri insanı taklit edebileceğine yönelik inancının temelsiz olduğu düşünülemez. Bu uğurda makinelerin insanı taklit edebileceğini düşündüğü en güçlü araç dildir. Bu uğurda kendine olan güvenini şu cümleleri ifade etmektedir.

"Bence elli yıl içinde bilgisayarlar  $10^9$  bellek kapasitesiyle programlanıp taklitçilik oyununu öylesine iyi oynayacaklar ki sıradan bir sorgucunun beş dakika sorguladıktan sonra doğru tanımlamayı yapma olasılığı %70'i geçmeyecektir. "Makineler düşünebilir mi?" sorusunun tartışılmayacak kadar anlamsız olduğunu düşünüyorum."<sup>163</sup>

Makinelerin düşünce eylemini gerçekleştirip gerçekleştiremeyeceğine ilişkin tartışmayı gereksiz gören Turing, makalesinde makinelerin düşünüp düşünmeyeceğini sormak yerine makinelerinin Turing testi adında davranışsal zeka testini geçip geçemeyeceğini sormamız gerektiğini önermektedir. Bu testte, bir bilgisayar programı sorgulayıcıyla çevrimiçi yazılı mesajlar yoluyla beş dakika süresince sohbet etmektedir. Sorgulayıcı daha sonra sohbet ettiği şeyin bir bilgisayar programı mı yoksa bir insan mı olduğunu tahmin etmek zorundadır. Eğer program, sorgulayıcıyı harcayan zamanın %30'unda kandırabilirse testi geçmektedir. Bu test ışığında Turing, 2000 yılına kadar  $10^9$  ünitelik depolama kapasitesine sahip bir bilgisayarın testi geçecek kadar iyi bir şekilde programlanabileceğini düşünmüştür. Ancak programlar henüz başarılı olamamıştır. Turing, makalelerinin yayınlanmasından bu yana ortaya çıkan zeki makinelerin olasılığına dair çeşitli itirazların neredeyse tamamını değerlendirmiştir.<sup>164</sup> Bu itirazlara alt başlıklarda kısaca değinilmiştir.

### 2.3.1 Teolojik İtiraz

Sayısal sistemlerin baş döndürücü gelişimine bağlı olarak makinelerin zeka sahibi olup olamayacakları sorusunu sormayı bile saçma bulan Turing'e (2005) birtakım eleştiriler yöneltilmiştir. Bunlardan ilki teolojik itirazdır. Tanrıbilimsel itiraza göre; insanın ölümsüz ruhunun bir işlevi de düşünme olarak kabul edilmektedir. Bu nedenle sadece insanların ruh taşıyabileceği gerekçesiyle makine ve hayvanlara Tanrı tarafından ruh bahşedilmediğine dikkat çekilmektedir. Turing bu iddiayı, insana ruhu bağışlayabilen ama hayvana bağışlayamayan Tanrı'nın yetersizliği olarak düşünmüş ve reddetmiştir. Bu itirazda, Tanrı'nın yüce saygınlığının ciddi şekilde kısıtlandığını göndermede bulunmuştur. Ayrıca Tanrı'nın biri ikiye

<sup>162</sup> Sapir, 2009: 92-95.

<sup>163</sup> Turing, 2005: 62-64.

<sup>164</sup> Russell ve Norvig, 2010: 1021.

eşitlemek gibi yapamayacağı bazı şeyler olduğunu belirten Turing, bir file ruhu bahsetme özgürlüğünden neden yoksun olacağını sormaktadır. Kilisenin geçmişteki bilimsel ilerlemeler karşısındaki tavrına göndermede bulunan Turing, teolojik tartışmaları umursamadığını ve geçmişte de sık görüşler olarak değerlendirildiğinin altını çizmektedir.<sup>165</sup>

Turing teolojik itirazını temellendirmekte çıkış noktası olarak benimsediği genel görüşünü akıl ve inanç kavramlarının ayrıklığına dayandırmaktadır. Yani bu türden bir sorgulamanın inançsal bakış açısıyla değerlendirildiği dönemlerde yaşanan sıkıntılara dikkat çekmektedir. Bunlara bağlı olarak öne sürülen itirazın temelsiz olduğunu konunun ruhsal unsurlardan öte maddeci bir anlayışı ilgilendirdiğiyle anlatmaktadır.

### 2.3.2 Kumun Altındaki Kafalar İtirazı

Düşünen makinelerin korkunç sonuçlarının olacağı kaygısıyla kaderci bakış tarafından bu tür bir gelişmenin yaşanmamasını umduğuna dikkat çekmektedir. Bu argüman Turing'e göre bundan daha aleni bir şekilde ifade edilemez. Turing, bu düşüncenin çoğu insanı etkisi altına aldığını savunmaktadır. İnsanoğlunun köktenci bir anlayışla, yaratılanların hepsinden daha üstün olduğuna inanmaktan hoşlandığına gönderme yaparak haliyle insanoğlunun yönetici konumunun tehlikeye düşeceğinden endişe duymaması gerektiğini ima etmektedir. Teolojik argümanın popülaritesinin bu duygu ile bağlantılı olduğunu açıkça belirtmektedir. Bu türden bakış açısının entelektüel insanlar arasında oldukça yaygın olduğuna değinerek bunun nedenini düşünme gücüne her şeyden daha fazla değer vermelerine bağlamaktadır. Bu argümanı yanlışlamanın gereksiz olduğunu düşünen Turing, bu hususa cevabın ruhların göçünde teselli aramanın daha uygun olacağını önermektedir.<sup>166</sup>

Söz konusu itirazı insanoğlunun kaderci bakışa teslim olması sonucunda ortaya çıkan bir kaygının bir eseri olduğu düşünen Turing, makinelerin düşünmesi durumunda dünyada telafisi güç olayların yaşanabileceği gereksiz bir endişe olarak görmektedir. Yani insan inançları dolayısıyla sabit bir bakış açısının esiri olmaktadır. Bu yönüyle teolojik itiraz ile bağlantılı gördüğü ve hafife aldığı bu itiraz karşısında Turing, insanoğlunun diğer canlılara kıyasla kudretinin ortada olduğunu ve bu gücü açığa çıkarabilmesiyle egemenliğini kaybetmeyeceğini ifade ederek bu itirazı çürütmeye çalışmıştır.

<sup>165</sup> Turing, 1950: 443.

<sup>166</sup> Turing, 1950: 444.

### 2.3.3 Matematiksel İtiraz

Turing alçak gönüllükle Gödel'e atfettiği ama aslında kendisinin de çoktan keşfetmiş olduğu teoremin (durma problemine olumsuz yanıt) makinelerin düşünebileceği bir argüman olup olmadığını tartışmıştır. Bunun yanı sıra Turing'in görüşlerini temelden sarsabilecek önemli bir itiraz da matematik cephesinden gelmiştir. Turing, tek başına herhangi bir algoritmayı hesaplayabilen evrensel soyut bir makinenin olasılığını, yani Turing makinesini tanıtmıştı. Turing'in bakış açısıyla bu makine o derece güçlüdür ki her türlü algoritmayı çözümleyebildiği gibi sonsuz belleğe de sahiptir. Ne var ki her türlü doğru önermeleri kapsayan kesin bir doğruluk sistemi olamayacağını ispatlayan Gödel Teoremi bağlamında Turing'i köşeye sıkıştıran nokta durma problemidir (*halting problem*). Basit işlemler bir yana, diğer yanda hiç duraksama yapmayan birtakım hesaplamalar verildiğinde durma probleminin bilişimsel olarak çözülemez olduğu anlaşılmıştır. Yani Turing makinesi bir anlamda çıkmaza girmiştir.<sup>167</sup> Bu hususa açıklamasıyla katkı sunan Say (2005), bazı programların durmaksızın sonsuz döngüde çalıştığı, bazılarının da sonlu dizi sonrasında durduğuna dikkat çekerek; Gödel Teoremi'nde programların “duranlar” ve “durmayanlar” olarak ikiye ayrıldığını vurgulamaktadır. Bu bağlamda Gödel Teoremi'nin tanımı “Programların durma ya da durmama özelliklerinin kanıtlanabildiği ve kanıtlanabilen tüm önermelerin doğru olduğu her biçimsel sistemde kanıtlanamayan doğru önermeler vardır.” şeklinde verilmiştir. Bu ifade itibarıyla doğru önermeler ve yanlış önermeler arasında kesin ayrım yapabilecek bir bilgisayar programının olmadığı sonucu çıkarılmıştır.<sup>168</sup> İnsanoğlunun algoritma ötesinde bazı kabullere sahip olduğu düşünüldüğünde böyle bir karmaşaya düşmesi beklenmez. Turing'in bu itiraza karşı cevabı “Bizler de bazen sorulara yanlış yanıtlar verdiğimiz için, makinelerin başarısızlık kanıtlarına sevinmemiz pek de haklı olduğunuzu göstermez.”<sup>169</sup> şeklindedir. Makine karşısında bu üstünlüğün ancak belirli koşullarda ve tek makine ile ilgili olduğunu savunan Turing, tüm makineler üzerinde aynı anda zafer kazanmanın söz konusu olamayacağını düşünmektedir. Kısacası herhangi bir makineden daha zeki insanların olabileceği gibi, ilerleyen süreçlerde başka makinelerin daha da zeki olabileceği kanısındadır.<sup>170</sup>

Aslında gerçek bir matematiksel itiraz sayılabilecek bir yaklaşım Penrose tarafından ileri sürülmüştür. Özetle Penrose'a göre sezgiler sistemleştirilemediklerinden ve biçimselleştirilemediklerinden herhangi bir algoritmik işlemin dışında kalmaktadır.<sup>171</sup> Bu görüşü Penrose iki farklı argümanla savunmaktadır. Birinci argüman insanlarda bulunan

<sup>167</sup> Penrose, 1994: 29.

<sup>168</sup> Say, 2005: 65-71.

<sup>169</sup> Turing, 2005: 64.

<sup>170</sup> Turing, 1950: 445.

<sup>171</sup> Penrose, 1997: 131.

sezginin sayısal sistemlerden farklı olarak yeni matematik kuramları kurma konusunda vazgeçilemez bir role sahip olduğudur. Ancak bazı önermelerin doğruluğuna *düşünce ilkesinin* bir örneği olan sezgiyle ulaşmanın mümkün olduğu düşüncesi de vurgulanmaktadır.

“Ne olursa olsun, bana öyle geliyor ki Gödel'in teoreminin açıkça sergilediği sonuç, matematiksel doğruluğun herhangi bir formalist çerçeveye içerisine sıkıştırılamayacağıdır. Matematiksel doğruluk, salt formalizmin çok ötesinde bir şeydir. Gödel'in teoremi olmaksızın da bunu anlamak olasıdır. Çünkü bir sistem inşa etmeye kalkışsak hangi aksiyomları veya hangi kuralları seçeceğimize nasıl karar verebiliriz? Hangi kuralları seçeceğimiz konusunda bize yol gösteren her zaman ve mutlaka, sistemin simgelerinin anlamı verildiğinde bunlardan içgüdüsel olarak ne anladığımız olmalıdır.”<sup>172</sup>

İkinci argüman ise şöyle özetlenebilir; Penrose'a göre hem Gödel'in eksiklik teoremleri hem de Turing'in durma problemine verdiği olumsuz yanıt insanların sezgi yeteneği sayesinde gereksiz bir döngüye düşmeyeceğini vurgulamaktadır.

“Hatırlayacağınız gibi, bir Turing makinesinin işleminin durup durmayacağına karar vermek için herhangi bir algoritma verildiğinde çalışacağını bizim anlayacağımız fakat algoritmanın anlamayacağı bir Turing makinesi işlemi üretebiliriz. (Algoritmanın, bir Turing makinesi işlemini üretebiliriz (Algoritmanın, bir Turing makinesi işleminin ne zaman duracağı hakkında bize doğru bilgi vermesi gerektiği konusunda ısrar ettiğimizi ama bazen algoritmanın kendisi sonsuza kadar işlediği için, Turing makinesinin ne zaman duracağını bize bildirmediğini anımsayınız.) Bu nedenle Gödel'in teoreminde olduğu gibi, verilen algoritmik işlemin başaramadığını gerçekleştiren bir önermeye sahip olarak sezgiden yararlanarak genelde neyin doğru olduğunu (Turing makinesinin işleminin durmayacağını) görebiliriz.”<sup>173</sup>

Her iki argümandan da anlaşılacağı üzere matematiksel itiraz altında değerlendirilmesi gereken en önemli yaklaşımlar Penrose'un kaleme aldığı *Matematiksel Sezgi* bölümünden anlaşılmaktadır. Burada vurgulanan esaslı düşünce sezgilerin biçimselleştirilemeyeceği esasına dayanmaktadır. Öyle ki semboller her ne kadar nesneyi temsil etmekte başarılı olsa da öznenin ya da benliğin temsili noktasında büyük şüphe bulunmaktadır.

### 2.3.4 Bilinçlilik Argümanı

Yapılan diğer bir itiraz da bilinçlilik argümanı başlığı altında Profesör Jefferson'un 1949 yılı *Lister Söylevi*'nde ifade ettiği, insan beyninin sanatsal ve duygusal temelde ortaya koyduğu şiir yazmak, konçerto bestelemek, başarılarından onur duymak, birden küplere binmek, iltifat karşısında keyiflenmek, yanlışlarından ötürü yüzü kızarmak, cinsel olarak tahrik olmak, beklentileri geçekleşmeyince hüsrana uğramak gibi eylemleri hiçbir makinenin gerçekleştiremeyeceği, gerçekleştirse dahi bu olguların farkında olamayacağıdır. Turing bilinç

<sup>172</sup> Penrose, 1997: 133.

<sup>173</sup> Penrose, 1997: 142.

itirazına cevaben, insanın düşündüğünü açıklayabilmenin tek yolunun insan olabilmekten yani insan perspektifinden bakabilmekten geçtiğini, dolayısıyla da makinenin nasıl düşündüğünü anlamaksızın, yani makinenin yerine kendimizi koymaksızın ancak tekbenci bir yaklaşım sergileneceğini öne sürmektedir. Profesör Jefferson'un şiir yazar bir makinenin teste tabii tutulmasıyla ortaya çıkan cevaplar karşısında şaşıracağını da eklemektedir. Bilinç konusunun her yönüyle açık bir konu olmadığını, hâlâ gizemini sürdüren pek çok paradoks barındırdığını hatırlatan Turing, öncelikle bu hususların açıklığa kavuşturulması gerektiğini savunmaktadır.<sup>174</sup>

Yapay zeka tartışmalarının kilitlendiği ya da düğümlendiği can alıcı nokta bilinç problemidir. Bir sayısal sistemin bilinç sahibi olması veya olmaması olasılığı arasındaki fark dünyanın olağan akışını kökten değiştirebilecek güçtedir. Bu noktada sorunun cevabına giden yolda çıkış noktası kavramsal düşünme yöntemi olarak bilinç kavramını değerlendirme zorunluluğu olarak işaret edilmektedir. Aslında Turing'e yapılan itirazda Jefferson tarafından bilincin kapsadığı birtakım tikel örneklerle değinilmiştir. Bu tikel örnekler bilinci salt maddi boyuta indirgemediği gibi özünde ne denli kapsayıcı olduğunu, bu kapsamın insan bilincini ne ölçüde yüce kıldığını ifade etmeye çalışmaktadır. Kaldı ki verilen tikel örneklerle makinenin karşılaşması durumunda makinenin şaşıracağını öne sürmektedir. Burası önemlidir çünkü böylesi bir senaryoda makinenin şaşırması, makineyi insana yaklaştıracak hatta bilincin bazı emarelerini kısmen yerine getirdiğini ispatlayacak bir unsur olabilir. Bu yönüyle sayısal sistemlerin insana dair kavramlarla açıklanmaması ya da bu hususta dikkatli olunması gerektiğidir. Öyle ki makinenin söz konusu tikel örneklerle karşılaşması durumunda işlev bozukluğuna uğrayabileceğini ifade etmek daha yerinde olacaktır.

Turing'in bu itiraza cevabına değinilecek olursa bilicin o zamanlardan bugüne dek hâlâ bir paradoks olduğu doğrudur. Öte yandan indirgemeci materyalist bakışı savunan pek çok araştırmacıya göre, öznel, benlik ya da kendilik gibi kavramların tamamen spekülasyondan ibaret olması nedeniyle bu dünyada karşılığı bulunmamaktadır. Bu bağlamda Turing'in itiraza ilişkin cevabından indirgemeci bakış açısını benimsediği çıkarılabilir.

### 2.3.5 Çeşitli Engellere İlişkin Tartışmalar

"Makinelerin hata yapamayacağı" iddiasını ilgin bulan Turing, bu eleştiriyi taklit oyunuyla açıklamayı mantıklı bulmuştur. Sorgulayıcı cevapların aritmetik doğruluğunu hesaba katarak makineyi insandan ayırt edebileceği varsayımıyla, makine hassas kesinliğiyle kendini ele verecektir. Dolayısıyla sorgulayıcı, makinenin doğası gereği vereceği bütün doğru cevaplardan maskesini düşürebilecektir. Buna cevabı hazır olan Turing, oyun oynamak için

<sup>174</sup> Turing, 2005: 66.

programlanmış makinenin aritmetik problemlere doğru cevapları vermeyeceğine, hatta sorgulayıcıyı şaşırtmak için hesaplanmış bazı kasıtlı hatalar yapacağına, bunun dışında ortaya çıkabilecek mekanik hataların ise farklı bir şekilde kendini göstereceğini belirtmiştir. Sistemin geliştirdiği kasıtlı hataları talihsiz eleştiriler olarak kabul eden Turing, “işleyiş hataları” ve “sonuç hataları” şeklinde iki ayrım yapmıştır. İşlev hatalarını, makinenin tasarlandığından farklı davranmasına neden olan bazı mekanik veya elektriksel arızalar olarak açıklamaktadır. Felsefi tartışmalarda, insanların bu tür mekanik hataların olasılığını yok saymaktan keyif duyacağını; bu nedenle de "soyut makinelerin" tartışılmasını eleştirmektedir. Dolayısıyla soyut makinelerin fiziksel nesnelerden ziyade matematiksel kurgular olduğu ve tanımları itibarıyla işlev hatalarından yoksun olduklarını savunarak bu anlamda “makinelere asla hata yapamaz” demektedir. Sonuç hatalarının, makineden gelen çıkış sinyallerine bir anlam eklendiğinde ortaya çıkabileceğine dikkat çeken Turing, bunun nedenini yazılan komuta eklenen yanlış bir edat ile makinenin bu hataya verdiği dönütle açıklamaktadır. Turing, bir makinenin hiç şüphesiz kendi kendinin öznesi olabileceğini, kendi programlarını oluşturmada veya kendi yapısındaki değişikliklerin etkisini tahmin etmekte kullanılabileceğini ve kendi davranışının sonuçlarını gözlemleyerek bazı amaçları bilfiil yapmak amacıyla kendi programlarını değiştirebileceğini savunmaktadır. Ayrıca bunları ütopya hayalleri değil yakın geleceğin olanakları olarak görmektedir. Bir makinenin çok fazla davranış çeşitliliğine sahip olamayacağı eleştirisini, sadece bir depolama kapasitesi sorunu olarak gören Turing, buradaki eleştirilerin de bilinç argümanı ile gizli ilişkili olduğunu düşünmektedir.<sup>175</sup>

Çeşitli engellere ilişkin itirazlar başlıktan anlaşılacağı üzere taklit oyununda makine ve insan ayrımının bazı ipuçlarına yer verilerek bu örneklerin Turing’i zora sokacağı eleştirisi yapılmaktadır. Bu hususta makinenin salt algoritma temelli çalıştığı düşünüldüğünde rasyonel çizgiden çıkmayacağı ama insanın irrasyonel söylemlerde bulunabileceği gerekçesiyle bu iddia öne sürülmüştür. Yani bazı muğlak sorulara makine su götürmez cevaplar vereceğinden kendini ele vereceği savunulmaktadır. Oysa Turing’e göre bu geçerli bir itiraz değildir. Kaldı ki Turing, makinelerin numara yaparak şaşmaz doğruluğunu gizleyebilecektir. Örneğin bazı aritmetik sorularına bile bile yanlış cevap verebilecektir. Öte yandan makinelerin hem işlevsel hem de sonuççu hatalarından ötürü birtakım irrasyonel davranışta bulunabileceğini söylemektedir. Yani makinelerin de insanlar gibi hata yapabileceği ifade edilmektedir. Yazılımsal ya da donanımsal zeminde ortaya çıkan bu hataları işlevsel hatalar olarak ifade edilirken, sonuç hatalarının tamamen dilsel kullanıma ilişkin küçük bir karakter hatasından kaynaklanan, anlama etkisi olan bir hatadan ileri gelmektedir.

---

<sup>175</sup> Turing, 1950: 447,448.

Makinelerin kendilik taşıyamayacağı yani kendi düşüncelerinin öznesi olamayacağı iddiasına ilişkin itirazı Turing reddetmektedir. Öyle ki bazı düşüncelerin makine tarafından gerçekleştirildiği gösterilebilirse bu itirazı çürütmenin mümkün olacağı ifade edilmektedir. Makinelerin kendi davranışlarının sonuçlarına bağlı olarak yeni kurallar geliştirebileceğinden söz eden Turing, bu tür eylemleri makinenin kendine konu edeceğini, bir anlamda makinelerin kendilik sahibi olacağını iddia etmektedir. Ayrıca makinenin davranış çeşitliliğinden yoksun olduğu iddiasını temelsiz bulan Turing, bir zamanlar depolamanın da sınırlı olduğu ancak çalışmalar sonucu depolamanın arttığını ifade etmektedir. Yani zaman içerisinde bu konuda da ciddi kazanımların elde edileceğini düşünmektedir.

### 2.3.6 Leydi Lovelace'ın İtirazı

Yapay zeka kısmi olarak 1840'larda Leydi Ada Lovelace tarafından öngörülmüştür. Ada, sinir ağları ya da dinamik yapay zekaya ilişkin hiçbir şey bilmeksizin sembollerin ve mantığın gücüne odaklanmıştı. Temel kaygısı teknolojik bir yaklaşım olup makinenin "çok boyuta sahip olan karmaşık, ayrıntılı ve bilimsel müzik parçaları oluşturabileceğini" aynı zamanda "tarihin şanlı bir dönemini mümkün kılmak için doğal dünyanın büyük gerçeklerini" ifade edebileceğini söylemişti. Aklından geçirdiği makine, yakın arkadaşı Charles Babbage tarafından 1834'te tasarlanan Analitik Motor idi. Ada Lovelace, Motorun potansiyel genelleme yeteneğini, "evrendeki tüm konuları" temsil eden sembollerini işleme yeteneğini fark etmişti. Ayrıca modern program karışımının çeşitli temellerini açıkladı: depolanmış programlar, alt yordamlar hiyerarşik olarak iç içe geçmiş adresleme, mikro programlama, döngü, koşullu ifadeler, yorumlar ve hatalar. Ancak müzik bestesinin veya bilimsel muhakemenin Babbage'ın makinesinde nasıl uygulanabileceği hakkında hiçbir şey söylememiştir. Yapay zeka, ona göre mümkündü ama nasıl başarılacağı hâlâ bir muammaydı.<sup>176</sup> Leydi Ada, makine öğrenmesine ilişkin önemli katkılarının yanı sıra Turing'e yönelik itirazıyla bu denli öne çıkmıştır.

Babbage'ın Analitik Motoruna ilişkin en ayrıntılı bilgiler, Leydi Lovelace'ın (1842) bir anısına dayanmaktadır. Bu anıda, Analitik Motorun hiçbir şeyi kendiliğinden meydana getiremeyeceği, onu nasıl komuta edeceğimizi bildiğimiz müddetçe her şeyi yapabileceği düşüncesi vurgulanmıştır. Yani özü itibarıyla Lovelace'ın itirazı makinelerin yeni bir şey yapamayacağı iddiasıdır. Ancak Hartree (1949) için bu ifade "kendi kendine düşünebilen" ya da "öğrenmeye" imkan tanıyan elektronik bir teçhizat tasarlamayı imkansız kılmamaktadır. Hartree ile tamamen hemfikir olan Turing, o dönemlerde belki de bu makinelerin bu özellikleri taşımış olabileceği ancak o dönemde açığa çıkarılamamış olabileceğiyle cevap vermektedir.

<sup>176</sup> Boden, 2016: 7, 8.

Turing, düşünen makineler iddiasının Analitik Makine'ye uygulanabileceğini savunmaktadır. Bunun gerçekleşmemiş olmasını da bu türden bir argümanın Kontes'in ya da Babbage'ın düşünememiş olmasına bağlamaktadır. Bu doğrultuda Lovelace'ın itirazının farklı bir yönü olarak makinelerin "asla yeni bir şey yapamayacağı" düşüncesi karşısında Turing, Lovelace'ın bu fikrini "artık keşfedilecek bir şey yok" demenin kaçamak bir hali olarak değerlendirmektedir. Ayrıca bilginin kümülatif olduğuna dikkat çeken Turing, kimsenin kendi adına "hakiki eser" diyeceği bir şeye sahip olmadığını ve her şeyin genel ilkelerden etkilendiğini savunmaktadır. Turing, bu itirazı bir yönüyle makinelerin bizleri "şaşırtamayacağı" iddiasına benzeterek makinelerin kendisini devamlı şaşırttığını ileri sürmektedir. Bu şaşırtmanın temel nedeninin sistemdeki birtakım hesaplama, voltaj gibi mekanik eksikliklerden kaynaklandığını düşünen Turing, itirafların kendi inandırıcılığını güçlendirdiğini de vurgular. Ayrıca bu tür yaklaşımın kendisini katiben yıldıramayacağını ve tartışmanın ısrarla bilinçlilik problemine çekilmesinin gereksizliğine değinmiştir.<sup>177</sup>

### 2.3.7 Sinir Sistemindeki Süreklilik Argümanı

Makinelerin 0 ve 1 bitlerine dayanan çalışma sistemi ayrık durumlu bir işlevi anlatmaktadır. Oysa insan sinir sisteminin ayrık durumlu bir makine olmadığına dikkat çekilen itirazda; bir sinirde gerçekleşen sinir dürtüsünün ayrık durumlu bir sistem tarafından taklit edilemeyeceği iddia edilmektedir. Ayrık durumlu bir makine ve sürekli bir makinenin farklılığını kabul eden Turing, taklit oyununun koşullarına bağlı kalındığında sorgulayıcının bu farkı hiçbir şekilde keşfedemeyeceğini iddia etmektedir. Turing, daha basit sürekli bir makine olan diferansiyel çözümleyiciyi ele alarak durumu belirginleştirmeye çalışmış ve bunlardan taklit oyununu uygun görmüştür. Bir dijital bilgisayarın diferansiyel çözümleyicinin karşılaştığı bir soruna ne gibi cevaplar vereceğini tam olarak tahmin etmesi mümkün olmasa da ancak doğru cevap türünü verebilecek düzeyde olduğunun altını çizmektedir. Netice itibarıyla bu koşullar altında sorgucunun türevsel çözümleyiciyi sayısal bilgisayardan ayırt etmesinin çok zor olacağını ifade etmektedir.<sup>178</sup>

Makinenin ayrık sistemi ve insanın sürekli bilinci kıyaslaması Turing tarafından cevaplanan önemli bir itirazdır. Bu itirazda makinenin 0 ve 1 sembollerinin ayrıklığına bağlı olarak çalıştığı gerekçesiyle insan bilincinin sürekliliğinden tamamen farklı olduğu vurgulanmıştır. Şöyle ki bilincin bütüncül bir olgu olduğu varsayılarak pek çok algısal reflekslerin eş zamanlı, sürekli ve kesintisiz olarak gerçekleşmesiyle sinir sisteminin yapısının

<sup>177</sup> Turing, 1950: 450,451.

<sup>178</sup> Turing, 1950:451.

buna imkan tanıdığı savunulmaktadır. Turing bu süreklilik ve ayrıklık farkını kabul ederken, taklit oyununun koşullarının tasarımına bağlı kalınmasıyla makine ve insan arasındaki farkın ortadan kalkacağını iddia etmektedir. Yani ayırık bir sisteme sahip olan makinenin bu sınırlılığı verilecek cevapların tasarımıyla ya da bazı kaçamak cevaplarla sürekliymiş gibi gösterilebileceği iddiasını taşımaktadır.

### 2.3.8 Davranışın Belirsizliğine İlişkin Argüman

Her koşulda yapılması gereken şeyler için akla gelebilecek bütün kuralları belirlemenin imkansız olduğu iddiasına dayanan bu itiraz karşısında Turing, bu düşüncelerin tümüyle hemfikir olup insanların makine olamayacağı iddiasını ironiyle karşılamıştır. Soruyu tersinden düşünerek insan hayatını düzenleyen belirli davranış kurallarının varlığı kabul edildiğinde insanın bir makineden daha iyi olamayacağını savunmaktadır. Ancak bu tür kurallar bulunmadığına göre insanların makine olamayacağı sonucunu çıkarmaktadır. Davranışların kurallı ya da kural dışı yönüne ilişkin bu itirazda sorunun içeriği önemli görülmektedir. Bu bağlamda “hareket yasaları” ve “davranış yasaları” arasındaki kafa karışıklığı giderilerek itiraza cevap verilmiştir. Hareket yasaları bir insanın kırmızı ışık görünce durması ya da yeşil yanınca geçmesi örneğiyle açıklanmıştır. Her iki ışığın birlikte yanması durumunda da insanın iradeyle karar alabileceği anlatılmaya çalışılmaktadır. Öte yandan insanın bedeninin çimdiklenmesiyle ortaya çıkan refleks davranış yasalarıyla anlatılmaktadır. Aslında davranış yasaları bir tür makine olduğumuz imasının ötesinde bu türden bir makine olma durumumuzda dahi benzer yasalara maruz kaldığımızı vurgulamaktadır. Yani insanın içsel süreçlerinin kurallardan ibaret olduğu düşünülürse ve bu kurallar bir makineye uygulanırsa tıpkı insanlarda olduğu gibi makinenin de belirsiz görüneceği anlatılmaya çalışılmaktadır.<sup>179</sup>

Turing’e karşı öne sürülen itirazlardan en önemlilerinden birisi olan insan davranışının belirsizliği Aristoteles’in irrasyonel yön ifadesiyle karşılaştırılabilir. En yüksek iyiye mutluluk olarak bildiren Aristoteles’e göre geviş getirme gibi araçsal amaçlara karşın bir de içsel amaçlar bulunmaktadır. İçsel amaçlar insanın nedensizce davranmasına sebep olan şeyin kendisidir.<sup>180</sup> Bu bakımdan, davranışın belirsizliği makinenin asla gerçekleştiremeyeceği bir nitelik olarak öne sürülmektedir. Diğer bir anlamda konu özgür irade ve determinizmle ilişkilendirilmektedir. Bu noktada insan davranışını belirleyen şeyin nedensellikten ibaret olduğu fikrine karşın, özgür iradenin davranışların doğasına olan etkisi vurgulanmaktadır.

<sup>179</sup> Turing, 1950: 452, 453.

<sup>180</sup> Stumpf, 1988: 389 – 393.

### 2.3.9 Duyuüstü Algıya İlişkin Argüman

Turing, duyuüstü algının ve onun bileşenleri olan telepati, sağgörü, önsezi ve psikokinezinin (maddeyi zihinle etkileme) herkesçe bilindiğini varsayarak bütün bu belirsizliklerin bilimsel düşünceleri tümünden inkar ettiğini düşünmektedir. Aslında bunların hepsini çürütmek istediğimizi ancak telepati alanındaki istatistiksel kanıtların ezici boyutta olmasının elimizi kolumuzu bağladığını belirtmektedir. Bütün bunların bizi bir defa etkilemesiyle hayalet ve hortlak gibi şeylerin ardı arkasının kesilmeyeceğini, bakış açımızı bedenimizin bilinen fizik yasalarına göre hareket ettiği fikrine göre değiştirmenin şart olduğunu öne sürmektedir. Bunu oldukça güçlü bir argüman kabul eden Turing, duyuüstü algı ile çatışmasına rağmen birçok bilimsel teoremin pratikte uygulanabilir olduğunu; hatta duyuüstü algı unutulursa her şeyin çok daha güzel yaşanacağı varsayımını paylaşmaktadır. Bunun bir avantaj olduğunu belirterek duyu ötesi algılaşmanın düşünme eyleminin sınırları içinde geçerli bir olgu olmasından korkulduğuna da dikkat çekmektedir.<sup>181</sup>

Bilgisayar ve telepati alanında uzman bir kişinin tanıklığında taklitçilik oyunu oynandığında telepati uzmanı daha fazla doğru tahmin yapsa bile deneyin daha sık aralıklarla gerçekleştirilmesiyle bunun önüne geçilebileceği önerilmektedir. Bu örnekte, “rastgele sayı üreticinin” sorgulayıcının psikokinetik yetilerine tabii olacağı belirtilmektedir. Bir ihtimal psikokinez makinenin olasılık hesaplamasında doğru tahmine neden olabileceğinden sorgulayıcı hâlâ doğru tanımlamayı yapamayabilir. Öte yandan herhangi bir sorgulamada falcılık usulüyle doğru tahminin gerçekleştirilebileceğini de düşünmektedir. Yani duyuüstü algı ile her şey olabilmektedir. Turing; eğer telepati fikri kabul edilirse 'telepati-geçirmeyen' bir oda düzenlemesiyle bunun da üstesinden gelinebileceğini düşünmektedir.<sup>182</sup>

Bu itiraz Turing Testine yönelik en zayıf argüman olarak düşünülebilir. Kaldı ki duyular üstü algı itirazı olarak öne sürülen şey telepatinin ya da beş duyunun da ötesinde altıncı bir duyunun fiziksel şeylere etkisi varsayımından yola çıkılarak sayısal sistemlerin bu türden bir duyuya sahip olamayacağı iddiasıdır. Oysa Turing bu itirazı da ciddiye almış ve sayısal sistemlerin de telepatik güçlerle başa çıkabileceğini 'telepati-geçirmeyen' bir oda örneğiyle cevaplamıştır.

<sup>181</sup> Turing, 1950: 453.

<sup>182</sup> Turing, 1950: 453, 454.

## 2.4 Güçlü ve Zayıf Yapay Zeka

Yapay zeka, güçlü ve zayıf olmak üzere iki başlık altında değerlendirilmektedir. Makinelerin zekiymiş gibi davranabileceği iddiasına “zayıf yapay zeka hipotezi”, makinelerin düşünceyi taklit etmelerinin ötesinde gerçekten düşündüklerine inanılan iddiaya da “güçlü yapay zeka hipotezi” denilmektedir.<sup>183</sup> Searle’e (1998) göre yapay zeka ve işlevselcilik birleşimi beynin akılla ilişkisiz olduğunu ortaya koymaktadır. Öyle ki zihin bir program olarak kabul edildiği için ve herhangi bir donanımda çalışabilir, yani beynin nasıl çalıştığına bakılmaksızın zihnin anlaşılabilirliği ifade edilmektedir. Yani materyalist bakışta dahi zihni incelemek için beyni incelemenin gerekmeyeceği savunulmaktadır. Bu fikrin “bilişsel bilim” disiplini yön verdiği, yapay zeka ve felsefi işlevselcilik teorisinin zihni sadece bilgisayar programı olarak kabul eden fikirde birleştiği belirtilmektedir. Bu görüşü Searle (1998) "güçlü yapay zeka" olarak ilan etmiştir. Aynı zamanda Dennett (1978) tarafından "bilgisayar işlevselciliği" olarak da adlandırıldığının bilgisi verilmektedir. Güçlü yapay zekaya yönelik itirazların genelinin sağduyu itirazlarına yönelik teknik zorlukları sergilediğine de dikkat çekilmektedir. Ayrıca güçlü ya da zayıf yapay zekanın teknik zorluklarından ve itirazlarından yakınan Searle, genelde hepsinin amacının Turing testini geçmek olduğunu ima etmektedir.<sup>184</sup>

Her şeyden önce kavramsal sınırları belirlemeye çalışan Coppin (2004), çalışmaları neticesinde yapay zeka için şu tanımın ortaya çıktığını ifade etmektedir; “yapay zeka, karmaşık sorunları çözmek için insanların ve diğer hayvanların zeki davranışlarına dayanan yöntemleri kullanmayı içerir.”<sup>185</sup> Kitabının son bölümlerinde elde ettiği çıkarımlar ışığında Coppin, insan konuşmasını “anlayabilen” veya en azından insan söylemlerinden “anlam çıkarabilen” ve bu sözlerle dayalı eylemler yapabilen sistemlere bakıldığında bu tür sistemlerin akıllıca davranacak şekilde tasarlanamayacağını, ancak bazı yararlı işlevler sağlayabileceğini bildirmektedir. Bu uğurda kullanılan yöntemlerin de insanların akıllı davranışlarına dayandığının altını çizmektedir. Güçlü yapay zeka ile zayıf yapay zeka arasındaki farklılığa bakıldığında bu ayrımın daha keskin bir tezat oluşturduğunu ifade etmektedir. Güçlü yapay zeka savunucularına göre, yeterli işlem gücüne sahip bir bilgisayar programı sayesinde yeterli zekanın sağlanabileceği, dolayısıyla bir insanın bilinçli olduğu gibi tam anlamıyla düşünebilen ve bilinç sahibi bir bilgisayar oluşturma mümkün olduğuna inanılır. Bu görüşün sahte ve hatta gülünç olduğunu düşünen birçok filozof ve araştırmacının var olduğu gerçeğine de değinilmektedir. Duyguları ve gerçek bilinci olan bir robot yaratma olasılığı, genellikle bilim kurgu alemlerinde keşfedilen ancak nadiren yapay zekanın hedefi olarak kabul edilen bir problem olarak

<sup>183</sup> Russell ve Norvig, 2010: 1020.

<sup>184</sup> Searle, 1992: 43, 44.

<sup>185</sup> Coppin, 2004: 4.

gösterilmektedir. Zayıf yapay zeka için, aksine, zeki davranışın bir modelinin çıkarılmasıyla karmaşık problemleri çözmek amacıyla bilgisayarların insan hizmetine sunulmasından ibaret olduğu vurgulanmaktadır. Bu bakış açısı bir bilgisayarın zekice davranması nedeniyle aslında insanın sahip olduğu türden bir akla erişmesini kanıtlamayacağını savunmaktadır.<sup>186</sup> Oysa Penrose'a (1999) göre güçlü yapay zeka sistemlerinin sadece zeki olmadığı, aynı zamanda us yani akıl kavramına karşılık gelebilecek yetilere de sahip olduğu savunulmaktadır. Sözü geçen akılsal etkinliğin temelini de yine algoritmaların oluşturduğu belirtilmektedir.<sup>187</sup>

Güçlü yapay zekaya olan inancıyla Alan Turing, yeterince karmaşık bir şekilde tasarlanmış sistemlerin gerçek anlamda zekice davranabileceğine kanaat getirmekte ve bunun başlı başına "kritik" seviyenin aşılmasıyla mümkün olduğunu düşünmektedir. Basit ve öngörülebilir eserler ötesinde makinelerin karmaşıklığı artırıldığında sürprizlerin olabileceğini öne süren Turing'e göre başarı bir nevi kritik boyut meselesi olarak düşünülmektedir. Kritik boyutun altında hiçbir şey olmazken kritik boyutun üstünde kıvılcımların uçuşabileceği, belki de beyinlerde ve makinelerde de bunun mümkün olabileceği iddia edilmektedir. Beyinlerin ve makinelerin çoğunun kritik boyutun altında olduğundan uyaranlara hantal bir şekilde tepki verdikleri ve kendi fikirleri olmadığı, sadece mevcut yanıtları verebildikleri düşünülmektedir. Ancak şu anda birkaç beyin ve muhtemelen gelecekteki bazı makinelerin çok kritik olduğuna ve kendiliğinden bu seviyelere erişebileceğine inanılmaktadır. Turing, bunun sadece bir karmaşıklık meselesi olduğunu ve belli bir kritik seviyenin üzerinde niteliksel bir fark ortaya çıktığında süper kritik makinelerin şimdiye kadar öngörülen basit makinelerden oldukça farklı olacağını öne sürmektedir.<sup>188</sup>

Russell ve Norvig'e (2010) göre yapay zeka araştırmacılarının çoğu zayıf yapay zekayı tabiri caizse çantada keklik görmektedirler. Sistem işlediği müddetçe buna zeka taklidi denilmesini ya da gerçekten zeki olduklarına inanılmasını çok da umursamamaktadırlar. Tüm yapay zeka araştırmacılarının esas amaçlarının, bu alandaki çalışmaların etik çıkarımları ile ilgili olması gerektiğinin altı çizilmektedir. Russell ve Norvig'in vurguladıkları bu düşünce esasında yapay zekanın gelecekte doğuracağı sonuçları doğrudan ilgilendiren hususlardır. Zira inanılmaz davranışlar sergileyen yapay zeka sistemlerini geliştirmek bir şeyken, bu cihazların kullanımıyla doğabilecek etik sonuçları öngörebilmek bambaşka bir şeydir. Aslında insanlığı bekleyen en mühim tehlike bu çerçevede değerlendirilmelidir. Yapay zekanın doğurabileceği olası etik sonuçlar ön görülmeksizin; teşhisler, tespitler veya iyileştirmeler için çok geç kalınmış olabilir. Bilindiği gibi olgulara ilişkin problemlerin çözümlenmesi içinden çıkılmaz

<sup>186</sup> Coppin, 2004: 4-5.

<sup>187</sup> Penrose, 1997: 18.

<sup>188</sup> Lucas, 2003: 125-126.

bir duruma dönüşebilir. Bu da insanlığın gidişatını olumsuz yönde etkileyebilecek sonuçlar doğurabilir. O halde olası sonuçların felsefi diyalektik söylem aracılığıyla değerlendirilmesi gerekmektedir. Bu doğrultuda özellikle zihin felsefecilerinin güçlü yapay zeka düşüncesinin temelsiz olduğuna ilişkin argümanlarına değinme gereksinimi doğmaktadır. Searle’ün Çince Odası Argümanı, Turing’in iddialarını çürütmek maksadıyla ortaya atılan bir düşünce deneyidir. Bu düşünce deneyinin amacı niceliksel sembollerin zihnin davranışlarını temsil edemeyeceği fikrini ispatlamaya çalışmasıdır. Searle herhangi somut bir deneye başvurmaksızın tamamen düşünsel bir deney örneği olan Çince Odası Argümanında sentaktik bir sisteminin ne insan sağduyusunu ne de akılsal süreçleri gerçekleştirebileceğini ifade etmektedir. Ancak güçlü yapay zekanın şu anda olmasa da gelecekte gerçekleşeceğine ilişkin fizikalist bakışın güçlü inancı görülmektedir. Bu inancın altında somut deneylerden ziyade kavramların anlamının daraltılması yatmaktadır. Diğer bir ifadeyle insan aklının sentaksa indirgenmesi her iki kavramı da aynı düzeleme oturtacaktır. Descartes ile başlayan kavramsal değişim tarihin akışıyla belirginleşmiştir. Bu öngörüü bir örnek üzerinden temellendirmek gerekirse literatürde “yapay zeka” kavramı kullanılırken, “yapay akıl” ifadesi yer almamaktadır. Akıl kavramının kapsamı ve sayısal sistemlerde taklit edilemezliği nedeniyle yerine “zeka” sözcüğü yüceltilerek kullanılmıştır. Bu yaklaşım “insan” ve “makine” arasındaki farkı ortadan kaldırmak üzere kurgulanmıştır. Güçlü yapay zeka için sonu görünmeyen bir yol olduğuna kanaat getiren araştırmacılardan anlaşılağı üzere insan doğasının gerçek anlamından uzaklaştırılarak mekanik bir temele yerleştirilmesi hedeflenmektedir. Özetle güçlü yapay zeka sistemleri gerçek zihni olan sistemler olarak tanımlanması ve insan kontrolünden çıkması olasılığı nedeniyle yıkıcı etkiler doğurabilecek sistemler olarak kabul edilmektedir. Bu sistemlerin en kaygı verici yönleri güçlü yapay zekanın iradeye sahip olması, sorumluluk sahibi olması gibi tartışmalarıdır. Oysa zayıf yapay zeka tamamen insanın kullanımı için bir aracı görevindedir. O nedenle güçlü yapay zekanın neden olacağı endişelerin zayıf yapay zekada gerçekleşmesi beklenmemektedir. Ancak, bir sistemin yıkıcı etkileri sadece irade sahibi olmasıyla açıklanamaz. Bazı zayıf yapay zeka örneklerinin kötü ellere geçmesiyle de felaketler yaşanabilmektedir. Bu bağlamda sonuçlar hem güçlü hem de zayıf (dar) yapay zeka örnekleriyle değerlendirilmelidir. Her iki türden yapay zekanın sorunları “otomasyon ve işsizlik, özgünlüğün kaybolması, istenmeyen sonuçlar, sorumluluk - gizlilik ve insan türünün sona erme sorunu” gibi başlıklar altında son bölümde incelenmiştir. Bu başlıklar felsefi incelemende ziyade toplumsal, hukuki çevresel, psikolojik vb. sorunları kapsamaktadır. O nedenle tezin esaslı hedefi olan felsefi değerlendirmeler son sonuç bölümünde incelenmiştir.

## 2.5 Çince Odası Argümanı ve İtirazlar

Alan Turing'in (1950), bilgisayarın insan zihnini taklit edebileceği hatta insan zihninin bütün işlevini yerine getiren bir bilgisayarın var olabileceği iddiasını bir önceki bölümde incelemiştik. Yapay zekayı güçlü ve zayıf olmak üzere ikiye ayıran Searle, Turing'in iddiasını güçlü yapay zeka iddiası olarak değerlendirmektedir. Fotion'un (2000) tanıtımıyla John Searle analitik felsefe geleneğinde büyüdüğü gibi bu geleneği aşmıştır. Bunun nedeni, bu geleneğin odağında dar bakış açıları olmasına rağmen Searle'ün çeşitli konularda yazılar yazması gösterilmektedir. Ayrıca uzun kariyeri boyunca Searle'ün dil felsefesi, zihin felsefesi, sosyal kurumların doğası ve yapısı, bağlam (Ağ ve Arka Plan dediği şey), ontoloji, bilim ve nedensellik gibi konularda geniş kapsamlı yazılar yazdığının bilgisi verilmektedir.<sup>189</sup> Öncelikle Searle tarafından Çince Odası argümanının neden ortaya çıktığını kimi veya neyi hedef aldığını özetlememiz gerekmektedir. Tanıtımdan da fark edileceği üzere Searle'ün, kolay kabullenen ya da itaatkar bir kişilik olmadığı ortadadır. Bu bağlamda Turingçi yapay zeka düşüncesini güçlü yapay zeka iddiası olarak kavramlaştırarak insan beyninin bilgisayar donanımıyla ve insan zihninin bilgisayar programıyla özdeşleştirildiği kanısına varmıştır. Bu özdeşlik şu şekilde ifade edilmektedir:

“Bu görüş, insan beyni konusunda temel olarak biyolojik hiçbir şeyin olmadığı sonucuna dayanmaktadır. Sonsuz sayıda farklı türleri olan bilgisayarların donanımları gibi, beyin de yalnızca insan zekasını oluşturan programları saklar. Bu bakış açısına göre, ne olursa olsun, doğru bir programı ve doğru ‘girdileri’ ve ‘çıkıtları’ olan herhangi bir fiziksel sistem, tam olarak sizin ve benim zekalarımızla aynı anlamda bir zekaya sahip olacaktır.”<sup>190</sup>

Alıntıdan açıkça anlaşılacağı üzere Searle (1996), güçlü yapay zeka iddiasını çürütmek amacıyla Çince Odası argümanını öne sürmüştür. Bu argümanda temel mantık programı çalıştıran bilgisayarın yerine bir insanı koymaktadır. Yani insanın bir bilgisayar gibi yazılımı çalıştırdığı farz edilmektedir. Kilitli bir odada anadili İngilizce olan bir kişi ile bir sürü Çince semboller ve bu Çince sembollerin ayrımını, tasnifini, sınıflandırmasını açıklayan İngilizce bir talimat kitabı bulunmaktadır. Odanın dışında anadili Çince olan kişilerce kapının altından Çince soruların yazıldığı kartlar atılmaktadır. Bu kartlardaki Çince semboller hakkında hiçbir bilgisi olmayan kilitli odadaki kişi elindeki kitap aracılığıyla sembolleri uygun yerlere yerleştirmeye çalışır, yerleştirdikten de kapının altından geri atması beklenmektedir. Kilitli odadaki kişinin Çince semboller hakkında ya da kartları atanlar hakkında hiçbir bilgisi bulunmamaktadır. Kapı altından atılan Çince kartlar birer soru, içerde düzenlenen semboller de bu soruların cevabı niteliğindedir. Öyle ki odadaki kişinin bunlardan hiçbir haberi olmadığı gibi, kurala göre

<sup>189</sup> Fotion, 2000: 1.

<sup>190</sup> Searle, 1996: 39.

sembolleri dizmekten başka görevi de yoktur. İçerideki kişinin elindeki kural kitabı bilgisayar programını temsil ederken, Çince sembollerin yazıldığı kartlar da veri tabanına tekabül etmektedir. Odanın içindeki kişinin kendisi de bilgisayar ya da merkezi işlem yürütme birimini temsilen görev yapmaktadır. Görevi sadece gelen kartları odaya alıp, Çince sembolleri kural kitabına göre işleyerek hazırladığı kartı dışarı vermek olan kişinin bütün bu süreçte dair hiçbir fikri yoktur.<sup>191</sup> Kısacası kilitli odadaki kişiden bir bilgisayarın bilgiyi işlemesi gibi sadece kendine verilen görevi yerine getirmesi beklenmektedir. Yani amaç, kartlardaki sembollerin anlamsal ya da başka bir boyutuna bakılmaksızın kurallar çerçevesinde gereken düzenlemelerinin yapılmasıdır.

Mükemmel derecede hazırlanmış bir program karşısında kartları karıştırarak çok iyi derecede eşleştiren kişinin, bir müddet sonra Turing testini geçtiği varsayılmaktadır. Çünkü odadaki kişinin cevapları (düzenlenmiş sembolleri) gerçek bir Çinlinin vereceği cevaplardan ayırt edilememektedir. Burada esas mesele şudur; program ne kadar etkili olursa olsun kişinin cevapları gerçek bir Çinliyi ne kadar taklit ederse etsin, sonuçta odanın içindeki kişi Çincenin tek bir sözcüğünü dahi anlamayacaktır. Kişinin Çinceyi anlamama nedeni bir programı çalıştırıyor olması ise bu temeldeki hiçbir dijital bilgisayarın hiçbir şey anlamayacağı ortaya çıkmaktadır. Çünkü bir bilgisayarın kapasitesi tamamen bu süreçle sınırlıdır. Yani bilgisayarlar sembollerle işlenebilen bir kurallar düzenine sahiptir. Burada dikkat çekici husus İngilizce bilen kişi için İngilizcedeki cevapların neye benzediği ile Çincedeki cevapların neye benzediğinin karşılaştırılmasıdır. İngilizce bilen kişi hiçbir şey anlamadığı halde, Çince sembolleri tıpkı bir bilgisayar gibi kurallara göre tasnif edilebilmiştir. Dolayısıyla odanın dışındakiler içeride bir Çinli varmış hissine kapılmaktadır. Bu durumda Turing testinin geçildiği ima edilmektedir. Ancak Searle'e göre içerideki kişinin anlamsal ya da semantik olarak bir bilinci bulunmamaktadır. Anadili İngilizce olan kişi İngilizce sembollere anlam yüklediğinden, İngilizce sözcükler anlamlıdır. Aynı kişi için Çince semboller sadece biçimsel, sentaksla ilgili nesnelerden ibarettir. O halde bilgisayar ve zihin arasındaki ayrım; bilgisayarın biçimsel sembollerle sınırlı kalması, zihnin ise bunun ötesinde anlamı taşıması noktasında yapılmaktadır. Searle, bunu bilgisayarın bir zayıflığı olarak görmeyip, bir bilgisayarın yapabileceği her şeyin bundan ibaret olduğunu iddia etmektedir. Buna dayanarak programların bir zihni olmadığını ifade eder.<sup>192</sup> Turing'in (1950) iddiasına çeşitli çevrelerden itirazlar olduğu gibi çeşitli itirazlar Searle için de yapılmış olup Searle bu itirazlara cevap niteliğinde açıklamalar yapmıştır.

<sup>191</sup> Searle, 1984: 30.

<sup>192</sup> Searle, 1984: 29-33.

### 2.5.1 Sistemler Yanıtı (Berkeley)

Çince Odası'na yönelik itirazların başında bütün sistemin Çince bildiği iddiası gelmektedir. Yani kilitli odanın, sembollerle dolu sepetlerin, kuralları içeren defterin ve belki de diğer öğelerin de dahil olduğu her şeyin bütün olarak Çince bildiği düşünülmektedir. Ancak burada gözden kaçırılan noktayı Searle, kilitli oda içindeki kişinin programı işleten mekanizmadan başka bir şey olmadığı savıyla, Çincenin kesinlikle semantik anlamda anlaşılamayacağını bilgisini vermektedir. Bu haliyle daha önce de Searle'ün belirttiği gibi bu sistemin söz diziminden semantiğe, yani anlama geçmesinin imkansız olduğunun altı çizilmektedir. Dolayısıyla ne sistemin bütününe ne de tek başına merkezi işlem biriminin bu semboller hakkında fikir sahibi olamayacağı vurgulanmaktadır.<sup>193</sup> Sistem teorisine verdiği yanıtın oldukça basit olduğuna dikkat çeken Searle, sistemin tüm bu öğelerini kişinin içselleştirmesine izin verilirse defterdeki kuralları ve Çince sembolleri ezberlerse ve tüm hesaplamaları kafasında yaparsa tüm sistemi dahil edeceğini düşünmektedir. Kilitli odanın da yok sayılması durumunda sistemin kapsadığı hiçbir şey olmayacağından, kişinin yine de Çince hiçbir şeyi anlamayacağını iddia etmektedir. İnsanın anlamadığı yerde sistemin anlayabilmesini mümkün görmemektedir. Sistem teorisine bu cevabı vermekten bile utandığını ifade eden Searle, teorinin başından beri çok mantıklı görünmediğini düşünmektedir. Buradaki mantıksızlığı kişinin Çinceyi anlamadığı halde, o kişinin, kağıt parçalarının ve diğer şeylerin birleşiminin Çinceyi anlayabileceği düşüncesi olarak göstermektedir. Yine de güçlü yapay zeka ideolojisine bağlı birçok insanın sonunda buna benzer şeyler söylemeye meyilli olduğunu düşünmektedir.<sup>194</sup>

Searle'e yöneltilen sistemler eleştirisine bakıldığında Çinceyi bilme eyleminin tikele indirgenemeyeceği ancak tümel olarak sistemin geneline haiz olduğu düşüncesi vurgulanmıştır. Oysa Searle bu eleştiriye ciddi bir cevap vermeye dahi yanaşmamaktadır. Kaldı ki ona göre Çince Odasındaki tek kişinin geçirdiği süreci de tüm sistem geçireceği için, odadaki kişinin sahip olduğu biliş ile sistemin genelindeki bilişin birbirinden farksız olduğu iddia edilmektedir.

### 2.5.2 Robot Yanıtı (Yale)

Çince Odası argümanına yöneltilen bir diğer eleştiri de robot cevabıdır. Searle, Çinceyi semantik olarak anlayabilen bir programın robota yüklenmesi fikrini sunmaktadır: “Bir bilgisayarı bir robotun içine koyduğumuzu ve bu bilgisayarın biçimsel sembolleri sadece girdi-çıkı olarak işlemekten öte algılamak, yürümek, dolaşmak, çivi çakmak, yemek, içmek gibi her

<sup>193</sup> Searle, 1984: 32.

<sup>194</sup> Searle, 1980: 419.

şeyi yapabildiğini varsayalım.”<sup>195</sup> Bu varsayımı sağduyu bağlamında çürütmeye çalışan Searle, robotun uzuvlarını hareket ettiren ve yön bulmasını sağlayan kameraların bir sandviç ile karşılaşınca sandviç sembolünü tetiklemesinin tek başına yeterli olmayacağını düşünmektedir. Zira elde olan tek şeyin bir sembolden ibaret olduğu düşüncesiyle sandviç hakkında hiçbir önsel (*a priori*) bilginin olmadığını iddia etmektedir. Sağduyu sahibi bir varlık olan insan düşünüldüğünde bu durumun farklı işlediğini ifade etmektedir. Ayrıca robot ile dünyanın geri kalanı arasındaki nedensel etkileşimlerin zihin ekseninde temsil edilmediği sürece anlamsız olacağının bilgisini vermektedir.<sup>196</sup>

Robot yanıtında Searle’ün amaçladığı şey, tüm fenomenlerin niceliksel semboller aracılığıyla temsil edildiği bir durum varsayıldığında ve bir robotun bu veriler ışığında işlediği düşünüldüğünde dahi başarısızlıkla sonuçlanacağına inandığını belirtmektedir. Kaldı ki fenomenler salt neyin bilgisinden öte nasılın bilgisini de içerdiğinden deneyimlerin niceliksel karşılığının sembollerle var olamayacağına inancını belirtmektedir.

### 2.5.3 Beyin Simülatörü Cevabı (Berkeley ve M.I.T.)

Bu yanıtta ileri sürülen fikir bir bilgisayar programı yerine ana dili Çince olan birisinin beyindeki sinirlerin çözümlenmesi ya da yapay beyin niteliğinde bir bilgisayar sistemi tasarlanması düşüncesidir. Kısacası girdi ve çıktı temelinde çalışan bu sistem verileri alarak yapay sinirlere yönlendirir, işlenen veriler yine makine aracılığıyla çıktı olarak sunulur. Searle’e göre bu bakış açısında irdelenmesi gereken en önemli nokta, beyindeki sinirlerin, hücrelerin, bunların bağlantı noktaları olan sinapslara benzer bir yapı kurmanın zihin sahibi olan bir beyin oluşturmak anlamına gelmeyeceğidir. Searle, beyin simülatörünün çıkmazını, sinir hücrelerinin uç noktalarındaki ateşlemelere indirgeyen bir anlayış taşıdığı gerekçesiyle hedef almaktadır. Bu bağlamda; beyin simülatörünün, beyin yalnızca biçimsel yapısını taklit edebildiğini ama beyine ait temel unsurlar olan nedensellik ve amaçlılık durumlarını göz ardı ettiğini belirtmektedir.<sup>197</sup>

Searle bu açıklamasıyla iddia edilen beyin simülatörünü iki yönden incelemektedir. Yapısal anlamda beyine benzer bir yapı kurulsa dahi, bu yapının nedensellik ve amaçlılık durumlarını gerçekleştiremeyeceği düşünmektedir. Öte yandan Searle bu yanıtında aklın nedenini beyin olarak nitelendiren bakışın karşısında nedensel ve ereksel niteliği olan bir olgu olarak kabul ederek bir nevi beyin ve zihne ilişkin kişisel görüşünü de belirtmektedir.

<sup>195</sup> Searle, 1980: 420.

<sup>196</sup> Searle, 1984: 32-34.

<sup>197</sup> Searle, 1980: 420-421.

#### 2.5.4 Birleşim Cevabı (Berkeley ve Stanford)

Sistem, beyin simülatörü ve robot iddialarının tümünü tek bir birimde birleştirilerek programlanan yeni bir bilgisayarın bir robota yerleştirilmesiyle; insan davranışlarından ayırt edilemeyen girdi ve çıktı mantığının ötesinde bir robotu tasarlama düşüncesi birleşim cevabıdır. Bu durumda sistemin yönelimsellik sahibi olacağına inanılmaktadır. Bu varsayım ile iddiaların tümünü çürütmeyi hedefleyen Searle, güçlü yapay zekanın mantığı itibarıyla çelişkiye düşeceğini iddia ederek şu ifadeyi kullanır:

"Güçlü yapay zekaya göre, biçimsel bir programın doğru girdi ve çıktı ile başlatılması, aslında niyetliliği teşkil etmenin yeterli bir şartıdır. Newell'in (1979) belirttiği gibi, zihinsel özün fiziksel bir sembol sisteminin işleyişine dayanır. Ancak bu örnekte robota yaptığımız kasıtlılık niteliklerinin biçimsel programlarla ilgisi yoktur."<sup>198</sup>

Yukarıdaki örnekte bir robota atfedilen biçimsel olmayan özelliklerin, güçlü yapay zekaya temel oluşturamayacağına değinen Searle, sadece bize benzediği için bir sistemin yönelimsellik sahibi olduğu iddiasına karşı çıkmaktadır. Kaldı ki bunun tersi düşünülürse yani bir robot tamamen biçimsel bir programla işlerse kukladan farkı kalmayacaktır. Bu durumda bir kuklanın zihin sahibi olabileceği hipotezini yersiz ve gereksiz gören Searle, artık robota veya bir parçası olduğu sisteme, yönelimsellik atfetmek için hiçbir neden olmadığını vurgulamaktadır.<sup>199</sup>

Önceki eleştirilerin her birini sistematik bir şekilde çürüten Searle'e tüm sistemlerin bir arada bulunmasıyla yeni bir hamle yapılmıştır. Ortaya konulacak yapının özne niteliklerine sahip olacağı inancıyla bir araya getirilen sisteme Searle umutsuz bakmaktadır. Salt benzeşimle yönelimselliğin gerçekleştirilemeyeceğini düşünmektedir. Bu yönüyle algoritmik düzlemin zihnin engebeli, düzensiz, uçsuz bucaksız yönlerini karşılayamayacağını ifade etmektedir.

#### 2.5.5 Diğer Zihinler Cevabı (Yale)

"Diğer insanların Çinceyi veya başka bir şeyi anladığını nasıl bilebiliriz?" sorusuyla öne çıkan itiraza "Sadece davranışlarıyla." diyerek davranışların ispatını öne süren bu iddiaya Searle, kısa bir yanıtla cevap vermektedir. Tartışmadaki sorunun diğer insanların bilişsel durumlara sahip olduklarını nasıl bildiğimizle ilgili olmadığını, atfedilen bilişsel durumların bizzat kendisiyle ilgili olduğunu öne sürmektedir. Argümanın itici gücünü Searle, bilişsel durumların çıktıları olmasa da hesaplama işlemlerinin var olabilmesiyle bağdaştırmaktadır. Bu bağlamda tıpkı fen bilimlerinde fiziksel cisimlerin gerçekliğinin ve bilinebilirliğinin önceden kabul edilmesi zorunluluğu gibi, "bilişsel bilimlerde" de kişinin zihinsel olanın gerçekliğini ve

<sup>198</sup> Searle, 1980: 421.

<sup>199</sup> Searle, 1980: 421.

bilinebilirliğini önceden kabul etmesi gerektiğini belirtmektedir.<sup>200</sup> Kısacası hesaplamalı süreçlerin sadece çıktılarla anlaşılamayacağı gibi, diğer insanların bir şeyi bilebilmesi de sadece davranışlarıyla açıklanamamaktadır.<sup>201</sup>

Bu itirazda Çinceyi bilmenin ötesinde pek çok şeyi bilmenin ispatının davranışsal yönüne dikkat çekilmektedir. Oysa Searle'ün tepkisi bireysel bilişin ne olduğuna bakılmaksızın, problemin bilişin kendi doğasıyla ilgili olduğuna dikkat çekmesidir. Bu yönüyle itiraz kısa bir cevapla geçirilmiştir.

### 2.5.6 Çoklu Konumlar Cevabı (Berkeley)

Başlıktan anlaşılacağı üzere burada Searle, farklı itirazlara ilişkin cevaplarını sunmaktadır. Searle'ün yapay zekanın olası ilerleyişini göz ardı ettiği iddia edilmektedir. Bu yönüyle argümanlarının yalnızca kısıtlı bir teknolojiyi kapsadığı gerekçesiyle Searle'e itirazlar yöneltilmektedir. Bu bağlamda teknolojik gelişmeler kapsamında tam olarak öne sürülen itiraz şöyledir:

“Tüm argümanınız yapay zekanın sadece analog ve dijital bilgisayarlarla ilgili olduğunu iddia ediyor. Ama bu sadece teknolojinin şu anki halinden ibaret. Yönelimsellik için gerekli olduğunu söylediğiniz bu nedensel süreçler ne olursa olsun, (haklı olduğunuzu düşünerek) eninde sonunda bu nedensel süreçlere sahip ve yapay zeka diyebileceğimiz cihazları inşa edebileceğiz. Bu nedenle argümanlarınız yapay zekanın bilişi üretme ve açıklama yeteneğine asla yöneltilmemiştir.”<sup>202</sup>

Bu cevabın kendi kendini çürüttüğünü iddia eden Searle, güçlü yapay zeka projesini gerçekten yapay olarak bilişi ürettiği ya da açıkladığı iddiasıyla yeniden tanımlayarak önemsizleştiren bu cevaba itirazının olmadığını belirtmektedir. Bu bakış açısı karşısında beynin işleyişiyle ilgili önemli olan şeyin, sinapsların dizilişindeki biçimsel gölge olmadığını ama dizilerin gerçek özellikleri olduğunu iddia eden Searle, yapay zekanın güçlü versiyonuna dair tüm argümanların biliş tarafından kullanılan gölgelerin etrafına bir taslak çizdiği ve ardından gölgelerin gerçekliği hususunda ısrarcı davrandığına dikkat çekmektedir. Yapay olarak sinir sistemine sahip bir makine, akson, dendritli nöronlar ve diğer her şeyin bizimkine benzer şekilde üretilebileceği varsayılırsa sorunun cevabının “evet” gibi görüneceğini de belirtmektedir. Nedenlerin birebir kopyalanmasıyla sonuçların da kopyalanabileceğine dikkat çeken Searle, bu yöntemle insanların sahip olduklarından başka bazı kimyasal prensipler kullanarak bilinç, yönelim ve diğerlerini üretmenin de mümkün olabileceğini varsaymaktadır. O halde deneysel bir soruya başvurarak neticeye ulaşmaya çalışır: “Tamam ama dijital bir

<sup>200</sup> Searle, 2005: 353.

<sup>201</sup> Searle, 2005: 353-354.

<sup>202</sup> Searle, 1980: 421.

bilgisayar düşünebilir mi?" Eğer "dijital bilgisayar" ile kastedilen şey, bir bilgisayar programı örneğinde olduğu gibi tanımlanan bir makine ise sorunun cevabının yine ‘evet’ olduğunu belirtmektedir. Zira bizler pek çok bilgisayarın örneklemi olduğumuzdan ve düşünebildiğimizden bunun böyle olduğunu düşünmektedir. Ancak sorulması gereken doğru soruların "Fakat bir şeyler sadece doğru programa sahip bir bilgisayar olduğu için düşünebilir, anlayabilir ve böyle yapabilir miydi? Doğru olan bir programın başlatılması, anlayışın yeterli bir şartı olabilir mi?" olduğuna inanan Searle bu sorulara ‘hayır’ cevabını vermiştir. Çince Odası örneğinde bir insanın yönelimsellik sahibi olabileceği ama bu insanı biçimsel bir program aracılığıyla programlayınca biçimsel programa ilaveten herhangi bir yönelimsellik katmadığını ispatlama çabası savunulmaktadır. Yani insanın Çinceyi bilme ya da anlama kabiliyetine herhangi bir şey eklememektedir.<sup>203</sup>

Çoklu konular başlığında değerlendirdiği diğer birkaç hususta bilgi işleme, düalizm, davranışçılık ve işlevselcilik konuları hedef alınmaktadır. “Bilgi işleme” kavramına ilişkin kafa karışıklığı sebebiyle zihinle insan beyninin “bilgi işleme” denilen bir şey yaptığına ve bilgisayarın da programlar vasıtasıyla bilgi işleme yaptığına inanıldığı düşüncesi vurgulanmaktadır. Yangınlar ve yağmur fırtınalarının bilgi işlemeyi hiç yapmadığına ironiyle değinen Searle, bu hususta bilgisayarla beyin arasında bilgi işlemenin benzerliği açısından hiçbir ilişki kurulmayacağını savunmaktadır.<sup>204</sup>

Diğer bir husus da davranışçılık ve işlevselcilik bakış açısını hedef almaktadır. Yapay zekanın çoğunda kalıcı davranışçılık ya da işlevselcilik olduğuna değinen Searle, uygun şekilde programlanmış bilgisayarlar insaninkine benzer girdi-çıkı düzenlerine sahip olabileceği için bilgisayardaki zihinsel durumları insanın zihinsel durumlarına benzer şekilde önermeye meyilli olabileceğimizi ifade etmektedir. Ancak bu dürtüden gerektiği zaman vazgeçebilmek gerektiğini de düşünmektedir.<sup>205</sup>

Searle, işlevselciliğin bir düalizm kalıntısıyla birleştiğine dikkat çekerek; güçlü yapay zekanın yalnızca zihnin söz konusu olduğu yerde beyin önemli olmadığı varsayımıyla anlamlı olacağını düşünmektedir. Güçlü yapay zekada ve işlevsellikte programların yani yazılımın önemli olduğu ve makinelerde uygulanmasına bakılmaksızın değerlendirildiği vurgulanmaktadır. Zihinsel olguları algoritmaya açıklama ve tasarlama gayesi bu projenin temelini düalizm ilkesiyle şekillendirildiğinin kanıtıdır.<sup>206</sup>

<sup>203</sup> Searle, 1980: 422.

<sup>204</sup> Searle, 1980: 423.

<sup>205</sup> Searle, 1980: 423.

<sup>206</sup> Searle, 1980: 424.

Çince Odası argümanında Searle'e kısmen katılan Churchland (1990) "sinir ağları"<sup>207</sup> argümanını öne sürmüştür. Verilerin belli kurallara göre işlenmesi için merkezi birime sahip olduğu bilinen klasik yazılım, hesaplamaları birbiri ardına adım adım gerçekleştirmektedir. Bu anlamda klasik bir yazılım, beynin nörofizyolojik yapısından muazzam ve karmaşık nöron ağlarından oluşan bir yapıdan farklıdır. Dolayısıyla Churchland klasik bilgisayarın insan beyni işlevlerini ve düşünmesini taklit edebilme şansının olmadığını iddia ederek Searle'ün görüşüne katılmaktadırlar. Ancak bu görüş birliğinin nedeni Çince Odası argümanının sonucu değil, beynin yapısı ile klasik yazılımın yapısı arasındaki uyumsuzluk olarak görülmektedir. O halde sinir ağı yazılımı insan beyninin yapısını ve faaliyetini taklit eden bir model oluşturma girişimi olarak düşünülmektedir.<sup>208</sup> Rakover (1999), ait sinir ağları yazılımının çalışma mantığını şu şekilde özetlemektedir:

“Birbirlerini karşılıklı ve paralel olarak etkileyen, dallanıp budaklanan bağlantıların kendi aralarında birleştiği, geniş ağlar biriminden oluşur. Bu etki ağıdaki birimlerin faaliyetinin çeşitli uyarım gücüyle ifade edilir. Bu güçler, bir bağlantının büyük bir ağırlığının ünitenin faaliyetinin uyarılması anlamına geldiği “ağırlıklar” olarak adlandırılır. Ağ, genel aktiviteyi ağırlıklardaki değişiklikler aracılığıyla değiştirir, bu ağın çevresiyle tekrar tekrar temas etmesi, ağın girdi ve çıktısıyla yaratılan temasın sonucudur. Ağın genel faaliyeti, kendisi ile çevre arasındaki karşılıklı ilişkileri öğrendiği ve hatırladığı şeyi temsil eder. Böylece Sinir Ağı, insan davranışını, ağırlıklarının desenindeki bir değişiklikte taklit eder, böylece ağın aldığı girdi ile çıktı arasındaki ilişki insan tepkileri ile insanın maruz kaldığı uyaranlar arasındaki ilişkiyi maksimum hassaslıkla eşleştirir.”<sup>209</sup>

Churchland'a göre, sadece klasik yazılımlı bilgisayarları kapsayan Çince Odası deneyi sinir ağı bilgisayarını yansıtmadığı için geçerli değildir. Bu nedenle beynin anlam yaratmada nasıl başarılı olduğunu bilmiyor olsak da oldukça karmaşık bir sinir ağı bilgisayarının ilke olarak düşünebileceğine inanmaktadırlar. Bu insanlar, sırasıyla sinir ağındaki birimleri ve bağlantılarını temsil ederek ağ modelindeki birimlerin işlemlerinin aynısını gerçekleştirmektedirler. Bu itiraza cevabını Searle, “Çin Stadyumu” örneğiyle vermektedir. Hedef, ağ bilgisayarının yapabildiklerini insan dolusu bir stadyumun gerçekleştirmesidir. Yani Çince Odası örneğindeki tek insan yerine binlerce insanın bu görevi yerine getirmesi beklenmektedir. Bu durumda düşüncenin gerçekleşeceğine yönelik iddia Churchland tarafından öne sürülmüştür. Ağ bilgisayarı ve insan stadyumu arasında fark olmadığına değinen Searle, ağ bilgisayarına belli bir uyaran modeli verildiğinde modeli hesaplayacağını ve belirli bir çıktı ile dönüt vereceğini ifade etmektedir. İnsan dolusu stadyumun da aynı şeyi yapacağını iddia ederek

<sup>207</sup> Bu kavram literatürde "bağlantıcılık" ve "Paralel Dağıtımli İşleme" (*Parallel Distributed Processing*) olarak da geçmektedir. Detaylı bilgi için bk. Rakover, 1999:59.

<sup>208</sup> Rakover, 1999: 59.

<sup>209</sup> Rakover, 1999: 60.

bu insan stadyumunun Çince bilip bilmediğini sormuştur. Soruyu "hayır" olarak cevaplayan Searle, ağ bilgisayarının düşünemeyeceğini bunun bizi şaşırtmaması gerektiğini, klasik ya da ağ bilgisayarı olduğuna bakılmaksızın her ikisinin de semboller üzerinde mekanik işlemler gerçekleştiren makineler olduğu iddiasını sürdürmektedir. Bu nedenle bilinçli anlam uğrunda herhangi zihinsel bir deneyim yaratma olasılıkları olmadığını savunmaktadır. İkna olmayan Churchland çifti; bağlantıcı görüşlerinin izinde her bir kişinin ayrı ayrı Çinceyi anlamadığını ama hepsinin birlikte aslında Çinceyi konuştuğunu ve anladığını iddia etmektedirler. Sistem argümanına oldukça benzeyen bu argümanda bir sistemin parçaları Çinceyi anlamadığı halde, tüm karmaşık sistemin bunu yapabileceğini iddia edilmektedir.<sup>210</sup>

Searle'ün izlediği yola ilişkin birtakım problemlere dikkat çekilmiş olup güçlü yapay zekanın sadece ikili bilgisayar sisteminden ibaret olmadığı düşüncesi de yapılan itirazlar arasındadır. Bu bağlamda Rey (1986), güçlü yapay zekanın normal insan ve girdi-çıkı denkliliğinden ibaret olmadığını öne sürmektedir. Ancak girdi ve çıktıların ötesinde güçlü yapay zekanın sistemin içsel durumlarıyla ilişkili "işlevselci" zihin kuramları olarak adlandırılan bir sürüm olduğu düşünülmektedir. Yani güçlü yapay zekayı davranışçı temelde inceleyen Searle'ü eleştirerek aslında güçlü yapay zekanın işlevselci çizgide incelenmesi gerektiği savunulmaktadır. Searle'e göre, kilitli odada bulunan kişinin biyolojik yapısı olmasına rağmen hâlâ Çinceyi anlamamaktadır. O halde ne biyoloji ne de program tek başına ya da birlikte yeterli değilse Searle için daha ne gerekli olabileceği sorusu öne çıkmaktadır.<sup>211</sup>

Ayrıca kilitli odadaki kişinin Çinceyi standart bir şekilde (bilinçli olarak) anlamadığını kabul eden güçlü yapay zeka savunucularına göre, Çincenin örtük bilinçle anlaşıldığı ve anlamsal Çince içeriğe sahip olduğu iddiası bulunmaktadır. Dillerin örtük bilinç bilgisini gösteren ve anladıklarından habersiz olan insanların durumlarında olduğu gibi kilitli odadaki kişinin, Çincenin anlamına bilinçli olarak erişemese de bilinçsizce böyle bir anlayışa sahip olduğu iddia edilmektedir. Searle'ün (1992) "bağlantı prensibi" burada devreye girmektedir. Searle'e göre içerik en azından ilkesel olarak bilinçli hale gelmedikçe kişinin gerçek anlamda bilinçli bir içeriğe sahip olamayacağını iddia etmektedir.<sup>212</sup>

Bu noktaya kadar öne çıkan itirazlara ve bu eleştirilere yönelik Searle'ün cevaplarına yer verdik. Konu hakkında genel bir değerlendirme yapmak gerekirse güçlü yapay zeka düşüncesini ve Turing'in iddialarını çürütme amacıyla Searle tarafından ortaya konulan "Çince Odası Argümanı", çalışmamızın en önemli dayanaklarından birisidir. Çünkü bu argüman, zihni birtakım biyolojik zemine indirgeyen davranışçı anlayışı temelden sarsmıştır. Searle, bunu

<sup>210</sup> Rakover, 1999: 60-61.

<sup>211</sup> Rey, 1986: 170.

<sup>212</sup> Wakefield, 2003: 300.

herkesçe anlaşılır bir forma soktuğu Çince Odası deneyiyle somutlaştırarak yapmıştır. En sade görünümüyle bu konuya felsefeyle değinmenin meyvelerini, yapay zeka hakkındaki yanlış ve asılsız iddiaları çürütmekle elde etmiştir ve bu savaşı zihin-beden tartışmasıyla hedef aldığı düalizmi temelden zayıflatarak yapmıştır. Yani Searle, Kartezyen düalist bakış açısının taçlandığı güçlü yapay zeka projesini kendine özgü felsefi bakışla çürütmüştür. Bunu da insanın zihinsel yönünün temelleri olarak öne sürdüğü; “bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik” kavramlarıyla gerçekleştirmiştir. Searle’e göre zihnin temel taşları olan bu kavramların gerek ikili bilgisayar sistemlerinde gerekse beynin bir kopyasını oluşturan sinir ağları bilgisayarında gerçekleştirilemeyeceği iddia edilmektedir. Aslında bu iddiasıyla, Searle’ün zayıf yapay zeka düşüncesiyle hemfikir olduğu hissedilmektedir. Kaldı ki konfor maksadıyla bu sistemlerin kullanılmasına herhangi bir itirazının olmadığını belirtmektedir. Searle savaşını, özünde insana ait olan zihinsel süreçlerin bilgisayar sistemleri tarafından uygulanabileceğini iddia eden görüşe karşı vermiştir. Günümüze dek tartışma konusu olan bu argümana yöneltilen itirazların ve eleştirilerin kapsamı düşünüldüğünde güçlü yapay zeka düşüncesine verdiği tahribat açıkça ortaya çıkmaktadır. Söz konusu itirazları değerlendirmek gerekirse bu itirazlardan pek çoğu Searle tarafından gülünç karşılanarak cevap verilmeye dahi layık görülmemiştir. Ancak özellikle “Sistem İtirazı” ile ilişkili olan “bağlantıcılık” itirazı tarafından ciddi eleştiriye maruz kalmıştır. Bu eleştirilerin ciddiyeti, Searle’ün itirazlara verdiği cevapların niteliğinden anlaşılmaktadır. Özellikle bağlantıcılık görüşüyle desteklenen sinir ağları bilgisayarının Çince Odası argümanını alt edebileceği düşünülse de Searle, bu itirazlara cevabını “Çin Stadyumu” gibi yeni örneklerle vermiştir. Öte yandan bağlantıcılık görüşünü savunanların, bu cevap karşısında yeni argüman sunmadıkları fark edilmektedir.

Literatür çalışmalarının genişleyen katmanları bağlamında bu argümanına yönelik eleştirilerin ardı arkası kesilmemektedir. Yepyeni bakış açılarının kimi zaman Searle’ü köşeye sıkıştırdığı da dikkat çeken bir gerçektir. Bunlardan birisi, Hannoeh Ben-Yami (1993) tarafından ortaya atılan “Çince Odası Üzerine Bir Not” başlıklı Çince Odasının geçersizliğine ilişkin iddiadır. Çince Odasının teoride bile imkansız olduğunu iddia eden Ben-Yami, kilitli odadaki kişinin hem söz dizimiyle sınırlı kalması hem de gerçek bir Çinlinin yetkinliğini taşımasının imkansızlığını öne sürerek bu benzetmeyi reddetmektedir. Çince yazılmış "Saat kaç?" sorusuna, odadaki kişinin söz dizimi talimatlarıyla cevap veremeyeceğine dikkat çekilmektedir. Ayrıca renkli bir kartın odaya atılmasıyla renginin ne olduğunun sorulabileceği ya da piyanoda iki nota çalınarak hangisinin daha yüksek olduğunun da sorulabileceği düşünülmektedir. Odadaki kişinin tüm bunlara tutarlı bir şekilde cevap veremeyeceği savunulmaktadır. Odadaki kişinin sadece söz dizimsel kurallara dayandığı için bir nevi kör,

sağır bir Çinli olacağının altı çizilmektedir.<sup>213</sup> Bu yönüyle Ben-Yami'nin eleştirileri, Searle'ün konumunu baltalayabilecek nitelikte bir eleştiri olarak öne çıkmaktadır. Ayrıca Ben-Yami makalesinin devamında bunların üstesinden gelebilecek alternatifleri de sıralamaktadır. Sadece söz dizimsel kuralların anlam için yeterli olmayacağını savunmakta Searle'e hak verirken, dil becerilerimizi başarıyla taklit eden bir bilgisayarın, yalnızca söz dizimsel kurallarla sınırlı olabileceğini varsaymakta yanlış olduğunu düşünmektedir. Böyle bir bilgisayarın algısal bilgilere erişebilecek tasarıma sahip olması, dış dünyayla belirli bir şekilde iletişim kurabileceği ve bu şekilde tıpkı dilimizin anlam kazandığı gibi anlamın oluşacağı düşünülmektedir. Ben-Yami, Searle'e yönelik eleştirilerin ötesinde yapay zekaya ilişkin görüşünü şu sözleriyle özetlemektedir: "Gerçekten de yapay zekada ürettiğimiz makineler, onları çevreleyen dünyayı ifade ediyor ve bu nedenle de anlayışları var."<sup>214</sup>

Her ne kadar Searle'ün argümanını zayıflatabileceği düşünülen eleştiriler olsa da bunlar çoğunlukla biçimsel eksikliklere işaret etmektedir. Kaldı ki zihin felsefesinde önemli bir yere sahip olan Çince Odası argümanı, yapay zekayı bütün boyutlarıyla temellendirmek amacıyla zihin felsefesinin nimetlerinden faydalanmaktadır. Bu yönüyle bir argümanın biçimsel eksikliği başka bir şey, kökten geçersizliğini iddia etmek farklı bir şeydir. Öte yandan Çince Odası argümanını eleştiren araştırmacılara bakıldığında eleştirilerinin yanı sıra çoğu konuda Searle'e hak vermekte oldukları, diğer bir deyişle ortaya konulan ispatlara kayıtsız kalmadıkları fark edilmektedir.

## 2.6 Sağduyu Tartışması

Güçlü yapay zeka fikrinin öne çıkan muhaliflerinden Dreyfus (1999), *Bilgisayarlar Ne Yapamaz?* başlıklı eserinde insana özgü sağduyu<sup>215</sup> (*common sense*) teriminin sayısal sistemlerde gerçekleşemeyeceğini belirtmektedir. Alan Turing'in kusursuzca programlanmış yüksek hızlı bilgisayarın zeki davranışlar sergileyebileceği iddiasının başarısız olduğuna değinerek tartışmasına başlamaktadır. Pek çok felsefi eleştiriye karşısına alan güçlü yapay zeka projesinin başını çeken Allen Newell ve Herbert Simon, ikili sistemlerin örnekleme ve eşleştirme kapasitesinin yükseltilmesi hedeflemiştir. Bu ifadeyi açmak gerekirse özellikleri aynı yapıya sahip olan fenomenin, hesaplamalı bir yapı temelinde tasarlanması hedeflenmiştir. Newell ve Simon'un, zekanın önemli yönlerini taklit etmek için bilgisayarların kullanılabileceğini iddia etmesiyle aklın bilgi işlem modeli ortaya çıkmıştır. İki araştırmacının

<sup>213</sup> Ben-Yami: 1993: 169, 170.

<sup>214</sup> Ben-Yami: 1993: 171.

<sup>215</sup> Sağduyu kavramına yönelik görüş bildiren Nilsson (1998); insanların doğmadan önce var olamayacağı, balıkların suda yaşadığı ve dışarı çıkarsa öleceği, insanların genellikle gece uyudukları gibi on yaşındaki birisinin bildiği şeylerin bilgisini sağduyu olarak yorumlamaktadır. (Nilsson, 1998: 302.)

çalışmalarıyla yapay zeka alanında doktora programları, profesyonel topluluklar, meslek grupları ve uluslararası toplantılar gibi faaliyetler de gelişmiştir. Baş döndürücü bu gelişmelere güveni tam olan M.I.T., yapay zeka projesi lideri Marvin Minsky, "Bir kuşak diliminde 'yapay zeka' yaratma sorunu büyük ölçüde çözülecek."<sup>216</sup> ifadesini kullanmıştır. Ancak bu çalışmada çocuk hikayelerini anlamaya çalışma girişimlerinin başarısızlığa uğramasıyla bir sorun ortaya çıkmıştır. Dört yaşındaki bir çocuğun sağduyusundan yoksun olan bu programların, en basit hikayeleri dahi anlaması için gereken temel bilgilerin nasıl verileceğini kimse bilememiştir. Çünkü yapay zeka girişimi zihinsel süreçlerin niceliksel karşılığının algoritma düzleminde tasarımlar oluşturabilmesi olarak kabul etmekteydi. Oysa Dreyfusçu sağduyu anlayışına göre fenomenlerin uygulanmasına ilişkin kurallardan oluşan devasa bir veri yapısı olarak temsil edilmesi gerekiyordu. Sağduyu problemi neticesinde ruh hali epey değişmiş olan Minsky, 1982'de bir gazeteciye şu ifadeyi kullanmıştır: "Yapay zeka sorunu şimdiye kadar yapılmış en zor bilimden biridir."<sup>217</sup>

1960'lı yıllarda iyimser olan araştırmacılar, 1970'lerin sonunda ve 1980'lerin başında yapay zeka üretmek için insanlara ait sağduyunun açıklığa kavuşturulması ve organize edilmesinin zorunlu olduğunu ve bunun çok büyük bir görev olduğunu isteksizce kabul etmişlerdir. Bu yaklaşımın en meşhur savunucusu Douglas Lenat, sağduyunun ansiklopedilerde bulunan bir bilgi türü olmadığını, daha ziyade ansiklopediyi yazarların verdiği bir bilgi olduğunu fark ederek yeni bir sağduyu temelli veri tabanı kurmayı hedeflemiştir. Bunu örneklemek amacıyla George Washington'un Capitol'de bulunması durumunda sol ayağının da onunla birlikte olduğu ve ölmesi durumunda da ölü olarak kalacağı gibi gerçeklere dikkat çekilmiştir. Yani zihin-beden bütünlüğünün ispatı yapılmıştır. Bu tür bilgiler sağduyu bilgisi kapsamında değerlendirilmektedir. Bu düşünceden hareketle Lenat, milyonlarca bilgi girdisi içeren tek bir zeki araç kurarak bu sağduyuyu yakalayacağını önermiştir. Ancak on beş yıllık araştırmanın sonucu hüsrana olmuştur. Bu duruma ilişkin değerlendirmede bulunan Dreyfus, sağduyu için tek başına insan zihninin sorumlu olmadığını, bedeninin de bir o kadar önemli olduğunu iddia ederek cevap vermiştir.<sup>218</sup> Burada statik bilgiden ziyade pratik beceri vurgusu yapılmaktadır. Bu konuyu Aristoteles'in pratik bilgi (*phronesis*) terimiyle bağdaştırmak mümkündür. Kaldı ki insanın bilgiye erişme yani edinme yollarından pratik akıl tıpkı sağduyu bilgisine sahip bir çocuğun bilgiyi edinmesiyle benzeşim kurulabilir.

Martin Heidegger, Maurice Merleau-Merleau-Ponty ve sonradan Wittgenstein'a atıfta bulunarak güçlü yapay zeka programının başarısız olacağı sezgisiyle Dreyfus, sağduyu

<sup>216</sup> Dreyfus, 1999: 9 - 10.

<sup>217</sup> Dreyfus, 1999: 11.

<sup>218</sup> Dreyfus, 2009: 18-19.

bilgisinin özünde bir tür 'yöntem bilgisi' ya da 'beceri' (*know-how*) olduğunu iddia etmektedir. Sağduyu bilgisinin zorunlu olarak bilgiyi temsil etmediğini, gündelik sağduyu anlayışının özünde bir tür 'beceri' olduğunu göstermeye çalışmıştır. Dreyfus buradaki problemi tam olarak bir becerinin, insanı insan yapan duyguların ve dürtülerin karmaşık bir inanç sistemi olarak bilgisayara aktarılması meselesi olarak değerlendirmektedir. Bu bağlamda sembolik bir tasarımda dilsiz, kavramsız bir insan olma düşüncesini umutsuz bir çaba görerek güçlü yapay zekanın sağduyu problemini çözebileceğinden şüphe duymuştur. Güçlü yapay zekanın sağduyuya bakışını eleştiren Dreyfus, problemin soy kütüğüne değinerek Descartes ve Leibniz gibi rasyonalistlerin akıl kavramına bakışını da eleştirmektedir. Bu bakış açısına göre akıl, bütün alanlarda tasarımlar oluşturan bir kapasite olarak kabul edilmektedir. Yani dünyada dolaşmak, bir şeylerle, birileriyle ilgilenmek anlamında kullandığımız genel 'becerilerimizin' bile, bu görüşe göre akılda önermeli bir biçimde yansıtıldığı iddia edilmektedir. Bu görüşün zihinle ve dünyayla olan ilişkisini Dreyfus, "tasarımcılık" olarak adlandırmaktadır. Bu bakış açısına göre gündelik anlayışın temeli, örtük bir inanç sistemi olarak kabul edilmektedir. Bu başlık altında öne çıkan problemler şu şekilde sıralanmaktadır. Birincisi, çıkarımlarda bulunabilmek için gündelik bilginin nasıl organize edilmesi gerektiğinin problemi. İkincisi, becerilerin nasıl temsil edilebileceğinin problemi. Üçüncüsü, özel durumlarda amaca uygun bilgilerin nasıl getirilebileceği problemi.<sup>219</sup>

Tasarımcılık görüşünü savunanların, bu sorunları çözmek amacıyla programlar geliştirmelerine rağmen kabul gören bir çözüm bulamadıklarına da dikkat çekilmektedir. Ayrıca bu çözümsüzlüğün yanı sıra önemli bir tespitte de yer verilmektedir. O da şudur; bir sisteme belirli bir durum hakkında ne kadar fazla veri yüklenirse ilgili veriyi çağırmak da o kadar uzun sürmektedir. Büyük bilgi yığınlarının neden olduğu bu sorun, insanlar açısından bir sorun değildir. Kaldı ki insanlar bir durum veya bir şey hakkında ne kadar çok bilgi sahibiyse ilgili diğer bilgileri o kadar kolay elde edebilecekleri savunulmaktadır. Bu tezatlık, insanların ve sistemlerin depolama biçimlerinin farkını ortaya koyan önemli bir unsur olarak öne sürülmektedir. Heidegger, Merleau-Ponty ve antropolog Pierre Bourdieu açısından gündelik becerinin fenomenolojisini önemli gören Dreyfus, becerilere ilişkin bilgiler vermektedir. Örneğin gerekli sosyal beceriyi bir kez edindikten sonra ikinci kez sorgulamaya gereksinim duyulmamaktadır. Kültürel olarak hediyeleşme düşünüldüğünde ne zaman ve nasıl bir hediye verileceği noktasında karar almaya gerek kalmayacaktır. Öyle ki kültürel bir çevrede hediyeleşmenin dinamikleri toplumun üyeleri arasında örtük olarak uygulanan bir pratiktir. İnsanlar tarafından uygun koşullar altında uygun bir hediye verilmesi ve hediyeinin uygun

---

<sup>219</sup> Dreyfus, 1999: 15-18.

olduğunun anlaşılması becerisi için geçerli olan bir durumdur. Bu durumda kişinin neyin uygun olduğunu rasyonel olarak tespit etmesi gerekmediği savunulmaktadır. Bütün sürecin, kültürel bir ortamda sağduyu tarafından yönetildiği fikri kabul görmektedir.<sup>220</sup> Kültürel bir eylem olarak karşılıklı hediyeleşmenin önemli olduğu toplumları örnek gösteren Bourdieu'nin (*know-how*) temelinde sağduyu düşüncesine şu ifadelerle yer verilmektedir:

“Geçmiş deneyimlerin aktif varlığı... her organizmada algı, düşünce ve eylem şemaları biçiminde biriktirildiğinde uygulamaların doğruluğunu ve zaman içindeki tutarlılığını tüm biçimsel kurallardan ve açık normlardan daha güvenilir bir şekilde güvence altına alma eğilimindedir.”<sup>221</sup>

Kültürel bağlamda hediye alışverişi örneğinin amacı; uygun zamanda uygun şekilde, uygun bir hediye için nasıl verilebileceğinin bilinmesi olarak ifade edilir. Ayrıca bir hediye için ne olduğunu bilmek ile hediyeleşme becerisine sahip olmanın farklı şeyler olduğu vurgulanmaktadır. Dreyfus, su gibi doğal şeylerde olduğu üzere hediyeler gibi sosyal türler için de sağduyunun, bir tür (*knowing-that*) değil (*knowing-how*)<sup>222</sup> bilgisi olduğuna inanmaktadır. Bu da sembolik deneyimlerimizi saklamanın bir yolu olarak gösterilmektedir. Kaldı ki sembolik yapay zeka için gerekli olan temsilin bambaşka bir şey olduğu vurgulanmaktadır.

Sağduyu problemini çözebilmek için geliştiriciler tarafında yoğun çalışmalar yapılmış ve kendi aralarında bazı ekoller dahi oluşmuştur. Tasarımcıların yanı sıra problemlere alternatif yollar arayan başka gruplara da dikkat çekilmektedir. Etkileşimciler grubunda amaç, sembolik tasarımlar ve içsel modeller olmaksızın mikro-dünyayla zekice etkileşim kurma çabasıdır. Sinir ağını modelleyenler olarak bilinen ikinci grup da tasarımcılığı terk ederek girdileri doğrudan haritalamaya çalışmaktadırlar. Güçlendirilmiş öğrenme (*reinforced learning*) olarak adlandırılan üçüncü bir grup ise uzmandan bağımsız olarak kendi başına başarılı bir girdi-çıkı kuralı bulmak için beceri alanındaki gerçek performansı kullanan bir program geliştirmeyi hedeflemektedir. Bu yaklaşımların her birinin avantajlarını ve kısıtlarını dikkate değer gören Dreyfus, her bir çalışma grubunun probleme ilişkin girişimlerini çok yönlü olarak ele almıştır. Detaylı olarak anlatılan bu girişimlerin problemlere çözüm sunma çabalarının, insana özgü ‘genelleme’ yeteneğinden yoksun olması nedeniyle çıkmaza girdiğini düşünmektedir. Bu bağlamda sistemin yeni kurallar oluşturmasını sağlayan bağımsız bir öğrenme yeteneğine sahip olmaması temel bir problem olarak gösterilmektedir. Yapay sinir ağlarının öğrenme yeteneğine sahip görünmesine rağmen; hangi vakaların zekayı sergilediğine karar verenin aslında yine bir insan olduğu savunulmaktadır. Ağın öğrendiği tek şeyin, bu zekayı bağlantı güçleri açısından

<sup>220</sup> Dreyfus, 1999: 18-23.

<sup>221</sup> Dreyfus, 1999: 23.

<sup>222</sup> Statik bilgi (*knowing-that*) ve bilme becerisi (*knowing-how*) ifadeleriyle deneyime bağlı bilme becerisinin, mekanik sistemlerdeki statik bilgi yığınlarından farklı olduğu anlatılmaya çalışılmaktadır.

yakalayabilmesi olduğu ifade edilmektedir. Dolayısıyla güçlü yapay zeka sistemlerinin öğrenilen şeylere uygun düşmeyen durumları tanıma yeteneğinden yoksun olduğuna dikkat çekilmektedir. Hataları tanımak, ağa öğretilen durumların çıktılarını değiştirmek veya uygun davranış değişikliğine yol açacak yeni durumlar sağlamak gibi eylemlerin yine insana bağlı olduğu vurgulanmaktadır. Bu açıdan bakıldığında sinir ağlarının neredeyse güçlü yapay zeka sistemleri kadar insan zekasına bağımlı olduğu gözlemlenmektedir. Ayrıca övülen öğrenme tekniklerinin yanıltıcı olduğuna dikkat çekilerek çevre ile nasıl başa çıkacağını öğrenen ve çevre değiştikçe değişebilen bir sisteme ihtiyaç duyulduğu sonucu çıkarılmaktadır.<sup>223</sup>

Dreyfus'un ifadelerinden insanın bilgiyi 'edinmesi' ile bir makineye bilginin 'öğretilmesinin' farklı şeyler olduğu anlaşılmaktadır. Bu ayrım, 'edinmek' ve 'öğrenmek' kavramları arasındaki farka benzetilebilir. Kaldı ki insanın bilgiyi zihinsel ve bedensel olgularının bütünü aracılığıyla edinmesi ile bir sistem tarafından olguların sembollerle temsilinin öğrenilmesinin farklı şeyler olduğu anlaşılmaktadır. Bu düşünceye dayanak oluşturacak çarpıcı bir izlenim Merleau-Ponty (2007) tarafından örneklendirilmiştir. Bu izlenimde bir çocuğun penceresinden dili edinme süreci olabildiğince doğal haliyle anlatılarak sürecin bir anlamda mekanik temellere dayanmadığı ispatlanmaktadır.

"Çocuğu çevreleyen konuşulan dilin bütünü, tıpkı bir kasırğa gibi çocuğu yakalar, içsel artikülasyonlarıyla onun aklını çeler ve öyle bir an gelir ki bütün bu gürültüler bir anlam ifade eder. Kelime trenlerinin yorulmaksızın bir oradan bir buradan geçişi, söylemlerin görünür bir biçimde belli bir fonemik ölçekte oluşturulmasının kusursuzca ortaya çıkmasıyla çocuk, nihayet konuşanlar tarafına bir eğilim gösterir. Ayrılmaz tek bir bütün olarak ele alındığında bir dilin, çocuğu kendisine nasıl çektiğini ve etki alanına sokmak için, kapısını içerden açmaya çalıştığını anlamaya olanak tanır. Bunun sebebi, işaretlerin başlangıçtan bu yana ses değiştirme özelliğinden kaynaklanır. Çünkü o kendi çapında oluşturulur ve organize edilir. Yani bir tür iç yönü vardır ve bir anlam iddiasıyla noktalanır."<sup>224</sup>

Merleau-Ponty'nin bu ifadesinde dikkat çeken en önemli unsur, deneyim ve zaman bağlamında zihinsel süreçlerin bütünsel faaliyeti olarak yorumlanabilir. Bunun da formel bir eğitimde olduğu gibi istendik yönde davranış değiştirmeye dayalı öğrenimden (*learning*) ziyade, olabildiğince sıradan ve doğal durumlar yoluyla bir edinim (*acquisition*) olduğu sonucu çıkarılabilir. Öğrenme ve edinme kavramları eğitim ve yapay zeka açısından değerlendirildiğinde çarpıcı bir sonuçla karşı karşıya kalınmaktadır. Bu minvalde, günümüz eğitim anlayışının daha çok öğretme-öğrenme odaklı olduğu düşünüldüğünde doğal insan ediniminden ziyade makine öğrenmesine benzer bir yöntem meylli olduğu ortaya çıkmaktadır. Yani bilgi işleme kuramında değindiğimiz eleştirilerde olduğu gibi, tıpkı bir

<sup>223</sup> Dreyfus, 1999: 25-38.

<sup>224</sup> Ponty, 2007: 241.

makineye yüklenen bilgiler gibi insana disiplinleri öğretme çabasının, insanı makine olarak görmenin başka bir yolu olduğu sonucu çıkarılabilir.

Tekrar Dreyfus'a dönecek olursak; algı, anlayış ve anlam gibi konuların felsefi bağlamda tartışılmasını günümüz öğrencilerinin eski moda olarak adlandırdığına dikkat çekilmektedir. Ayrıca Minsky liderliğindeki robot projesinin algılama ve anlama kapasitesiyle makineyi hareket ettirebileceğine yönelik öğrencilerin inancına da göndermede bulunmaktadır. Hatta filozofların iki bin yıldır boşa kürek çektiğini, ciddi araştırmacıların nihayet görev başında olduklarını ironiyle anlatmaktadır.<sup>225</sup> Tarih boyunca matematiğin kesinliğine duyulan güven ve teknolojinin yaydığı büyüünün, bu tartışmayı insanın makinelerle kıyaslanması noktasına taşıdığı anlaşılmaktadır. Ayrıca mekaniğin ötesinde insan niteliklerini, becerilerini, akılsal fakültelerini gerçekleştirme uğrunda sayısal bilgisayarlara yönelik inanç ve çabaların boyutu ortadadır. Bu denli yoğun çabalar zamanla dogmatik inançlara dönüşerek fosilleşme eğilimi taşıyabilir. İnsanları kör edebilecek böylesi bir fosilleşme neticesinde algı, anlayış ve anlam gibi konuları felsefeciler yeniden canlandırma mecburiyetinde kalacaklardır. Ayrıca makinelerin akıl sahibi olamayacağının felsefi temellerini tartışmak bir yana, dogmatik inançlardan kurtulmaya ve insan doğasına ilişkin bilgileri yeşertmeye çalışacaklarını ön görebiliriz.

## 2.7 Sonuç

Birinci bölümde zihnin fakülteleri olan akıl, zeka, bilinç ve algı gibi olguların kavramsal yönlerini ortaya koyarak ikinci bölümde bu kavramların karşılaştırılmasının yapay zekanın işlevsel yönleriyle yapılması hedeflenmişti. Bu iki bölümün ortaya koyduğu sonuç sayısal sistemlerin algoritmik hesaplama kapasitesi sadece hız, güvenilirlik, hassaslık gibi kolaylaştırıcı niteliklerin ötesine geçememesidir. Yani semantik anlamda bir öz düşünüm mümkün değildir. Bu nedenle insanın sahip olduğu akıl olgusuyla kıyaslamak güçtür. Bunun nedenleri Zihin-Beden Problemi, Turing Makinesi, Taklitçilik Oyunu, Çince Odası Argümanı, Güçlü Yapay Zeka ve Sağduyu tartışmalarında temellendirilmiştir. Sayısal sistemlerde öz düşünümü gerçekleştirme çabası, bilinç tartışmasıyla bir geleneğe dönüşmüştür. Bu geleneğin çıkışı kuşkusuz Turing'in "Makineler düşünebilir mi?" sorusunun tartışılmayacak kadar anlamsız olduğunu düşünüyorum." iddiasıdır. Algoritmanın gücüne taparcasına inanan Turing'in iddialarını destekleyenler kadar eleştirenler de olmuştur. Öyle ki taklitçilik oyununa itirazlar edilmiş olup, Turing bunların bir kısmına cevap vermiştir. Bu itirazların başını çeken düşünür John Rogers Searle'dür. Turing'in argümanını çürütmeyi hedefleyen Searle, düşünme

<sup>225</sup> Dreyfus, 1986: 1.

eyleminin sayısal sistemlerdeki karşılığına atfen insanı yücelttiği dört ana özellik bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik ile cevap vermektedir. Sayısal sistemlerin bu nitelikleri yerine getirememesi sorununun “içkinlik” olarak belirtmiştir. Searle, sözdiziminin anlambilim için tek başına yeterli olamadığı gibi ve anlambilim için temel oluşturmadığı ifadesiyle argümanını temellendirmeye çalışmıştır.

Öte yandan insan doğası ve algoritma farkı Dreyfusçu bir anlayışla değerlendirildiğinde sayısal sistemlerin durağan verileri “öğrenebilmesi” ancak insanın Aristotelesçi *phronesis* sayesinde olguları “edinmesi” durumu dikkat çekmektedir. Bu hususu Dreyfus, sağduyu kavramıyla açıklamaktadır. Bilginin doğasına *know-how* ve *know-what* kavramlarıyla değinilen çalışmada, sayısal sistemlerin biçimlendirilmiş öğrenme modelleri sonucunda salt *know-what* temelinde durağa bilgiden öte gidemeyeceği iddia edilmektedir. Oysa dinamik süreçlerle etkileşime giren bilinç sahibi insanın bir öğrenmeden öte, sağduyu temelinde bir edinme eylemine sahip olduğu belirtilmektedir. Bu durumda Dreyfus’un bize *know-how* ile anlatmak istediği şey akışkan olguların bilgisiyken, *know-what* ile anlatmak istediği şey durağan bilgilerdir. Bu bağlamda Dreyfus, bir bilgisayarın Türkiye’nin başkentini tanıma konusunda herhangi bir aksaklık yaşamamasına rağmen Ankaralılık kavramının *know-how* bilgisine erişemeyeceğini iddia etmektedir.<sup>226</sup> Farklı bir bakış açısıyla Dreyfus’un görüşleri ve Aristoteles’in *phronesis* kavramı arasında ilişki kurmak mümkün olabilir. Öyle ki *phronesis* pratik akıl ekseninde gündelik bilgiyle ilgili deneyime dayalı bir akıl fakültesi olarak anlatılmaktadır. Bu bağlamda deneyim sözcüğü, sayısal sistemlerin deneyimle öğrenebileceği sonucunu zorunlu olarak çıkarmamaktadır. Kaldı ki makine öğrenmesinin gerçekleştirdiği işlev tamamen istatistik temelinde büyük veri kümelerinin girilmesiyle gerçekleşmektedir. Öte yandan bu durağan verilerin ilişkilendirilmesiyle birtakım yeni modeller çıkarılmaktadır. Ancak bu süreçler salt sentaks ve *know-what* kıskacında kaldığından yeni olduğu düşünülen modellerden hiçbirisi sağduyu bilgisiyle ilişkilendirilmemektedir. Dreyfus’un tespiti öylesine önemlidir ki akıllara Aristoteles’in *phronesis* kavramını getirmektedir. Kaldı ki her iki düşünür de deneyime vurgu yaparak deneyimsel davranışın niceliksel sembollere indirgenemeyeceği gibi bu sürecin statik bir öğrenmeden ileri gelemeyeceğini vurgulamıştır. Genel olarak bakıldığında yapay zeka taraftarlarının ve karşıtlarının argümanlarının epey güçlü olduğu söylenebilir. Felsefi söylemler kapsamında değerlendirildiğinde gerek ilk bölümde aklın nitelikleri üzerine gerekse ikinci bölümde Searle ve Dreyfus’un kavramları üzerine düşünüldüğünde Turing’in iddialarının büyük ölçüde zayıfladığı görülmektedir.

<sup>226</sup> Dreyfus, 1999: 15-18.

Öte yandan sayısal sistemlerin insani nitelikleri gerçekleştirmenin ötesinde bu sistemleri kurabilmenin önünde daha ciddi engeller bulunmaktadır. Başlıca; dakik, sağlam ve güvenli kararlar alabilen dinamik ortamlarla etkileşime girerek sürekli öğrenen yapay zeka sistemlerini geliştirmenin zorluğu. Kişiselleştirilmiş uygulamalara ve hizmetlere olanak sağlayan ancak kullanıcıların gizliliğini ve güvenliğini tehlikeye atmayan yapay zeka sistemlerin geliştirilme zorluğu. Olanakları köreltmeksizin farklı kuruluşların veri kümeleri üzerinde kendini eğitebilen yapay zeka sistemleri kurabilme zorluğu gibi problemler bulunmaktadır. Donanımsal problemlere ilişkin yapılan bir araştırmada<sup>227</sup>; yapay zekanın (1) büyük miktarda veri, (2) ölçeklenebilir bilgisayar ve yazılım sistemleri ve (3) bu teknolojilerin erişilebilirliği gibi unsurlardan ibaret olduğu vurgulanmaktadır. Yine de hâlâ halihazırda pek çok donanımsal ve yazılımsal sorunlar nedeniyle yapay zekanın ilerleyişinde bazı engeller öngörülmektedir. Bu problemlerden birisi, büyük miktarda veriyi işleme ve saklama ihtiyacına cevap verebilme problemidir. Yani katlanarak artan verilere ev sahipliği yapabilecek kapasitede depolama alanı sorunudur. İkincisi değişebilen, hızlı ve beklenmedik bir şekilde ortaya çıkan ve tekrarlanamayan ortamlarda yapay zekanın taleplere cevap verme olasılığıdır. Örneğin, bir ofis binası için güvenlik sağlayan bir grup robottan birisi arızalandığında veya gruba yenisi eklendiğinde, diğer robotlar koordineli bir şekilde yönlendirme, planlama ve kontrol stratejilerini güncellemesi zorunluluğu öngörülmektedir. Benzer şekilde ya robotların kendi eylemleri ya da dış koşullar nedeniyle ortam değiştiğinde, tüm robotların, ışık hızında değişiklikleri yeniden kalibre etmesi gerekliliğidir. Bu tür ortamların işlenmesi, daha önce karşılaşılmamış senaryolara bile hızlı ve güvenli bir şekilde tepki verebilen yapay zeka sistemleri gerektirecektir. Yapay zekanın tüm vaadini gerçekleştirmek için üstesinden gelmesi gereken göz korkutucu zorluklar bulunduğu ve bu zorlukların çoğunun sistemsel ve altyapıdan kaynaklandığına dikkat çekilmektedir. Bu zorlukların, yapay zeka sistemlerinin daha hızlı, daha güvenli ve daha açıklanabilir kararlar alması gerektiğinin farkına varılması, bu kararların yanı sıra öğrenme süreçlerini her zamankinden daha karmaşık saldırı türlerine karşı güvence altına alarak hesaplama yeteneklerini sürekli olarak artırmasıyla ortaya çıktığına değinilmektedir.

Yapay zekanın gerek yazılım gerekse donanım tarafında yaşanan engeller, teknolojinin bilince ilerleyişte hızını yavaşlatan unsurlardır. Ancak, insan doğasının günümüz teknolojisi karşısında dahi büyük tahribata uğradığı bilinen bir gerçektir. Bu bağlamda tekrar ifade etmek gerekirse teknoloji, süratle doğasına yabancılaşan insanı hem özneliğinden hem de doğasından koparmaktadır. Bu tahribat, insanın sınırsız bilincini algoritmaya sığdırma çabası olarak

<sup>227</sup> Stoica vd., 2017: 3-6.

özetlenebilir. Yapay zekanın felsefî sonuçlarına giden yolda bu tür sistemlerin bilince sahip olup olmamasından bağımsız olarak sonuçları olacaktır. Bu bağlamda en önemli tartışma alanı etikdir. Kaldı ki algoritma temelli nesne ile özne arasında gerek kural temelli gerekse etik yaklaşımlar yaygın bir biçimde belirlenmediği gibi, yeni dünya düzeni bu türden bir ihtiyaca günümüz şartlarında tam olarak cevap verebilecek nitelikte değildir. Bu yüzden, üçüncü bölümde yapay zeka sistemlerine yönelik uygulamalı etik bağlamında bir yaklaşım ortaya koyma gereği duyulmuştur. Kavramsal karmaşanın önüne geçmek amacıyla, öncelikle etiğe ilişkin ön bilgi verilmiştir. Daha sonra makinelerin etik yargılarla donatılmasına ilişkin bugüne dek öne çıkan çabalara dikkat çekilerek bu hususa ilişkin değerlendirmelerde bulunulmuştur.



*“Sayısal sistemler zihinsel durumlara sahip değildirler. Yapıları zihinsel durumlar olarak yorumlansa bile kendi özgürlüklerinden kaynaklanan bir yönelimselliğe sahip değildirler. Bu nedenle asla (özerk, bağımsız) ahlaki birimler olamazlar.”* Deborah G. Johnson

## ÜÇÜNCÜ BÖLÜM

### ETİK KURAMLARIN SINIFLANDIRILMASI

#### 3.1 Etik Kavramına Genel Bakış

Tarihsel olguları temellendirmek ne kadar güçse olguların geleceği hakkında öngörülerde bulunmak da o derece güçtür. Öyle ki bu amaca hizmet edecek olan en güçlü dayanak tarihsel olgular ve kavramlardır. Bu nedenle yapay zekanın Eski Çağ’lardan bu yana gelişimi etik, risk faktörleri ve kavramların anlamsal değişimi gibi hususları yansıtmaktadır. Bu sebeplerden dolayı öncelikle etik terimine ardından uygulamalı etiğe değinmek yerinde bir başlangıç olacaktır. Feldman (2012) *Etik Nedir?* eserinde “ahlak” ve “etik” sözcüklerinin analizini “Ahlak Nedir?” ve “Felsefi inceleme nedir?” sorularıyla tartışmaktadır. Felsefe aracılığıyla ahlakı tartışmanın etiği anlama noktasında doyurucu cevaplar sağlayacağını değerlendirmektedir. Bu hususta etik; ahlakın felsefi zeminde ele alınmasından ibaret olduğu vurgulanmaktadır. Ahlak isim kökeninden bir sözcükken, “ahlaki” sıfat kökeninden bir sözcüktür. Bu yönüyle Feldman, “ahlaki” sıfatını incelemenin doğru bir başlangıç noktası olacağı inancıyla, bazı yargıların “ahlaki” olduğuna ilişkin kanaatimizi açıklayan şeylerin, “doğru, yanlış, zorunlu, iyi, kötü, şer, -meli, -malı, -mak zorunda” gibi değer bildiren terimler veya kipler olduğunu belirtmektedir. Bu noktada “ahlaki” sıfatına tam olarak cevap bulunamayacağını gibi “Ahlak nedir?” sorusunun da net bir cevabı bulunmadığına da dikkat çekilmektedir. Oysa insanın yegane özelliklerinden birisi olan sezgisel kavrayış Feldman’a göre ahlaki olan ile ahlaki olmayanın ayrımını yapan şeydir. Toplumsal ahlakı oluşturan şeyleri de toplumun benimsediği ahlak ilkelerinin bütünü olarak görmektedir. Bu çıkarımla Feldman, ahlak kavramını genel anlamıyla “doğru ahlak ilkeleri bütünü” olarak değerlendirmiştir.<sup>228</sup> Buraya kadar anlaşılacağı üzere etiği ahlaktan ayıran belirleyici unsur etiğin felsefi düzlemde konumlanan bir kavram olması, ahlakın ise özneyi yoğunlukla toplumsal dinamikler çerçevesinde etkileyen bir kavram olmasıdır. Kendi içinde çeşitli alanlara ayrılan etiğin

<sup>228</sup> Feldman, 2012: 11- 23.

konumuzla ilgisi alanı uygulamalı etikdir. Diğer bir deyişle ş emsiye kavram olan etiğin önemli alanlarından biri de uygulamalı etikdir. Uygulamalı etik, etik yaklaşımların pratik alanlara uyarlanmasıyla olası problemlerin önüne ilkesel olarak önüne geçmeyi amaçlamaktadır. Bu anlamda pratik yaşamın pek çok alanında ihtiyaç duyulan bir alandır. Uygulamalı etik yapay zeka açısından düşünüldüğünde teknolojinin bir uzantısı olan yapay zekanın pratikte olduğu kadar teoride de etik kuramlara ihtiyaç duymaktadır. Öncelikle öznenin nesneyle etkileşiminin, nesnenin özneye ilişkisine dönüştüğü çağımızda ilişkilerin düzenlenmesi ya da kontrol edilmesi zorunluluğu doğmuştur. Bu düzenlemelerin gereği kimi zaman otonom bir aracın neden olacağı kaza, kimi zaman ise yapay zeka destekli askeri bir teçhizatın tehdidiyle ilgili olabilir. Dahası yapay zeka sistemlerinin en üst seviyeye ulaşmasıyla insan aklının yapabileceği işlerin bu sistemlere teslim edilmesi gibi sorunlara neden olabilir. İşte bu tür kaygıların genel adı olan özne nesne ve nesne özne etkileşiminin güncel problemi bu meselenin teknik tarafıyla, gelecekte olası sonuçların önüne geçilmesi ise felsefi yanıyla ilgilidir. Felsefenin bu noktadaki rolü geleceğe ilişkin olası sorunların uygulamalı etik bağlamında belirlenmesi ve önlenmesi için kurallar bütünü ortaya koymasıdır. Bu uğurda araştırma geliştirmeye katkı sağlayacak bazı görüşler bulunmaktadır. Bu görüşler özellikle yapay zeka sistemlerinin tasarımında önemli bir yere sahiptir. Kaldı ki yapay zeka etiği, yapay zekanın en yeni ve kısırlı alanı olarak bilinmektedir.

Yapay zeka hakkında yapılan etik tartışmalar uygulamalı etik kapsamında değerlendirilmektedir. Bu yönüyle etiğin bir kurallar sistemi olduğunu düşünenler “deontolog” olarak adlandırılmaktadır. Amaçları birbirleriyle çatışmayan karmaşık ve spesifik kuralları tespit etmek ya da aralarındaki çatışmaları çözmek için kuralları hiyerarşik bir yapıda sıralamaktır. Diğer taraftan sonuççu görüş ahlaki kurallarla değil, hedeflerle başlarlar. Eylemleri, bu hedefleri ne ölçüde ilerlettiklerine göre değerlendirirler. Sonuççu teorilerin en iyi bilineni faydacılıktır. Klasik faydacı, bir eylemi, etkilenen herkes için herhangi bir alternatif eylemden daha fazla mutluluk üretiyorsa doğru, yoksa yanlış olarak görür. Bu ifadenin iki yönüyle önemlidir: Burada ‘daha fazla mutluluk’, eylemden de kaynaklanabilecek herhangi bir ıstırapı veya sefaleti çıkardıktan sonra net mutluluk anlamına gelir ve eğer iki farklı eylem, en fazla mutluluğu üretme unvanına bağlanırsa, bunlardan herhangi biri haklıdır. Bir eylemin sonuçları gerçekleştirildiği koşullara göre değişir. Dolayısıyla, bir faydacı hiçbir zaman gerçekçilik eksikliğinden ya da pratik deneyime meydan okuyarak ideallere sıkı sıkıya bağlı kalmakla suçlanamaz. Faydacı, sonuçlarına bağlı olarak bazı durumlarda yalan söylemeyi kötü, diğerlerinde iyi olarak değerlendirecektir.<sup>229</sup>

<sup>229</sup> Singer, 2011: 3.

Yapay zekanın mühendislik kökleri düşünüldüğünde net parametrelerle belirlemek mümkün olabilir. Oysa etiğin disipline sığdırılması bir yana yapay zeka gibi geleceği ilgilendiren bir alana ilişkin tespitlerde bulunmak oldukça zahmetli bir süreçtir. Bu güçlüğü bazı nedenleri bulunmaktadır. İlk olarak yapay zeka sistemlerinin geliştirilmesi pratik bir uygulamayken, etik yönlerinin tartışılması düşünsel bir faaliyettir. İkinci olarak, sistemlerin geliştirilmesi doğrudan gözlemlenebilirken, etik kaygıların mahiyetinin gelecekte ortaya çıkacak olmasıdır. Örneğin, ahlakın bireysel ve alıcı bir konumu işaret ettiği düşünüldüğünde yapay zeka ahlakından söz etmek zorunlu olarak yapay zekanın irade sahibi özne olabileceğinin kabulü anlamına gelecektir. Bu hususta kavramların doğru ve yerinde kullanılması tartışmanın güncelliğini ve yapay zekanın halihazırdaki durumunu belirtmek adına önemlidir. Bu sebepler açısından bakıldığında sayısal sistemlerin geliştirilmesi kolay bir problemken, felsefi yönlerinin etik bağlamda değerlendirilmesi zor bir problemidir. Bu zorluğun belirleyici faktörü kavramların mahiyeti hususunda ortak bir fikir yürütülememesidir. Bu noktada etik kavramlarının tanıtılması ve çözümlemesi kapsamında amacımız, özne – nesne ilişkisi bağlamında ortaya çıkması olası sorunların belirlenmesi, bu sorunların yol açabileceği sonuçların ön görülmesi ve bir sınıflandırma yöntemiyle önlemlerin sunulmasıdır.

O halde konumuz itibarıyla “etik” ifadesinden anlayacağımız; (1) felsefi bir kavram olarak olması ya da olmaması gerekenleri belirlemesi, (2) ahlakın pratik, etiğin ise kuramsal bir olgu olması, (3) etiğin geleceğe yönelik rehberlik, kılavuzluk ya da bir ana yol sağlaması gibi hususları kapsadığıdır. Bu bağlamda ahlak ve etik yapay zeka temelinde ele alındığında yapay zeka ahlakından ziyade yapay zeka etiğinden söz etmek yerinde olacaktır. Öyle ki hem yapay zekanın felsefi boyutu hem de geleceğe ilişkin pek çok belirsizliğin varlığı nedeniyle etiğin kılavuzluğuna ihtiyaç duyulmaktadır. Birkaç örnekle bu izleği somutlaştırmak faydalı olacaktır. Bunlar başlıca *netizen*, *posthümanizm*<sup>230</sup>, *transhümanizm*<sup>231</sup>, *küresel felaket riskleri*, *sanal yaşamlar* gibi kavramlarla örneklendirilebilir. Etik kaygıları örnekleyebilecek pek çok kavramsal gelişmeye tanıklık edeceğimizin işaretlerini bugünden görmek mümkündür. Örneğin kendi ülkelerinin mensubu olan insanlar “vatandaş” sözcüğüyle nitelendirilir. Gelecekte

<sup>230</sup> Hayles (1999); hümanizmin insana, doğal benliğe ve özgürlüğe dair konuları kapsamasına rağmen, posthümanizmin insan doğasından uzak zekanın, zeki makinelerin üretilmesini yücelten bir anlayışı kapsadığını ileri sürmektedir. Bu bağlamda posthüman sözcüğü ‘insan ötesi’ olarak yorumlanmaktadır. Detaylı bilgi için bk. Hayles, 1999: 22.

<sup>231</sup> Kökeni belirsiz olan “Transhümanizm” ve “Posthümanizm” kavramlarından, çağdaş transhümanistler Aydınlanma düşüncesindeki kökenlerine ve laik hümanizme bağlılıklarını vurgulamaktadırlar. Ayrıca, “transhümanizm” kavramını sıklıkla Julian Huxley (1887-1975) ile ilişkilendirmektedirler. 1957 senesinde “transhümanizm” teriminin Julian Huxley tarafından ortaya atıldığının bilgisi verilmektedir. Detaylı bilgi için bk.

internet ağının gelişmesiyle insanların *netizen*<sup>232</sup> “ağdaş” olarak nitelendirilebileceğini Hauben (1997) şu ifadelerle açıklamaktadır.

“21. yüzyıla hoş geldiniz. Siz bir Netizensiniz (Ağdaş) ve Ağ’ın mümkün kıldığı küresel bağlantı sayesinde dünya vatandaşı olarak varınız. Herkesi hemşehri olarak görüyorsunuz. Fiziksel olarak bir ülkede yaşıyorsunuz, ancak küresel bilgisayar ağı üzerinden dünyanın çoğu yerindesiniz. Neredeyse dünyadaki diğer tüm Netizenlerin bitişiğinde yaşıyorsunuz. Coğrafi bölünme yerini artık, sanal birlikteliğe bırakmış durumdadır.”<sup>233</sup>

Hauben’in bu kavramı tanıttığı yıllarda internet kullanımının yaygın olmadığı bilinmektedir. Yine de gelecekte böyle bir kavramın sanal bir toplum ortaya koyabileceği öngörüsü, insana ilişkin kavramların ekseninin sayısal sistem temelli kavramlara doğru kayması olarak nitelendirilebilir. Hatta günümüzün önemli gelişmelerinden *metaverse* düşünüldüğünde insanların gelecekte sanal bir yaşam üzerinde netizenler olarak var olacağını ifade etmek mümkündür. Öte yandan insanlığa ilişkin dinamiklerin hümanizm kavramıyla açıklandığı çağımızda posthümanizm ve transhümanizm gibi kavramlara ilişkin öngörüler de dikkat çekicidir. Her iki kavram da hümanizm sözcüğünün başına *post* ve *trans* ön eklerinin getirilmesiyle oluşturulmuştur.

Posthüman kavramı yeni bir biyolojik tür, sibernetik bir organizma ve dijital, bedensiz bir varlık olarak da açıklanmaktadır. Öte yandan hümanizm ve posthümanizm arasındaki bağlantı ya da geçişi sağlayan kavram, transhümanizm olarak gösterilmektedir. Bu bakımdan transhümanizm bir geçişi temsil etmektedir. Aynı zamanda kendi anlayışı bağlamında hümanizmin çağdaş bir formu olarak düşünülmektedir. Bu yönüyle transhümanizm akıl, bireycilik, bilim, ilerleme, kendini mükemmelleştirme gibi aydınlanmacı hümanist düşüncüyü kapsamakta ve güçlendirmektedir.<sup>234</sup> Söz konusu kavramlardan netizen, posthümanizm ve transhümanizmin anlamına ilişkin kısa bir bilgi verdik. Daha kapsamlı inceleme küresel felaket riskleri ve sanal yaşamların ele alındığı noktalarda gerçekleştirilecektir.

Yapay zekanın risk faktörlerinden kasıt gelecekte sosyolojik ya da çevresel anlamda teknolojinin riskleri ya da olası felaketlerin neler olabileceğine ilişkin öngörülerdir. Bu hususta ilk eleştirel görüş Bostrom ve Ćirković’in (2011) transhümanizmin insanlık açısından ciddi etkileri olacağına ilişkin düşünceleridir. Transhümanistler ve biyo-muhafazakarlar arasında yapay zeka, biyolojik silahlar, gelişmiş nanoteknoloji ve diğer teknolojilerin küresel felaket riskleri doğurabileceği hakkında ortak bir anlayış bulunmaktadır. Yani insana ciddi zarar verme

<sup>232</sup> *Netizen* sözcüğü için Ünalı (2003) tarafından “İnternet Vatandaşı” kavramı önerilmiş olsa da kavramın “*internet + citizen*” birleşimiyle ortaya çıktığı düşünüldüğünde “ağdaş” sözcüğünün amaca hizmet edeceği kanaatindeyiz. *Netizen* kavramına ilişkin detaylı bilgi için bk. Ünalı, 2003:4.

<sup>233</sup> Hauben ve Hauben, 1997: 3.

<sup>234</sup> Ranisch ve Sorgner 2014: 8.

potansiyeli taşıyan risklerin getirdiği endişe hakkında her iki tarafın da hemfikir olduğuna dikkat çekilmektedir. Örneğin nükleer terörizm, hastalıklar, genetik mühendislik, gök taşlarının hareketleri ve yapay zeka gibi gelişmelerden kaynaklanan risklerin nasıl ele alınacağını, gelecekte nasıl değişeceği değerlendirilmiştir. Bu gibi örnekler “Küresel Felaket Riskleri”<sup>235</sup> kavramı altında konumlandırılarak nanoteknoloji ve yapay zeka, bu tür muhtemel risk analizlerine örnek olarak gösterilmiştir. Fark edileceği üzere yapay zekanın yaygın olduğu transhümanist bir toplum düzeninde etik kaygılar endişe yaratmaktadır. Bu tür gelişmelerin Küresel anlamda ciddi felaketleri yani etik sorunları ortaya çıkaracağı öngörülmektedir.

Diğer risk faktörü de Schneider (2016) tarafından yorumlanmıştır. Epistemik zihin kavramının metafizik ve etik gibi birçok felsefi boyutları olduğuna dikkat çekerek ileri teknoloji toplumlarının etik değerlendirmesini, Aldous Huxley'in *Cesur Yeni Dünya* eserinde ileri sürülen sonuçlarla özetlemektedir. Bu toplumda doğal üremenin aksine çocuklar genetik mühendislik yoluyla beş ayrı sınıfta doğurtulmaktadır. Toplumun bütün üyeleri ait oldukları sınıfın rollerini benimsemek, toplum için iyi olanı onaylamak ve özellikle herkesi mutlu eden hafif halüsinojen Soma'ya minnet duymak için eğitilmektedir. Eserde öne çıkan genetik bilimi "aya çıkmak" kadar yüceltilen bir gelişme olarak sunulmaktadır. Oysa tarihsel deneyimlerin referansıya bu genetik mühendisliğinin sınıflar arası katliama neden olacağı ve soykırımı beraberinde getireceği iddia edilmektedir. Transhümanist insanların bilim kurgu öykülerinde tasvir edilen sanal yaratıklara benzeyecekleri ifade edilmektedir. Belirli “tür değiştirme” tekniklerini yasaklayan uluslararası bir antlaşma savunulurken pek çok transhümanist, bireyin entelektüel ve fiziksel yaşamını ilerlettiği sürece tür değişikliğinin makul olduğuna inanmaktadır. Gerçekten de transhümanizme göre gelecekteki bazı insanların “yüklemeler” şeklinde var olabileceği, bilgisayarlarda ölümsüz ve sanal yaşamlar yaşayabilen, birçok yönden yapay zekaya benzeyen süper zekalar olabileceği ön görülmektedir.<sup>236</sup> Schneider’ın ifadeleri düşünüldüğünde doğrudan yapay zeka terimi göze çarpmasa da söz konusu genetik, nanoteknoloji, nükleer tıp vb. teknolojik ilerlemelerin altyapısını genel anlamda yapay zeka sistemleri sağlamaktadır. Dolayısıyla risk faktörleri yapay zeka konusunda etik boyutta bağlayıcı bir role sahiptir. Ölüm sonrası sanal yaşamlara dair “yüklemeler” görüşü günümüzde sinema yapımlarında da vurgulanmaktadır. Söz konusu kavramların insan zihni bağlamında karşılaştırılması, gelişim süreçleriyle birlikte doğuracağı sonuçlara Rubin (2003) örneğinde okumak mümkündür.

<sup>235</sup> Bostrom ve Ćirković, 2011: 5.

<sup>236</sup> Schneider, 2016: 225-227.

Rubin (2003), elektro-kimyasal bir makine olarak beynin fizik yasalarıyla bağdaştığını savunmaktadır. Bu yüzden tüm fonksiyonların anlaşılabilirliğine ve çoğaltılabileceğine yönelik görüş bildirilmektedir. Bilgisayarlar zaten beyinden çok daha yüksek hızlarda çalıştığından, bilgiyi depolamak ve işlemek için beyne fazlasıyla rakip olabileceği hatta onu aşabileceği savunulmaktadır. Bu gerçekleştiğinde bilgisayarın en azından uyaranlara insan tepkilerinden ayırt edilemeyecek şekilde cevap verebileceği kabul edilmektedir. Bu noktada makineyi zeki olarak adlandırmanın da yerinde olduğu belirtilmektedir. Bu hikayenin doğru olması; insanlığın yok oluşunun, kendimizi gönüllü olarak makinelere dönüştürmenin ve makinelerle evrimsel rekabette kaybedişin kaçınılmazlığı vurgulanmaktadır. Dolayısıyla insan türünün bazı hayvanat bahçesi benzeri veya ayrılmış bölgelerde hayatını sürdürebileceği ihtimali öngörülmektedir. Makine çocuklarımız tarafından ebeveyn olarak alınacağımız durumların da gerçekleşeceği işaret edilmektedir. Sonuç olarak nanoteknoloji yoluyla atom düzeyinde yeniden yapılanan ve inşa edilen; örgütü tüm insani kavrayışı aşan bir zeka ile şekillendirilecek bir dünya tasarımı olarak ifade edilmektedir. Kısacası eğer insanlar en yalın haliyle geliştirilebilecek mekanizmalarsa parçalarımız başkaları tarafından değiştirilebiliyorsa o halde biyolojik ya da tasarlanmış olmalarının pek önemi olmadığına değinilmektedir. Bu yön beden için geçerliiyken zihnin durumunun ne olacağı sorulmaktadır. Burada sadece beynin biyolojik mekanizmasından ziyade aynı zamanda bu mekanizma aracılığıyla deneyimlediğimiz şeyin zaten sanal bir terim olduğu savunulmaktadır. Gerçek dünya hakkında hiçbir bilginin olmadığı; sadece beynimizin vücudumuzun duyusal girdilerini işlediğini bildiğimize inanılmaktadır. Bilinç de tamamen öznel ve zorunlu olarak tekil bir olgu olarak kabul edilmektedir. Alan Turing'in, taklit oyununda insanı kandırabileceği iddia edilen ünlü makine zekası testinden bu yana bilgisayarların bu tür dışa dönük işaretleri göstermesi yapay zekanın mihenk taşı olarak gösterilmektedir. Turing sonrası düşünürler, her ne kadar yapay yaşam hakkında sofistike bir resim geliştirmiş olsalar da ana hat Turing'in iddiasına dayanmaktadır. Turing'in düşüncelerine paralel bir yaklaşımla zihnin, bilgisayarların işlevinden neredeyse ayırt edilemez noktalara gelmesi için bir engel olmadığı savunulmaktadır. Her şey yerine getirildiğinde insanın ayırt edici özelliği belirli bir biyolojik konfigürasyondan başka bir şey olarak anlaşılamayacağı savunulmaktadır.<sup>237</sup>

Yapay zekaya ilişkin çeşitli kavramların ileri sürülmesi, risk faktörlerinin ne ölçüde önemli olduğunu ortaya koymaktadır. Kaldı ki *netizen*, *posthüman* ve *transhüman* kavramlarının evrimi insan bağlamında düşünüldüğünde kavramsal indirgeme veya daralma açıkça fark edilmektedir. Söz konusu kavramlar aslında insana ya da insanlığa dair sınırsız

<sup>237</sup> Rubin, 2003: 88-89.

bilinci ve özgür iradeyi mekanize etme gayesindedir. Diğer bir taraftan yine zihin-beden problemi ekseninde yapay zekanın insanlığı ne tür durumlara sürükleyebileceği hakkında bir öngörüü Fuller’de (2011) görebiliriz.

“Şimdi, 21. yüzyılın başlarında nano, biyo, bilgi ve bilişim teknolojilerindeki hızlı ve yakınsak ilerlemelerin ardından, zihin-beden sorunu giderek daha işlevsel hale geliyor. Karbon temelli bedenlerimizdeki insanlığı korumaya devam edecek olanlar bunu büsbütün değiştirmezlerse diğerleri, insanlığı daha dayanıklı silikon esaslı kaplara kaldıracaktır.”<sup>238</sup>

İnsanlık kavramının makineye göç ettirilmesi riskiyle Fuller, düalizme karşı bir duruş sergileyerek kavramların özünün korunmasına dikkat çekmektedir. Aksi halde, zihin-beden düalizmine dayalı kavramların insanı yapaylaştırarak doğasından koparacağını ve insanlığı makineye teslim edeceği belirtilmektedir. Geleceğe ilişkin kavramların dününü, bugününü ve yarınını öngörmek felsefenin görevlerindendir. Bu nedenle yapay zeka ifadesi felsefe söylemle ihtiyaç duyar. Kaldı ki yapay zeka sözcüğünün mucidi John McCarthy, Stanford Üniversitesi’nde şu ifadeleri kullanmıştır.

“Yapay zekanın şimdiye kadar yalnızca felsefeciler tarafından araştırılmış olan pek çok fikre gereksinimi var. Çünkü bir robot, insanlar kadar akıllı olabilmek ve deneyimlerinden bir şeyler öğrenebilmek için, birbirinden bağımsız olguları derleyip toparlayacak genel bir dünya görüşüne gereksinim duyacaktır.”<sup>239</sup>

İlk paragrafta sıraladığımız açıklamalara bakıldığında hem zihin felsefecileri hem de yapay zeka savunucuları tarafında pek çok farklı argüman bulunduğu fark edilecektir. Oysa risk faktörleri bu argümanların neredeyse hepsi için geçerli olup yoğunlukla insan doğasında düğümlenmektedir. Sayısal sistemlerin tam anlamıyla insana dair nitelikleri yerine getirmesi ya da getirememesi problemi kavramsal bir alana taşınmaktadır. Yapay zeka savunucuları bu durumda insanı sayısal sistemlere benzetme yoluyla farklı bir strateji izlemektedirler. Daha açık ifade etmek gerekirse felsefi derinliği olan akıl sözcüğünü zekaya indirgeyerek zihinsel süreçleri salt biyolojik, sayısal ya da davranışçı bir boyutta değerlendirmektedirler. O halde açıkça ifade edebilir ki yapay zeka ya da genel anlamda sayısal sistemler hız ve güvenilirlik ötesinde insanın zihinsel süreçlerine ilişkin ciddi bir katkı sağlamayacaktır. Bu çıkarım bizlere şu noktayı düşündürmelidir; kusursuz sayısal sistemleri inşa edebilmek çok basit bir süreçken, bu sistemlerin gelecekte doğurabileceği küresel felaket risklerinin önüne geçebilmek güç olabilir. Bu nedenle etik yaklaşımlar bütün sayısal sistemlerin tasarımında ön koşul olarak değerlendirilmelidir. Aksi takdirde felsefeden yoksun eylemlerin insan doğasını ortadan kaldırarak robot türevi insanlar ortaya koyması kuvvetle muhtemeldir.

<sup>238</sup> Fuller, 2011: 5.

<sup>239</sup> Güzeldere, 1998: 41.

### 3.2 Yapay Zeka Sistemlerine Yönelik Etik Yaklaşımlar

Bu bölümde yapay zeka sistemlerinin ahlaki değerler ve normlarla donatılabilmesinin olanakları üzerine görüşler, tartışmalar ya da en genel anlamıyla yaklaşımlar sınıflandırılacaktır. Söz konusu ahlaki değerler insani bir değer mekanik bir sistem üzerine inşası olasılığına yöneliktir. Dolayısıyla makinelerin ahlaki değerler ışığında değerlendirilmesi zorunlu olarak onların ahlakı temsil edebilecek sistemler olduğu sonucunu doğurmamaktadır. Bu konunun aydınlatılması amacıyla makineler için tartışılan etik yaklaşımlarının sınıflandırılması yapılarak hangi yaklaşımın hangi felsefi görüş bağlamında desteklendiğine ilişkin tartışmalara yer verilecektir.

Yapay zeka etiği incelemesinde bu konunun öncelikle bir tür insan etiği mi yoksa makine etiği mi olduğu tartışılmaktadır. Anderson ve Anderson (2011) tarafından kaleme alınan *Makine Etiği*'nde makineler için etik geliştiren yeni bir araştırma alanına dikkat çekilmektedir. Bu ayrımın hem pratik hem de teorik önemine değinilerek teoride makinelere etik ilkeleri ya da olası etik ikilemleri çözmenin bir yolunu bulma gayesindedirler. Kendi etik karar verme süreçleriyle sorumlu bir şekilde çalışmalarını sağlamak için bir ana yol tanımlama çabası gözlemlenmektedir. Bu noktada uygulamalı insan etiği, makine kullanımında uygunsuz insan davranışlarını önlemeyi içeren alanlardan biridir. Oysa makineler nasıl kullanılması ya da kullanılmaması gerektiği konusunda etik kurallara gereksinim duyulan araçlar olarak kabul edilmektedir. Anderson ve Anderson, esas amacın otonom makinelerin kendi etik kararlarını verebilmelerine yönelik güven sorunu olduğunu, bu nedenle de makineler için etik geliştirme çabasında olduklarını ifade etmektedirler.<sup>240</sup> Bu ifadeden anlaşılacağı üzere makine etiğinde konu yine iradeye ve sorumluluğa gelmektedir. Bu bölümde etik kuramlar çerçevesinde sayısal sistemlerin tasarlanması için benimsenin yaklaşımların sınıflandırılması yapılmıştır.

#### 3.2.1 Ahlaki Araç olarak Robot Düşüncesi

Robotlara ahlaki faaliyetlerin yüklenmesini anlamak için ayırt edilmesi gereken iki farklı robot teknolojisi bulunmaktadır. Sullins (2006) “telerobotlar” ve “otonom robotlar” olarak iki grupta bu teknolojiyi incelemiş olup telerobotların yalnızca asgari düzeyde özerk kararlar verebilen uzaktan kumandalı makineler olduğuna dikkat çekmektedir. Muhtemelen karmaşık yapay zekaya ihtiyaç duymadıkları için şu anda robotun en başarılı dalı olduğu öne sürülen telebotlara operatörleri tarafından gerekli zeka sağlanmaktadır. NASA Mars Rovers ve birçok dip deniz keşif robotlarının bu şekilde kontrol edildiğinin bilgisi de verilmektedir. Ayrıca telerobotik hemşireler gibi telerobotik cerrahinin de gerçekleşmesine ramak kaldığının

<sup>240</sup> Anderson ve Anderson, 2011: 1.

sinyalleri verilmektedir. Bu makinelerin aynı zamanda arama ve kurtarma çalışmalarının yanı sıra *Predator Drone* ve *SWORD* gibi uzaktan kontrol edilen silah platformlarını da içeren ve muhtemelen piyade gibi destek sunan ilk robot olarak savaş alanında uygulanacağını bilgisi verilmektedir. Bu makineler ahlakın söz konusu olduğu durumlarda kullanıldığına ve "Operatör - Robot - Kurban" şeklinde etkileşime girdiğine dikkat çekilmektedir. Telerobotların etik analizinde bu teknolojinin tasarımcısı, programcısı ve kullanıcısı kanadında ahlaki övgü veya suçlamanın yoğunlaştığına dikkat çekilmektedir. Neticede insanlar makinenin aldığı tüm önemli kararlarda belirleyici olduklarından makinenin ahlaki gerekçesini de sağlamaktadırlar. Yine de daha fazla araştırılması gereken bir soruna değinilen makalede, robotun uzaktan komuta edilmesiyle sağlanan eylemin niteliği düşünülmektedir. Uzakta bulunan robotu komuta eden operatörün belirli ahlaki kararlar almasının kolay olabileceği olasılığıdır bu husus. Örneğin bir telerobotik silah platformu operatörün başkalarına zarar vermek amacıyla makineyi tetiklemeye karar vermesini kolaylaştırmak için operatörünü savaş ortamından uzağa konumlandırabilmektedir. Robotun ahlaki bir araç olması için makinenin bu nedenlerle akıl yürütme ve hareket etme konusunda önemli derecede özerk bir yeteneğe sahip olması gerekmektedir. Bunu başarmaya çalışan makineler bu noktada devreye girmektedir.<sup>241</sup>

Ahlaki kuramların otonom robotlardan, yapay zeka destekli sayısal sistemlere yönelik kullanımı düşünüldüğünde ilginç problemler ortaya çıkmaktadır. Dikenli bir felsefi konu olarak değerlendirilen "özerklik" bağlamında "otonom robotlar" terimi, robot uzmanlarınca kullanıldığından Sullins bu hususu değerlendirmemiştir. Otonom robotlar, en yalın haliyle kendi programlarını kullanarak eylemleri hakkında en azından bazı kararları alabilmesi gereken robotlar olarak tanımlanmaktadır. Otonom makineleri tasarlayanların ahlaki karar verme sürecindeki sorumluluğuna dikkat çekilen makalede; tasarımcıların ahlaki kuruluşun tek sorumlu merci olmadıkları da kastedilmektedir. Bu durum, robotun kendini ahlaki eylemi için olası bir konumu işaret etmektedir. Ahlaki faaliyet bir ilişkiler ağında bulunduğundan makinelerin tasarımcıları, yapımcıları ve pazarlamacıları gibi diğer paydaşların yanı sıra diğer robotik ve yazılım paydaşları ve bu makinelerin kullanıcılarının hepsinin bir etkileşim topluluğu meydana getirmektedirler. Ahlaki öznenin gerçek kişi olması gerektiğine vurgu yapılarak bugünün robotlarının kesinlikle kişi olmadığı vurgulanmaktadır. Ayrıca bu teknolojinin gelecek bir süre için kişilik anlayışımıza meydan okumasının pek mümkün olmadığı da vurgulanmaktadır.

Yine de robotların ahlaki araçlar olabileceği iddiasını sürdürmek için Sullin, ahlaki tasarım için kişiliğin (*personhood*) gerekli olmadığını savunduğu makalede, *Robotların Ahlaki*

---

<sup>241</sup> Sullins, 2006: 25-26.

*Araçlığı Üzerine Felsefi Görüşler* başlığıyla bu konuda olası dört görüşten söz etmektedir. Birincisi, robotların ahlaki araçlar olmadıkları, ancak gelecekte olabileceğine ilişkin olumlu bakıştır. Daniel Dennett bu görüşü “HAL Cinayet İşlediğinde Kim Suçlanacak?”<sup>242</sup> başlıklı makalesinde tartışmaktadır. Kurgusal HAL gibi bir makinenin katil olarak kabul edilebileceğinden makinenin kasıt durumlarını, bilişsel inanç durumlarını veya zihinsel olmayan bir ihmal durumunu içeren suç işleme vakasıyla değerlendirmektedir. Bu örnek üzerinden robotların; inanç üzerine inanç, arzu üzerine arzu, korkular, düşünceler ve umutlar hakkında kanaat getirebileceği kastedilerek yine de ahlaki açıdan suçlu olmak için “daha üst düzey yönelime” sahip olmaları gerektiği belirtilmektedir. Yani pek çok soyut eylemde bulunması olasılığına rağmen irade sahibi olamayacağına dikkat çekilmektedir. Dennett, bugün böyle makinelerimiz olduğunu öne sürmezken, gelecekte bunlara sahip olmamak için hiçbir neden de görmemektedir.

Bu konuda alınabilecek ikinci konum, robotların şu anda veya gelecekte ahlaki değerleri gerçekleştirebilmesinin muhtemel olmadığı yönündedir. Selmer Bringsjord (2007) bu mevkide güçlü bir duruş sergilemektedir. Bu iddiaya ilişkin onun tartışması robotların asla özerk bir iradeye sahip olmayacağı gerçeğinden kaynaklanmaktadır, çünkü hiç programlanmamış bir şeyin eyleminden söz edilemeyecektir. Bringsjord bunu, PERI adlı bir robot deneyiyle göstermektedir. Bettina Schimanski ile laboratuvarındaki bir robot olan PERI’nin bir küreyi düşürüp düşürmeme durumu denenmiştir. Ahlak dışı bir durumda topu bırakacak, ahlaki değer taşıyan iyi durumda ise küreye tutması beklenmektedir. PERI’nin küreyi tutması ya da onu tamamen bırakması çalıştığı program tarafından kararlaştırılacağından ve bu da insan programcılar tarafından yazılmış olacağından Bringsjord görüşünü ısrarla sürdürmektedir. Bringsjord’a göre PERI’nin programcılar için şaşırtıcı olan her şeyi yapabilmesinin tek yolu, programa rastgele bir faktör eklenmesiyle gözlemlenmesini önermektedir. Ancak bu durumda da eylemlerin makine tarafından özgürce seçilen faktörlerden kaynaklanmayacağını ileri sürerek tercihlerin birtakım rastgele faktörler tarafından belirleneceğini düşünmektedir. Bu nedenle PERI ahlaki bir özne olarak kabul edilmemektedir.<sup>243</sup> Bu argümanla ilişkili diğer bir örnek de Bernhard Irrgang’a (2006) aittir. Ona göre “ahlaki açıdan sorumluluk taşıyabilmek için kişilik veya öznellik ile karakterize edilen bir katılıma gereksinim duyulmaktadır.”<sup>244</sup> Sadece bir insanın ahlaki araç olabileceği iddiasıyla insana çok yakın özellikler sergileyen bir robotun bile özneliği başaramayacağı öne sürülerek bu türden bir ahlaki davranışın imkansızlığını işaret etmektedir. Bu argümanla ilgili bir sorun olduğunu düşünen Anderson ve

<sup>242</sup> Dennett, 2014: 203-214.

<sup>243</sup> Bringsjord, 2007: 6.

<sup>244</sup> Irrgang, 2006: 12.

Anderson'a göre, hepimiz sosyalleşmenin ürünüyüz ve bu kültürel inanışlar da algoritmadan ileri gelmektedir. Bu nedenle PERI'den daha iyi olmadığımız sonucu çıkarılmaktadır. Ayrıca Bringsjord'un haklı olması durumunda bile bizlerin de ahlaki öznelere olamayacağına vurgu yapılmaktadır. Kaldı ki inançlarımızın, hedeflerimizin ve arzularımızın kesinlikle özerk olmadığı düşünülmektedir. Zira bütün bunlar kültür, çevre, eğitim, beyin kimyası gibi şeylerin ürünü olarak kabul edilmektedir. Bu noktada Bringsjord'un robotların ahlaka sahip olamayacağına ilişkin görüşü, Anderson ve Anderson tarafından bahsedilmeye değer görülmemektedir.<sup>245</sup>

Üçüncü olası duruş bizlerin ahlaki öznelere olmadığımız, ancak robotların esasen ahlaki öznelere olduğu görüşüdür. En azından bir kişinin bile bu görüşe sahip olması oldukça ilginç karşılanmaktadır. Joseph Emile Nadeau'nun (2006), bir eylemin özne tarafından tamamen rasyonel nedenlere dayanması kaydıyla özgür bir eylem olabileceği iddiası tartışılmaktadır. Dahası yalnızca salt mantık temelinde işleyen bir öznenin gerçekten özgür olabileceği iddiasıdır. Eğer ahlaki özne için özgür irade gerekliyse ve biz insanlar olarak beynimizde çalışan böyle bir aparat yoksa o zaman Nadeau'nun mantığına göre özgür özler olmadığımız sonucu çıkarılmaktadır. Nadeau'ya göre bu türden bir tasarım gerçekleştirilebilirse dünyadaki ilk gerçek ahlaki öznelere onlar olacağı savunulmaktadır.<sup>246</sup>

Bu konuda dördüncü duruş, Oxford Üniversitesi Bilgi Etik Grubu'ndan Luciano Floridi ve J. W. Sanders (2004) tarafından tartışılmaktadır. Ahlak teorisindeki birçok belirgin paradoksun özgür irade ve yönelimsellik gibi sorunları ortadan kaldıran “zihin-siz bir ahlak” anlayışını benimseme iddiası tartışılmaktadır. Zira bunların hepsi zihin felsefesinde robotlar gibi yapay birimlere olur olmadık yerde uygulanan çözülmemiş konular olarak kabul edilmektedir.<sup>247</sup> Bu anlayışın temelini soyutlama seviyesinin dengelenmesi oluşturmaktadır. Bu yönüyle Floridi ve Sanders, makine etiğini zihin felsefesindeki soyutlamalardan bağımsız bir şekilde gerçekleştirmeyi hedeflemektedirler.

Floridi'nin savunduğu görüşe göre aracın davranışı belirli soyutlama düzeyinde; etkileşim (durum değişikliğiyle uyarana tepki), özerklik (uyarıcı olmadan durumu değiştirme yeteneği) ve uyarlanabilirlik (geçişli değiştirme yeteneği) sergilediğinde "tasarımcılık" kavramının ortaya çıktığını vurgulamaktadır. Ahlaki bir aracı, bir kişiye iyi veya kötü tepki verebilecek araç olarak görmektedir. Ne yönelim ne de özgür iradenin ahlaki tasarım için şart

<sup>245</sup> Anderson ve Anderson, 2011: 156.

<sup>246</sup> Nadeau, 2006: 241-247.

<sup>247</sup> Anderson ve Anderson, 2011: 157.

olmadığını belirtmektedir. Floridi'ye göre bu ahlaki kuruluşun yapay araçlara atfedilmesi mümkündür.<sup>248</sup>

Sullins, görüşlerini dördüncü konuma dayandırarak söz konusu teknolojinin "Robot önemli ölçüde otonom mu? Robotun davranışı niyetli mi? Robot sorumluluk düzeyinde mi?" şeklinde üç soruya indirgemıştır. Makul bir soyutlama düzeyinde yaklaşılmaması gereken bu soruların üçüne de "evet" cevabı verilebilirse robotun ahlaki bir ajan kabul edilebilecektir. Gayet güvenilir ahlaki değerlendirmelere layık araçlar oldukları iddia edilebilse de robotlar, diğer iki gereklilikten yoksunlarsa yetişkin insanlardan talep ettiğimiz ahlaki değerleri bekleyebileceğimiz sağlam ahlaki birimler olamayacaklardır.<sup>249</sup>

Robotların özerkliği hususunda öne çıkan düşünce; robotun önemli ölçüde özerk algılanıp algılanmayacağıdır. 'Özerkliğin' felsefi olarak tespiti zor olduğu düşüncesiyle herhangi bir tür robotun radikal ölçüde özerkliğe sahip olacağı fikrinden kaçınılmaktadır. Hatta insanların bile bu niteliğe sahip olduğundan şüphe duyulmaktadır. "Özerklik" terimini mühendislik anlamında kullanmak, makinenin başka bir aracı veya kullanıcının doğrudan kontrolü altında olmaması olarak kabul edilmektedir. Robotun bir telebot olmaması veya geçici olarak bir robot gibi davranmaması gerektiğini vurgulayan Sullins, robotun bu düzeyde bir özerkliği varsa nesnel bağımsız bir temsile sahip olacağını düşünmektedir. Bu özerk eylem, robotun hedeflerine ve görevlerine ulaşmada etkiliyse robotun etkili bir özerkliğe sahip olabileceğini savunmaktadır. Makinenin etkin özerkliği, yani hedeflerine ve görevlerine ulaşmadaki ustalığı ne kadarsa tasarım düzeyinin de o ölçütte olduğunu belirtmektedir. Bu tasarım ahlaki anlamda zarar veya iyiliğe neden olduğunda makinenin ahlaki temsilinden söz edilebileceğini bildirmektedir. Açıklandığı üzere özerkliğin kendi başına ahlaki kurumu tanımlamak için yeterli olmayacağı vurgulanmaktadır. Dolayısıyla bakteri veya hayvanlar, ekosistemler, bilgisayar virüsleri, basit yapay yaşam programları veya basit otonom robotlar gibi özerklik sergileyen varlıkların sadece bu yönlerden sorumlu ahlaki unsurlar olarak değerlendirilmemesi gerektiğinin altı çizilmektedir.<sup>250</sup>

İkinci husus, makinenin 'kasıtlı olarak' hareket etme yeteneğidir. Yani en güçlü robotların dahi yönelimselliğe sahip olup olmadığı sorunu. Bu bölümde Sullins, bunun ispatına gerek olmadığını çünkü insanlar için de geçerli bir argüman olmaksızın bunu kanıtlamanın imkansız olduğunu vurgulamaktadır. Davranışlar, standart halk psikolojisine yatkınlığı ya da "niyeti" sağlayabilecek düzeyde karmaşık olduğu sürece, bu soruya olumlu cevap alabilmenin yeterli olduğu belirtilmektedir. Robotun yazılımı ve ortama dair karmaşık etkileşimler, ahlaki

<sup>248</sup> Anderson ve Anderson, 2011: 85.

<sup>249</sup> Sullins, 2006: 27.

<sup>250</sup> Sullins, 2006: 28.

açından zararlı veya faydalı bir şekilde hareket etmesine neden oluyorsa ve eylemler görünürde niyetli ve hesaplanmış ise makinenin ahlaki bir unsur olduğu iddia edilmektedir. Eylemlerin gerçekten felsefi titizlikte niyetli olmalarına veya tüm soyutlama seviyelerinde özgür olan bir iradeden türetilmesine gerek olmadığı düşünülmektedir. Gerekli olan tek şeyin, ilgili araçlar arasındaki etkileşime dahil olan tüm araçların kıyaslanabilir seviyede bireysel yönelim ve özgür irade taşıması olduğuna inanılmaktadır.<sup>251</sup>

Nihai olarak robotun başka bir ahlaki araca karşı sorumlu olduğu varsayıldığında bir robota ahlaki araçlığı yüklemenin mümkün olabileceği iddia edilmektedir. Makinenin bilinç, ruh, zihin ya da felsefi anlamda şüpheli diğer varlıklardan herhangi biri hakkında hiçbir iddiası olmayabilir. Bu inançların veya bunu sağlayan programların, makinenin karşılaştığı ahlaki soruları ve muammaları çözmede motive edici olması gerekliliğine dikkat çekilmektedir. Bir insan hemşirenin ahlaki bir araç olduğu düşünüldüğünde bir makinenin de aynı görevleri özerk bir şekilde yerine getirmesiyle ahlaki bir araç kabul edilebileceği öngörülmektedir. Bu makinenin içinde çalıştığı sağlık sisteminin sorumluluğunu, doğrudan bakımı altındaki hastaya yönelttiği gerekçesiyle sorumlu davrandığı düşünülmektedir. Bundan daha az kabiliyete sahip herhangi bir makinenin tam bir ahlaki araç olmayacağı, yani ahlaki tasarımcılık açısından özerkliğin, yönelimselliğin ve sorumluluğun zorunlu kavramlar olduğu düşünülmektedir. Sonuç itibarıyla Sullins'e göre; robotlar otonom niyetleri ve sorumlulukları makul bir soyutlama seviyesinde taşıdığına ahlaki araçlar olarak kabul edilebilmektedir. Robot birçok açıdan özerk görünüme sahipse muhtemelen insanın ahlaki durumuna yakın veya onu aşan sağlam ahlaki bir sistem olarak kabul edilmektedir. Bu nedenle gelecekte oldukça karmaşık etkileşimli robotların ilgili haklara ve sorumluluklara sahip ahlaki araçlar olacağına kesin gözüyle bakılmaktadır. Kaldı ki günümüz robotlarının ahlaki tasarım örneği olabileceğine dikkat çekilmektedir.<sup>252</sup>

### 3.2.2 Felsefi Yaklaşımlar

İlk olarak Christopher Grau'nun faydacılık temelinde makine etiği değerlendirilmiştir. Felsefi bir kaynak olarak "Ben, Robot" filmini kullanmıştır. Bu filmde, makine etiğinin hem büyüleyici hem de rahatsız edici yönlerini sunarak bazı sorunlara dikkat çekmiştir. Hikaye, Isaac Asimov'un Ben, Robot adlı kısa öyküler koleksiyonunda sunduğu orijinal robot etiğinin "üç yasasını"<sup>253</sup> içermektedir.

- İlk yasa; "Bir robot bir insana zarar veremez veya zarar görmesine müsaade edemez",

<sup>251</sup> Sullins, 2006: 29.

<sup>252</sup> Sullins, 2006: 28.

<sup>253</sup> Detaylı bilgi için bk. *I, Robot* (2004)

- İkinci yasa; “Bir robot birinci kanunu ihlal eden emirlerin haricinde insanların verdiği emirlere uymak zorundadır.”,
- Üçüncü yasa; “Bir robot kendi varlığını korumak zorundadır, ancak bu koruma birinci ve ikinci kanunu ihlal etmemelidir.”

Bu yasaların robotlara yüklenmesiyle robotların zarar verme olasılığını ortadan kaldıracığı hatta insanların güvenliği için kendilerini feda edebileceği savunulmaktadır. Filmde de bu düşünce gayet açık bir şekilde sergilenmiştir. Ancak beklenmedik bir müdahale sonucu üç yasanın dışına çıkan son derece karmaşık robotlar sürü zekaya<sup>254</sup> dönüşerek felakete yol açmıştır.

Grau, faydacılığı bir kimsenin kayıplarının başkalarının kazançlarıyla dengelenmesine izin veren ahlaki bir felsefe olarak yorumlamaktadır. Faydacılığı, çoğunluğun kazanımlarını sağlamak uğruna birey tarafında yaşanan acıları meşrulaştıran ahlaki bir felsefe olarak görmektedir. Çoğunluğun ıstıraplarını ve zevklerini kümeleyen anlayış karşısında bireysel kimlik önemsizdir. “Bireysel kimliğe” sahip olmanın olası kriterleri; öz-farkındalık ve öz-yönetim kapasitesi, nedenleri tanıma ve bunlara cevap verme yeteneği, özgür ve sorumlu seçim kapasitesi gibi ifadeleri öne çıkarmaktadır. Yeterli koşulların bir listesini ortaya koymaksızın, psikolojik yapı ve kapasite açısından insana benzeyen bir robotun muhtemelen bir tür bireysel kimliğe sahip olacağını varsaymanın nispeten güvenli olduğunu düşünmektedir. Buna göre, böylesi bir robotun faydacı hatlar boyunca ele alınmaması gerektiğinin de altını çizmektedir. Sadece sezgi özelliğini taşıyan hafif sofistike robotların faydacı tavrı engellemeyeceğini, bu tür robotların ahlaken birçok hayvandan farksız olacağını vurgulamaktadır. Bu tür bir varlığın faydacılık açısından irrasyonel olmayacağını belirtmektedir. Kaldı ki güçlü bir benlik duygusundan yoksun ancak sezgiye sahip olan bir varlığı daha geniş kitlelerin çıkarları uğruna feda edilebileceğini öngörmektedir. Bireysel bir kimliğe sahip olmayan, ancak yine de sezgiye sahip olan robotları ahlaki olarak orta bir konumda değerlendiren Grau, ahlaki hiyerarşide sezgisi olmayan bir nesne ile insan arasında bir yerde konumlandırmaktadır. Grau’ya göre hayvanların eyleme geçmek için gerekli düşünce kapasitesinden yoksun olduğuna dikkat çekmektedir. Buna karşın robotlar için bu sınırlamanın söz konusu olmadığını, çünkü sezgi sahibi robotların tamamen faydacı hesaplamalar yapabilmesinin mümkün olabileceğini iddia etmektedir. Dolayısıyla bu varlıkların birbirleriyle faydacı ilkeler çerçevesinde etkileşimde

<sup>254</sup> 1980’lerde hücresel bir robotik sistem sınıfını belirtmek amacıyla bir “vızıltı kelimesi” olarak tasarlanan sürü robotiği başlıca karıncalar, eşekarısı ve termitler gibi sosyal böceklerin gözlemlenmesiyle çok sayıda robotun koordinasyonuna yeni bir yaklaşım getirmeyi hedeflemiştir. Daha sonraları sürü robotiği terimi, sürü (*swarm*) zekasının fiziksel olarak sistemlere uygulanması olarak kullanılmaya başlanmıştır. Öyle ki sürü robotiği, çok sayıda aracın kendi aralarında ve çevreleriyle etkileşimlerden beklenen kolektif bir davranış tasarımının incelenmesi olarak tanımlanmaktadır.. Detaylı bilgi için bk. Sahin E. vd. 2008: 87,88.

bulunması ve robotların da bu bakış açısıyla dikkate alınması gerekmektedir. Grau'ya göre faydacı ilkelere yönelik davranan bu sistemler benlik duygusundan yoksundur. Başka bir deyişle robotlar tarafından filmlerde ve edebiyatta sıkça gösterilen toplu davranışlar ve bireysel fedakarlıklar, robotların ilgili benlik duygusundan yoksun olduğu göz önüne alındığında anlamlı hale geldiğini öne sürmektedir. Daha büyük yeteneklere sahip robotlar oluşturmak için zorlayıcı nedenler ortaya çıkması durumunda farklı ahlaki standartlara gereksinim duyulabileceği de vurgulamaktadır. Bu bağlamda insan ahlakına benzer ahlaki bir statüye sahip robotlar yaratmaya çalışmadan önce çok dikkatli olunması gerektiğini düşünmektedir. Sonuç olarak böyle büyük güçlere sahip makinelerin oluşturulması sadece yaratılan robotlar için değil, bizim için de büyük sorumluluklar getireceği fikrini öne sürmektedir.<sup>255</sup>

İkinci olarak Powers (2006), makine etiğini kötülükten uzaklaşarak ve belki de iyiliği teşvik edecek bilgisayarları nasıl programlayabileceğimiz sorununa taşımıştır. Makinelerin kendi içinde etik olabilmesini, ahlaki olarak zarar veya faydaya neden olan veya bunlardan kaçınan pratik akıl yürütmeye bağlamaktadır. Dolayısıyla makine etiğinin temel sorunu makinenin etik bir düşünüp taşınma simülasyonunu gerçekleştirme sorunu olarak öne çıkmaktadır. İnsanlar arasında uygun etik düşüncüyü neyin oluşturduğu konusunda anlaşmazlığın süregeldiği bilgisine yer veren Powers, faydacı gelenekteki etik düşüncenin temelini aritmetiğin oluşturduğunu iddia etmektedir. Bu yaklaşımla, bir dizi olası eylemden etkilenecek tüm bireyler için muhtemel faydaları hesaplayarak ve sonra toplam faydayı en üst düzeye çıkarmayı vaat eden eylemi seçerek doğru etik sonuca ulaşılmaktadır. Fakat farklı bireyler üzerindeki faydaları nasıl ölçtüğümüz ve etik sonuçlar hakkında yeterli bilgiye sahip olup olamayacağımız faydacılığın problemleri olarak gösterilmektedir. Oysa deontolojik<sup>256</sup> gelenek, diğerlerini nasıl etkileyebileceklerine bakılmaksızın bazı eylemlerin yapılması veya yapılmaması gerektiğini savunmaktadır. Deontoloji, eylemler ve bunların mantıksal (deneyselin aksine) sonuçları hakkında karmaşık akıl yürütmeyi vurgulayan bir yaklaşımdır. Eylem kurallarına; hangi kuralları benimseyeceğimizi nasıl bildiğimize, kural sistemlerini nasıl oluşturabileceğimize ve olası bir eylemin kuralın kategorisine katılıp katılmadığını nasıl bildiğimize odaklanır. En ünlü deontolog Immanuel Kant (1724-1804), eylem kurallarının (yani kategorik buyruğun) oluşturulması için bir prosedürün var olduğunu ve kategorik buyruğun bir türünün tamamen formel şekilde işlediğini iddiasına Powers vurgu yapmaktadır. İnsanın pratik

<sup>255</sup> Grau, 2006: 52-55.

<sup>256</sup> “Kant ahlakında, yani deontolojik ahlakta, insanı bir araç olarak değil, daima bir amaç olarak görmek ve buna göre eylemde bulunmak söz konusudur. Örneğin hayatın elemelerinden kurtulmak için intihar etmek istiyorsunuz. Bu durumda, elemenden kurtulmak hissine, yani bencilliğe dayanarak intiharın meşruiyetini ve ahlaken makbul olduğunu doğal bir yasa olarak isteyebilir misiniz? Hayır! Çünkü nefsimiz hakkındaki sevgimiz, başka deyişle, kendimizi beğenmek mümkün merteye ömrümüzü sürdürmeyi gerektirir.” Detaylı bilgi için bk. Öktem: 2019: 46.

akıl yürütmesi öncelikle olguları dikkate alması ve bir sonraki eyleme geçmesiyle bağdaştırılırken bu ilişki veri tabanındaki bir dizi bildirim biriminden çıktıya dönüşmesiyle makinede gerçekleşen değişikliklere benzetilmektedir. Ancak insanın pratik muhakemesi; neye izin verileceği, ne yapması gerektiği, neyin ahlaki olduğu gibi makinelerin en azından şimdiye kadar yetersiz kaldığı her şeyi bir araya getirme aşamasını içerdiği belirtilmektedir. Aristoteles'in, insanlar "iradenin zayıflıklarından" mustarıptir ifadesine göndermede bulunan Powers, bunun bir makine için sorun olmayacağı kanısındadır. Makine bir kez ne yapması ya da yapmaması gerektiği konusunda sonuca ulaştığında çıktının otomatik olarak peşinden gideceğini vurgular. Ancak makinenin orta aşamaya yani gerçekleri kurallarla eyleme bağlayan normatif sonuçlara nasıl ulaşacağı sorusunu sorar. Kural tabanlı bir etik teorinin eyleme yönelik görevleri veya kuralları hesaplanabilir şekilde izlenebileceği inancı savunulmaktadır. Kural tabanlı teoriler arasında Kant'ın kategorik buyruğunun formel bir prosedür sunduğuna dikkat çekmektedir.<sup>257</sup>

Kant'ın kategorik buyruğunda ileri sürdüğü "Ancak aynı zamanda genel bir yasa olmasını isteyebileceğin maksime göre eylemde bulun."<sup>258</sup> ifadesi dışarıdan bir zorunluluk bulunmaksızın kişinin kendiliğinden, içten bir ödevle eyleme geçmesi ve her bir eylemin bütünü oluşturmasıyla evrensel bir ahlakın ortaya çıkmasını kastetmektedir. Kant'ın maksimlerden<sup>259</sup> görev türetme prosedürü, insanlara özgü özel bir ahlaki veya entelektüel sezgi gerektirmediğine vurgu yapılmaktadır. Formalist bir Kantçı için, bir maksimin evrensel bir kural olup olmadığı, bir insan ya da makine için de aynı ölçütte karar problemi doğurmaktadır. Kant'ın bir makineye yönelik fikir verici bir çözümü şu ifade üzerinden örneklenmektedir:

"Çelişki ve karışıklık ortaya çıkarsa eylem reddedilir; uyum ve uygunluk ortaya çıkarsa kabul edilir. Ahlakî tutum sahibi olma yeteneği sezgisel bir araç olarak bundan gelir. Çünkü doğamız gereği sosyal varlıklarız ve başkalarında kabul etmediğimiz şeyleri, kendimizde de içtenlikle kabul edemeyiz."<sup>260</sup>

Makine için evrensel ahlaki yasalar sınıfı şart koşulmadığından bu durum bir makine etiği değil, araçla çalışan bir insan etiği olarak kabul edilmektedir. Makinenin kendisi, evrenselleştirme adımını bireysel maksimlere uygulayıp sonra haritalama yaparak etik bir teoriyi oluşturabileceği sonuçlara göre bunları "yasak, izin verilebilir, zorunlu eylemler" gibi

<sup>257</sup> Powers, 2006: 46.

<sup>258</sup> Kant, 2002: 38.

<sup>259</sup> "Bu kelime, Kant'ın terminolojisinde, bir istekteki sübjektif niyet ve ilkeyi ifade eder. Örneğin bir kimse, herkes tarafından beğenilmek için bazı davranışlarda bulunuyorsa, ün kazanmak istemek bu davranışın maximidir. Fakat bu maxim, herkesi içine alabilecek bir şekilde genişletilebilir; bir ilke, bir kanun haline getirilebilir. O zaman maxim, insanın sübjektif sferinden, niyet ve isteklerinden, insanlar arasındaki ilgilere, bağlara yükseltilmiş olur. Demek ki, akıl sahibi bir varlık, kendi öznel pratik ilkelerini, yani maximlerini ya aynı zamanda genel yasalar olarak düşünemez ya da onları genel bir yasa koymaya uygun kılan biçimler olarak kabul etmek zorundadır." Detaylı bilgi için bk. Öktem: 2019: 45.

<sup>260</sup> Powers, 2006: 47.

geleneksel deontik kategorilere ayırabileceğini açıklamaktadır. Daha sonra makinenin herhangi bir olası maksimin bir elemanı olup olmadığını kontrol edecek, sonlu bir aksiyom setinin sonuçlarıyla özdeş ise bunun nihai bir aksiyom haline dönüştürülebileceği ifade edilmektedir. Dolayısıyla yasaklı maksimler teorisinin makinenin yapmaması gereken şeylerden kaçınmasının mümkün olabileceğini savunmaktadır.<sup>261</sup>

Tartışılan diğer bir konu da Susan Leigh Anderson'un (2011) *Makine Etiği ile Felsefi Endişeler* eserinde felsefi bir bakış açısıyla makine etiği projesine yönelik yedi zorluğu ele almasıdır: (1) Etik, hesaplanabilecek türden bir şey değildir. (2) Makine etiği, etiğe erdeme dayalı yaklaşımla bağdaşmaz. (3) Makineler etik davranamazlar çünkü özgür irade, yönelim, bilinç ve duygulardan yoksundurlar. (4) Etik göreceliler, etik ikilemlerde makinelere programlanacak tek bir doğru eylem olmadığını savunurlar. (5) Bir makine etik davranmaya başlayabilir, ancak daha sonra kendi çıkarlarını gözeterek etik olmayan davranmaya dönüşebilir. (6) Makineler etik davranamazlar çünkü kendi çıkarları doğrultusunda davranamazlar ve bu nedenle asla gerçek etik ikilemlerle karşılaşmazlar. (7) Bir makinenin karşılaşabileceği her türlü etik ikilemi öngöremeyebiliriz, bu nedenle eğitiminin eksik olması muhtemeldir, böylece makinenin bazı durumlarda etik dışı davranmasına izin verilir.

Anderson'ın bu zorluklara cevabı şöyledir: Eylem Faydacılığı teorisi, etiğin prensipte hesaplanabilir olduğunu göstermektedir. Daha tatmin edici bir teorinin, Eylem Faydacılığında eksik olan deontolojik görevleri de içeren birinci dereceden görev (*prima facie*)<sup>262</sup> yaklaşımı olduğunu ve bu yaklaşımı desteklemek için gereken karar ilkelerinin bir makine tarafından keşfedilebileceğini savunmaktadır. (2) Yalnızca makinelerin eylemleriyle ilgilendiğimiz için eyleme dayalı etik yaklaşımın benimsenmesi uygundur. (3) Bir makineyi eylemlerinden sorumlu tutmak için özgür irade, yönelimsellik ve bilinç gerekli olabilir, ancak yalnızca makinenin ahlaki açıdan doğru eylemleri gerçekleştirmesini ve istenirse bunları haklı çıkarması önemsenmektedir. Başkalarının acılarını hesaba katabilmek için makinelerin kendi duygularına sahip olmaları şart olmayabilir. Dahası insanlar genellikle duygularına o kadar kapılırlar ki etik olmayan bir şekilde davranırlar, bu yüzden makinelerin duyguları olmamasını tercih edilebilir. (4) Birçok etik ikilemde, etik rölativistleri çürüten doğru eylem konusunda etikçiler arasında bir anlaşma bulunmaktadır ve makinelerin yalnızca kabul edilebilir davranışın ne olduğu konusunda anlaşmanın olduğu alanlarda çalışmasına izin verilmelidir. “Etiğin uygulanması, kabul edilen etik teoriden daha eksiksiz olamaz.” (5) İnsanlar, başkalarıyla rekabet halinde olan

<sup>261</sup> Powers, 2006: 47.

<sup>262</sup> Birinci dereceden bir görev, daha güçlü bir görevle çelişmediği sürece yapılması gereken bir şeydir, bu nedenle istisnaların olmadığı mutlak bir görevden farklı olarak istisnalar olabilir. Detaylı bilgi için bk. Anderson ve Anderson, 2011: 23.

biyolojik varlıklar olarak, kendi çıkarlarını destekleme eğiliminde olan varlıklara dönüşmüş olabilirler; ancak bu yatkınlıktan yoksun makineler yaratmamız mümkün görünmektedir. (6) "Etik bir ikilem paradigması, kişinin ahlaki açıdan doğru eylemin ne olduğunu bildiği ancak yapmayı zor bulduğu bir durum değil, ahlaki açıdan doğru eylemin ne olduğu belli olmayan bir durumdur. İdeal olarak belirlenmiş bir ahlaki ilke veya ilkeler kullanılarak belirlenmelidir." Ayrıca neden sadece etik davranabileceğinden emin olmak yerine, bir makinede iradenin zayıflığını yeniden yaratmak isteyelim? (7) Makine genel etik ilkeleri izleyecek şekilde eğitilmişse, bunları beklenmeyen durumlara bile uygulayabilmelidir. Ayrıca "Bir makinenin aldığı etik eğitimi güncellemenin bir yolu olmalı."<sup>263</sup> ifadesiyle makinelerin sadece etik davranmasını sağlamak yerine, onun irade yönünden zayıflığını yeniden canlandırmanın anlamsız olacağını ifade etmektedir. Anderson'ın göz önünde bulundurduğu kaygıların en ciddişi muhtemelen etiğin hesaplanabilecek türden bir şey olup olmadığı ve bir makineye programlanacak tek bir doğru ve yanlış standardı bulunup bulunmadığıdır. İkincisi göz önüne alındığında farklı etik inançların farklı toplumlarda işleyen makinelere programlanabileceğini, böylece etik rölativistlerin hala makine etiği üzerinde çalışabileceği iddia edilebilir.<sup>264</sup>

Söz konusu sayısal sistemlerin etik bir yaklaşımı ortaya koyması faydacılık açısından değerlendirilmiştir. Faydacılık, çoğunluğun kazanımlarını sağlamak adına azınlığın acılarını meşrulaştıran ahlaki bir felsefe olarak görülmektedir. Faydacılık bu çerçevede düşünüldüğünde sayısal bir sistemin ikileme düştüğü durumda karar verirken çoğunluğun çıkarını gözeterek karar vermesi beklenmektedir. Örneğin günümüzde yaygınlaşan otonom araçlar düşünüldüğünde olası bir aksaklık sonucu kırmızı ışıktaki geçmekte olan bir yayanın mı yoksa araç içindeki birkaç kişinin mi kurban edileceği sorusu akla gelmektedir. Bu durum "vagon ikilemi" olarak da bilinen (*trolley paradox*) düşünce deneyidir. İlk tartışılan konu faydacılık ilkesiyle sayısal sistemlerin kararlarında çoğunluğun çıkarlarını gözetmesidir.

Öte yandan diğer bir yaklaşım da kötülükten uzaklaşarak iyiliği hedefleyen sistemlerin tasarlanması dikkat çekmektedir. Bu noktada deontik bir bakışla sistemlerin kendi içinde karar vermesi "yasak, izin verilebilir, zorunlu eylemler" gibi kategoriler altında belirlenmeye çalışılmaktadır. Kant'ın ödev ahlakını mekanik sistemlere yerleştirme çabasında olan bu görüş nihayetinde bazı eleştirilere maruz kalmıştır. Kaldı ki irade sahibi olduğundan kuşku duyulan bir sistemin ödev ahlakını gerçekleştirme olasılığı başlı başına şüphelidir. Bu noktada Powers'ın Kant'a ilişkin bir mantığı savunan makalesini eleştiren Anderson ve Anderson'a göre, Kategorik Buyruğun izinden gitmek, bir dizi maksim inşa etmeyi ve her yeni maksimin

<sup>263</sup> Anderson ve Anderson, 2011: 83-84.

<sup>264</sup> Anderson ve Anderson, 2011: 84.

kabul edilmiş olanlarla tutarlılığını tasarlamayı içermektedir. Ancak bu durumda kabul edilmeyen bir maksimin süreci bozacağı gerekçesiyle insanların etik düşünce süreciyle başa çıktığı gibi bir makine kendini düzeltmeyeceği ve süreci çökerteceğine dikkat çekilmiştir. Ayrıca eleştirilerin yanı sıra etik problemlerde yedi farklı zorluklara yer verilmiştir. Bu zorluklar; etiğin niceliksel bir şey olmadığı, makinenin erdemi olamayacağı, özgür irade, yönelim, bilinç ve duygulardan yoksun olduğu, etik ikilemlerde tek bir programlama yolunun olmadığı, etik davranan makinelerin etik olmayan davranışlara dönüşebileceği, makinelerin hiçbir zaman gerçek etik davranış taşıyamayacağı ve etik olmayan davranışlara kimi zaman izin verileceği gibi hususlardır.

### 3.2.3 Etiğe Yönelik Tepeden ve Temelden Yaklaşımlar

Wallach ve Allen (2008) insan ahlakını modellemek amacıyla tepeden ve temelden<sup>265</sup> olmak üzere iki yaklaşımla görüşlerini bildirmişlerdir. Bu anlayışta ahlaki karar verme özelliğinin yapay zekaya uygulanması otonom robotların doğal ve gerekli bir uzantısı olarak kabul edilmektedir. Ahlaki açıdan zeki araçlar geliştirmenin mimarisini iki yaklaşımla açıklayan Wallach ve Allen, etik teorilerini tepeden ve temelden yaklaşımlarla açıklamaktadır. İnsanlardaki ahlaki yargının karmaşık bir faaliyet olduğuna dikkat çekilerek pek çok insanın bu beceriyi gerçekleştirmekte başarısız olduğunun altı çizilmektedir. Kültürel farklılıkları aşan ortak değerler olsa da kültürlere ve bireylere yönelik etik kuramların ve geleneklerin ayrıntılarında farklılık gösterdiği belirtilmektedir. Dolayısıyla ahlaki kararların yeterliliğini kriterler üzerinden belirlemek son derece zor kabul edilmektedir. Bu yönüyle etik, yöntemleri ve uygulamalarıyla kafa karıştırıcı zorluklarla dolu olan bulanık bir disiplin gibi görülebilmektedir. Etiğin bir bilime indirgenebilmesi iddialı bir girişim olsa da otonom yazılım araçlarının ahlaki tasarımlarının önemli ölçüde insan etiği bağlamında gerçekleşeceği savunulmaktadır. Ayrıca bu projenin felsefi etik çalışmasında devrim yapma potansiyeli taşıdığı da belirtilmektedir. "Tepeden" ifadesiyle hem mühendislik hem de etik anlamlar kastedilmektedir. Mühendislik açısından tepeden temele doğru yaklaşımla, bir görevin doğrudan uygulanabildiği ve istenen sonucu elde etmek amacıyla hiyerarşik olarak düzenlenebildiği daha basit alt görevler gerçekleştirilmeye çalışılmaktadır. Etiğe yukarıdan aşağıya yaklaşım, faydacılık gibi felsefi olarak türetilmiş veya "Altın kural" gibi dini anlamda gerekçelendirilmiş, belirli genel bir etik teoriyi alan ve belirli durumlar için sonuçlarını elde eden yaklaşımdır. Sistemlerin yukarıdan aşağıya bir yaklaşımla belirli bir etik teoriyi alarak bu

<sup>265</sup> Burada sözü geçen tepeden ve temelden yaklaşımlar *top-down* ve *bottom-up* sözcüklerinin yerine kullanılmıştır. Sözlük anlamı itibarıyla yukarıdan-aşağıya ve aşağıdan-yukarıya anlamlarına gelen yaklaşımlar kastedilmektedir.

teoriyi uygulayabilen algoritmaların kompütasyonel gereksinimlerini analiz eden ve alt sistemlerin tasarımını yönlendirebilen bir yaklaşım olarak özetlenmektedir. Buna karşılık, temelden yani aşağıdan yukarıya yaklaşımlar sadece sistemin görevinden ibarettir. Yani bir uygulama yöntemini veya kontrol yapısını yoklayan bir yolu izlemezler. Aşağıdan yukarıya mühendislikte, görevler bir çeşit performans ölçeği (örneğin satranç oyunlarını kazanmak, Turing testini geçmek, bir odada tökezlemeden yürümek gibi) teori ile ilgisi olmayan şeylerle de belirtilebilir. Mühendisler için sistemlerin performansını, kriterleri aşacak şekilde ayarlamak için çeşitli deneme yanılma teknikleri mevcuttur. Etik anlamda, etiğe yönelik temelden bir yaklaşımın normatif değerlerini açıkça ifade etmekten ziyade, araçların eylemlerini örtük gören bir anlayış olarak değerlendirilmektedir. Burada amaç, ajanların kendi ahlakını ve başkalarının ahlakını doğru bir şekilde tanımlayabilmesi olarak düşünülmektedir. Otonom sistemlerin görevlerini esnek bir şekilde yerine getirebilmesi bir tür seçimle mümkün olduğuna vurgu yapılmaktadır. Bu seçimlerin bazıları da insanlar için potansiyel olarak zararlı sonuçlara yol açacağı sonucu çıkarılmaktadır. Nihayetinde halkın otonom sistemlerle hissedeceği rahatlık derecesi, bu sistemlerin insanlara ve diğer varlıklara zarar vermeyeceği, temel insani değerleri veya normları onurlandıracağı inancına bağlı kabul edilmektedir.

Karmaşık kararlar veren yapay sistemlerin tasarlanması olasılığına yönelik derin felsefi itirazların bu ilerleyişi yavaşlatabileceği olası görülmemektedir. Bu bağlamda güçlü yapay zekanın mümkün olup olmadığı hâlâ önemli bir tartışmayken, bu sistemlerin seçim yapma eyleminde bir dereceye kadar ahlaki hassasiyet gerektireceği ön görülmektedir. Kusursuz işleyen bir sisteme yönelik ahlaki yargı için, birçok farklı becerinin bütünleşmesine gereksinim duyulacağı belirtilmektedir. Yapay zekada ahlaki karar verme fakülteleri tepeden ve temelden yaklaşımlarla gerekçelendirilmeye çalışılarak bu zorlu girişimin ne kadar karmaşık olduğunu ima etme gayesindedir. Öyle ki insanların ahlaki yargı kapasitesi hem evrim ve öğrenme ile şekillenen aşağıdan yukarıya mekanizmaların hem de teori güdümlü akıl yürütme yeteneğine sahip yukarıdan aşağı mekanizmaların bir melezi olarak ifade edilmektedir. Buna bağlı olarak zeki robotların, aşağıdan yukarıya eğilimlerinin ve ayrık becerilerinin yukarıdan aşağı değerlendirilmesinin gerektiği vurgulanmaktadır. Nihayetinde er ya da geç farklı girdileri barındıran aşağıdan yukarıya sistemlerin dinamik ve esnek ahlakını korurken, seçimlerin ve eylemlerin değerlendirmesini karşılamaya yönelik yukarıdan aşağı ilkelere tabii tutulan sistemlere ihtiyaç duyulacağı belirtilmektedir. Ancak bu öğelerin nasıl bütünleştirileceğinin değerlendirilmesi mevcut çalışmanın kapsamı dışında olan ahlaki kurumun doğası ile ilgili önemli metaetik sorunlara yol açacağı da ön görülmektedir. Yine de bu girişimin zorluklarıyla

yüzleşmeye hazır olan bilim adamları tarafından derhal konuşlandırılabilir olacak önsezilerin ortaya konulmasının önemine dikkat çekilmektedir.<sup>266</sup>

Wallach ve Allen'in *Makine Ahlakı: İnsan Ahlakı Faktörlerinin Modellenmesinde Aşağıdan Yukarıya ve Yukarıdan Aşağıya Yaklaşımlar* çalışması yapay zeka etiği bağlamındaki önemli bir tartışma alanı olarak öne çıkmaktadır. Yapay zeka temelli sistemlerin etik değerleri ne ölçüde sağlayabileceğine ve ölçütü gerçekleştirme uğrunda iki yaklaşıma yer verilmiştir. İnsanın ahlaki yönlerinin temelden yani aşağıdan yukarıya deneyimsel gelişerek aynı zamanda tepeden yani yukardan aşağıya teori temelinde akıl yürütmeye gerçekleştiğinin önemine yer verilmiştir. Bu iki yönün bütünlüğü, şimdiye kadar yaptığımız sorgulamalarda öne çıkan zihin-beden birliği ya da ayrıklığını hatırlatmaktadır. Wallach ve Allen bu bütünlüğü savunuyor olacaklar ki elektriksel tepkimelerden ibaret sistemlerde de bunun her ikisinin de sağlanmasını zorunluluk olarak görmektedirler. Aksi halde, ahlaki değerlerin gerçekleştirildiği eylemlerin tam anlamıyla ortaya çıkarılamayacağı anlaşılmaktadır. Özetle bir sistemin tepeden bir yaklaşımla, temelden inşa edilen eylemleri teorik olarak değerlendirilmesi bu mücadelenin en önemli iki hususu olarak gösterilmektedir. Her iki yaklaşımı da gerçekleştirebilecek sistemlerin tasarlanmasının bir zorunluluk olduğu düşüncesiyle bilim insanlarının bu konu hakkındaki donanımlarını sergilemesi gerektiğinin önemine de vurgu yapılmıştır. Ancak ahlaki değerlerle başarılı bir sistemin inşa edilmesinin koşulunu zihinsel ve bedensel bütünlüğü olan insan ahlakına benzetmeyle mümkün olacağı iddiası bir anlamda kendini çürütmektedir. Kaldı ki çalışmamızın esaslı düşüncesi yapay zeka sistemlerinin arka planını Kartezyen düalizmin oluşturduğu, oysa insan zihninin Aristotelesçi zihin-beden birliğinde anlaşılabileceğidir.

### 3.2.4 Sonuççu Kuramlar

Sonuççu kuramlarda eylemlerin sonuçlarıyla değerlendirilmesi esastır. En iyi eylem gelecekte en iyi duruma yol açan eylem olarak kabul edilmektedir. Etik akıl yürütmeyi sonuççu düzlemde açıklamak amacıyla örneklenen bir robot; dünyadaki durumu tanımlama, olası eylemler üretebilme, mevcut durumda eylemin gerçekleşmesiyle ortaya çıkacak sonucu tahmin edebilme, bir durumu iyilik ya da arzu edilebilirlik açısından değerlendirebilme gibi özellikleri taşımaktadır. Robotun görevi mümkün olan en iyi duruma yol açacak eylemi bulmaktır. Bu sistemin “etik” bileşeni bir durumu iyilik veya arzu edilebilirlik açısından değerlendirme yöntemidir. Bu durumda arzu edilebilir bir durumu belirlemek nasıl mümkün olabilir? Genel olarak bu planlar, her bir kişi için bu durumlara karşılık olan zevk, mutluluk veya iyilik miktarının ölçülmesini ve daha sonra bu miktarların toplanmasını içerir. Bu planlardan en iyi

<sup>266</sup> Wallach and Allen, 2008: 565-582.

bilinenleri faydacılıktır. Bentham'ın XVIII. yüzyılın sonlarında önerdiği gibi, faydacılıkta ahlaki hareket, acıdan zevk için en büyük dengeyi üreten eylemdir. Bir eylemin iyiliğini ölçmek için sonuçlanacak duruma bakın ve her kişi için zevk ve acıyı özetleyin. Faydacılıkta her insan eşit kabul edilmektedir. Sonuççular kötülük karşısında en büyük iyilik dengesini kurmaya çalışmaktadırlar. Öne çıkan bir başka önemli soru da kimin (ya da neyin) kişi olarak sayılacağı sorunudur. Kimin refahına değer verileceği sorunudur. Tarih boyunca “kişi” fikrinin izini sürmek gerekirse: “Kadınlar kişi sayılır mı? Yabancılar kişi sayılır mı? Başka ülkelerden insanlar kişi sayılır mı? Başka ırklardan insanlar kişi sayılır mı? Dininize inanmayan insanlar kişi sayılır mı? Komada bulunan insanlar kişi sayılır mı? Fetüsler kişi sayılır mı?” gibi sorular öne çıkmıştır. Diğer bir problem de aşırı nüfus artışı olarak tartışılmaktadır. Yaşayan kişi sayısının artırılması değerlendirme formülünün değerini arttırıyorsa o zaman mümkün olduğunca çok insanın hayatta kalması arzu edilen bir kriter olarak kabul edilmektedir. Tabii ki başka bir kişinin dünyaya gelmesi, bu gezegendeki diğer insanların refahını azaltabilecek olasılığı düşünülmektedir. Dolayısıyla sonuççu hatlar ekseninde etik olarak mantık yürütme hedefindeki bir robotun olası eylemlerin listesini oluşturması ve daha sonra her eylemin neden olduğu durumu, eylemin insanlarda neden olduğu iyi veya kötü toplamına göre değerlendirmesi gerekir. Bu durumda robotun, dünyanın en çok fayda sağlayan eylemini seçeceği bildirilmektedir.<sup>267</sup>

Faydacılıkta her insanın eşit olduğu varsayımıyla çoğunluğun korunması esastır. Ancak sonuççu kuramda bir durum sonuç itibarıyla çıktılarına bakarak her bir birey açısından olumlu ya da olumsuz etkileri değerlendirilmektedir. Sonuççu görüşe göre tüm olumsuzluk karşısında en yüksek olumlu durum dengelenmelidir. Bu yönüyle faydacılıkta olduğu üzere çoğunluğun kurtarıldığı ancak azınlığın kaybedildiği anlayışın ötesinde her bir kişi açısından olumlu olumsuz durumların tespiti yapılarak genel bir dengenin kurulması amaçlanmaktadır. Bu bağlamda sayısal bir sistemin olası eylemleri sınıflandırması, bu eylemlerin etkilerini tespit etmesi ve insanlardaki yansımalarını ölçmesi hedeflenmektedir. Bu türden bir yapıda sayısal sistemin en makul eylemi tercih edeceği savunulmaktadır.

### 3.2.5 Deontolojik Kuramlar

Deontolojik etik teoride eylemler ürettikleri sonuçlara göre değil kendi içinde değerlendirilmektedir. Eylemlerin doğurabilecekleri sonuçlardan bağımsız olarak doğuştan ahlaki veya ahlak dışı olduğu düşünülebilir. Önerilen deontolojik ahlaki sistemlerden öne çıkan çağdaş bir deontolojik ahlaki sistem örneği Bernard Gert'in sistemi olarak belirtilmektedir. Gert

<sup>267</sup> Ford, Kenneth, Glymour ve Hayes, 1995: 243-252.

(1988) on ahlaki kural önermektedir: “1. Öldürme. 2. Acıya yol açma. 3. Sakatlama. 4. Özgürlükten yoksun bırakma. 5. Memnuniyetten yoksun bırakma. 6. Aldatma. 7. Sözünü tut. 8. Hile yapma. 9. Yasalara uy. 10. Görevini yap.” Çok kurallı bir sistem önerildiğinde kurallar arasında çatışma olasılığı hep vardır. Diyelim ki robot bir söz veriyor ancak sözün yerine getirilmesinin insanlarda acıya neden olabileceğini anlıyor. Robot sözünü tutmak zorunda mıdır? Bu tür durumlarda kural çatışması ile başa çıkmanın en etkin yaklaşımı, kuralları öncelik sırasına göre belirlemek olarak ifade edilmektedir. Asimov'un, *Üç Robotik Yasa* adlı eserinde kuralları doğrudan metnin içinde düzenlediği belirtilmektedir. Ahlaki sistemlerdeki çatışmalar sorunuyla başa çıkmanın yaygın bir yolu, kuralları birinci dereceden görevleri (*prima facie duties*) dayatmak anlamında kabul etmektir (Ross 1930). Bu yöntemle farklı kuralların kendi içinde ve karşılıklı olarak dengelenmesi sağlanmaktadır. Örneğin "Her durumda söz tutmak bir zorunluluktur. Geri kalan her şey doğrulandığında sözünü tutmalısın." gibi önceliklerin ya da koşulların belirlenmesiyle bu çatışmanın önüne geçilebileceği bildirilmektedir.<sup>268</sup>

Gerçek ahlaki çıkmazların mümkün olup olmadığına ilişkin güncel tartışmaların varlığına da dikkat çekilen makalede; bir kişinin bazı eylemleri yapmak zorunda olduğu ya da olmadığı birtakım durumlar var mıdır? Ya da her ikisini gerçekleştirmenin fiziksel olarak imkansız olduğu durumlarda iki eylemin her birini gerçekleştirebileceği durumlar var mıdır? Özü itibarıyla çözülemeyen kural çatışmaları var mıdır? (Gowans 1987).<sup>269</sup>

Faydacılar bazen faydacı hesaplamalar için deontolojik sistemlerin sırf buluşsal (*heuristic*) ve kestirme tahminlerden ibaret olduğunu iddia etmektedirler. Deontologlar, eylemlerin gerçek sonuçlarından bağımsız olarak doğuştan yanlış olabileceği iddiasıyla bunu inkar etmektedirler. Deontolojik ahlaki sistemin en eski örneklerinden bir tanesi "On Emir" olarak belirtilmektedir. Eski Ahit Tanrısı faydacı değildir. Tanrı “Zina yapmanın sonucu acıdan daha büyük bir zevk vermedikçe zina etmeyeceksin” yerine, zina eylemin doğası gereği ahlak dışı olduğu gerekçesiyle yasaklanmıştır.<sup>270</sup>

Deontolojik yani görev ahlakının belirleyici özelliği birtakım olmazsa olmaz kurallar dayatmasında robotik sistemlerin ahlaki davranışlarda bulunmasını sağlama gayesidir. Ancak kültürel yaşamın dinamikliği düşünüldüğünde prensip niteliğinde pek çok kurallar öne çıkabilmektedir. Bu durumda kuralların hangi durumlarda, hangi sırayla, hangi ilişkiye göre belirlenmesi deontolojik bir sorun olarak ortaya çıkmaktadır. Bu sorunun çözümünü sağlamak amacıyla bazı yöntemler sözü edilen düşünürlerin penceresinden açıklanmıştır. Gerek kuralların kapsamı gerekse sayısal bir sistemin çalışma mantığı düşünüldüğünde Hubert

<sup>268</sup> Gips, 2008: 248.

<sup>269</sup> Gips, 2008: 248.

<sup>270</sup> Gips, 2008: 248-249.

Dreyfus'un dediđi türden bir tıkanıklık mümkün görünmektedir. Şöyle ki Dreyfus'a göre sayısal sistemler sağduyudan yoksun oldukları için her sentaksı sırayla işleyerek yanıřları ve doğruları bu sıraya göre eleyerek vakit kaybedecektir. Yani işlenmesi gereken ne kadar çok satır varsa cevaplar o da derecede geç çağırılacaktır. Sağduyu sahibi olan insan içinse durum tam tersidir. İnsan ne kadar çok şey bilirse o derecede hızlı cevap verebilmektedir.

### 3.2.6 Psikolojik / Sosyolojik Yaklaşımlar

Ahlaki karar vermeye ilişkin psikolojik bulguları modellemek amacıyla çeřitli yapay zeka tekniklerini entegre eden MoralDM adında bir hesaplama modeli Dehghani, Tomai, Forbus ve Klenk'in (2008) ortak çalışmasında örneklenmiştir. Bu makalede, iki karar alma biçimini içeren faydacı ve deontolojik olan MoralDM adlı bilişsel temelde gerekçelendirilmiş bir ahlaki karar alma modeli açıklanmaktadır. Hataları azaltmak için, basitleştirilmiş İngilizcede yeniden oluşturulan psikolojik uyaranlardan biçimsel tasarımları üretmek amacıyla doğal bir dil anlama sistemi kullanılmıştır. Seküler ve kutsal değerlerin farklı etkileri, bir büyüklük gösterimi sırası kullanılarak nitel akıl yürütme yoluyla modellenmektedir. MoralDM, ahlaki karar alma kurallarını uygulamak ve daha önce alınan kararları kullanmak için ilk ilkeler muhakemesini ve analog muhakemeyi birleştirir. Ahlaki karar alma modeli MoralDM, birbirini dışlayan iki konum içerir: faydacı ve deontolojik. Analiz edilen vakada herhangi bir kutsal değer yoksa MoralDM en yüksek sonuç faydasını sağlayan eylemi seçerek geleneksel faydacı karar alma kurallarını uygular. Öte yandan MoralDM ilgili kutsal değerlerin varlığını tespit ederse deontolojik konumda çalışır ve eylemlerin sonuçlarına daha az duyarlı hale gelerek eyleme karşın eylemsizliği tercih etmektedir. Ahlaki karar verme teorilerinin, kültüre göre deđişen bağlamsal faktörlere dayandığı düşüncesiyle salt faydacı modellerin ötesine uzandığı bildirilen makalede MoralDM; psikolojik uyaranlardan biçimsel tasarımlar üreterek olası hataları azaltmak için doğal bir dil sistemini kullandığı ifade edilmektedir. Seküler ve kutsal değerlerin etkileri, bir tür büyüklük temsili kullanılarak nitel akıl yürütme yoluyla modellenir. MoralDM, ahlaki yargılarda bulunurken sonuçları ve faydaları belirlemek için ilk ilkeler muhakemesini ve analogik akıl yürütme kombinasyonundan faydalanmaktadır. Bu çalışma MoralDM'nin nasıl işlediğini, örnekleri toplayarak psikolojik sonuçları nasıl modellediğini ve performansını artırabileceğini göstermek üzerine kurgulanmıştır. Yapay zekadaki geleneksel karar verme modelleri faydacı teorilere odaklanmış olsa da bu teorilerin insana ait karar verme sürecine dair bütün yelpazeyi yakalayamadığına dair önemli psikolojik kanıtların varlığı bildirilmektedir. Özellikle ahlaki akıl yürütme üzerine yapılan arařtırmalar, normatif sonuçlar

ile sezgisel kararlar arasındaki çatışmayı ortaya çıkarmıştır. Bu durum, deontolojik ahlaki kuralların öne çıkmasına yol açmıştır.<sup>271</sup>

Baron ve Spranca'ya (1997) göre rasyonel yaklaşımların karar verme sürecinde baskın olduğu dönemlerin ardından, normatif sonuçlar ve sezgisel yargılar arasındaki çatışma, kutsal değerlerin öne çıkmasına neden olmuştur. Bu kutsal değerlerin, deontolojik ahlaki kuralları ortaya çıkararak faydacı motifleri engellediği bildirilmektedir. Yaşamın kutsal bir değer olduğu düşünüldüğünde insanlar genellikle yaşam kaybıyla sonuçlanacak bir eylemde bulunmayı reddetmektedirler. Dolayısıyla bu çalışmada yapay zekanın etik temellerinde en önemli yönlerin kutsal ve kültürel alanlar olduğu vurgulanmaktadır.<sup>272</sup>

Kutsal ve kültürel alanları da içerebilen böylesi bir yapay zeka örneğinin çalışma sistemine bakılacak olursa temelde büyüklük veya küçüklük değerlerine dayalı sayı sisteminde işlediği fark edilmektedir. Ayrıca bu türden kararları otomatik olarak almak için doğal dili anlama, nitel akıl yürütme, analog akıl yürütme ve öncül prensipler dahil olmak üzere bütünleşik bir yaklaşımı işletmesi hedeflenmektedir. Bu bağlamda MoralDM'nin işleyişine yönelik ifadeler şöyledir:

“Ahlaki karar verme araştırmasıyla desteklenen ilk prensipler akıl yürütme modülü, yardımcı yazılımlar, kutsal değerler ve eylem ile eylemsizlik arasındaki büyüklük ilişkisinin derecelerine göre kararlar alırken üç yöntem kullanılmaktadır. İlk olarak tercihi en yüksek fayda ile seçen faydacı yöntem, seçim kutsal bir değer içermediğinde çağrılır. İkincisi, kutsal değerleri olan ve sonuçlar arasında bir büyüklük sırası olmayan durumlarda, saf-deontolojik yöntem, kutsal bir değeri ihlal etmeyen seçimi tercih eder. Üçüncüsü, senaryo kutsal değerler ve sonuçlar arasındaki büyüklük sırasını içerdiğinde faydacı-deontolojik yöntem en yüksek fayda ile tercihi seçer. Bu nedenle saf-deontolojik yöntem, faydacı normları ihlal eden kararlar veren tek yöntemdir.”<sup>273</sup>

Sonuç olarak ahlaki bir karar alma modelinin nicelik duyarlılığını esas alan doğal bir dil mantığının gerekli olduğu görülmektedir. Ayrıca yeni durumlarla karşılaşılması halinde var olan depolanan kararlardan ya da örneklemelerden faydalanarak akıl yürütmesi gerektiği bildirilmektedir. Dolayısıyla insana ait ahlaki karar alma sürecini modellemek amacıyla çoklu yapay zeka tekniklerini entegre eden MoralDM, büyüklük temsili yoluyla kutsal ve seküler değerler arasındaki farklılıkları yakalayarak nitel modellemeyi kullandığı anlaşılmaktadır. Girişilen deneylerin neticesinde MoralDM araştırmacıları bu alanda ilerlenmesi gereken uzun yolun olduğunu düşünmektedirler. Bunlardan ilki; değerlendirme kurallarını, bilişsel psikologlar ve antropologlarla mevcut iş birliklerine dayanarak farklı kültürleri modellemek gereksinimi, ikincisi; çok çeşitli kültürel öykülerle başa çıkmak amacıyla kapsama alanının genişletilmesi,

<sup>271</sup> Dehghani, Tomai, Forbus, Klenk, 2008:1280.

<sup>272</sup> Dehghani, Tomai, Forbus, Klenk, 2008:1281.

<sup>273</sup> Dehghani, Tomai, Forbus, Klenk, 2008:1283.

üçüncüsü; ana akıl yürütmeyi hikaye kütüphanesi büyüdükçe daha ölçeklenebilir hale getirmek için bilişsel olarak akla yatkın bir benzerlik temelli geri alma modelinin eklenmesi olarak sıralanmıştır. Bütün bu öncüllerin yerine getirilmesinin ardından çok daha geniş bir yelpazede kültürel dinamikler temelinde testlerin yapılması gerektiği de vurgulanmaktadır.

Psikolojik ve sosyolojik yaklaşımlar başlığı altında örneklenen bir model aracılığıyla yapay zeka sistemlerinin kültürel ve kutsal değerler kapsamında etik eylemleri gerçekleştirme kapasitesi ölçülmüştür. Sunulan deneylerin sonuçları ve sistemin çalışma mantığı düşünüldüğünde John Searle'ün muhalifi olduğu güçlü yapay zeka projesine sunduğu eleştiriler akla gelmektedir. Kaldı ki sunulan model faydacı ve deontolojik kuralları eş zamanlı olarak işleterek kültürel ve kutsal dinamikleri gerçekleştirme hedefindedir. Faydacı yaklaşım bu sistemin gerçekleştirebilmesi çoğunluğun çıkarlarını korumak amacıyla zorunlu olarak azınlığın yıkımıyla sonuçlanacaktır. Ayrıca sayısal bir sistemin bu işlemi tamamen nicel olan büyüklük ya da küçüklük temsiliyle yerine getirmesi en küçük hatada çoğunluğun yıkımına neden olması olasıdır. Deontolojik açıdan düşünüldüğünde kültürel ve kutsal yaklaşımların sayısal bir sistem tarafından gerçekleştirilmesi için henüz can alıcı bir işaret olmadığı anlaşılmaktadır. Ayrıca araştırmacılar bu uğurda çok uzun bir yolun görüldüğü hususunda hemfikirdirler. Bütünüyle değerlendirmek gerekirse tek tek bireylerin izole bulunduğu bir ortamdaki, toplumun bir arada yaşadığı ortamlara dek yapay zekaya etik değerlerin nicel sembollerle yüklenmesi fikri, Aristotelesçi mutluluğa yaklaşma hususunda etik bir teorisinin algoritmada karşılık bulması olası görünmemektedir.

### 3.2.7 Sonuç

Tezin ilk bölümünde insana, ikinci bölümde nesne olan sayısal sistemlere ve üçüncü bölümde de bu ikisinin ilişkisini değerlendirmek amacıyla etik kuramlara yer verilmiştir. Etik kuramların incelenmesine gereksinim duyulmasının nedeni bir sonraki bölümde tartışılan sorunlara temel oluşturmaktır. Örneğin bir robotun ahlaki araç olarak değerlendirilmesinde etik yaklaşımların rolü, kitle imha silahlarının kötüye kullanılmasının önüne geçmektir. Öte yandan etik meselesine felsefi yaklaşımlardan bakışın nedeni kötülükten uzaklaşarak ve iyiliği teşvik edebilecek sistemleri inşa etme arzusudur. Bu başlık altında özellikle felsefi bir bakış olan faydacılığın referansında sistemlerin istendik yönde işlemlerini sağlama girişimleri incelenmiştir. Bu bağlamda sayısal sistemlerin etik düzlemde üretilmesinin yollarından temelden ve tepeden yaklaşım olarak iki bakış incelenmiştir. Temelden yaklaşım, sayısal sistemlere etik kuramların tasarım sürecinde entegre edilmesi düşüncesidir. Tepeden yaklaşım ise belirli genel bir etik teorisinin tüm sisteme uygulanabildiği ve alt sistemlerden istenen sonucu

elde etmek amacı olarak ifade edilmiştir. Öte yandan bir sistemin en iyi duruma yol açacak eylemi gerçekleştirmesi ve bu eylemlerin sonuçlara göre değerlendirildiği yaklaşım, sonuççu kuramlar başlığı altında incelenmiştir. Deontolojik kuramlar başlığında ise eylemler sonuçlardan bağımsız olarak kendi içinde ahlaki veya ahlak dışı olabilmektedir. Son olarak kültürel ve kutsal değerlerin gerçekleştirilmesinde etik yaklaşımların rolü yapay zeka modellerinde açıklanmaya çalışılmıştır.

Yer verilen etik kuramlar, son bölümde tartışılan sorunların önüne geçebilmek amacıyla sayısal sistemlerin istendik işlevler sergilemesini amaçlamaktadır. Bu nedenle disiplinler arası bir yaklaşımla etik kuramlar incelenmiştir. Etik kuramların bu sistemlere entegre edilmesi ve sonuçlarının doğrudan gözlemlenmesi için henüz erken bir çağda yaşamaktayız. Çünkü bu konuda hem etiğin hem de hukukun kısır kaldığı bazı noktalar fark edilmektedir. Sullins'in önerdiği yönelimsellik, özerklik, sorumluluk ve irade kavramlarının sayısal sistemlerdeki imkansızlığı bu türden bir kısırlıktır. Buna karşın yine de bu sistemlerin özne – nesne ilişkisini düzenleyebileceği, pek çok olumsuzluğun önüne geçebileceği ifade edilebilir. Anlaşılan o ki bu tür felsefi kavramlara bir makinenin, robotun ya da sayısal sistemin sahip olamayacağı endişesiyle konuyu felsefeden uzak bir zeminde şeklen incelemiştir. Sullins'e göre bir yapay zekanın ahlaki değerleri uygulayabilmesi onun ahlaki eylemi gerçekten yapmak yerine yapar gibi görünmesiyle yeterlidir. Bu düşünce, sayısal sistemlerin felsefi kavramlar karşısında neler yapamayacağını kanıtı olabilir. Kaldı ki Alan Turing'in iddiası tamamen zihni olan, kendi kendine düşünebilen, “yönelimsellik, özerklik, sorumluluk, irade” gibi insana ait pek çok özelliği yerine getirebilen sayısal sistemlerdi. Ancak makaledeki ikinci görüş olarak öne çıkan Bernhard Irrgang'ın (2006) iddiası felsefi açıdan desteklenmeye değerdir. Kaldı ki Irrgang “ahlaki açıdan sorumluluk taşımak için, kişilik veya öznellik ile karakterize edilen bir katılımcıya ihtiyacı vardır.” ifadesiyle doğrudan irade sahibi bir özneyi işaret etmektedir. Bunun yanında Bringsjord (2007) bu görüşü destekler nitelikte bir modeli uygulamalı olarak modellemiş, bir makinenin özerk olarak etik değerlere sahip olma yeterliliğinden yoksun olduğunu ispatlamaya çalışmıştır. Ayrıca Anderson ve Anderson'un özellikle bu iki görüşün felsefi temellerini görmezden gelmesi ve Floridi ve Sanders'ın soyutlamadan uzak mekaniğe yakın yaklaşımı yüceltmesi gayet manidardır. Kaldı ki Anderson ve Anderson tarafından tam olarak gerçekleştirilmeye çalışılan şey, imkansızlığı hissedilmesine rağmen makineye hakikati mekanik görünümü felsefi olabilecek etik değerleri yükleyebilme çabasıdır.

Makinelere etik değerler yükleyebilme çabasının deontolojik bağlamda Kant'ın Kategorik Buyruğuyla ilişkilendirilmesi teorik anlamda gayet ilgi çekici görünmektedir. Ancak ödev ahlakında insanın “zorunluluk” yerine sahip olması beklenen “gereklilik” anlayışı

makinelere uyarlanmaya çalışılmıştır. Hatta bu sistemin bir modeli dahi verilmiştir. Nitekim, olguların zaman, mekan, yönelim gibi dinamikleri bu türden ödev ahlakını insanlarla özgü bir husus olarak öne çıkarmaktadır. Kaldı ki evrensel bir ödev anlayışının insanlar için bile yaygınlaştırılması ve ahlak için ön koşul olan sarsılmaz bir yapının oluşturulması pek çok açıdan güçtür. Bunun gerçekleştirilmesi demek doğrudan kusursuz ahlaki bir yapının inşa edilebilmesi anlamına gelecektir. Ancak insan doğası düşünüldüğünde “kusursuzluk” sözcüğü boşa çıkmaktadır. Çünkü insan rasyonel olduğu kadar irrasyonel bir varlıktır. Bir makinenin gerçekleştiremeyeceği en önemli şeyler kusurları, eksiklikleri ya da belirsizliğidir. Diğer bir ifadeyle makinelere etik düşünceler yükleme girişimlerinden anlaşıldığı kadarıyla insan etiğinin, makine etiğine bir yönlendirilmesi ya da arasındaki farkın gölgelenme çabasıdır. Bu görüş doğrudan güçlü yapay zeka projesinin paraleli niteliğindedir. Bu hususa örnek teşkil edebilecek güçlü yapay zeka projesine Searle’ün getirdiği eleştiriler hatırlanmalıdır. Powers’ın makinelere uyarlamaya çalıştığı Kantçı ahlak anlayışını Searle açısından düşünürsek. Searle’ün bu konuya cevabı muhtemelen, Kant’ın ahlak anlayışını robotlar için değil insan nesli için ortaya attığı yönünde olacaktır. Bu görüşe cevaben; insansı robotların gelecekte başarılmasıyla Kant’ın ödev ahlakının robotlar açısından tartışılmasının gündeme gelebileceği iddia edilebilir. Ancak bu türden düşünebilen irade sahibi robot düşüncesi Searle’e göre olası değildir. Bu tür girişimlerin, insan doğasının irrasyonel yönü nedeniyle başarıya ulaşmasının mümkün olmadığı öne çıkmaktadır. Bu meseleyi değerlendirmek gerekirse Bernhard Irrgang’ın görüşlerine ihtiyaç duyarız. Bernhard Irrgang’ın iddiası bizim için önemli olmasının ayrıca bir nedeni daha bulunmaktadır. Irrgang’ın görüşlerini dayandırdığı “kişilik veya öznellik” kavramları önemlidir. Öznellik, özünde bir tür “keyfiyeti” barındırmaktadır. Yani makinenin tam anlamıyla etik bir araç olduğunu iddiasının temellendirilebilmesi için makinenin yapay zeka yazılımları aracılığıyla özerk bir şekilde ahlaki kararlar alabilmesi gerekmektedir. Ancak böylesi kararları sayısal sistemlerin alması güçtür. Bu başarılsa dahi ikici kısım olan özneliliğin algoritmaya sağlanması felsefi açıdan mümkün görünmemektedir.

*“Otuz yıl içinde insanüstü zeka yaratabilecek teknolojik araçlara sahip olacağız. Kısa bir süre sonra da insanlık dönemi sona erecek.” Vernor Vinge*

## **DÖRDÜNCÜ BÖLÜM**

### **TARTIŞMA**

Günümüzde insanlığın geleceğini belirleyen en önemli unsurlardan birisi şüphesiz yapay zekadır. Gelişen teknoloji özne-nesne temelinde düşünüldüğünde sayısal sistemlerin hızla ilerlediği, insan doğasını anlama gayretinin nispeten yavaş olduğu dikkat çekicidir. Bu nedenle tezin ilk bölümünde insan doğasının önemli unsurlarından zihne ve bileşenlerine yer verdik. Öte yandan ikinci bölümde yapay zeka tartışmalarına ve sayısal sistemlerin teknik yönlerine değindik. Bu sayede insanın referans alınmasıyla gelişimi belirlenen yapay zekanın ne olduğu veya ne olamadığının çerçevesini çizmeye çalıştık. Bu doğrultuda tezin asıl hedefi felsefi sonuçlara ışık tutacak olan yapay zekaya ilişkin etik kuramların sınıflandırmasına yer verdik. İnsan zihni, yapay zeka tartışmaları ve etik kuramlar incelendiğine göre bu noktada amacımız özne-nesne çerçevesinde bu teknolojilerin sosyolojik, psikolojik, hukuki, tıbbi, özel yaşam vb. alanlarda doğurabileceği sorunları incelemenin yeridir. Bu tartışma kapsamında Russell ve Norvig (2010) tarafından sınıflandırılan yapay zekanın etik ve risk faktörlerin çalışmamıza kılavuzluk edecektir. Fark edileceği üzere yukarıdaki alanlarda felsefi sonuçlar kapsama alınmamıştır. Zira felsefi sonuçlar kavramların anlaşılması, karşılaştırılması ve düşünce deneyi süzgecinden geçirilmesi neticesinde ortaya atılabilecek bir iddia ya da tezdur. Bu noktada öncelikle tartışma bölümünde çeşitli başlıklar değerlendirilecek, sondaki sonuç bölümünde felsefi olarak iddialar anlatılacaktır.

Yapay zeka sistemlerini geliştirme çalışmalarının mühendislikle ilgili olduğu, ancak böyle bir teknolojiyi kullanımına sunma izni ya da yasağı konunun etik, risk ve hukuki tarafıyla ilgilidir. Öyle ki nükleer bombalamalar, Çernobil faciası, çevre ve hava kirliliği gibi benzer kitlesel felaketler Yeni Çağ’ın tarihinde kaygı yaratan gelişmeler olarak bilinmektedir. Bu türden gelişmeler karşısında bilim insanlarının çalışmalarında nasıl davranmaları gerektiği, hangi projelerin yapılması veya yapılmaması gerektiği gibi hususların insanlığa etkisi anlamında ciddi sorumluluklar doğmaktadır. Bu sorumluluklara bir önceki bölümde yer vermiştik. Bu noktada otomasyon ya da makineleşmenin üst bir formu olan yapay zekanın

insanlığa olumsuz etkileri aşağıda sıralanmıştır. Bu sınıflandırma Russell ve Norvig<sup>274</sup> tarafından gerçekleştirilmiş olup, halihazırda yapay zeka etiğinin öne çıkan tartışma alanlarıdır.

- Otomasyon ve İşsizlik Sorunu
- Özgünlüğün Kaybolması Sorunu
- İstenmeyen Sonuçlar Sorunu
- Sorumluluk ve Gizlilik Sorunu
- İnsan Türünün Sona Erme Sorunu

Yer verilen başlıklar yapay zekanın ortaya koyabileceği genel sorunlar olup felsefi sonuçlar tezin sonucunda değerlendirilmiştir. Burada amaç tezin başlığında yer alan felsefi sonuçlara giden yolda karşılaşılan etik problemlere değinmektir. Öyle ki etik sorunlar felsefi sonuçlarla doğrudan ilgilidir. O halde listenin ilk başlıklarından otomasyonun iş alanlarını devralmasıyla insanların kaçınılmaz olarak yüzleşeceği işsizlik problemine değinelim. Bu sorunun felsefi bir sonuç olarak doğrudan “deneyimsellik” problemiyle ilgili olduğunu şimdiden belirtmenin yararı vardır.

#### 4.1 Otomasyon ve İşsizlik Sorunu

Bu çalışmada sayısal sistemlerin iş gücünü devralmasının herhangi bir kaygısı oluşturmadığına ilişkin görüş belirtilmiştir. Hatta yapay zekanın daha fazla ve daha yüksek ücretli işler yarattığı iddiası savunulmaktadır. Burada sözü edilen yapay zeka programları insana yardımcı olmak için tasarlanmış araçlar olduğundan, insanların yerini alacak şekilde tasarlanan "uzman sistemlere" nazaran daha masum kabul edilmektedir. Ayrıca mesleklerin kaybolması ya da yenilerinin ortaya çıkmasına bağlı olarak diğer bir husus da insanların daha çok ya da az boş vakit geçirebilmesi olarak ifade edilmektedir. Bu hususta herkesin üzerinde daha çok çalışma baskısı olacağı düşüncesiyle yapay zekanın teknolojik yenilenme hızını artıracığı ve otomatik araçların gereken işleri devralarak kendimize vakit ayırmamıza izin vereceği vurgulanmaktadır. Yani gelecekteki toplumların haftalık sadece dört saat çalışma imkanına otomasyon ve dış kaynak sayesinde sahip olabileceği vurgulanmaktadır.<sup>275</sup>

Yapay zeka insanların yaptığı işleri devralacak demek yerine yapay zeka insan görevlerinin yerini alacak ifadesinin doğru görüldüğü çalışmada, bazı mesleklerin basitliğine bağlı olarak tek bir göreve indirgenebileceği ve tamamen yok olma riski altında olduğu öngörülmektedir. Kamyon şoförleri ve otonom araçlar buna örnek gösterilmektedir. Öte yandan nükleer santralleri teftiş etmek veya açık deniz petrol kulelerini izlemek gibi tehlikeli işleri

<sup>274</sup> Russell ve Norvig, 2010: 1034-1040.

<sup>275</sup> Russell ve Norvig, 2010: 1034.

devralmak için makinelerin konuşlandırılması bu türden amaçlara hizmet edeceğinden insanların yerini alabileceği ifade edilmektedir. 2010 yılında Çin'in Shenzhen kentindeki Foxconn elektronik fabrikasında 14 işçinin bir pencereden atlamasına değinilerek bu tür işlerin ruhsuz makineler tarafından devralınmasına olumlu bakılmaktadır. Yapay zeka bazı işleri devralırken başka işleri ortaya çıkaracağı düşüncesiyle insanları verileri etiketleme konusunda makine öğrenimi için eğitmenin fayda sağlayabileceği önerilmektedir.<sup>276</sup>

Toplumsal bir sorun örneği olarak nitelendirebileceğimiz otomasyon ve işsizlik problemi gelecek nesillerin önemli kaygılarından biridir. Öyle ki olası bir işsizlik durumunda devletlerin vatandaşlarına asgari yaşam giderini sağlama zorunluluğu doğacaktır. Yani sayısal sistemlerin insan gücünü gerektiren işleri devralması zorunlu olarak işsizliği doğurabilecektir. Ancak uzun vadede bu sürecin devletler tarafından bir şekilde düzenlenmesi mümkün olabilecektir. Kaldı ki işlerin sayısal sistemler tarafından yapılması sonucu daha da özgürleşen insanın dilediği etkinlikleri yapabilmesi, kendini yaşayabilmesi anlamına gelecektir. Öte yandan işsizliğin maddi kaybın ötesinde diğer bir olumsuzluğu insanın sosyalleşme eksikliği ya da duygu paylaşım ortamından uzaklaşma olarak açıklanabilir. İnsan için çalışma ortamı sadece gelir kapısı olmayıp sosyalleşme, iletişimde bulunma, duygu ve düşünce paylaşımı gibi imkanları sağlamaktadır. Bu yönüyle insanlar için işsizlik, temel gelirin devlet tarafından sağlanmasına bakılmaksızın psikolojik, bedensel ve sosyolojik açılardan ciddi olumsuz etkilere neden olabilecektir. Zira insanın mesleğini icra etmesi sadece hayatın kazanılmasından ibaret olmadığı gibi, kişiye kendi doğasını yaşama imkanı tanıyan bir yön sunmaktadır. Bu yönlerden düşünüldüğünde yapay zekanın neden olacağı işsizlik sorunu, insanlara farklı alanlardan yüksek gelir imkanı tanısa da nihayetinde Aristotelesçi bir mutluluk anlayışından yoksun bırakacaktır.

#### 4.2 Özgünlüğün Kaybolması Sorunu

ELIZA programının yazarı Weizenbaum (1976), yapay zekanın toplum için yarattığı potansiyel tehditlerden bazılarına dikkat çekmiştir. Weizenbaum'un argümanlarından birisi yapay zeka geliştirme çalışmalarında insanların otomat olduğuna yönelik yaygın inancın temelsiz olduğudur. Bu fikir doğrudan özerklik kaybına ve hatta insanlığın özgünlük kaybına yol açan bir fikirdir. Bunun örneklerini tarihteki bazı önemli gelişmede görmek mümkündür. Örneğin, La Mettrie'nin (1748) *L'HOMME Makinesi*'nde bu türden bilginin olduğunu, en azından tekrar gündeme geldiğini belirtmektedir. Kopernik'in (1543) *De Revolutionibus Orbium Coelestium* eserinde Dünya'yı güneş sisteminin merkezinden uzaklaştırması insanlığın

<sup>276</sup> Mauro ve Valigi, 2020: 214 – 215.

benzersizliğini ortadan kaldıran bir gelişme olarak görülmektedir. Öte yandan Darwin (1871) tarafından insanın *homo sapiens*'e indirgenmesiyle diğer türlerle aynı seviyeye getirilmiştir. Bu bağlamda yapay zeka, eğer geniş çapta başarılı olursa en azından 21. yüzyıl toplumunun ahlaki varsayımlarını, Darwin'in evrim teorisinin 19. yüzyıllıklar için olduğu kadar tehdit edici olabilecektir.<sup>277</sup>

Özgünlüğün kaybolması sorunda öne çıkan fikir anlaşılacağı üzere insanın farklı türlerle eşit düzleme indirgenmesidir. Bunun detaylarına Weizenbaum'un (1976) *Computer Power and Human Reason* eserinde verdiği bilgilerle daha iyi anlamak mümkündür. yapay zekanın toplum için oluşturduğu potansiyel tehditlerden bazılarına dikkat çekmiştir. Oysa Russell ve Norvig yapay zekanın fikir, özerklik ve hatta insanlık gibi kavramları olumsuz etkilemeyeceğini düşünmektedirler. Weizenbaum'un ifadesine bakıldığında bilgisayarların adli kararlar verebileceği, psikiyatrik yargılarda bulunabileceği, hatta madeni paraları en sabırlı insandan daha sofistike şekilde çevirebileceği kanısı hakimdir. Ne var ki, ona göre asıl mesele sistemlere bu tür görevlerin verilmemesidir. Kaldı ki bazı durumlarda doğru kararlara varsalar bile bu türden görevlerin verilmesini onaylamaması gerektiğini düşünmektedir. Bilgisayarlar ve zihin üzerine gerçekleşen pek çok tartışmada Weizenbaum, bu konuların ne teknoloji ne de matematikle ilgili olmadığı dile getirmektedir. Konunun etik bir mevzu olduğu ve "olabilir mi?" ile başlayan sorularla çözülemeyeceğine dikkat çekmektedir. Neticede bilgisayarları akıllı hale getirmek için herhangi bir yolumuz olmadığına göre, bilgisayarların bilgelik gerektiren görevlerden muaf tutulması tavsiyesinde bulunmaktadır..<sup>278</sup>

Tarih boyunca insan neliğine ilişkin yapılan tespitlerden de anlaşılacağı üzere insanın özgünlüğünü farklı bir türe indirgeyerek ortadan kaldırma çabalarının son durağı yapay zeka olarak işaret edilebilir. Bu bağlamda Descartes düalizminin büyük etkisi vardır. O nedenle yapay zeka tartışmalarında insan daima referans noktası olmuştur. Sayısal bir sistemin akıllı, zeki, bilinçli, algılayan nitelikleri taşıyıp taşıyamayacağı sorunu zorunlu olarak insanın indirgemeci bir bakışa kurban edildiğinin ispatıdır. En geniş anlamıyla insan özellikle zihne ilişkin yönleriyle sayısal sistemlerin kıyasına sunulmaktadır. Bu bağlamda yapay da olsa insan benzeri otomatların ya da sistemlerin varlığı insanın biricikliğini zedeleyebilecek gelişmeler olarak değerlendirilmelidir. Nihayetinde algoritmanın kusursuzlaştığı bir çağda her ne kadar irade, bilinç, yönelimsellik gibi niteliklerden yapay zeka uzak olsa da insan özgünlüğünü en azından kavramsal düzeyde olumsuz etkileyebilecektir.

<sup>277</sup> Russell ve Norvig, 2010: 1035.

<sup>278</sup> Weizenbaum, 1976: 227.

### 4.3 İstenmeyen Sonuçlar Sorunu

İleri teknolojilerin, güçlüler tarafından rakiplerini bastırmak amacıyla kullanıldığı bilinen bir gerçektir. Hardy'nin (1940) ifadesinde belirttiği gibi, "Bir bilimin, zenginlik dağılımındaki mevcut eşitsizlikleri belirginleştirmesi veya daha açık bir ifadeyle insan yaşamının yok edilmesini teşvik etmesiyle yararlı kabul edildiği söylenir." Bu hem yapay zeka hem de tüm bilimler için geçerlidir. Otonom yapay zeka sistemlerinin yaygınlaşmasıyla dünyanın çeşitli bölgelerinde kullanılmaya başlandığının bilgisi verilmektedir. Örneğin, Singer (2009) ABD ordusunun Irak'ta 5.000'den fazla otonom uçak ve 12.000 otonom kara aracı konuşlandırıldığının bilgisini vermektedir. Bu tür hamlelerin insanın insanla olan savaşına kıyasla, robotun insanla olan savaşında felaketler yaratabileceğinin işareti verilmektedir. İnsan karar verme sürecinin çıkarıldığı takdirde robotların masum sivillerin öldürülmesine yol açacak kararlar alabileceği ön görüşünde bulunmaktadır. Öte yandan güçlü robotlara sahip olmak uluslara aşırı güven verebilir ve gerekenden daha cüretkar savaş girişimlerine neden olabilmektedir. İnsanın insanla olan savaşında cüretkar girişimlerden ziyade barışçıl çözüm çoğu zaman her iki taraf için de olumlu bir gelişmedir. Kaldı ki, savaşın kazananı olmamaktadır. O nedenle, yapay zeka destekli otonom sistemlere duyulan aşırı güven çoğu zaman toplumların tahribatına neden olabilmektedir.<sup>279</sup>

İstenmeyen sonuçlar başlığı altında değerlendirilmesi gereken bir konu da etnik ve dilsel ayrımcılıktır. Yapay zekanın doğal dili anlama konusunda yetersizliğine ilişkin bir tespit algoritma geliştiricileri Mauro ve Valigi (2020) tarafından yapılan çalışmada sunulmuştur. Burada sözü geçen problem algoritmanın farklı aksan modellerinin seslerini tanımada sorunudur. Bu algoritmaların göçmenin bir kişinin aksanlı İngilizcesini anlayamamasıyla ayrımcılığa uğrayabilecekleri ve bilgiye erişimlerinin ortadan kalkacağı değerlendirilmektedir. Öte yandan insan kaynaklarının işe alma kararlarının algoritmayla geliştirildiği bir örnekte kadın erkek eşitsizliğinin ortaya çıktığı anlatılmaktadır. Diğer bir örnekte de makine öğrenimi temelli sistemde sözde sanığın gelecekte tekrar suç işleyeceği riskini tahmin eden ve olasılığı daha yüksek olanlara daha sert cezalar öneren bir algoritma değerlendirilmiştir. Karar verme mekanizması daha önceki yargı kararlarına dayandığı için bu sistemin siyahi ve renkli insanlara karşı ön yargılı olmasının muhtemel olduğu düşünülmektedir.<sup>280</sup>

Bishop (2022) tarafından kaleme alınan *Yapay Zeka Aptaldır ve Nedensel Akıl Yürütme Bunun Üstesinden Gelemez* başlıklı makalede disiplinler arası bir yaklaşımla yapay zekanın ırkçılığına ilişkin örnekler verilmiştir. Örneğin, Microsoft tarafından geliştirilen sohbet robotu

<sup>279</sup> Russell ve Norvig, 2010: 1035.

<sup>280</sup> Mauro ve Valigi, 2020: 210 - 214.

Tay'ın Adolf Hitler'den övgüyle söz etmesi bunlardan biridir. Tay'ın işlediği sözcüklerin anlamını gerçekten anlayamadığının tespiti yapay sinir ağlarının sınırlılığını da ortaya çıkaran bir gelişmedir. Ayrıca Tay'ın kalıpları tanımasını ve ilişkilendirmesini sağlarken uyguladığı algoritmalar Tay'a bir epistemoloji veremediğinin de altı çizilmektedir. Yani Tay isimleri, fiilleri, zarfları ve sıfatları tanımlayabilmiş, ancak 'Hitler'in kim olduğu' veya 'soykırım'ın gerçekte ne anlama geldiği' hakkında hiçbir fikir edinememiştir.<sup>281</sup> Ayrıca yapay zeka algoritmalarına ve sentaktik kısıtlılara değinen Bishop, insan zihninin teleolojisini açıklamakta yapay zekanın yetersiz kalacağını düşünmektedir.

Yukarıdaki örneklerden anlaşılacağı üzere teknolojinin hem ayırıcı hem de yıkıcı etkilerine değinilmiştir. Öte yandan karmaşık yapay zeka sistemlerini elinde bulunduran güçlerin beklenmedik felaketlere neden olabileceği potansiyele dikkat çekilmiştir. Kaldı ki nükleer füzelerden, biyolojik salgınlardan çok daha fazla olumsuzluklara neden olabilecektir. Bu yönüyle bu tür teknolojilerin geliştirilmesi bir yana, nasıl ve kimler tarafından kontrol edilmesi gerektiği ayrı bir etik mevzudur. Giriş bölümünde yer verdiğimiz ilaç geliştirme masadıyla modellenen yapay zeka sistemini hatırlayalım. Bu sitemde farklı bir algoritmanın çalıştırılmasıyla 6 saatten daha az sürede 40.000'den fazla biyolojik silah niteliğinde molekül sentezi sunulmuştur. Böylesine çarpıcı bir örnek etik çalışmaların özellikle yapay zeka sistemlerini kontrol eden kişi, kurum ya da kuruluşlar açısından ne derece önemli olduğunu ortaya koymaktadır.

#### 4.4 Sorumluluk ve Gizlilik Sorunu

Amerika Birleşik Devletleri'nde yapay zeka sistemlerinin kullanımına ilişkin yasal sorumluluğun önemli hale gelmesi tıp uzmanı sistemler üzerinden örneklendirilmektedir. Bir doktor teşhis için uzman sistemin kararına güvenmesi durumunda teşhis yanlışsa kim suçludur? Neyse ki hekimin faydacı yaklaşım çerçevesinde tıbbi prosedürleri uygulamasıyla doktorun sorumlu tutulamayacağı kabul edilmektedir. Bu nedenle soru şu olmalıdır: "Teşhisin mantığa aykırı olması durumunda kim suçludur?" Bu hususta mahkemeler, tıbbi uzman sistemlerinin rolünü tıbbi ders kitaplarıyla eşit kabul etmektedir. Dolayısıyla doktorların herhangi bir kararın arkasındaki mantığı anlamaktan ve sistemin tavsiyelerini kabul edip etmeyeceklerine karar vermede kendi kararlarını kullanmaktan sorumlu olduğu kanaati yaygındır. Bu nedenle tıbbi uzman sistemlerin tasarımı hastadan çok hekimin davranışını nasıl etkileyeceğine odaklanmalıdır. Uzman sistemler insan teşhisçilerinden daha güvenilir hale gelmesi durumunda doktorların uzman bir sistemin tavsiyelerini kullanmamaları halinde yasal olarak sorumlu

<sup>281</sup> Bishop, 2021: 2-17.

olabileceklerdir. Bu önermeyi araştıran Gawande (2002) “Belki makineler karar verebilir ama yine de iyileşmek için doktorlara ihtiyacımız var.”<sup>282</sup> ifadesiyle değerlendirmektedir.

Yapay zeka araçlarının sorumluluğuna ilişkin diğer bir sorun da mali işlemlerin “insan adına” yapılması durumunda ortaya çıkan borçlardan kimin sorumlu olacağı ve bu tür araçların varlık edinebilmesi, kendi adına elektronik işlemler yapabilmesi gibi hususlar da tartışılan diğer konulardır. Sorumluluk hususunda zeki araçlara finansal işlemlerde herhangi bir yasal statü verilmediği, bu tür yazılımların karayollarında “sürücüler” olarak kabul edilmediğinin bilgisi de verilmektedir. Russell ve Norvig’e göre bu tür faaliyetleri gerçekleştirebilecek sistemlerin hukuki sorumluluğu için yoğun çalışmalara ve çok zamana ihtiyaç duyulmaktadır.<sup>283</sup>

Öte yandan, gizlilik konusunda Weizenbaum’un konuşma-tanıma teknolojisi hakkındaki görüşleri öne çıkmaktadır. Konuşma-tanıma sistemleri işleyiş itibarıyla düşüncelerin niceliksel sembollere dökülmesi anlamında geldiğinden bir takım gizlilik ihlaline ilişkin kaygılar oluşturmaktadır.<sup>284</sup> Öyle ki bu teknolojinin kötüye kullanılması mahremiyet kaybına yol açabileceği, küresel çapta devletlerin anlaşmazlığa düşebileceği gibi kaygılara neden olmaktadır.

Sorumluluk ve gizlilik yapay zeka sistemlerinin neden olacağı sorunların başında gelmektedir. Kaldı ki irade taşımayan bir sistemin, irade sahibi insana ilişkin ve onun yerine karar vermesinin bazı sonuçları vardır. Bunlar özetle özel hayatın ihlali, yaşanabilecek felaketlerden kimin sorumlu tutulacağı gibi sorunlar olarak ifade edilmiştir. Bu çerçevede sayısal sistemler sınırları aşarak bazı gizli bilgilerin ya da mahremiyet ilgilendiren hususların ihlaline sebep olabilmektedir. Bu türden bir sorun sadece bireysel insanın mahremiyetine saldırı olmayıp, özellikle topluma ve devlete ait gizli bilgilerin bu sistemleri yönetenler tarafından kötüye kullanılması anlamına gelecektir. Nihayetinde pek çok masum insanın etkilenebileceği savaflara neden olabilecektir. Bu bağlamda yapay zeka konusunda karşılaşılabilecek problemlerden birisi de sorumluluk ve gizlilik hakkındadır. Bu sorunun insanlık açısından ciddi problemler doğuracağını ifade etmek abartı olmayacaktır. Dolayısıyla bu tür sistemlerin yasal sorumluluklar açısından düzenlenmesi için önemli çalışmalara ve uzun vadeye ihtiyaç duyulduğu anlaşılmaktadır.

<sup>282</sup> Gawande, 2002: 46.

<sup>283</sup> Russell ve Norvig, 2010: 1036.

<sup>284</sup> Weizenbaum, 1976: 270-273.

#### 4.5 İnsan Türünün Sona Erme Sorunu

Yapay zekanın doğurabileceği sorunlardan insan türünün yok olması sorunu öne çıkan bir problem olarak tartışılmaktadır. Fakat bazı görüşlere göre algoritmanın doğası gereği yapay zeka hiçbir zaman insan türünü tehlikeye sokamayacaktır. Bu görüş çerçevesinde Bishop, yapay zekanın şimdiye dek ortaya koyduğu ilerlemeyi hem kısır görmekte hem de insanlığın sonunu getirmek adına doğrudan bir etkisinin olmayacağı kanaatini taşımaktadır. Söz konusu birkaç yoruma yer vermek gerekirse Turing'in iddiasından bu yana yetmiş yıldan fazla zaman geçmesine rağmen sinir ağları sayesinde sürücüsüz araçlardan "Go" oyununa kadar çeşitli alanlarda başarıya ulaşan yapay zekanın semantik düzeyde bir anlam bilgisine ulaşmadığının altını çizmektedir. Öte yandan yapay zekanın doğuracağı problemleri sıralayan Bishop, yapay zeka sohbet robotu Siri'den alışveriş listesine "bir litre kitap" eklemesi istenince; kitapların ve sıvıların ontolojisini anlamadığını delil göstermektedir. Dahası yapay zeka aygıtlarının, gündelik dilin nüanslarını anlamada feci şekilde başarısız olan Apple ile sınırlı bir sorun olmadığını belirtmektedir.<sup>285</sup> Ne var ki yapay zekanın insanlığı kökten etkileyebileceğine ilişkin görüş bildiren araştırmacılar ve düşünürlerin sayısı epey fazladır.

İlk olarak yapay zeka sisteminin yanlış tahmin sonucunda yanlış olan şeyi yapmasıyla ortaya çıkabilecek felaketlere bakıldığında örneğin otonom bir otomobilin hava ve yol şartlarını hatalı tespit etmesi ölümlere neden olabilecektir. İnsanlığı tehdit edebilecek daha ciddi bir örnek de füze savunma sistemindeki olası hata nedeniyle milyarlarca insanın ölmesi ve neslin tükenmesi olasılığıdır. İkinci olarak bir yapay zeka sistemini en üst düzeye çıkarmak için doğru işlevi belirlemenin güç olduğu tartışılmaktadır. Yani insanların acılarını dindirebilecek bir algoritma geliştirilmesi durumunda yapay zekanın insanlığın kökünü kazıyacağı öngörülmektedir. Çünkü insan, doğası gereği en rahat koşulda dahi acıyı da arzulayan bir varlıktır. Bu durumda algoritma, insanlığın sonunu getirerek acıları dindirecektir. Bu sebeplerle yapay zeka tasarımının çok önemli olduğunun altı çizilmektedir. Kaygı verici olası sonuçlar dolayısıyla bu sistemlerinden tam olarak ne beklenildiği önemle vurgulanmaktadır. Ayrıca insan türünün nasıl sona erdirileceğini anlayacak kadar zeki bir robotun, bu fonksiyonun gerçek maksadının katliam yapmak olmadığını anlayacak kadar zeki olmasının da umulabileceği belirtilmektedir.<sup>286</sup>

Yapay zeka tekniklerinden makine öğrenmesinin üst düzey seviyelere ulaşmasıyla doğabilecek bir soruna dikkat çekilmektedir. Bu da Good'un (1965) insana dair bütün akılsal faaliyetlere sahip ultra zeki bir sistemin insandan çok daha karmaşık sistemler üretebileceği

<sup>285</sup> Bishop, 2021: 2-17.

<sup>286</sup> Russell ve Norvig, 2010: 1037.

düşüncesidir. Bu türden bir faaliyetin “zeka patlaması” kavramına neden olacağı gerekçesiyle insan zekasının geride kalacağı vurgulanmaktadır. Dolayısıyla makinenin bize onu nasıl kontrol altında tutacağımızı söyleyecek kadar uysal olması koşuluyla ilk ultra zeki makinenin, insanın yapması gereken son icat olduğu iddia edilmektedir.<sup>287</sup> Bu bağlamda Vinge (1993) şu ifadeyi kullanmıştır. “Otuz yıl içinde insanüstü zeka yaratabilecek teknolojik araçlara sahip olacağız. Kısa bir süre sonra da insanlık dönemi sona erecek.”<sup>288</sup>

Diğer yandan robotik ve biyoteknolojik icatların birleştirildiği veya değiştirildiği bir gelecek olarak ifade edilen transhümanizmin, insan yaşamını ve türünü korumayı amaçlayan çoğu ahlak teorisyenini zorlayacağı ifade edilmektedir. Söz konusu zorluk, tekillik görüşünün savunucusu Ray Kurzweil’in *The Singularity Is Near: When Humans Transcend Biology* eserindeki ifadelerine dayandırılmaktadır. Tekilliğin biyolojik bedenin ve beynin sınırlarını aşmaya imkan tanıyacağını iddia eden Kurzweil; insanın kaderi üzerinde güç kazanacağı ve ölüm hakkında kontrol sahibi olacağı iddiasında bulunmaktadır. İnsanın düşüncesinin tam olarak anlaşılacağını ve zekanın biyolojik olmayan kısmının trilyonlarca kat daha güçlü hale getirileceğini müjdelemektedir.<sup>289</sup>

Yudkowsky (2008), insanlara zarar vermeme arzusuyla donatılmış bir yapay zeka tasarımının imkanından söz etmektedir. Yapay zeka sistemlerini geliştirme sürecinde dostluk yaklaşımının tasarım sürecinin en başında benimsenmesi gerektiğini düşünmektedir. Yani yapay zeka sistemlerinin eğitilmesi için bir mekanizma tanımlamak ve süreçteki değişiklikler karşısında yapay zekanın dostça davranışta bulunacağı faydalı işlerle eğitilmesi esasına dayanmaktadır.<sup>290</sup>

Satranç oynayabilen yapay zekanın bile toplum için oluşturabileceği risklere değinen Omohundro (2008), bu sistemlerin yapılarındaki değişikliklerle amaçları dışında hareket edebileceği ve zararlı işlere kalkışabileceğini ifade etmektedir. Gelişmiş yapay zeka mimarilerinin kendi kendilerini değiştirebilmek, dahili ölçüm sistemlerinin değiştirilmesine engel olmak ve herhangi bir kısa devreyi önlemek gibi isteklerde bulunabileceği savunulmaktadır. Alt bileşenlere ihtiyaç duyacakları, aynı zamanda tutarlı ve uyumlu bir bütün olarak hareket etmek isteyecekleri bildirilmektedir. Kendi iç “polis güçlerine” veya “bağışlılık sistemlerine” ihtiyaç duyacak olmaları bu tür sistemlerin geleceğini işaret etmektedir. Omohundro, bu konuların daha derin anlaşılmasının insan ruhunun yapısına bile ışık tutabileceği kanısındadır. Sonuç olarak zeki araçların kendisini tasarlayanın yanı sıra işlev

<sup>287</sup> Good, 1965: 33.

<sup>288</sup> Vinge, 1993: 11.

<sup>289</sup> Kurzweil, 2005: 9.

<sup>290</sup> Yudkowsky, 2008: 312, 313.

göreceklere sosyal bağlamı da tasarlamanın önemine dikkat çeken Omohundro, bireylerin temel haklarını belirleyen bir "evrensel anayasa" tasarlama fikrini öne sürmüştür.<sup>291</sup>

#### 4.6 Sonuç

Yapay zeka gibi yeni teknolojilerin sosyal bilimlerde, popüler kültürde ve sinemada görülebileceği üzere potansiyel etkilerine yönelik ilgiler artmaktadır. Bu teknolojinin geleceği, insan yaşamında ve faaliyetinde bir takım önemli değişikliklere yol açacaktır. Çeşitli yapay zeka teknikleri aracılığıyla gerçekleştirilecek olan bu süreç, günlük hayatı ve insan eylemlerini tartışma bölümündeki işsizlik, özgünlük, ayrımcılık, sorumluluk, gizlilik ve neslin yok olması başlıkları altında değerlendirilmiştir. Bu sorunlardan bazıları sayısal sistemlerin tasarım süreciyle, bazıları da insanın kötüye kullanım olasılığıyla ilgilidir. Birinci durumda algoritmaların ön yargıyla programlamasına bağlı olarak sistemlerin etik davranıştan yoksun olmasından kaynaklanan problemlere yer verilmiştir. İkinci durumda ise bu sistemleri elinde bulunduranların kötüye kullanılmasıyla kitle imha silahlarına dönüşmesi ve insan neslinin tehlikeye girmesi olasılığı değerlendirilmiştir. Her iki durumda da hem öznenin hem de nesnenin hatalı işleminden sorumlu olduğu anlaşılmaktadır. O halde yapay zeka sistemlerinin insana kıyasla neden olacağı kaygının temel nedenini irdelemek gerekmektedir. Şöyle ki sayısal sistemler yüksek güvenlik, hız ve hassaslık sağlayan yapılar olması nedeniyle işlemleri insandan çok daha hızlı ve güvenilir şekilde gerçekleştirmektedir. İşte bu kapasitenin ufak bir hataya düşmesi kitlelerin etkilemesine neden olacaktır. Oysa hata ya da etik dışı davranış durumunda insanın etki alanı sınırlı çevreyi, sınırlı sayıda canlıyı kapsayacaktır. Öte yandan yapay zeka durumunda kitlelerin hatta küresel olarak insanlığın etkilenebileceği sonuçlarla karşılaşmak mümkün olabilir. O halde yapay zekanın olumsuz sonuçlarının temelinde yüksek hız, güvenlik ve hassasiyet olduğu anlaşılmaktadır. Söz konusu ifade, insan bedeninin çevresel aygıtlarla geliştirilmesine olarak bilinen transhümanizmle karıştırılmamalıdır. Burada herhangi bir yapıdan bağımsız olan doğal insan referans alınmaktadır.

Bu çerçevede ortaya çıkan sonuç yapay zekanın geleceğe ilişkin kitlesel sonuçlarının olacağıdır. Bu sonuçlardan birisi yapay zekanın yazılımında gerçekleştirilecek bir değişiklik beklenen işlevin dışına çıkması ve olumsuz sonuçlara neden olması olabilir. Örneğin, sürü zekası ya da sürü robotiği olarak bilinen teknolojinin Omohundro'nun iddia ettiği türden yazılım değişikliğiyle örgütlenmesi, kendi alt birimlerini gruplandırmasıyla saldırıya geçmesi teknik açıdan mümkün görünmektedir. Böyle bir durumda kontrolden çıkan sürü karşısında insanın bu duruma cevabı yetersiz kalabilir. Bu tür sonuçlar karşısında insanın kontrol gücünün

<sup>291</sup> Omohundro, 2008: 489-491.

yetersizliğinin yanında bu sistemlerin gerçek kişi olmadıkları için hukuki sorumluluktan yoksun da ayrı bir problemdir. Yani yasal sorumluluğu olan insan için yasalar çoğu zaman caydırıcı olabilirken bu sistemler en azından yakın gelecekte hukuki bağlayıcılıktan yoksun olacaktır. Hukuki sorumluluğun irade sahibi gerçek kişiyi kapsadığı düşünüldüğünde irade sahibi insanın da bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik gibi zihinsel niteliklere sahip olduğu anlaşılmaktadır. Sayısal sistemlerde bu niteliklerin olanağına bakıldığında yakın gelecekte yasal sorumluluk alabilen irade sahibi sistemlerin olası görünmediği ortaya çıkmaktadır.

Yapay zekanın toplumsal sonuçlarının yanında konun felsefi bakışla değerlendirilmesi insanın makineden bağımsız düşünülmediği bir durumu ortaya çıkmaktadır. Bu noktada öznenin nesneyle ilişkisinden, nesnenin özneyle ilişkisine evrilen bir sıraya dönüşür. Öyle ki bu sıranın değişmesi insanın madde karşısında arka plana atıldığı bir anlayışı ima eder. Yani söz konusu ileri teknolojilerin toplumsal açıdan yıkıcı sonuçlarının ötesinde insanın özneliği tartışmalı hale gelecektir. Öz bir ifadeyle doğanın bahsettiği dolaysız, yalın ve doğal eylemlerin sıradan kabul edildiği çağda, yapay zeka gibi suni gelişmelerin işlevlerine duyulan hayranlık doruk noktasına ulaşacaktır. Bu da insanın zorunlu olarak göz ardı edilmesine, öznenin ancak nesne güdümünde anlaşıldığı bir anlayışı doğuracaktır. Bu tür bir gelişmeye imkan vermemenin en önemli şartı, insan aklının ürünü olan bu sistemlerin yalnızca mühendislik faaliyeti olarak kabul edilmemesidir. Kaldı ki neyin icat edilmesi ya da edilmemesi gerekliliği felsefi söylemde etiğin konusudur. O nedenle yapay zeka potansiyel tehlikelere gebe bir alanın felsefi temellerinin tartışılması gereklidir. Felsefi temellerden kasıt bakış açısıdır. Felsefi açıdan yapay zekaya yaklaşım, Searle'ün önerdiği gibi zayıf yapay zeka türünde olmalıdır. Bu bakışa göre insan, zayıf yapay zeka sistemlerini yönlendirebilen konumdadır. Oysa güçlü yapay zeka, kendi kendine karar verebilen yani gerçek zihinleri olan sistemlerdir. Bu durumda güçlü yapay zekanın kabulü zorunlu olarak insanın nesne karşısında geri plana itilmesine neden olacaktır. Zayıf yapay zeka ifadesinin kavram olarak benimsenmesi ve sistemlerin tasarımının bu bakışa göre şekillendirilmesi, yapay zeka karşısında insanı yüceltecektir. Yer verdiğimiz yorumlar güçlü yapay zekanın gerçekleşip gerçekleşmediğinden bağımsızdır. Şöyle ki bu türden bir teknolojinin insan aklının sınırlarını geçebileceği ve felaketlere neden olacağı söylemi dahi insanlar tarafında korku ve endişeye neden olacaktır. Bir yönüyle amaçlananın da bu olduğunu ifade etmek yanlış olmayacaktır. Bu sayede küresel çapta gücünü korumaya ve sürdürmeye çalışan oluşumların kavramsal zafer peşinde koşması söz konusudur.

## SONUÇ

Yapay zeka meselesinde belli başlı görüşlerin incelendiği ve temel argümanların belirlendiği çalışmaya zihninin ortak kavramlarıyla başlanmış, tarihte yapay zeka tartışmalarıyla devam edilmiştir. Bu sıra insanın nesne karşısında öne çıkarıldığı bir anlayışı ima etmektedir. Dolayısıyla her şeyden önce insana özgü kavramları incelemek, ardından bu kavramların yapay zekadaki karşılığını tartışmak amaçlanmıştır. Yapay zeka meselesinde kavramsal muğlaklığın başat problem olduğu düşünüldüğünde bu sorunların giderilmesi, etik sorunları somutlaştırmayı ve olası çözüm önerilerini sunmayı hedeflemektedir. Çünkü yapılan incelemede ilk olarak akıl ve zeka terimlerinin benzer kullanımına ilişkin sorunlar tespit edilmiştir. Bu tespit insanın aklından ziyade zekasıyla öne çıkarıldığı hatta insan doğasının zeka üzerinden belirlendiği bir anlayışı doğrulamıştır. Bu gelişmenin temel nedeni yapay zeka çalışmalarında karşılaşılan engellerin araştırmacıları farklı arayışlara zorlamasıdır. Şöyle ki yapay zekanın hayatımızda olumlu bir güç olarak kullanılması için üstesinden gelinmesi gereken zorluklara tartışma bölümünde yer vermiştik. Bu zorlukların çoğu sistemlerin altyapısından kaynaklanmaktadır. Tezin genelinde yer verildiği üzere başlıca sentaks ve semantik problemi, büyük miktarda verinin saklanması problemi, enerji tüketimi problemi, kritik görevleri gerçekleştirebilen sistemler problemi, farklı iklim koşullarına ayak uydurma problemi gibi hususlar öne çıkmıştı. Bu engellerin üstesinden gelebilmenin uzun yıllar gerektireceğinin fark edilmesiyle farklı bir çözüm yoluna gidildiği anlaşılmaktadır. Bu çözüm yolu da insana özgü kavramların sayısal sistemlere indirgenmesiyle, sayısal sistemlerin insanlaştırılmasıdır. Daha açık bir ifadeyle sayısal sistemlerin insana özgü nitelikleri gerçekleştirdiğinin ispatını kavramsal düzlemde gerçekleştirme arzusu fark edilmektedir. Kısacası sayısal sistemlerle hedeflerin gerçekleştirilememesi, kavramların gerçek anlamının dışında ve farklı bağlamlarda kullanılmasına neden olmuştur. Bu da etik problemlerin kaynağını oluşturmaktadır. Akıl ve zeka başta olmak üzere zihninin ortak kavramlarının sayısal sistemlere atfedilmesinin felsefi sonuçlara etkisi şu açıdan önemlidir. Sayısal sistemlerin tasarımı ve geliştirilmesi mühendislik faaliyetinin ötesinde özne-nesne ilişkisi bağlamında felsefi söylemi gerekli kılmaktadır. O nedenle yapay zekanın teknik anlamda modellenmesi mühendislik açısından büyük bir başarı olarak görülürken felsefi açıdan insanlığa etkilerinin olumsuz olacağı çıkarımı yapılabilir. Bu çıkarım olumsuz etkilerin teşhis edilmesinden ziyade önlenmesine yönelik bir tutum olmalıdır. Çünkü bu tür sistemlerin sonuçlarını teşhis etmekten çok geç olabilir. O nedenle felsefi bakışın misyonlarından birisi geleceğe ilişkin yansımaları bugünden belirlemektir. Sözü geçen sebeplerden ötürü yapay zekaya eleştirel bir yaklaşım söz konusudur.

Bu eleştirel tutum teknoloji karşıtlığı olmayıp, insana özgü kavramların içinin boşaltılması ve aklın yeteneklerine zekanın göz dikmesi olarak yorumlanabilir. Bu bağlamda doğan etik problemler özne ve nesne ilişkisini yeniden değerlendirmeyi zorunlu kılmaktadır. Bu değerlendirme kapsamında etik kuramların sınıflandırılması savunulan eleştirel yaklaşıma katkı sağlamaktadır. Bu gereksinim yapay zekanın uygulamalı etik bağlamında sorunlarını ima etmektedir. Söz konusu sorunlar her ne kadar tartışma bölümünde toplumsal açıdan incelenmiş olsa da kavramsal çerçevede felsefi sonuçları müjdelemektedir. İki temelde değerlendirilen sonuçlar başlıca deneyimin ortadan kalkması problemi ve nesnenin özneyi belirleme problemidir.

Bu durumda tezin ilk sonucu, yapay zeka çağında insanın hem bedensel hem de zihinsel yönlerinin körelecek olmasıdır. Bedensel yöne ilişkin etkiler zayıf yapay zekayla bağdaştırılırken, zihinsel yöne ilişkin etkiler güçlü yapay zekayla bağdaştırılmalıdır. Bu bağdaşımalar kuramların ve ekollerin gelişimleriyle anlaşılabilir. Bilindiği üzere bilgi işleme kuramında<sup>292</sup> insan zihni, algoritmanın işleyişine benzer bir süreçle tanıtılmaktadır. Öte yandan Howard Gardner'in çoklu zeka kuramında beynin farklı bölümlerinin çeşitli zeka türlerine karşılık geldiği ifade edilmektedir. Bu görüş bedensel ve zihinsel süreçleri sınıflandırarak, algoritmik dizilime indirgeyen bir anlayış olarak düşünülebilir. Aslında Turing'in (1950) *Evrensel Turing Makinesi*'ni<sup>293</sup> tanıtmasının ardından John McCarthy'nin (1956) "yapay" ve "zeka" kelimelerinin birleşimi olan "yapay zeka"<sup>294</sup> terimini ortaya atması, daha sonra Miller'in (1956) bilgi işleme kuramını tanıtması birbirini izleyen ve destekleyen gelişmelerdir. Öyle ki Turing'in neden olduğu heyecana bağlı olarak böyle bir gelenek ortaya çıkmış olabilir. Anlaşılacağı üzere hem bedenin hem de zihnin çeşitli kuramlar ve ekoller kapsamında algoritmik süreçlere indirgendiği tarihsel olarak ortadadır. Bu bağlamda gerek zihin gerekse beden tarafında gerçekleştirme potansiyeli olan deneyim yoksunluğunun felsefi temellerine değinmek gerekmektedir.

Aristoteles'te insan akli deneyim ve tarihsellikten bağımsız düşünülemez. Bu hususta *phoronesis*'in önemi yeniden öne çıkmaktadır. En somut ifadeyle yapay zeka çağında hız, güvenilirlik ve hassaslığın sağladığı konfor, insan tarafında deneyimi büyük ölçüde gereksiz hale getirecektir. Öte yandan felsefi bir ifadeyle Platon'da, insan ve evren hem temel bir bedenden hem de rasyonel bir ruhtan oluşmaktadır. Mikro kozmos ve makro kozmos ifadeleri insan bedeninin evren bedeninden türediği gibi, insan ruhunun da evrenin ruhundan türediği anlamına gelmektedir. Dolayısıyla evren sadece düzenli bir sistem olmayıp, aynı zamanda akıllı bir

<sup>292</sup> Miller, 1956: 63, 81–97.

<sup>293</sup> Turing, 1950: 433 – 460.

<sup>294</sup> Russell ve Norvig, 2010: 18.

organizmadır.<sup>295</sup> Bu durumda akıllı bir organizma olduğu düşünülen evrenin minyatür suretleri insanlardır. İnsanın yaşantılardan, deneyimlerden, evrenle temastan yoksun olması düşünülemez. Özne ve nesne ilişkisi bağlamında otomasyon ve işsizlik, istenmeyen sonuçlar, özgünlüğün kaybolması gibi kaygıların yanında insanın evrenle etkileşiminin büyük ölçüde azalacağı sonucu çıkmaktadır. Bu da ilerleyen süreçlerde bedensel ve zihinsel yönlerin körelmesine bağlı olarak doğayla etkileşimin kopmasına neden olacaktır. *Phronesis*'i etkileyen deneyim yoksunluğu, yalnızca birey olarak insanı değil toplumsal anlamda insanlığı yani tarihsel birikim olarak aklı da olumsuz etkileyecektir. Söz konusu sonuç büyük ölçüde zayıf yapay zeka sistemleri çerçevesinde bedensel deneyime yönelik olumsuzlukları kapsamaktadır. Diğer taraftan aklın olanaklarının körelmesine ilişkin etkiler güçlü yapay zekayla ilgilidir.

Güçlü yapay zekanın doğuracağı olası tehlikelerinden en önemlisi insan aklının potansiyelinin zayıflamasıdır. Aklın tarihselliği düşünüldüğünde uzun zaman ve birikim en önemli unsurlardandır. Oysa güçlü yapay zekanın amaçlarından en önemlisi insan aklının çözmekte yetersiz kaldığı soruları daha kısa sürede doğru bir biçimde çözmektir. Yapay zekanın karmaşık felsefi problemleri çözmeyi hedefleyen modellerine yönelik çalışmalar buna örnek gösterilebilir. Charniak ve McDermott'ın (1985) zihnin fakültelerinin bilişim modelleri aracılığıyla incelenebileceğine olan inançlarıyla yapay zekaya böyle bir görev yüklemişlerdir. Unutulmamalıdır ki öz düşünümüne sahip olması beklenen güçlü yapay zeka sistemlerinin gerçekleşme olasılığı farklı bir tartışma konusu olduğu gibi bu sistemler için Searle'ün ifade ettiği bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik gibi unsurlar önkoşuldur. Ancak yapay zekanın güçlü seviyeye ulaştığını varsayarak geleceğe ilişkin sonuçlardan söz etmek gerekirse öznenin etkileneceği bazı olumsuzluklar öne çıkmaktadır. Öncelikle Aristotelesçi açıdan insanın rasyonel ve irrasyonel yönlerinden insanın özgünlüğü, belirsizliği ve soyut yönüyle ilgili olan irrasyonel yönün yapay zeka egemen çağda zarar göreceği muhtemeldir. Yani irrasyonel yönü zayıflayan öznenin mekanikleşmesi, özgünlük, yaratıcılık, belirsizlik vb. pek çok soyut yönleri tehlikeye atacaktır. Diğer bir deyişle insanoğlu rasyonel yönüyle bilimi, irrasyonel yönüyle de sanatı ortaya koymuştur. Doğayı referans alarak bilimsel faaliyetlerde bulunan insan, irrasyonel yönüyle doğasını bir bütün olarak koruyabilmiştir. Halbuki yapay zeka, irrasyonel yönü yok sayarak insanı nesne referansında yeniden tanımlamaya zorlayabilir. Bu durumda nesneleşen insanın irrasyonel yönünü ortadan kalkacak, yapay zeka güdümünde aklın araçsallaştığı kitle kültürünün yükseldiği bir anlayış doğabilecektir. Horkheimer'ın ifadelerine bakıldığında bu yorum daha da belirginleşmektedir. Zihin veya bedenden birinin

<sup>295</sup> Platon, 2019: 28d–30d.

kontrolüyle ötekinin tahakkümünün gerçekleşmesini Horkheimer, iç doğa ve dış doğa<sup>296</sup> kavramlarıyla açıklamaktadır. Bu denetim, insanın dış doğasının baskılanmasıyla iç doğasının kaçınılmaz olarak egemenlik altına alınması anlamına gelmektedir. O halde zayıf yapay zeka bedeninin, güçlü yapay zeka da zihnin eylemlerini devralacaktır. Bu da insanın doğayla etkileşimi olan deneyimin her iki anlamda azalacağı anlamına gelmektedir. Bu türden bir sonucun gerçekleşmesine ilişkin belirli bir tarih öngörüsünde bulunmak güçtür. Ancak yapay zekanın insan seviyesine erişme süresinin, insanın yapay zeka seviyesine indirgeneceği süreden çok daha uzun olacağı sonucu çıkarılabilir. Bu gelişme yapay zekanın başarısından ziyade insanın kendi doğasına karşı başarısızlığı olarak nitelendirilebilir.

İkinci sonuç, doğanın gözlemiyle bilimsel keşiflerini gerçekleştiren insanın yapay zeka çağında yeni bir tanıma zorlanacak olmasıdır. Şöyle ki tarih öznenin nesneyi belirlediği bir akışla ilerlemiştir. Bu akış sürecinde insanın nesne karşısında yüceltiği bilinmektedir. Oysa gelecekte insanın modellemesi beklenen ileri teknoloji yapay zekadır. Bu teknolojinin tarihteki bilimsel keşiflerin aksine bazı kavramsal sonuçlar doğurabileceği kuvvetle muhtemeldir. Ancak, yapay zekanın bu sürece etkisinin doğrudan olmayacağını altını çizmek gerekir. Kaldı ki bu ifade insanın doğrudan yapay zekanın kölesi olacağı anlamına gelebilir. Oysa yapay zekanın bu süreçteki rolü hız, güvenilirlik ve hassaslık gibi kapasitelerle insanın en önemli yönü akla olan inancın ya da bakışın değişmesine sebebiyet verebilecektir. Temelde bu süreç yapay zekanın sunduğu imkanlar sayesinde akla olan inancı zayıflatarak, hesaplamalı sistemlere hayranlığı doruk noktasına çıkaracaktır. Bu durumda insan, aklından ziyade hesaplamalı niteliklere sığdırılarak yeni bir tanıma mecbur bırakılacaktır. Bu sonuç siborg<sup>297</sup> anlamda insanın fiziksel tasarımıyla ilgili olmayıp daha çok insanın öz nitelikleri bağlamında doğasına ilişkin yeni bir tanımlama süreciyle ilgilidir. Şöyle ki tarihte insan doğasına ilişkin pek çok tanım getirilmiştir. Söz gelimi, Aristoteles insanı *rasyonel hayvan*<sup>298</sup> olarak tanımlamıştır. Ayrıca nihai hedefi yuva, aile ve topluluk kurmak olan insanın diğer bir tanımı da siyasal hayvandır.<sup>299</sup> Öte yandan Aristoteles'in tanımlarını yetersiz bulan Cassirer bir adım ileri giderek simgeleştiren hayvan<sup>300</sup> olarak insanı nitelendirmiştir. Dahası Mill (1995) homo eco-

<sup>296</sup> Horkheimer, 1986: 37.

<sup>297</sup> Kısmen insan kısmen de mekanik etkenler olarak bilinen sibernetik organizmalar (siborglar), kabaca canlı organizmaların ve makinelerin fiziksel birleşimi olarak ifade edilmektedir. Genel bir ifadeyle, birçok insan, vücut bölümlerinin protez cihazlar, takma dişler, mekanik kalp kapakçıkları, kontakt lensler veya yapay kemiklerle değiştirilmesi veya güçlendirilmesi veya kalp pili ve otomatik biyokimyasal mekanizmalar gibi mekanizmalara bağlı olarak "siborg" olarak nitelendirilebilir. Ancak kurgusal bir siborg, merkezi sinir sistemi mekanik veya dijital olarak değiştirilmiş bir ajandır. Teoride, beynimize yerleştirilen ve doğrudan "içeriden" kontrol edilebilen bellek çipleri olduğu hayal edebilir. Detaylı bilgi için bk. Floridi, 1999: 209.

<sup>298</sup> Aristotle, 1998: 293.

<sup>299</sup> Aristoteles, 1975: 9.

<sup>300</sup> Cassirer, 1980: 15.

nomicus<sup>301</sup> ifadesiyle insan doğasını ekonomi çerçevesinde anlamaya çalışmıştır. Tarihsel örneklere bakıldığında insan doğasının tanımı zorunlu olarak yaşadığı zamanla, ortamla ve deneyimlerle ilgilidir. Geleceğin ileri teknolojisi yapay zekanın dolaylı olarak insan doğasına yeni bir tanımı dayatması kaçınılmazdır. Ancak yapay zekanın bilinçlilik, yönelimsellik, öznellik ve akılsal nedensellik gibi niteliklere sahip olmasının önündeki engeller düşünüldüğünde bu sürecin doğrudan gerçekleşmeyeceği ortaya çıkmaktadır. Yine de yapay zekanın pek çok şeyi belirlediği, yaşamın her alanına nüfuz ettiği algoritma egemen bir çağda hız, güvenirlik, hassaslık gibi kavramlar akıl, bilinç, algı, yönelim, benlik gibi kavramları gölgede bırakacaktır. Söz gelimi kavramların belki de en önemlisi Aristoteles’in akıl terimine ve olanaklarına ilişkin ifadelerinde okunabilir. *Logos*’un, *nous* ve diğer fakültelere ev sahipliğini belirttiği argümanda özellikle günümüzdeki “zeka” teriminin *nous* ile “akıl” teriminin ise *logos* ile özdeş olduğunu göstermektedir. Bu özdeşlik de yapay zekanın insan akıyla aynı şey olmadığını bir çıkarımını sunmaktadır. Burada *logos*’un belirleyici unsurlarından *phronesis*’e gönderme yapılmaktadır. O halde insan akı yani ihtiyatlı, akı selim ve sağduyulu olabilme durumu zorunlu olarak *phronesis* ile mümkün olan bir şeydir. Öyle ki Aristoteles, insanoğlunun mücadelesini çetrefilli bir sürece benzeterek geometride uzman bir gencin akıllı olamayacağını savunmaktadır. İşte bu noktada akıl yılların birikimiyle mümkün olan bir olguyken, zekanın matematiksel işlemlerle sınırlı bir yapıya sahip olduğu vurgulanmaktadır. Dahası akıl tarihsel ve deneyimsel bir yaşantıyla görmüş geçirmiş, yılların meyvesini tatmış tecrübe sahibi insanlara özgü bir nitelik olarak ifade edilmektedir. Aristoteles’in ifadesinden aklın tarihselliği ve zekanın dar açısı arasındaki fark anlaşılabilir. Öte yandan *techne* ve *logos* kavramlarının birleşiminden meydana gelen teknoloji kelimesi bile aklın ortaya koyduğu kavramlardan biridir. Bu bağlamda öznenin nesneyi keşfi, onu tanımlaması, belirlemesi gibi gelişmeler sıradan süreçlerdir. Nesnenin özneyi belirleme olanağının gerçekleşeceği yapay zeka çağı ise yeni bir gelişme olacaktır.

Sonuç olarak tarihsel aklın olanaklarına rağmen yapay zeka çağında insanın modası geçmiş kavramlardan ziyade sözde çağdaş, keskin, hassas terimlerle anılması muhtemeldir. Bu sonuç bilimin felsefeyle çatıştığı bir argümanla ilişkilendirilebilir. Harman, (2018) felsefenin bilgiye “sahip olmaktan” ziyade bilgeliği sevmeye gayesinde olduğunu ifade etmiştir. Oysa bilimin, bilgiyi sadece sevmek yerine onu “elde etmeyi” arzulayan farklı bir tutumu bulunduğu dikkat çekmektedir.<sup>302</sup> O halde felsefi yaklaşım aklın olanaklarını açıklamada saydam bir tutum benimserken, bilimin gayesinin zekanın çıktılarını elde etmek olduğunu ifade

<sup>301</sup> Mill, 1995: 41.

<sup>302</sup> Harman: 2018, 6.

edebiliriz. Nihayetinde bu türden yaklaşımlar insan doğasına ilişkin yönleri sembollerin dar açısına hapsedecektir. Bu yönüyle yapay zeka; insanı hesaplamalı, sınırlı, donuk, tatsız ve tekdüze anlayışın baskın olduğu ifadelerle tanımlama mecbur bırakacaktır. Bu tanım insanın engin aklını görmezden gelen, zekanın hesaplama ve güvenilirlik niteliklerine sıkıştırıldığı bir yapıda gerçekleşecektir. Bu sayede yapay zekanın insan seviyesine eriştiği hatta onu geçtiği yanılsaması ortaya çıkacaktır. Özetle gerçek dışı bir durumun zaferi kavramsal müdahaleyle ilan edilecektir. Yapay zeka güdümünde insanın yeniden tanımlanması olasılığı doğrudan insanın kendi düştüğü bir hatadan kaynaklanacaktır. Bunun nedeni zihnın ortak kavramlarının sentaks düzeyinde gerçekleşmesinin önündeki engellere bağlı olarak yapay zekanın insan seviyesine erişmesinin güçlüğüdür. Bu yüzden robotların insanlaşması ihtimali, insanın robotlaşma ihtimalinden çok daha düşüktür. Nihayetinde öznenin nesneyi, nesnenin de özneyi belirleme yarışında insanın bir adım önde olacağı ortadadır. Zira yapay zekanın önündeki engeller sayısal sistemlerin zihnın süreçlerine ulaşmada çok yetersiz kaldığını kanıtlamaktadır. Diğer taraftan insan doğasının süratle sayısallaşarak mekanizmaya indirgenmesi kavramların daralan anlamlarından anlaşılmaktadır.

## KAYNAKÇA

- Anderson, M. ve Anderson, S.L. (2011). *Machine Ethics*. Cambridge University Press, NY. USA.
- Arf, C. (1959). *Makine Düşünebilir Mi ve Nasıl Düşünebilir?* Atatürk Üniversitesi Yayınları. Erzurum.
- Aristotle (1893). *The Nichomachean Ethics Of Aristotle*. (Çev. F.H. Peters). M.A. 5th Edition. Kegan Paul, Trench, Truebner & Co. London.
- Aristotle, (1998). *Metaphysics*. (Çev. H. Lawson-Tancred). Penguin Books, NY. USA.
- Aslan, H. (2003). *Algoritma*. Felsefe Ansiklopedisi, Cilt I, Etik Yayınları, İstanbul.
- Audi, R. (2003). *Contemporary Introductions To Philosophy*. Routledge, NY.
- Audi, R. (2003). *Epistemology A Contemporary Introduction To The Theory Of Knowledge*. Routledge, NY. USA.
- Ayer, A.J. ve Cohen L. J. (1977). "The Causal Theory Of Perception". *Proceedings Of The Aristotelian Society*, Supplementary Volumes. Cilt (51): 105.
- Aygün, Ö. (2017). *The Middle Included: Logos In Aristotle*. Northwestern University Press, Evanston, Ill, USA.
- Bellman, R. (1978). *An Introduction To Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company, San Francisco.
- Ben-Yami, H. (1993). "A Note On The Chinese Room". *Springer*. Synthese Cilt. 95, No. 2, Pp. 169-172.
- Berkeley, G. (2002). *A Treatise Concerning The Principles Of Human Knowledge*. Jacob Tonson. London.
- Bishop, J. M. (2022). *Artificial Intelligence Is Stupid And Causal Reasoning Will Not Fix It*. *Frontiers In Psychology*, 11, 513474.
- Boden, M. A. (2016). *AI: Its Nature And Future*. Oxford University Press, Oxford, UK.
- Bose B. K. (1994). "Expert System, Fuzzy Logic, And Neural Network Applications In Power Electronics And Motion Control," *Proceedings Of The IEEE*, Vol. 82, No. 8, Pp. 1303-1323, Aug. 1994, Doi: 10.1109/5.301690.
- Bostrom N. Cirkovic M. (2008). *Global Catastrophic Risks*. Oxford University Press.
- Bush, V. (1945). *As We May Think*. The Atlantic Monthly. 176 (1): 101–108.
- Cassirer, E. (1945). *An Essay On Man An Introduction To A Philosophy Of Human Culture*. Yale University Press. USA.

- Chalmers, D. J. (1996). *The Conscious Mind: In Search Of A Fundamental Theory*. Oxford University Press, New York.
- Chalmers, D.J. (1995). "Facing Up To The Problem Of Consciousness" *Journal Of Consciousness Studies*, 2(3):200-19.
- Charniak, E. ve Mcdermott, D. (1985). *Introduction To Artificial Intelligence*. MIT Press Cambridge, MA, USA.
- Churchland, P. M. (2013). *Matter And Consciousness*. The MIT Press, Cambridge. USA.
- Coleman, F. (2019). *A Human Algorithm: How Artificial Intelligence Is Redefining Who We Are*. Counterpoint, Berkeley, USA.
- Copeland, J. (2004). *The Essential Turing Seminal Writings In Computing, Logic, Philosophy, Artificial Intelligence, And Artificial Life Plus The Secrets Of Enigma*. Oxford University Press. NY, USA.
- Coppin, B. (2004). *Artificial Intelligence Illuminated*. Jones And Bartlett Publishers, Inc. Boston, USA.
- Cottingham, J. (2003). *Akılcılık*. Doruk Yayıncılık. İstanbul.
- Crane, T. (2000). *Dualism, Monism, Physicalism*. *Mind And Society* 1 (2):73-85.
- Cudworth, R. (1837). *The True Intellectual System Of The Universe*. Gould And Newman. NY, USA.
- Deleuze, G. ve Guittari, F. (2001). *Felsefe Nedir?* (Çev T. Ilgaz. YKY: İstanbul.
- Dennett, D. C. (1998). *Brainchildren Essays On Designing Minds*. A Bradford Book The MIT Press, Cambridge, Massachusetts. USA.
- Dennett, D. C. (2014). *When HAL Kills, Who's To Blame?: Computer Ethics*. Pisa University Press.
- Descartes, (2006). *A Discourse On The Method Of Correctly Conducting One's Reason And Seeking Truth In The Sciences*. (Çev. I. Maclean). Oxford University Press, UK.
- Descartes, R. (1994). *Metot Üzerine Konuşma*. (Çev. K. Sahir Sel). Sosyal Yayınları. İstanbul
- Descartes, R. (1998). *Yöntem Üzerine Konuşma*. (Çev. T. Afşin). Cumhuriyet Yayınları, İstanbul.
- Descartes, R. (2013). *Metot Üzerine Konuşma*. (Çev. A. Altınörs). Paradigma Yayıncılık. İstanbul.
- Descartes, R. (2013). *Ruhun İhtirasları*, (Çev. M. Özdil). Sayfa Yayınları. Topkapı/İstanbul.
- Dreyfus, H. (1999). *What Computers Still Can't Do: A Critique Of Artificial Reason*. The MIT Press. USA.
- Dreyfus, H. (2009). *On The Internet*. Routledge. London And New York.

- Dreyfus, H. ve Dreyfus S. E. (1986). *Mind Over Machine*. The Free Press. NY, USA.
- Etzioni, A. ve Etzioni, O. J Ethics (2017). "Incorporating Ethics Into Artificial Intelligence. The Journal Of Ethics, An International Philosophical Review". *Springer*. 21: 403
- Feuerbach, L. (1866). *Ludwig Feuerbach's Sämmtliche Werke. Gottheit, Freiheit Und Unsterblichkeit Vom Standpunkte Der Anthropologie*. Cilt (10). Leipzig.
- Floridi, L. (1999). *Philosophy And Computing*. Routledge. NY, USA And Canada.
- Ford, K.M. Glymour, C. Hayes, P.J. (1995). *Android Epistemology*. The MIT Press.
- Fotion, N. (2000). *John Searle*. Acumen Publishing Limited. Teddington, UK.
- Freud, S. (1962). *Civilization And Its Discontents*. (Çev. J. Stracher). W.W. Norton Company Inc. NY, USA.
- Fuller, S. (2011). *Humanity 2.0 What it Means To Be Human Past, Present And Future*. Palgrave Macmillan. UK.
- Furley. D. (1997). *From Aristotle To Augustine*. Routledge. NY.
- Gawande A. (2002). *Complications: A Surgeon's Notes On An Imperfect Science*.: Metropolitan Books, New York.
- Gips, J. (2008). "Towards the Ethical Robot" S. Anderson ve L. Anderson (Ed.), *Machine Ethics*. Cambridge University Press, NY, 244 – 253.
- Girard, D. (1965). *Cassell's New French-English Dictionary*. Cassell & Co. Ltd. London.
- Goldstein. B.E. (2010). *Encyclopedia Of Perception*. SAGE Publications, Inc. California.
- Good, I. J. (1965). *Speculations Concerning The First Ultraintelligent Machine*. *Advanced In Computers*, 6: 31-88.
- Grau, C. (2006). *There Is No I In Robot: Robots And Utilitarianism*. Intelligent Systems, IEEE. 21. 52 - 55. 10.1109/MIS.2006.81.
- Guttenplan, S. (1995). *A Companion To The Philosophy Of Mind*. Blackwell Publishers Inc. Cambridge, USA.
- Güzeldere, G. (1998). —*Yapay Zekanın Dünyü, Bugünü ve Yarını*. Cogito. 13. Yapay Zeka. 27-41.
- Harman, G. (2018). *Object-Oriented Ontology: A New Theory Of Everything*. Pelican Books, London.
- Hauben, M. & Hauben, R. (1997). *Netizens: On The History And Impact Of Usenet And The Internet*. Los Alamitos, Calif: IEEE Computer Society Press.
- Haugeland, J. (1985). *Artificial Intelligence: The very Idea*. MIT Press Cambridge, MA, USA.
- Haugeland, J. (1997). *Mind Design II*. The MIT Press, Cambridge, Massachusetts.

- Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies In Cybernetics, Literature, And Informatics*. The University Of Chicago Press, Chicago, USA.
- Hegel, G. W.F. (1991). *Tarihte Akıl* (Çev. Ö. Sözer). Ara Yayıncılık, İstanbul.
- Heinämaa, S. Lähteenmäki V. ve Remes P. (2007). *Consciousness From Perception To Reflection In The History Of Philosophy*. Springer, Dordrecht, The Netherlands.
- Hobbes, T. (2013). *Leviathan*. (Çev. S. Lim). Yapıredi Yayınları. İstanbul.
- Homer, (1957). *The Iliad Of Homer*. (Çev. R. Lattimore). University Of Chicago Press, USA.
- Horkheimer, M. (2010). *Akıl Tutulması*, (Çev. O. Koçak). Metis Yayınları, İstanbul.
- <http://www.ectaco.co.uk/english-multilanguage-dictionary/> (erişim tarihi: 04.05.2019).
- <http://www.tdk.gov.tr> (erişim tarihi: 10.02.2018).
- In Gove, P. B. & Merriam-Webster, Inc. (1986). *Webster's Third New International Dictionary Of The English Language Unabridged*.
- Irrgang, B. (2006): "Ethical Acts In Robotics". *Ubiquity*, Volume 7, Issue 34  
Www.Acm.Org/Ubiquity.
- Kaku, M. (2011). *Geleceğin Fiziği*. (Çev. Y. Saraç OYMAK ve H. OYMAK). ODTÜ Geliştirme Vakfı Yayıncılık ve İletişim A.Ş. Ankara.
- Kant, I. (1998). *Critique Of Pure Reason* (Çev. P. Guyer ve A. W. Wood). Cambridge University Press, USA.
- Kant, I. (2002). *Ahlak Metafiziğinin Temellendirilmesi* (Çev. İoanna Kuçuradi). Türkiye Felsefe Kurumu. Ankara.
- Kelly, D. R. (2017). *The Descent Of Ideas: The History Of Intellectual History*, Routledge, New York.
- Kerferd, G. B. (1967). "Logos" In *Encyclopedia O F Philosophy*, Volume 5.
- Kurzweil, R. (1990). *The Age Of Intelligent Machines*. MIT Press Cambridge, MA, USA.
- Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*, Viking, London.
- Langer, S. K. (1942). *Philosophy In A New Key: A Study In The Symbolism Of Reason, Rite And Art*. Harvard University Press. Cambridge, Mass.
- Laurie, J. (1983). *Better Mind The Computer* [Horizon]. BBC. İngiltere.
- Legg, S. ve Hutter, M. (2006). *A Collection Of Definitions Of Intelligence*  
<https://arxiv.org/abs/0706.3639>: (erişim tarihi: 13.01.2019)
- Long, A. A. (1986). *Hellenistic Philosophy Stoics, Epicureans, Sceptics*. University Of California Press, 2nd Edition. Berkeley And Los Angeles.
- Lucas, J. (1961). "Minds, Machines And Gödel". *Philosophy*, 36(137), 112-127.

- Mauro, G. ve Valigi, N. (2020). *Zero To AI: A Nontechnical, Hype-Free Guide To Prospering In The AI Era*. Manning Publications Co. NY. USA.
- Mccarthy, J. (2007). *What Is Artificial Intelligence?*  
[Http://Jmc.Stanford.Edu/Articles/Whatisai/Whatisai.Pdf](http://Jmc.Stanford.Edu/Articles/Whatisai/Whatisai.Pdf) (Eriřim Tarihi: 10.02.2018)
- Mccorduck, P. (2004). *Machines Who Think*. A K Peters, Ltd. Natick, MA. USA.
- Mclaughlin, B. P. (2019). *Blockheads!: Essays On Ned Block's Philosophy Of Mind And Consciousness*. The MIT Press. Cambridge, Massachusetts. USA.
- Merleau-Ponty, M. M. (1958). *Phenomenology Of Perception*. Routledge & Kegan Paul. NY. USA.
- Merleau-Ponty, M. M. (2005). *Algılanan Dünya* (Çev. Ömer Aygün). Metis Yayınları. Topkapı, İstanbul.
- Merleau-Ponty, M. M. (2007). *The Merleau-Ponty Reader*. Northwestern University Press. Evanston, Ill. USA.
- Mill, J.S. (1995). *On The Definition And Method Of Political Economy*, Ed. Daniel M. Hausman, The Philosophy Of Economics, Cambridge Univerity Press, Cambridge.
- Miller, G.A. (1956). *The Magical Number Seven, Plus Or Minus Two: Some Limits On Our Capacity For Processing Information*. Psychological Review, 63, 81-97.
- Moore, J. D. (2009). *Visions Of Culture*. Altamira Press. USA.
- Murphy R. R. (2000). *Introduction To AI Robotics*. The MIT Press. Cambridge, Massachusetts.
- Myers, D.G. (2004). *Psychology* 10th Edition. Worth Publishers. NY. USA.
- Nadeau, J. E. (2006). *Only Androids Can Be Ethical*. The MIT Press, Cambrige, England.
- Nagel, T. (1974). *What Is It Like To Be A Bat?* Philosophical Review 83 (October):435-50.
- Nilsson, J. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
- Omohundro, S. M. (2008). "The Basic AI Drives". *Proceedings of the First AGI Conference*. IOS Press, NLD, 483–492.
- Öktem, Ü. (2019). *Evrensel Bir Etik Mümkün Müdür?* Temařa Felsefe Dergisi 11: 44-51.
- Penrose, R. (1991). *The Emperor's New Mind: Concerning Computers, Minds And The Laws Of Physics*. Penguin Books. USA.
- Penrose, R. (1994). *Shadows Of The Mind A Search For The Missing Science Of Consciousness*. Oxford University Press. USA.
- Penrose, R. (1997). *Bilgisayar ve Zeka Kralın Yeni Usu I*. (Çev. T. Dereli). Tübitak Yayınları.
- Penrose, R. (1997). *Us Nerede? Kralın Yeni Usu III* (Çev. T. Dereli). Tübitak Yayınları.

- Pepperell, R. (2003). *The Posthuman Condition, Consciousness Beyond The Brain*. Intellect Books, Bristol, UK.
- Plato, (1888). *The Timaeus Of Plato*. (Çev. R.D. Archer-Hind). Macmillan And Co. NY, USA.
- Platon, (2015). *Phaidon* (Çev. F. AKDERİN). Say Yayınları, Topkapı, İstanbul.
- Platon, (2019). *Philebos Ya Da Haz Duygusu Üzerine*, (Çev. Furkan Akderin). Say Yayınları.
- Poole, D. Mackworth, A. ve Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press, Inc. New York, USA.
- Powers, T. (2006). *Prospects For A Kantian Machine*. IEEE Computer Society.
- Premebida, Cristiano & Ambrus, Rares & Marton, Zoltan. (2018). *Intelligent Robotic Perception Systems*. 10.5772/Intechopen.79742.
- Rakova, M. (2006). *Philosophy Of Mind A–Z*. Edinburgh. Edinburgh University Press.
- Rakover, S. S. (1999). *The Computer That Simulated John Searle In The Chinese Room*. New Ideas In Psychology 17. 55-66.
- Raynor, W. J. (1999). *The International Dictionary Of Artificial Intelligence*. The Glenlake Publishing Company, Ltd. USA.
- Reid, T. (1941). *Essays On The Intellectual Powers Of Man*. Macmillan And Co. Limited. London.
- Rey, G. (1986). *What's Really Going On In Searle's 'Chinese Room.'* Philosophical Studies: An International Journal For Philosophy In The Analytic Tradition, Cilt. 50, No. 2, Springer, 169–85.
- Rich, E. Knight, K. ve Nair, S. B. (1990). *Artificial Intelligence*. Tata Mcgraw-Hill Publishing Company Limited. New Delhi, India.
- Rosenthal, D.M. (2005). *Consciousness And Mind*. Clarendon Press. Oxford.
- Ross, S. D. (1995). *Aristotle*. Routledge. New York, NY.
- Ross, W. D. (2011). *Aristoteles*. Kabalcı Yayınevi. İstanbul.
- Ross, W. D. Smith J. A. Mure ve Diğerleri. (1908). *The Works Of Aristotle*. The Clarendon Press, Oxford.
- Rothman, D. (2020). *Artificial Intelligence By Example: Acquire Advanced AI, Machine Learning, And Deep Learning Design Skills*. Packt Publishing.
- Rubin, T. C. (2003). *Artificial Intelligence And Human Nature*. The New Atlantis, No. Center For The Study Of Technology And Society.
- Russell, B. (1912). *The Problems Of Philosophy*. Oxford University Press, London.

- Russell, S.J. ve Norvig P. (2010). *Artificial Intelligence A Modern Approach*. Pearson Education, Inc. New Jersey.
- Sahin E., Girgin S., Bayindir L. ve Turgut A.E. (2008) . “Swarm Robotics, in Swarm Intelligence - Introduction and Applications” (Ed: C. Blum ve D. Merkle), *Springer*, pp. 87-100.
- Sartre, J. P. (2011). *Varlık ve Hiçlik F enomenolojik Ontoloji Denemesi* ( ev. Turhan Ilgaz ve Gaye  ankaya Eksen). İthaki Yayınları. İstanbul.
- Say, C. (2005). *G del’in Eksiklik Teoremi*. Matematik D nyası, Kış, Kadık y Matbaa, İstanbul.
- Schade, J.P. ve Smith, J. (1970). *Computers And Brains*. Elsevier Publishing Company. Amsterdam, The Netherlands.
- Schneider, S. (2016). *From Time Travel To Superintelligence*. John Wiley & Sons, West Sussex, UK.
- Searle, J. (1984). *Minds, Brains And Science*. Harvard University Press, Cambridge.
- Searle, J. (1992). *The Rediscovery Of The Mind*. The MIT Press.
- Searle, J. (1996). *Akıllar Beyinler ve Bilim*. ( ev. K.Bek). Say Yayınları. Sirkeci/İstanbul.
- Searle, J. (2004). *Zihn n Yeniden Ke fi*. ( ev. M. Macit). Litera Yayıncılık. İstanbul.
- Searle, J. R. (1980). *Minds, Brains, And Programs*. Behavioral And Brain Sciences 3 (3): 417-457.
- Searle, J. R. (2006). *Zihin Dil ve Toplum, Ger ek D nyada Felsefe*. ( ev. A. Tura). Litera Yayıncılık Ltd.  ti. Topkapı, İstanbul.
- Shaffer, J. A. (1991). *Bilin , Ruh ve  tesi* ( ev. Dr. T. Ko ). İz Yayıncılık. İstanbul.
- Singer, P. W. (2009). *Wired For War: The Robotics Revolution And Conflict In The Twenty-First Century*. Penguin Press, New York.
- Singer, P. (1999). *Practical ethics*. Cambridge Univ. Press. Cambridge.
- Sloman, A. (1971). “Interactions Between Philosophy And Artificial Intelligence: The Role Of Intuition And Non-Logical Reasoning In Intelligence”. *Artificial Intelligence 2*. North-Holland Publishing Company. (209-225)
- Smith, R. N. (1990). *Collins Dictionary Of Artificial Intelligence*. Collins. London.
- Steiglitz, K. (2019). *The discrete charm of the machine: Why the world became digital*. Princeton University Press. New Jersey.
- Sternberg, R.J – Kaufman, S.B. (2011). *The Cambridge Handbook Of Intelligence*. Cambridge University Press, New York.

- Stoica, I. Song, D. Popa, R. A. Patterson, D. Mahoney, M. W. Katz, R. & Abbeel, P. (2017). *A Berkeley View Of Systems Challenges For Ai*. Arxiv Preprint Arxiv:1712.05855.
- Stumpf, S.E. (1988). *Elements Of Philosophy*. McGraw-Hill Book Company, NY, USA.
- Sullins, J. P. (2006). "When Is A Robot A Moral Agent?" *International Review Of Information Ethics* Vol. 6 (12/2006).
- Toulmin, S. (1990). *The Hidden Agenda Of Modernity*. The University Of Chicago Press.
- Tulley, R. (2008). "Is There Techne In My Logos? On The Origins And Evolution Of The Ideographic Term-Technology". *International Journal Of Technology, Knowledge And Society*. 4. 93-104. 10.18848/1832-3669/CGP/V04i01/55813.
- Turing, A. M (2005). "Bilgişlem Makineleri ve Zeka". (Çev. F. Doruker). *Aklın G'özü Benlik ve Ruh Üzerine Hayaller ve Düşünceler*. Hofstadter D. R. ve Dennett D. C. Boğaziçi Üniversitesi Yayınevi. İstanbul.
- Turing, A. M. (1950). *Computing Machinery And Intelligence*. *Mind A Quarterly Review Of Psychology And Philosophy*. Cilt (59). 433-460.
- Unat, Y. (2006). "Artuklular Donemi'nde Bir Türk Muhendis; Cezeri". *I. Uluslararası Mardin Tarihi Sempozyumu Bildirileri*. 26-27-28 Mayıs 2006, Mardin, 223-238.
- Urbina, F., Lentzos, F., Invernizzi, C. Vd. *Dual Use Of Artificial-Intelligence-Powered Drug Discovery*. *Nat Mach Intell* 4, 189–191 (2022).
- Uriarte, C. Cañeque, S.T. ve Sunal, C.B (2010). *Diccionario Turco- Español*. Akal, S. A. Navalmcarnero, Madrid.
- Ülkü, V. (2005). *Büyük Almanca- Türkçe Sözlük*. Inkılap Yayınevi. Bağcılar, İstanbul.
- Vinge, Vernor (1993). *The coming technological singularity: How to survive in the post-human era*. Whole Earth Review.
- Wakefield, J. C. (2003). *The Chinese Room Argument Reconsidered: Essentialism, Indeterminacy, And Strong AI*. *Minds And Machines* 13: 285–319, 2003. Academic Publishers. Netherlands.
- Wallach, W. ve Allen C. (2008). *Machine Morality: Bottom-Up And Top-Down Approaches For Modelling Human Moral Faculties*. *AI & Soc* 22:565–582.
- Webster, M. (1986). *Webster's Third New International Dictionary*. Merriam-Webster Inc. Springfield, Massachusetts, USA.
- Wedin, M. V. (1988). *Mind And Imagination In Aristotle*. Yale University Press. New Heaven And London.
- Weizenbaum, J. (1976). *Computer Power And Human Reason: From Judgment To Calculation*. W. H. Freeman & Co.

- Whorf, B. L. (1959). *Language, Thought, And Reality Selected Writings Of Benjamin Lee Whorf*. Massachusetts Institute Of Technology And John Wiley & Sons, Inc. New York, London.
- Winston, P.H. (1992). *Artificial Intelligence*. Addison-Wesley Publishing Company, USA.
- Yudkowsky, E. (2008). "Artificial Intelligence as a Positive and Negative Factor in Global Risk." "Bostrom & M. M . Cirkovic (Eds.), *Global catastrophic risks* (pp. 308-345). Oxford University Press.
- Yule, G. (2006). *The Study Of Language*. Cambridge University Press. NY, USA.



## ÖZGEÇMİŞ

|                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>Adı ve SOYADI</b>                                                                                                                                                                                                                                                                                                                                                                                  | Ferhat BAYIK                                                                |
| <b>EĞİTİM DURUMU</b>                                                                                                                                                                                                                                                                                                                                                                                  |                                                                             |
| <b>Mezun Olduğu Lise</b>                                                                                                                                                                                                                                                                                                                                                                              | Serik Anadolu Lisesi                                                        |
| <b>Lisans Diploması</b>                                                                                                                                                                                                                                                                                                                                                                               | Anadolu Üniversitesi, İngilizce Öğretmenliği Lisans Programı, 2007          |
| <b>Yüksek Lisans Diploması</b>                                                                                                                                                                                                                                                                                                                                                                        | Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü Felsefe Ana Bilim Dalı, 2014 |
| <b>Tez Konusu</b>                                                                                                                                                                                                                                                                                                                                                                                     | Kültür Kavramının Tarihsel ve Felsefi Yönlerden İncelenmesi                 |
| <b>Yabancı Dil</b>                                                                                                                                                                                                                                                                                                                                                                                    | İngilizce                                                                   |
| <b>BİLİMSEL FAALİYETLER</b>                                                                                                                                                                                                                                                                                                                                                                           |                                                                             |
| <p><b>Aldığı Destekler/Burslar:</b><br/> Fulbright FLTA Bursu, Florida Uluslararası Üniversitesi, 2014-2015, Miami, ABD.</p> <p><u>Comenius</u>, “Language and Methodology for Teachers of English - A Blended Approach”, 5-21 Nisan, 2013, Scarborough, İngiltere.</p> <p>IREX “Öğretimde Mükemmellik ve Başarı Programı” Kuzey Dakota Üniversitesi, 3 Temmuz - 26 Ağustos 2012 Grandforks, ABD.</p> |                                                                             |
| <p><b>Hakemli ve Indexli Dergilerde Yayımlanan Özgün Makaleler ve Çeviriler</b><br/> Bayık, F. (2019). “Aristoteles ve Descartes Bağlamında Akıl ve Zekâ Kavramlarının Farkları”. <i>Kaygı</i>. Bursa Uludağ Üniversitesi Fen-Edebiyat Fakültesi Felsefe Dergisi, 18 (1), 172-187. DOI: 10.20981/kaygi.529827</p>                                                                                     |                                                                             |
| <p><b>Uluslararası Sempozyum ve Kongrelerde Sunulan, Basılan Tam Metin Bildiriler:</b><br/> Bayık, F. (2016). “Culture Concept and Perceptual Dynamics in Second Language Teaching” 2. Uluslararası Dil, Kültür ve Edebiyat Çalıştayı. 8-10 Haziran 2016, Antalya, Türkiye.</p>                                                                                                                       |                                                                             |
| <p><b>Kitap Bölümü</b><br/> Bayık, F. (2017). “Kültür Kavramının Felsefi Açılardan Çağdaş Dünyaya Etkileri”. C, Ergun. – S, Öğrekçi. <i>Sosyal Bilimlerde Kültür Tartışmaları</i>. Gece Kitaplığı, Ankara, Türkiye. (319-329).</p>                                                                                                                                                                    |                                                                             |
| <p><b>Uluslararası Sempozyum ve Kongrelerde Sunulan, Basılan Özet Bildiriler:</b><br/> Bayık F.( 2017) "Doğruluk Kavramının Doğası", 2. Uluslararası Felsefe, Eğitim, Sanat ve Bilim Tarihi Sempozyumu, Bildiri 3-7 Mayıs 2017, Muğla, Türkiye</p>                                                                                                                                                    |                                                                             |
| <p><b>Ulusal Dergilerde Yayımlanan Çeviriler:</b><br/> Bayık F., (2016) “Foucault ve Lacan: Efendi Kimdir” <i>Düşünbil Felsefe Dergisi</i> sayı 53.<br/> Bayık F., (2016) “Wilhelm Reich ve Cinsel Devrim” <i>Libido Psikanaliz Dergisi</i> sayı 19.</p>                                                                                                                                              |                                                                             |

Bayık F., (2015) “Bilimin Ruhu İçin Mücadele: Popper Vs Kuhn” Düşünbil Felsefe Dergisi sayı 50.

Bayık F., (2015) “Özgür İrade ve Determinizm” Düşünbil Felsefe Dergisi sayı 49.

Bayık F., (2015) “Danah Zohar’ın Kuantum Benliği” Düşünbil Felsefe Dergisi sayı 47.

### İŞ DENEYİMİ

#### Çalıştığı Kurumlar

(2006 – Devam) Milli Eğitime Bağlı Okullarda İngilizce Öğretmeni