



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Neslihan Verda ÖZMEN

ÜNİVERSİTE ÖĞRENCİLERİNİN YAPAY ZEKÂ ETİĞİNE İLİŞKİN TUTUMLARI
ÜZERİNE BİR KARMA YÖNTEM ARAŞTIRMASI

Yönetim Bilişim Sistemleri Anabilim Dalı

Yüksek Lisans Tezi

Antalya, 2024



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Neslihan Verda ÖZMEN

ÜNİVERSİTE ÖĞRENCİLERİNİN YAPAY ZEKÂ ETİĞİNE İLİŞKİN TUTUMLARI
ÜZERİNE BİR KARMA YÖNTEM ARAŞTIRMASI

Danışman

Doç. Dr. Tayfun YÖRÜK

Yönetim Bilişim Sistemleri Anabilim Dalı

Yüksek Lisans Tezi

Antalya, 2024

Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Neslihan Verda ÖZMEN'in bu çalışması, jürimiz tarafından Yönetim Bilişim Sistemleri
Anabilim Dalı Yüksek Lisans Programı tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Sezgin IRMAK (İmza)

Üye (Danışmanı) : Doç. Dr. Tayfun YÖRÜK (İmza)

Üye : Doç. Dr. Mehmet Özer DEMİR (İmza)

Tez Başlığı: Üniversite Öğrencilerinin Yapay Zeka Etiğine İlişkin Tutumları Üzerine Bir
Karma Yöntem Araştırması

Tez Savunma Tarihi : 15/02/2024

Mezuniyet Tarihi : 14/03/2024

AKADEMİK BEYAN

Yüksek Lisans Tezi olarak sunduğum “Üniversite Öğrencilerinin Yapay Zekâ Etiğine İlişkin Tutumları Üzerine Bir Karma Yöntem Araştırması” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımdan yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir, bunu şerefimle doğrularım.

İmza

Neslihan Verda ÖZMEN





AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



TEZ ÇALIŞMASI ORJİNALLİK RAPORU BEYAN BELGESİ

Öğrenci Bilgileri	
Adı-Soyadı	Neslihan Verda ÖZMEN
Öğrenci Numarası	202152096006
Anabilim Dalı	Yönetim Bilişim Sistemleri
Programı	Tezli Yüksek Lisans
Danışman Öğretim Üyesi Bilgileri	
Unvanı, Adı-Soyadı	Doç. Dr. Tayfun YÖRÜK
Yüksek Lisans Tez Başlığı	Üniversite Öğrencilerinin Yapay Zekâ Etiğine İlişkin Tutumları Üzerine Bir Karma Yöntem Araştırması
Turnitin Bilgileri	
Ödev Numarası	2314090487
Rapor Tarihi	07.03.2024
Benzerlik Oranı	Alıntılar hariç: %2 Alıntılar dahil: %3
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE, Yukarıda bilgileri bulunan öğrenciye ait tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 122 sayfalık kısmına ilişkin olarak Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarında belirlenen filtrelemeler uygulanarak yukarıdaki detayları verilen ve ekte sunulan rapor alınmıştır. Danışman tarafından uygun olan seçenek işaretlenmelidir: ☒ Benzerlik oranları belirlenen limitleri aşmıyor ise: Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporunun 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 Raporunun doğruluğunu onaylar ve Uygulama Esaslarında öngörülen yüzdelik sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.	
Gerekçe:	
Benzerlik taraması yukarıda verilen ölçütlere uygun olarak tarafımda yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim.	
Danışman Öğretim Üyesi Doç. Dr. Tayfun YÖRÜK	
İmza	

İÇİNDEKİLER

ŞEKİLLER LİSTESİ.....	iv
TABLolar LİSTESİ.....	v
KISALTMALAR LİSTESİ.....	vi
ÖZ.....	vii
ABSTRACT.....	ix
TEŞEKKÜR.....	xi
GİRİŞ.....	1

BİRİNCİ BÖLÜM

YAPAY ZEKÂ VE KAPSAMI

1.1. Yapay Zekâ Nedir?.....	5
1.2. Yapay Zekâ Araştırmalarına ve Felsefî Argümanlara Kronolojik Bir Bakış.....	7
1.3. Makine Öğrenmesi ve Yapay Zekâ Alt Alanları.....	10
1.3.1. Makine Öğrenimi.....	11
1.3.2. Bilgisayarlı Görü.....	13
1.3.3. Doğal Dil İşleme.....	13
1.3.4. Üretken Yapay Zekâ ve Temel Modeller.....	14
1.3.5. Robotik Sistemler.....	14
1.4. Sektörel Uygulamalar.....	16
1.4.1. Sağlık Hizmetleri.....	17
1.4.2. İnsan Kaynakları.....	18
1.4.3. Güvenlik.....	18
1.4.4. Tavsiye Sistemleri.....	19
1.4.5. Bankacılık, Finansal Hizmetler ve Sigortacılık.....	19

İKİNCİ BÖLÜM

YAPAY ZEKÂ VE ETİK

2.1. Etik.....	20
2.1.1. Metaetik.....	20
2.1.2. Normatif Etik.....	21

2.1.3. Uygulamalı Etik.....	22
2.2. Yapay Zeka Etiği.....	22
2.2.1. Yapay Zekâ Etiği Çerçevesinin Gündemi.....	26
2.2.1.1. Hükümetler ve Uluslararası Girişimler.....	28
2.2.1.2. Sivil Toplum.....	30
2.3. Yapay Zekânın Etik Boyutları Üzerine.....	32
2.3.1. Adalet.....	33
2.3.2. Şeffaflık (Açıklanabilirlik).....	37
2.3.3. Kötü Niyetli Olmama (Zarar Vermeme).....	40
2.3.4. Gizlilik (Mahremiyet).....	44
2.3.5. Sorumluluk (Hesap Verebilirlik).....	47
2.4. Toplumsal ve Etik Çıktılar.....	48

ÜÇÜNCÜ BÖLÜM

ÜNİVERSİTE ÖĞRENCİLERİNİN YAPAY ZEKÂ ETİĞİNE İLİŞKİN TUTUMLARI ÜZERİNE BİR KARMA YÖNTEM ARAŞTIRMASI

3.1. Araştırma Tasarımı ve Metodoloji.....	55
3.2. Nicel Araştırma.....	56
3.2.1. Veri Toplama Aracı ve Türkçeye Çevrilmesi.....	56
3.2.2. Veri Toplama Süreci.....	58
3.2.3. Geçerlilik ve Güvenilirlik.....	58
3.2.4. Verilerin Analizi.....	60
3.2.5. Bulgular.....	61
3.2.5.1. Cinsiyet Değişkenine Yönelik Farklılık Analizleri.....	61
3.2.5.2. Akademik Disiplin Değişkenine Yönelik Farklılık Analizleri.....	63
3.3. Nitel Araştırma.....	64
3.3.1. Araştırma Tasarımı.....	64
3.3.2. Araştırma Alanı ve Çalışma Grubu.....	66
3.3.3. Veri Toplama Süreci.....	67
3.3.4. Verilerin Analizi.....	68
3.3.5. Geçerlilik ve Güvenilirlik.....	68

3.3.6. Bulgular.....	69
3.3.6.1. Adalet.....	70
3.3.6.1.1. Adalet Algısı.....	70
3.3.6.1.2. Adalet Etkisi.....	73
3.3.6.2. Kötü Niyetli Olmama.....	74
3.3.6.2.1. Kötü Niyet Algısı.....	75
3.3.6.2.2. Kötü Niyet Motivasyonu.....	77
3.3.6.2.3. Kötü Niyet Çıktıları.....	78
3.3.6.3. Gizlilik.....	81
3.3.6.3.1. Gizlilik Algısı.....	81
3.3.6.3.2. Gizlilik Toleransı.....	84
3.3.6.4. Şeffaflık.....	86
3.3.6.4.1. Şeffaflık Algısı.....	87
3.3.6.4.2. Şeffaflık Toleransı.....	89
3.3.6.5. Sorumluluk.....	90
3.4. Tartışma.....	96
3.4.1. Adalet.....	97
3.4.2. Kötü Niyetli Olmama.....	100
3.4.3. Gizlilik.....	104
3.4.4. Şeffaflık.....	110
3.4.5. Sorumluluk.....	113
SONUÇ.....	116
KAYNAKÇA.....	124
İnternet Kaynakları.....	146
EK 1 – Anket İzin Maili.....	155
EK 2 – Çevirileri Değerlendirme Formu.....	156
EK 3 – Anket Formu (AT-EAI Ölçeği).....	159
EK 4 – Yarı Yapılandırılmış Mülakat Formu.....	160
EK 5 – Gönüllü Katılım ve Bilgilendirilmiş Onam Formu.....	162
ÖZGEÇMİŞ.....	163

ŞEKİLLER LİSTESİ

Şekil 1.1. Tarihsel YZ Tanımlarının Taksonomisi.....	6
--	---



TABLolar LİSTESİ

Tablo 3.1. DFA Sonucunda Ulaşılan Uygunluk İstatistikleri Değerleri ve Referans Aralıkları	60
Tablo 3.2. Katılımcıların Betimsel Bilgileri.....	60
Tablo 3.3. Cinsiyet Değişkenine İlişkin T-Testi Sonuçları.....	62
Tablo 3.4. Cinsiyet Değişkenine İlişkin U-Testi Sonuçları.....	62
Tablo 3.5. Akademik Disiplin Değişkenine İlişkin T-Testi Sonuçları.....	64
Tablo 3.6. Katılımcılara Ait Demografik Bilgiler.....	66
Tablo 3.7. Adalet Boyutuna İlişkin Ana ve Alt Temalar.....	70
Tablo 3.8. Kötü Niyetli Olmama Boyutuna İlişkin Ana ve Alt Temalar.....	75
Tablo 3.9. Gizlilik Boyutuna İlişkin Ana ve Alt Temalar.....	81
Tablo 3.10. Şeffaflık Boyutuna İlişkin Ana ve Alt Temalar.....	87
Tablo 3.11. Gözden Çıkarılan İlk Boyuta İlişkin Görüş Bildiren Katılımcılar.....	94

KISALTMALAR LİSTESİ

AB	: Avrupa Birliği
ABD	: Amerika Birleşik Devletleri
AGI	: Artificial General Intelligence - Yapay Genel Zekâ
AR	: Augmented Reality - Artırılmış Gerçeklik
AV	: Autonomous Vehicle - Otonom araç
BM	: Birleşmiş Milletler
DARPA	: Defense Advanced Research Projects Agency - Amerikan Savunma Bakanlığı İleri Araştırma Projeleri Ajansı
DFA	: Doğrulayıcı Faktör Analizi
EPSRC	: Engineering and Physical Sciences Research Council - Mühendislik ve Fiziksel Bilimler Araştırma Konseyi
IEEE	: Institute of Electrical and Electronics Engineers - Elektrik ve Elektronik Mühendisleri Enstitüsü
ISS	: International Space Station – Uluslararası Uzay İstasyonu
ML	: Machine Learning - Makine Öğrenmesi
OECD	: Organisation for Economic Co-operation and Development - Ekonomik İş Birliği ve Kalkınma Teşkilatı
UNESCO	: United Nations Educational, Scientific and Cultural Organization - Birleşmiş Milletler Eğitim, Bilim ve Kültür Teşkilatı
UYZS	: Ulusal Yapay Zekâ Stratejisi
VR	: Virtual Reality - Sanal Gerçeklik
XAI	: Explainable Artificial Intelligence - Açıklanabilir Yapay Zekâ
YZ	: Yapay Zekâ

ÖZ

Bu araştırma ile “teknik sistemlerin toplumsal sistemlerle uyumsuz hizalanmalarını” insan lehine değiştirmenin olanaklarını düşünmeye ahlâk felsefesi vasıtasıyla yönelerek, günümüz kullanıcılarını ve yapay zekâ ekosisteminin potansiyel paydaşlarını temsilen farklı akademik kökenlerden üniversite öğrencilerinin yapay zekâ etiğine yönelik tutumlarını ortaya çıkarmak, yaşamsal etkileri olan bu sistemlerin etik boyutlarına ilişkin görüşlerimizi nasıl şekillendirdiğini ele almak amaçlanmıştır. Yazında mikro-analiz düzeyinde çalışmaların temsiliyetini artırmak da hedeflenmiştir. “Faillik alanını insanlardan makine nesillerine kaydıran” makine failliğine ilişkin söylemlerin aksine, insan ve tekniğin birlikte evrimini tanımlayan *epifilogenez* fikrinden güç alarak insanların ve artefaktların karşılıklı bağımlılıklarını kabul eden bir anlayış benimsenmiştir.

Karma metodolojilerden sıralı açıklayıcı desenin izlendiği iki aşamalı araştırma, yapay zekâ için etik ilkeler ya da kılavuzlar içeren belgelerin “adalet”, “şeffaflık”, “kötü niyetli olmama”, “gizlilik” ve “sorumluluk” ilkeleri etrafında yakınlığı tespitine dayanarak bu ilkeler ile temellendirilmiştir. Kantitatif aşamada veri toplama aracı olarak “Yapay Zekâ Etiğine Yönelik Tutum Ölçeği”nden faydalanılmış, araç Türkçeye çevrilerek çevrimiçi anket aracılığıyla 395 katılımcı yanıtına ulaşılmıştır. Geçerlik ve güvenilirlik analizlerini takip eden farklılık analizleri ile “Adalet”, “Kötü niyetli olmama” ve “Gizlilik” faktörlerine yönelik *cinsiyet* ve “Şeffaflık” faktörüne yönelik *akademik disiplin* temelli istatistiksel olarak anlamlı farklılıklar keşfedilmiştir. Kalitatif araştırma yoluyla ise fenomenolojik bir yaklaşımla veri edinme tekniği olarak yarı yapılandırılmış görüşmeler gerçekleştirilmiş, ilgili grupların birincil aşamada saptanan farklılık seviyelerinin yapısını irdelemek ve anlayışlarının altında yatan nedenlere ulaşabilmek maksadıyla 16 katılımcının bu sistemler ile etkileşimleri, deneyimleri ve ele alınan ilkeler ile ilişkiselliklerine dair öznel bakış açılarına yoğunlaşmıştır. Görüşme metinlerinin tematik analizi neticesinde belirlenen “Adalet algısı”, “Adalet etkisi”, “Kötü niyet algısı”, “Kötü niyet motivasyonu”, “Kötü niyet çıktıları”, “Gizlilik algısı”, “Gizlilik toleransı”, “Şeffaflık algısı” ve “Şeffaflık toleransı” ana temaları ile kadınların ve İnsan ve Toplum Bilimleri disiplinlerinden katılımcıların ilgili boyutlarda daha yüksek hassasiyet seviyelerinin özünü betimleyen bir izah ortaya konulmaya çalışılmıştır. Kantitatif çözümlemeler ile gruplar arası anlamlı bir farklılığa ulaşılamayan “Sorumluluk” boyutuna yönelik ise kalitatif süreç ile katılımcıların çoğunluğunun ‘döngüde’ bir insanın varlığını daimî gözetken ve bu ilkelere riayet

edildiğini temin edecek çok paydaşlı bir sorumluluğu tayin eden düzenleyici ve yasal çerçevelerin beklentisinde oldukları sonucuna ulaşılmıştır.

Araştırma gruplar arası tutum farklılıklarını açıklarken, yapay zekâ yönergelerini şekillendireceği öngörülen ilkelerin Türkiye’de öğrenim gören öğrencilere yansımalarının keşfedilmesine, ülkeye özgü nüansların yakalanarak anlayışın derinleştirilmesine ve meşru kaygıların tüm çeşitliliğiyle ortaya konulmasına imkân tanımıştır. Bulguların, yerel değerler ve tercihler dikkate alınarak merkezi etik sorunlara ortak yanıtlar sağlayacak kapsayıcı yapay zekâ geliştirme ve ilgili yönetim ve gözetim çabalarına yönelik içgörüler sağlayacağı düşünülmektedir. Çalışmanın, yapay zekâ etiği müfredatı ve eğitim materyallerinin geliştirilmesine ve yapay zekâyâ yatırım yapmayı hedefleyen bilhassa yerli işletmelerin rekabet avantajlarını artırabilmeleri yönünde faydalı olacağı da umulmakta, aynı zamanda pek çok potansiyel araştırma istikametine işaret edilmektedir.

Anahtar Kelimeler: Yapay Zekâ (YZ), Yapay Zekâ Etiği, Karma Yöntem Araştırması, Sıralı Açıklayıcı Tasarım

ABSTRACT

A MIXED METHODS STUDY OF UNIVERSITY STUDENTS' ATTITUDES TOWARDS ARTIFICIAL INTELLIGENCE ETHICS

This research aims to reveal the attitudes of university students from different academic backgrounds, representing today's users and potential stakeholders of the artificial intelligence ecosystem, towards artificial intelligence ethics, and to address how artificial intelligence systems which have vital effects shape our views on its ethical dimensions, by turning towards thinking about the possibilities of changing the "incompatible alignments of technical systems with social systems" in favor of human beings through moral philosophy. It is also aimed to increase the representation of micro-analysis level studies in the literature. In contrast to contemporary discourses on machine agency “that shift the sphere of agency from humans to machine generations”, an understanding that acknowledges the interdependence of humans and artifacts is adopted, drawing strength from the idea of *epiphylogenesis*, which describes “the co-evolution of humans and technology”.

This two-stage research, which followed a sequential explanatory design from mixed methodologies, was based on the determination that documents containing ethical principles or guidelines for artificial intelligence converge around the principles of "fairness", "transparency", "non-maleficence", "privacy" and "responsibility". In the quantitative phase, the "Attitude Towards the Ethics of Artificial Intelligence" was used as a data collection tool. The instrument was translated into Turkish and 395 participant responses were obtained through an online survey. Following the validity and reliability analyses, statistically significant differences based on gender for the "Fairness", "Non-maleficence" and "Privacy" factors and academic discipline for the "Transparency" factor were discovered through difference analyses. Through qualitative research, semi-structured interviews were conducted as a data acquisition technique with a phenomenological approach, focusing on the subjective perspectives of the 16 participants on their interactions with these systems, their experiences and their relationality with the principles discussed, in order to examine the structure of the levels of difference identified in the primary phase of the relevant groups and to reach the underlying reasons for their understanding. With the main themes of "Fairness perception", "Fairness effect", "Malicious intent perception", "Malicious intent motivation", "Malicious intent outcomes", "Privacy perception", "Privacy tolerance", "Transparency perception" and "Transparency

tolerance", which were determined as a result of the thematic analysis of the interview texts, an explanation that describes the essence of the higher sensitivity levels of women participants and the participants from the disciplines Humanities and Social Sciences in the relevant dimensions was tried to be put forward. Regarding the dimension of "Responsibility", for which no significant difference was found between the groups through quantitative analysis, the qualitative process revealed that the majority of the participants expect regulatory and legal frameworks that assign a multi-stakeholder responsibility to ensure that these principles are respected and that a human being is always in the 'loop'.

While explaining the differences in attitudes between the groups, the research has allowed exploring the reflection of the principles that are envisioned to shape AI guidelines on Turkish students, deepening understanding by capturing country-specific nuances, and revealing legitimate concerns in all their diversity. The findings are expected to provide insights for inclusive AI development and related governance and oversight efforts that will provide common responses to central ethical issues, taking into account local values and preferences. It is hoped that the study will be useful for the development of AI ethics curricula and training materials, and for enhancing the competitive advantage of domestic businesses, especially those aiming to invest in AI. This research also points to many potential research directions.

Keywords: Artificial Intelligence (AI), Artificial Intelligence Ethics, Mixed Methods Research, Explanatory Sequential Design

TEŞEKKÜR

Öncelikle, tez danışmanlığımı yürüten ve İnsan-Bilgisayar Etkileşimi çalışma alanına hevesli bir Siyaset Bilimi ve Uluslararası İlişkiler mezunu olarak bana istediğim alanda araştırma imkânı tanıyan Doç. Dr. Tayfun YÖRÜK'e süreç boyunca sabırlı rehberliğinden ve emeğinden ötürü teşekkürlerimi dile getirmek isterim. Disiplinlerarası yaklaşıma her zaman desteğini gösteren, bu doğrultuda çalışmaları teşvik ederek değerli tavsiyelerini ve hoşgörüsünü hiçbir zaman esirgemeyen bölüm başkanımız Prof. Dr. Sezgin IRMAK'a, tanıma fırsatı bulduğum için kendimi şanslı hissettiğim, yüksek lisans eğitimim süresince güvenle bilgi ve tecrübelerinden istifade ettiğim Doç. Dr. Nuray YAPICI AKAR'a, Dr. Öğr. Üyesi Güray TONGUÇ'a, yapay zekâ algoritmalarıyla keyifle uğraşmama yardımcı olan Doç. Dr. Ömür TOSUN'a, biyolojiye olan sevgim ve merakımla yöneldiğim Evrimsel ve Sezgisel Algoritmalar'ı daha yakından tanımama vesile olan Doç. Dr. Kerim Kürşat ÇEVİK'e ve Dr. Öğr. Üyesi Hande ERDOĞAN'a teşekkürlerimi sunarım. Tez savunma jürimde yer alan Doç. Dr. Mehmet Özer DEMİR'e değerli katkı ve önerilerinden dolayı ayrıca teşekkür ederim.

Yapay zekâ etiği üzerine düşünmek için ankete ve özellikle mülakatlara, kendi zamanından feragât ederek dahil olan tüm katılımcılara ve veri toplama sırasında iş birliği göstererek daha fazla katılımcıya ulaşmama katkı sağlayan herkese de gönülden teşekkür ederim, onların desteği olmadan bu araştırmayı yürütmek mümkün olmazdı.

Ve tabii beni her konuda koşulsuz destekleyen ve cesaretlendiren aileme minnettarım.

Neslihan Verda ÖZMEN

Antalya, 2024

GİRİŞ

Bir organizmaya dair sahip olduğumuz temsilin, bir makineye dair sahip olduğumuz temsil ile izomorfik olarak yaklaşık bir ilişkiye sahip olduğu fikrine işaret eden, organizma ve makine arasındaki bağlantının tarihinde en temsili pozisyon olarak nitelendirilen *mekanosentrik* (mekanik-merkezli) ilişkinin aksine, Esposito (2019: 119), öne sürdüğü *organosentrik ilişki* hipotezinde, makine ya da çeşitli yapay mekanizmaların aslında biyolojik dünyanın bir izdüşümü ya da uzantısı olduğunu ileri sürer. Bu esasen aletler, makineler ve insan anatomisi arasındaki temasa yönelen ve özellikle çevrenin farklı sosyo-kültürel oluşumları nasıl şekillendirdiği ve beslediğiyle ilgilenen Ernst Kapp'ın önerdiği, varlık ve türevi arasındaki ilişkiyi tanımlayan *organik yansıtma* hipoteziyle ve mekanik eserleri biyomimetik varlıklar olarak gören tüm hipotezler ile bağlantılı bir genellemedir (Esposito, 2019: 136). Kapp, eser üretiminin, insanların çevreleri ve kendileri hakkında bilgi edindikleri evrimsel bir faktör olduğunu savunur; öz farkındalık, pratik faaliyetlerin üretken bir şekilde tekrarlanmasıyla ortaya çıkan bir olgudur ve teknik aletler ya da artefaktlar aynı zamanda adaptasyon ve anlama araçlarıdır (Esposito, 2019: 122-123). Bu noktada, paleontologların ve antik çağ tarihçilerinin çalışmalarına da dayanarak (Latour ve Venn, 2002: 248), Stiegler'in insan ve tekniğin birlikte evrimini tanımlamak için kullandığı *epifilogenez* kavramına değinmek gerekir. Epifilogenez yoluyla kültür, “hafızanın inorganik organizasyonu” olarak mümkün hale gelmektedir (Stiegler, 1998: 174; Introna, 2017; Stiegler, 2018: 9); böylece “insan, aleti icat ederek, teknolojik olarak dışsallaşarak, kendini teknikte icat eder” ve zamanda var olur (Stiegler, 1998: 141). Bu birlikte evrim süreci, organ ve teknik alet arasında giderek daha derin bir eklemlenme meydana getirirken, eserlerin ve makinelerin evrimi zaman içinde belirli bir özerklik de kazanmaktadır. Yine de Kapp, “bir teknoloji ne kadar sofistike olursa olsun ve bir makine ne kadar insan yahut hayvan fizyolojisine benzerliğini kaybederse kaybedsin, tüm eserlerin nihai kökeninin bir projeksiyonda olduğunu” vurgular (Esposito, 2019: 124).

Şu var ki, Günther Anders, *İnsanın Eskimişliği (Die Antiquiertheit des Menschen)* başlıklı iki ciltli eserinde, sunduğu olanaklar artık insanın kavrayabilme yetisinin çok üzerinde olan ürünlerin üretildiğini, bu eksikliği dengeleyebilmek için de insanın pasifleştiğini ve teknolojinin ham maddesi, bir parçası haline geldiğini ileri sürmüştür (Kornwachs, 2021: 61). Zira, “fiziksel ve sinirsel sistemlerimizin uzantıları” olan makinelerin aşırı hızı karşısında “sürekli yeni dengeler aramak zorunda kalınan bir biyokimyasal etkileşimler dünyasında”

(McLuhan, 2013: 174), “aklın karar verme ve sınır koyma işlevi zayıf kalır ve bu orantısızlık yüzünden biz teknik bir sistem karşısında karar verme sürecini başlatamadan, söz konusu sistem bizim üzerimizde etkide bulunur” (Şan, 2022: 129). Teknolojinin, tarihin öznesi haline gelmesi ve insanlara salt “tarihe iliştirilmişlik” kalması durumunda, “söz konusu değişim insanın tüm pratik ve teorik etkinlikleri için öylesine yoğundur ki artık eskisi gibi insan doğasından söz etmek sorunludur” (Anders, 2018: 9; Şan, 2022: 109). Teknolojinin, insan faaliyetlerinin tüm yönleri üzerindeki bu kademeli tahakküm sürekliliğinde, insanlığın, atom bombası aracılığıyla “endüstriyel ölçekte kitlesel ölüm üretimiyle” kendi kendini yok etme koşullarını yaratması (Borowski, 2022) ile başlayan “araç-amaç” kavram ikilisinin geçirdiği dejenerasyon sürecinde (“Efekt amacı aşar” – “*effectus transcendit finem*”), günümüzde “amaçların amacı araçlara araçlık etmeye” dönüşmüştür (Anders, 2018: 23-302). Stiegler’e göre numerik bir episteme evresi olan, içinde bulunduğumuz dijital çağda “teknik sistemlerin toplumsal sistemlerle uyumsuz hizalanmaları” (Şan, 2022: 126-129) ya da Anders’in bu tutarsızlığı betimlediği insan ve teknoloji arasında giderek artan bir asenkronizasyona ilişkin *Prometheusçu uçurum* tabiri (Anders, 2018: 28) dikkate alındığında ivedilik arz eden, “insanların ve insan olmayanların iç-işiksel edimlerinin, asimetritlerinin ve farklılıklarının eş zamanlı bir açıklaması”na ihtiyacımız olduğudur (Barad, 2007: 141; Suchman, 1998: 12). Burada, bu araştırmanın tecessüm etmesine ilham olan Lucy Suchman’ın, “etkileşim metaforunun yeniden keşfi”ne dikkat çektiği çalışmasına yönelmek yerinde olacaktır. Suchman (1998: 11), insanların ve artefaktların karşılıklı bağımlılıklarını kabul etmekten çok uzak, makine failliğine ilişkin günümüz söylemlerinin, faillik alanını insanlardan makine nesillerine kaydırıldığı endişesini dile getirmiştir. Halbuki makinelerin tarihi, “düzenleyici (iktidar/bilgi) pratiklerin ve bunların bedenlerin üretimindeki katılımcı rollerinin tarihine yakından ve girift bir şekilde” bağlıdır (Barad 1998: 23). Suchman (1998: 12), insanlar ile makineler arasındaki sınırların, özellikle tarihsel yollarla ve belirli tarihsel sonuçlarla nasıl inşa edildiğinin hatırlanması gerektiğine dayanarak, bu sınırların “tarihsel ve kültürel olarak özgül bir yeniden biçimlendirme eylemi mahiyetinde restorasyonunu” savunur. Teknolojik sistemler, her biri öncekileri daha belirsiz hale getiren katmanların art arda birikmesiyle bir o kadar opaklaşırken (Latour ve Venn, 2002: 251), bilgilerin “tasfiye ve otomatize edildikleri için bilgi olmaktan çıkıp kapalı sistemler haline geldikleri” *Antroposen* döneminde (Stiegler, 2019: 200; Şan, 2022: 129), **yapay "zekâ"** veya **makine "öğrenmesi"** gibi metaforlar da toplumsal söylemi sürdürülebilir bir şekilde yönlendirmekte ve uzman topluluklarda olduğu kadar kamuoyunda da tahayyülleri ve gelecek vizyonlarını beslemektedir (Bareis ve Katzenbach, 2022: 857). Bu sistemlerin girdi ve çıktılarını sağlıklı bir şekilde analiz etmek ise dolayısıyla güçleşmektedir. Yapay zekânın (YZ)

“öğrenmeyle meşgul ve 'kalbi' olan bir fail olarak tekilleştirilmesi” gizemleştirilmesine katkıda bulunur; “**derin öğrenme**” gibi tekniklerde doğruluk ölçümleri, sezgisel olmayan özelliklerin ya da yorumlanamayan mekanizmaların ardında yatan sosyal ve siyasi riskleri değerlendirmeksizin bu tekniklere sosyal tahmin ve kategorizasyon gücü atfeder; bu teknikler sosyal alanlarda kullanılmaya başlandığında ise hiyerarşileri ve farklılıkları yoğunlaştırırken bunları görünürliğe veya hesap verebilirliğe kapatma potansiyeline de sahiptir (Suchman, 2018).

Söz konusu şartları ve böylelikle Anders’in ifade ettiği “teknoloji tarihi karşısındaki tarihsel rolümüzü” değiştirmenin olanaklarını düşünmeye ise felsefe uğraşısıyla, bilhassa ahlak felsefesi ile, yaşam idealini gerçekleştirmeye yönelik sorgulayıcı bir düşünme etkinliği olarak “ihtimam gösterme” vasıtasıyla yönelinebilir (Anders, 2018: 357; Patočka, 2022: 110; Şan, 2022: 109-130-132). Gerçekte biyomimetik varlıklar olan işlevsel eserlerin yaratılması ve öznenin, eserin aslında çevreyle daha iyi entegrasyon sağlayan organik bir taklit olduğunu algılamaya başlaması ve bir yansıtma olduğunun bilincine varması; bu bilinçli tanıma sayesinde ise eserin nihayet organik süreçleri anlamak için etkili bir epistemik aygıt haline gelmesi evrelerinden (Esposito, 2019: 126) ve *epifilogenez* fikrinden güç alarak, **günümüz ahlaki failinin, tasarlanacak teknolojik sistemlerin gelecek nesillere varoluş olanaklarını sağlama hususunda sorumluluk üstlenmesinin aciliyeti** ileri sürülebilir (Introna, 2017).

İlgili literatürün mevcut ve gelecekteki sosyo-teknik yolları şekillendirmede kritik bir rolü olduğu kabulünden yola çıkarak (Bareis ve Katzenbach, 2022: 858) böyle bir girizgâhın amacı; 21. yüzyılın önemli bir sosyo-teknik fenomeni olarak beliren YZ teknolojilerinin toplumumuzun işleyişine nasıl entegre edileceğinin önemine ilişkin bütüncül bir bakış açısı ortaya koymaktır. Bu araştırma ile ise geleceğin veri güdümlü toplumunun yaşam biçimlerinin bu sistemleri yönlendiren değerler ve motivasyonlar tarafından nasıl etkileneceği ve “sıcakkanlı öznelerin kimliklerinin nasıl dönüşeceği” (Leslie, 2019: 73) sorularının muhatabı olan **üniversite öğrencilerinin YZ etiğine yönelik tutumlarını keşfetmek** amaçlanmıştır. YZ teknolojileri, görünmez bir altyapı olarak kurumsal yönetime ve kurumlar ile toplum arasındaki ilişkiye, mikro-politik düzeyde bedeni etkileyerek öznelliklerin üretimine, benlikle olan ilişkimize, özneler arası ilişkilerimize, gerçeklik temsillerimize ve şimdiki zaman ile geleceğin paylaşılan tahayyüllerine aracılık etmekte iken (Ricaurte, 2022: 728), literatürde mikro analiz seviyesinde çalışmaların temsili dikkat çekmemektedir. Dolayısıyla Latour’un ‘*ahlakın kayıp kütlesi*’ olarak adlandırdığı, gündelik ahlakımızın önemli bir kısmının teknolojik aygıtlara

dayandığı fikrine de istinaden (Latour ve Venn, 2002: 253) bu süreci yakından gözlemlemek, YZ'nin toplumsal ve etik boyutlarına ilişkin görüşlerimizi nasıl şekillendirdiğini ele almak zamanıdır. Zira, yalnızca iş gücünün gelecekteki bir parçası olarak değil, “tam teşekküllü bireyler” ve vatandaşlar olarak da yaşamsal etkileri olan bu sistemler hakkında eleştirel bir şekilde düşünebilmemiz ve onları uygun şekilde bağlamsallaştırma gereğinin ayırında olmamız gerekir (Pötzsch, 2019: 221).

Açıklayıcı sıralı karma yöntem araştırma stratejisine dayandırılan bu çalışmanın ilk bölümünde öncelikle YZ kavramına ve araştırmalarına genel bir bakış sunulmuştur. Takip eden bölümde çok sayıda etik görüşe kulak veren çeşitlendirilmiş bir etik yaklaşım için temel oluşturma amacı güdülerek araştırma hedefi doğrultusunda Anna Jobin ve ekibinin (Jobin vd., 2019: 395), YZ için etik ilkeler ya da kılavuzlar içeren 84 belgeyi içerik analizine tabi tuttuğu çalışmalarında küresel bir yakınsama tespit ettikleri beş ilkeye (adalet, şeffaflık, kötü niyetli olmama, gizlilik ve sorumluluk) ve bu ilkelere ilişkin toplumsal ve etik çıktılara yer verilmiştir. Üçüncü bölüm ise benimsenen nicel ve nitel araştırma metodolojilerine yönelik ayrıntıları, bulguları ve bulguların değerlendirildiği, aynı zamanda potansiyel araştırma yönlerine dikkat çekildiği tartışmayı ayrı birer ana başlık altında içermektedir. Sonuç kısmında, çalışmaya genel bir bakış sağlanarak yazına katkılara ve araştırma kısıtlarına yer verilmiştir.

BİRİNCİ BÖLÜM

YAPAY ZEKÂ VE KAPSAMI

1.1. Yapay Zekâ Nedir?

YZ'ye farklı yaklaşımlar bulunmaktadır ve bu yaklaşımların amaçlarına göre kavramın nasıl tanımlandığı da değişecektir (Russell ve Norvig, 2009: 29; Sayler, 2020: 3; Bringsjord ve Govindarajulu, 2022). YZ'nin bir kavramsal kargaşa olarak, eksik ya da yanlış tanımlanması veya Suchman'ın (2018) altını çizdiği gibi YZ'den tekilmiş gibi bahsedilmesi, bu sistemlerin gerçekliğini gizlemektedir; dolayısıyla etik zorluklarına ilişkin bu doğrultuda herhangi bir değerlendirme de ancak “sorumsuzca bir dikkat dağıtıcı” olacaktır (Taddeo ve Floridi, 2018: 751; Larsson, 2020: 439). Bu son derece tartışmalı tanım meselesi, az evvel vurgulandığı üzere, tanımı yapan kişinin dayandığı disipline bağlı olabileceğinden muhtemelen yönetim üzerinde de kuvvetli bir etkiye sahiptir (Larsson, 2020: 439).

YZ bilimi “sentetik psikoloji”, “deneysel felsefe” veya “hesaplamalı epistemoloji” olarak tanımlanabilir; zira “bilginin ve zekânın doğasını, daha önce mevcut olandan daha güçlü deneysel araçlarla incelemenin bir yolu” olarak görülebilir (Poole ve Mackworth, 2023). Daniel Dennett (1979: 64), YZ'nin, “zekâ ya da bilgi olasılığına dair en soyut araştırma” olarak görülmesi gerektiği görüşünü ortaya koymuştur. Suchman (2018), YZ'yi, “ilgili parametreleri dahili veya harici olarak üretilen geri bildirimlere göre ayarlanabilen bir dizi veri analizi ve işleme tekniği” için kullanılan bir terim olarak tanımlamaktadır. Hatta “yapay zekâ” yerine “bilgi teknolojileri” veya “yazılım” kelimelerinin tercih edilebileceğini ifade eder; bireyler ve türler arasında farklılık gösteren doğal zekâyı dair yanlış kanıya dikkat çekmek gerektiğini savunur. Çoğunluk, zekâyı “artan genlikli, tek boyutlu, doğrusal bir grafik olarak çizme eğiliminde” olsa da (Stiegler, 2018: 6), M. Minsky, J. McCarthy, D. Lenat, E. Feigenbaum ve B. Goertzel gibi bazı YZ araştırmacılarının tanımları da dahil olmak üzere, yetmişten fazla zekâ tanımının bir araya getirildiği çalışmada da görülebileceği üzere (Legg ve Hutter; 2007: 2-8) zekâ, tek bir fenomen değildir; aksine algılama, öğrenme, muhakeme, problem çözme, hafıza gibi çeşitli bilişsel kapasiteleri içerir ve “varlığını kabul etmemize yol açan da bu kapasitelerin üst düzey etkileşimidir” (Zambak ve Vergauwen, 2009: 57).

YZ'yi teknik bir bakış açısıyla tanımlamanın güçlüğüne bir nedeni ise bu sistemlerin tek bir teknolojiden değil, konuşma tanıma ve bilgisayarlı görü gibi alanlardan dikkat ve belleğe kadar değişen bir dizi alt teknik ve disiplinden meydana gelmesidir (Gasser ve Almeida, 2017:

59). Aslında YZ, silah teknolojileri de dahil olmak üzere çok sayıda teknoloji türüne “bir tür bilişsel yetenek” kazandırmak için kullanılabilecek çok çeşitli uygulamaları kapsayan bir “portföy teknolojisi”dir (Boulain vd., 2020a: 8). Bu noktada, Russell ve Norvig (2009: 2)’in, “YZ nedir?” sorusuna literatürün önemli bir kısmını yansıtan bir dizi olası yanıt sunduğu çalışmalarına da başvurulabilir (Bkz. Şekil 1.1). Samoili vd. (2021: 45) de 1955 ile 2021 arasında geliştirilen YZ tanımlarını bir araya getirmişler ve bu kavramın temel özelliklerini ilgili literatüre yansıdığı şekliyle özetlemişlerdir. YZ sistemleri, dayandığı hesaplama modelleri veya teknolojinin mimarisi gibi özellikler açısından daha ayrıntılı da tanımlanabilir (Taddeo ve Floridi, 2018: 751).

	(İnsan performansı bazlı)	(İdeal rasyonalite)
(Muhakeme temelli)	İnsan gibi düşünen sistemler <i>"Bilgisayarların düşünmesini sağlamaya yönelik heyecan verici yeni çaba... tam ve gerçek anlamda zihinleri olan makineler."</i> (Haugeland, 1985) <i>"İnsan düşüncesiyle ilişkilendirdiğimiz, karar verme, problem çözme ve öğrenme gibi faaliyetlerin otomasyonu."</i> (Bellman, 1978)	Rasyonel düşünen sistemler <i>"Hesaplamalı modellerin kullanımı yoluyla zihinsel yetilerin incelenmesi."</i> (Charniak ve McDermott, 1985) <i>"Algılamayı, akıl yürütmeyi ve eylemde bulunmayı mümkün kılan hesaplamaların incelenmesi."</i> (Winston, 1992)
(Davranış temelli)	İnsan gibi hareket eden sistemler <i>"İnsanlar tarafından yapıldığında zekâ gerektiren işlevleri yerine getiren makineler yaratma sanatı."</i> (Kurzweil, 1990) <i>"Şu anda insanların daha iyi olduğu işlerin bilgisayarlara nasıl yaptırılacağına araştırılması."</i> (Rich ve Knight, 1991)	Rasyonel hareket eden sistemler <i>"Hesaplamalı zekâ, akıllı ajanların tasarımı üzerine yapılan bir çalışmadır."</i> (Poole vd., 1998) <i>"Yapay zekâ ...yapay nesnelerdeki akıllı davranışlarla ilgilenir."</i> (Nilsson, 1998) <i>"Bilgisayar biliminin akıllı davranışların otomasyonu ile ilgilenen dalı."</i> (Luger ve Stubblefield, 1993)

Şekil 1.1. Tarihsel YZ Tanımlarının Taksonomisi

Kaynak: (Bringsjord ve Govindarajulu, 2022; Russell ve Norvig, 2009: 2; Sayler, 2020: 3) çalışmalarından uyarlanmıştır.

Bu teknolojiye dair temel kavramlara yönelik ortak bir sınıflandırma sistemine ulaşma fikri önem arz ettiği için ve uluslararası arenada anlamsal uyumu sürdürebilmek amacıyla YZ’nin sorumlu ve güvenli bir şekilde uygulanmasını teşvik etmeyi amaçlayan düzenleyici

girişimler de çeşitli tanımlar yayımlamayı sürdürmektedir. Genellikle bu sistemlerin farklı düzeylerde özerkliğe sahip olduğu konusunda hemfikir olunsa da ve insanların bir sistemin hedeflerinin tanımlanmasında ve veri veya girdi sağlanmasında rol oynadığı kabul edilse de hangi sistemlerin YZ kapsamına girdiği konusunda bir uzlaşma eksikliği bulunduğu ifade edilebilir. Avrupa Birliği Yapay Zekâ Yasası, YZ'nin veya bir YZ sisteminin “Farklı seviyelerde özerklikle çalışmak üzere tasarlanmış ve açık ve örtülü hedefler için fiziksel veya sanal ortamları etkileyen tahminler, öneriler veya kararlar gibi çıktılar üretebilen makine tabanlı bir sistem” anlamına gelmesini öngörmektedir (European Parliament, 2023: 137). Yine de “Kaliforniya Meclis Yasası 331”, “OECD YZ İlkeleri”, “ISO JTC 1/SC 42 teknik komitesi tarafından yayınlanan ISO/IEC 22989:2022”, “NIST YZ Risk Yönetimi Çerçevesi”, “Yapay Zekânın Güvenli, Emniyetli ve Güvenilir Şekilde Geliştirilmesi ve Kullanılmasına ilişkin 75191 sayılı yürütme emri”, “Kanada Yapay Zekâ ve Veri Yasası”, “Brezilya'nın 2238 Sayılı Yasa Tasarısı”, UNESCO Yapay Zekâ Etiği Tavsiye Kararı”, “Connecticut Senato Tasarısı 1103” dahil olmak üzere toplam 17 düzenleyici girişimin YZ tanımlamaları değerlendirildiğinde;

- içerik, tahminler, öneriler ve kararlar olarak sistem çıktıları,
- modeller için veri (ve girdiler) sağlamak ve modellerin hedeflerini tanımlamak başta olmak üzere insanların rolü,
- otonomi (özerklik) ve
- YZ'yi karakterize eden teknolojinin türü olmak üzere, dört temel temanın öne çıktığı dikkat çekmektedir (Gulley ve Hilliard, 2023).

1.2. Yapay Zekâ Araştırmalarına ve Felsefi Argümanlara Kronolojik Bir Bakış

Aslında tarih boyunca her yeni teknoloji hem zekâ için teknolojik metaforlar hem de zihni modelleme mekanizmaları olarak önerilmiş, “düşünmenin yüksek sesle konuşmak veya kalem ve kağıtla bir cevap üzerinde çalışmak gibi sembolik akıl yürütme olduğu” görüşünü benimseyen Hobbes'ı takiben sembolik akıl yürütme fikri, Descartes, Pascal, Spinoza, Leibniz (1646-1716) ve diğer Avrupa zihin felsefesinin öncüleri tarafından daha da geliştirilmiş ve sembolik işlemler, bilgisayarların gelişmesiyle daha somut hale gelmiştir (Poole ve Mackworth, 2023). İnsan benzeri makinelerin revaçta olduğu, “mistisizm ve büyü korkusunun” yerini Sanayi Devrimi'ne bırakacağı zamanlarda, üzerinde satranç tahtası olan büyük bir dolabın arkasında oturan, oryantal giysili tahta bir adam olan “Mekanik Türk” veya “Türk”, aslında bir aldatmaca olsa da oynadığı oyunların çoğunu kazanarak insan zekâsının sınırları ve insan düşüncesini kopyalayabilen bir makine inşa etme olasılığı hakkında tartışmaları

hararetlendirmiştir (Ward, 2023). Amazon'un "yapay yapay zekâ" ya da "insan bulutu" olarak lanse edilen web sitesi MTurk'e adını dahi veren bu otomat satranç oyuncusunun 1819'da seyircilerinden olan Charles Babbage da bir makinenin gerçekten satranç oynayacak şekilde nasıl yapılabileceği üzerine düşünmek için ilham almış ve genellikle ilk dijital bilgisayar olarak kabul edilen ve "Analitik Motor" olarak adlandırılan genel amaçlı otomatik mekanik bir hesap makinesini icat etmiştir (Hook, 2016; Kaganer vd., 2012; Schwartz, 2019).

YZ araştırma alanı ise nörofizyolog Warren McCulloch ve matematikçi Walter Pitts'in "yapay sinir ağları" mimarisini tanımladığı makaleye kadar uzanabilir (McCulloch ve Pitts, 1943; Russell ve Norvig, 2009: 16). Alan Turing'in "Makineler düşünebilir mi?" sorusunu ortaya attığı ünlü "Computing Machinery and Intelligence" makalesi ve devamında önerdiği "Turing Testi"ni müzakere edebilecek makineler inşa etme çağrısı ile de YZ fikri belirgin hale gelmiştir (Turing, 1950). Resmi olarak 1956 yılında New Hampshire'daki Dartmouth Koleji'nde, DARPA tarafından desteklenen; John McCarthy, Claude Shannon, Marvin Minsky, Arthur Samuel, Trenchard More, Ray Solomonoff, Oliver Selfridge, Allen Newell ve Herbert Simon olmak üzere on düşünürün katıldığı bir yaz çalıştayının ardından başlamıştır ve "yapay zekâ" terimi de burada icat edilmiştir (Bringsjord ve Govindarajulu, 2022). Üç yıl sonra Cahit Arf da Erzurum Atatürk Üniversitesi'nde 1958-59 Halk Konferansı bildirisinde "Makine Düşünebilir mi ve Nasıl Düşünebilir?" çalışmasıyla konuyu ele almış ve düşüncelerini sunmuştur (Arf, 1959). Söz konusu çalıştay önerisinde "Öğrenmenin her yönü veya zekânın diğer herhangi bir özelliği o kadar kesin bir şekilde tanımlanabilir ki bir makine bunu simüle edebilir." iddiasında bulunulmuş ve araştırma alanı, "Zayıf" YZ'nin mümkün olduğu varsayımı üzerine inşa edilmiştir (McCarthy vd., 2006: 12; Russell ve Norvig, 2009: 1020).

"Zayıf" YZ, "Güçlü" YZ ("Genel" ya da "Yapay Genel Zekâ" (AGI)) ve "Süper" YZ hedeflerine de dikkat çekmek şu noktada yerinde olacaktır. Turing Testi'ni geçen bir makinenin yine de gerçekten düşünemeyeceği, sadece düşünmenin bir simülasyonu olacağı iddiası filozoflar tarafından "**Zayıf YZ**" hipotezi olarak adlandırılmıştır; bu, daha önce de belirtildiği üzere günümüzde yaygın olan YZ türüdür (Samoili vd., 2021: 11). Zayıf YZ sistemleri, görüntülerden nesneleri veya insanları tanımak, dili tercüme etmek veya oyun oynamak gibi ayrı 'akıllı' görevleri yerine getirebilen karmaşık yazılım programlarıdır; ancak doğaları gereği programlamaları ile sınırlıdırlar ve yalnızca amaçlanan görevler ve çalışma ortamı için güvenilir bir şekilde çalışırlar (Boulain vd., 2020a: 7). Bir insandan daha iyi performans göstermese de bir insanın dünyayı anlamlandırma ve çevresine dair bir anlayış geliştirme

becerisine eşdeğer nitelikte YZ ise “**Güçlü YZ**” olarak ifade edilmektedir (Boulanin vd., 2020a: 7). Birinci dereceden mantık ve uzantıları; doxastic tutumların ve deontik muhakemenin modellenmesine uygun intansiyonel mantıklar; tümevarım mantığı, olasılık teorisi ve olasılıksal muhakeme gibi YZ’de kullanılan temel formalizmlerin ve tekniklerin çoğu felsefeden çıkmış, felsefede çokça kullanılmakta ve geliştirilmekte olsa da çoğu YZ araştırmacısı ve geliştiricisi, soyut tanımlar üzerinde düşünmekten ziyade aslında yararlı ya da kârlı eserler inşa etmekle uğraşmaktadır (Bringsjord ve Govindarajulu, 2022). Yine de “Güçlü” YZ ile fenomenal bilinç de dahil olmak üzere sahip olduğumuz tüm zihinsel güçlere sahip makineler yaratmanın amaçlandığı, “Zayıf” YZ ile de insan zihninin tüm repertuarına sahipmiş gibi görünen bilgi-işlem makineleri inşa etmenin amaçlandığı söylenebilir (Searle, 1980: 417; Flowers, 2019: 1).

Bazı ilk YZ araştırmaları, “yüksek yetileri ve yoğunlaştırılmış bilinçli düşünceyi içeren” satranç oynama ya da matematiksel alıştırmalar çözme gibi kombinatoryal görevlere odaklanmıştır ve saniyede milyarlarca hesaplama hızında birçok alternatif hamle hesaplayabilen bilgisayarlar için bu görevlerin çok uygun oldukları görülmüştür (Elements of AI, 2023). Diğer yandan, küçük bir organizma olan “*C. Elegans nematod*” solucanının algıları ve eylemleri arasında hesaplamalı haritalama yapmak için on yılı aşkın bir süredir yürütülen “OpenWorm” projesinde henüz sürünmek ve yemek için simüle edilmiş solucanlar elde edilememiştir; emekleme ve sürünme, basit ama hala zorlayıcı gösterge görevleridir (Allyn-Feuer ve Sanders, 2023: 15). Hangi görevlerin kolay hangilerinin zor olduğunu bilmenin güçlüğü YZ’nin anlamı konusundaki kafa karışıklığının bir diğer kaynağı iken bu sistemleri, Yiu vd. (2023: 2)’nin önerdiği gibi söz konusu yetenekleri kazanana kadar “genel” zekâlar olarak değil, kültürü taklit eden ve aktaran araçlar olarak görmek daha uygun olacaktır. YZ felsefesi bünyesinde argümanlar burada sıralanamayacak kadar çeşitli olsa da Güçlü YZ’ye karşı iyi bilinenler; John Searle’ün Çince Odası Argümanı, Gödelci Argüman ve Amerikan felsefi bölümlerine hâkim analitik felsefeden uzak bir şahsiyet olarak YZ alanını “simyaya” benzeten Hubert Dreyfus’un ve kardeşi Stuart Dreyfus’un argümanları olarak sayılabilir (Astakhov, 2020: 161; Bringsjord ve Govindarajulu, 2022).

Araştırma tarihine dönecek olursak, bazı problemlerin beklenenden daha zor ve diğerlerinin de zamanın teknolojileri ile aşılmaz olduğu ortaya çıktıkça 1965-1980 ve 1985-1997 arası birer duraklama dönemi yaşanmıştır (Boulanin vd., 2020a: 10). İlk olarak Marvin Minsky ve Roger Schank’in “Yapay Zekâ Kışı” olarak adlandırdığı kavram, 1984 yılında Amerikan Yapay Zekâ

Derneği'nde (şimdiki adıyla Yapay Zekâ İlerleme Derneği ya da AAAI) ortaya atılmıştır (Umbrello, 2021: 7). Floridi'nin (2020: 1) ifadeleriyle bu kavram, “gerçekleri süslemeksizin YZ'nin bir teknoloji olarak hakikaten ne yapabileceği veya yapamayacağı konusunda uzlaşmaya varıldığı, yatırımcıların daha seçici hale geldiği ve gazetecilerin bir sonraki trendi körüklemek için YZ hakkında yazmayı bıraktığı” sürece karşılık gelir. YZ'nin alt sorunlarına ve görüntü tanıma veya tıbbi teşhis gibi gerçek dünya meselelerine uygulanmasına daha fazla odaklanıldıkça 1990'ların sonlarına kadar araştırmalarda ilerleme hızlanmaya başlamıştır. IBM'in özel olarak üretilmiş benzersiz bir süper bilgisayarında çalışan bir satranç oynama uzman sistemi olan Deep Blue'nun 1997 yılında Sovyetler doğumlu dünya satranç şampiyonu Garry Kimovich Kasparov'u alt etmesi (Britannica, 2023), ABD Savunma Bakanlığı'nın deneysel DARPA bölümünün bağımsız YZ disiplinlerini bir araya getirdiği ve Apple Inc.'in Siri'sine öncülük edecek CALO (Cognitive Assistant that Learns and Organizes) (Nusca, 2013), 2000'li yıllarda DARPA Grand Challenge yarışmalarında sürücüsüz arabaların başarısı ve yine IBM'in doğal dil işleme, bilgi çıkarımı, bilgi temsili ve makine öğrenimi dahil olmak üzere YZ'nin birçok farklı alanının oldukça karmaşık bir uygulaması olan Watson bilgisayar sisteminin “Jeopardy!” adlı TV yarışma programında 2011 yılında insan rakiplerine elde ettiği zafer, önemli atılımlardan birkaçıdır (CBDDO, 2021: 93). 2010 yıllarına gelindiğinde ise e-ticaret ve sosyal medya benzeri çeşitli kaynaklardan elde edilebilen ve hem insanlar hem de makineler tarafından üretilen "büyük veri" setlerinin sürekli genişleyen kullanılabilirliği (Knight, 2023), bilgisayar işlem gücündeki artışlar, bulut bilgi işlem platformlarının yükselen hızı ve kapsamı, nörobilişimsel ve olasılıksal tekniklerdeki yeniden canlanma ve giderek daha karmaşık hale gelen algoritmalar ile derin öğrenme gibi önemli ölçüde geliştirilmiş makine öğrenimi yaklaşımları YZ alanına olan ilgiyi yoğunlaştırmıştır (NSTC, 2016: 6; Bringsjord ve Govindarajulu, 2022).

1.3. Makine Öğrenmesi ve Yapay Zekâ Alt Alanları

YZ yetenekleri;

- “**Etkileşim zekâsı**” altında sırasıyla görüntülerin, ses verilerinin ve doğal dilin işlenmesiyle ilgili beceriler olan bilgisayarlı görü, bilgisayarlı işitme, bilgisayarlı dilbilim,
- “**Analitik zekâ**” altında zaman serisi verilerinin analizi ve ileriye dönük düşünülmesi olan tahmin, kümelere dayalı tüm analizleri ve veri kümelerindeki yapıları bulmayı içeren keşif, sorunları çözmek için uzun vadeli stratejiler ve çok adımlı planlar

geliştirmeye olanak tanıyan makine öğrenimi yaklaşımlarını tanımlayan planlama ve arama,

- “**Hareket ve yaratıcı zekâ**” altında ise fiziksel araçların yönlendirilmesi ve kontrolü ile ilgili yetenekleri tanımlayan robotik ve kontrol, makine öğrenimi sistemleriyle yeni görüntülerin veya ses verilerinin oluşturulmasını ya da “insanların büyük olasılıkla düşüneyemeyeceği” yeni yapılar veya yaklaşımlar bulmak için mühendislik gibi disiplinleri artırmak için makine öğreniminin uygulanmasını tanımlayan yaratıcılık olarak sekiz geniş kategoriye ayrılabilir (European Institute of Innovation and Technology, 2021: 13).

YZ sistemlerinin pek çok alt alanda çok sayıda uygulaması bulunmaktadır. İnsan konuşmasını yazılı formata dönüştürmek için doğal dil işlemeden (NLP) faydalanan ve birçok mobil cihazın sistemlerine dahil ettiği *otomatik konuşma tanıma* (ASR), e-ticaret sitelerinde sanal temsilcilerle mesajlaşma botları gibi görevlerde karşımıza çıkan *müşteri hizmetleri*, sosyal medyada fotoğraf etiketlemeden sürücüsüz araçlara çeşitli uygulamaları bulunan *bilgisayarlı görü*, çevrimiçi perakendecilerin ödeme işleme esnasında müşterilere ilgili ek öneri yapmak için kullandıkları *tavsiye motorları* ve insan müdahalesi olmadan günde milyonlarca işlemin gerçekleştiği *otomatik hisse senedi ticareti* YZ sistemlerinin en yaygın kullanım alanlarından sadece birkaçıdır (IBM, 2023a). Sektörel uygulamalardan bahsetmeden önce makine öğrenimi ve ilgili YZ alt alanlarını detaylandırmak gerekir.

1.3.1. Makine Öğrenimi

Bir YZ dalı olarak “**makine öğrenimi**” (ML), ideal performans örnekleri verildiğinde veya tekrarlanan deneyimler aracılığı ile bir görev üzerindeki performanslarını geliştiren sistemler oluşturmakla ilgilidir (Bringsjord ve Govindarajulu, 2022). YZ araştırmalarının başlangıcından beri var olmakla beraber, belirli dönemlerde popülerleşse de pratik kullanımı sınırlı olduğu için ya da finanse edilemez hale geldiğinden uzun süre marjinal bir alt alan olarak kalmıştır. 2000’li yıllara gelindiğinde, bilişsel psikolog ve bilgisayar bilimci Geoffrey Hinton’ın farklı sinir ağları katmanlarını ve ‘geriye yayılma’ olarak bilinen süreci birleştirmeyi öngören, yapay sinir ağlarının görüntü veya konuşma tanıma, verileri sınıflandırma, davranışları tahmin etme, dili tercüme etme veya satranç gibi oyunları oynama gibi örüntü tanımayı içeren herhangi bir görevde geleneksel YZ programlama tekniklerinden büyük bir farkla daha iyi performans göstermesini sağlayan teknik bir değişiklik öne sürmesi ile ilgili yeniden canlanmıştır. Öğrenme çözümleri, insan beyninin çalışma şekline göre modellenmiş

olan, üç veya daha fazla katmana sahip sinir ağlarının kullanılmasıyla ayırt edilen, ML'nin bu bir alt kategorisi “**derin öğrenme**” (DL) ise daha büyük veri, daha fazla depolama ve hesaplama gücü, yüksek performanslı sunucu kümeleri gibi ciddi altyapılar gerektirir (AWS, 2023). Özellik mühendisliği sürecini asgarî insan müdahalesi ile gerçekleştirmesi ile öne çıkmaktadır; tarihi belge görüntülerini daha yüksek çözünürlüğe kavuşturma (Süper-Çözünürlük – Super Resolution: SR) ve belge görüntülerindeki metinleri parçalara ayırma gibi kullanım alanlarında tercih sebebidir (Crabtree, 2023).

ML, gelişmiş örüntü tanıma gerektiren bilgisayarlı görü (nesneleri, insanları veya durumları tanıma), doğal dil işleme (ses ve konuşma tanıma) ve diğer her türlü sinyal tanıma (akustik veya elektromanyetik izleri) gibi makine algılama görevlerinden, veri sınıflandırma, tahmine dayalı analiz, anomali tespiti, sentetik veri manipülasyonu ve üretimi, optimizasyon gibi veri yönetimi görevlerine kadar her türlü görevi otomatikleştirmek için güçlü bir araçtır (Boulanin vd., 2020a: 11). Konuşma tanıma sistemlerinde, spam filtrelerinde, çevrimiçi dolandırıcılık tespit sistemlerinde, ürün tavsiye sistemlerinde ve daha birçok mecrada kullanılmakta olan ML algoritmaları;

- algoritmanın hem girdisini hem de çıktısını belirten örnek veriler üzerinde eğitilen ve ML alanını domine eden “**denetimli öğrenme algoritmaları**”,
- etiketlenmemiş veriler kullanılarak eğitilen, öğrenilmesi gereken girdi ile ilişkili herhangi bir fonksiyonun bulunmadığı, özellikle veri madenciliğinde faydalanılan “**denetimsiz öğrenme algoritmaları**”,
- sürekli olarak hareket ettiği ve algıladığı bir ortamda serbest bırakılarak davranışları hakkında ödül veya ceza şeklinde geri bildirim alan “**takviyeli**” ya da “**pekiştirmeli öğrenme algoritmaları**” olarak farklı kategorilerde incelenir (Bringsjord ve Govindarajulu, 2022).

ML algoritmaları, büyük hacimli yapılandırılmamış verilerin işlenmesini içeren görevler için çok uygun olsa da bu algoritmaların yalnızca üzerinde eğitildikleri veriler kadar iyi olacaklarının da altı çizilmelidir; zira eğitim veri seti temsili değilse sistem başarısız olabilir, kötü performans gösterebilir ya da çalışmanın ilerleyen bölümlerinde daha detaylı ele alınacağı üzere mevcut önyargıları güçlendirerek veya yenilerini yaratarak insanların karar ve eylemlerini etkileyebilir (Boulanin vd., 2020a: 12). Açık kurallara ve mantığa göre çalışan geleneksel elle kodlanmış YZ sistemlerinin aksine, makinelerin veri içinde istatistiksel bir ilişki bularak öğrenmesi de işleyişlerini anlaşılmaz ve davranışlarını öngörülemez kılabilmektedir ve

bu da özellikle arabalar, uçaklar ve silahlar gibi güvenlik açısından kritik sistemler ya da muharebe operasyonlarında kullanılmak üzere otomatik hedef tanıma gibi askeri uygulamalar söz konusu olduğunda benimsenmelerini sorunlu hale getirebilecek birçok teknik zorluğa ve istenmeyen sonuca işaret etmektedir.

1.3.2. Bilgisayarlı Görü

Dijital görüntülerin işlenmesi gibi görüntü tanıma görevleri ile yakından ilgili olan **bilgisayarlı görü** (computer vision), "işlevsel bir birimin görüntü veya videoyu temsil eden verileri alma, işleme ve yorumlama yeteneği" olarak tanımlanabilir (ISO/IEC 22989: 2022(E): 51). Derin evrişimli sinir ağlarının (Deep Convolutional Neural Network: DCNN) kullanımı, görüntü sınıflandırma görevlerindeki yüksek doğrulukları ve eğitim ve tahmin performansları nedeniyle son yıllarda popülerlik kazanmıştır. Bir dizi görüntüden belirli görüntülerin tanımlanması (örneğin, bir dizi hayvan görüntüsünden köpek görüntüleri), otonom araçlarda nesne algılama ve tanımlama ve trafik sinyali, tıbbi görüntüleme hastalık ve anomali tespiti, kalite kontrol (bir montaj hattındaki kusurlu parçaların tespit edilmesi gibi) ve yüz tanıma, bilgisayarlı görü ve görüntü tanımaya dayalı YZ uygulamalarına örnek olarak verilebilir.

1.3.3. Doğal Dil İşleme

Makinelerin düşünüp düşünemeyeceğinin bir ölçütü olarak bir imitasyon oyununa dayandırılan Turing Testi, dili temel alır. Bilgisayar ajanlarının, insanlarla iletişim kurmak ve yazılı dilden bilgi elde etmek için doğal dilleri işleyebilmesi ve insanların kullandığı muğlak, dağınık dilleri kısmen de olsa anlaması gerekir (Russell ve Norvig, 2009: 860). Zira, "Makine dil yeteneğinin sınırları, bir makinenin aracı dünyasının sınırları anlamına gelir." (Wittgenstein, 1922: 74; Zambak ve Vergauwen, 2009: 57). Hesaplamalı dilbilimi ile istatistiksel, ML ve DL modellerini birleştiren **doğal dil işleme** (NLP), bilgisayarların insan dilini metin veya ses verisi şeklinde işlemesini ve konuşmacının veya yazarın niyeti ve duygusuyla birlikte anlamasını sağlamaktadır. Joshua Weizenbaum tarafından 1966'da oluşturulan ELIZA, NLP aracı olan dil modellerinin en eski örneklerindendir; belirli anahtar kelimelere ağırlık uygulayarak ve kullanıcıya buna göre yanıt vererek bir bilgisayarın Rogerian bir psikoterapisti nasıl taklit edebileceğini göstermiştir (Weizenbaum, 1966: 42; Bostrom, 2014: 6; Tarnoff, 2023). Modern büyük dil modelleri (LLM) ise yazım ve dilbilgisi denetiminden (Hu vd., 2021: 1) metin özetlemeye, duygu analizinden çeviriye, doğal dilde yazılmış bir görev için program kodu oluşturmadan bir programın yeniden programlamasına ya da diğer programlama dillerine çevrilmesine (Federal Office for Information Security, 2023: 7) kadar çeşitli uygulamaları

desteklemek için uyarlanabilir. En iyi bilinen dil modellerinden bazıları, Google’ın 2017 yılında “Attention Is All You Need” (“Tüm İhtiyacınız Olan Dikkat”) başlıklı makalesinde tanıtılan “**transformatör modeli**”ne dayanır (Vaswani vd., 2017; IBM, 2023b). Piyasada OpenAI ile Google, Microsoft, Meta gibi büyük şirketlerin LLM’lerinin yanı sıra Falcon 40B ve StableLM gibi açık kaynak birçok rakip model bulunmaktadır.

1.3.4. Üretken Yapay Zekâ ve Temel Modeller

Çeşitli formları ve kullanım alanları bulunan “**üretken YZ**”, belirli bir veri kümesinin altında yatan kalıpları veya dağılımları yakalamayı amaçlayan, böylelikle bu veri kümesiyle benzer özellikleri paylaşan çıktılar üretebilen makine öğrenmesi modellerini ifade etmektedir. Son yıllara kadar YZ araştırma alanına hâkim olan göreve özgü modellerin aksine, önceden bir görev tanımlama ihtiyacını ortadan kaldıran transformatörler sayesinde minimum ayar ile farklı görevler için kullanılabilen modellere ise “**temel modeller**” adı verilir (IBM, 2023b). Genellikle öz-denetimli öğrenme yoluyla çok büyük miktarlarda etiketlenmemiş veri üzerinde eğitilen bu modeller, bir durum hakkında öğrendiği bilgileri diğerine uygulayabilir ve uygulamaları güçlendirebilecek çeşitli yetenekler kazanır; farklı modaliteleri işleme (dil, görü), fiziksel dünyayı etkileme (robotik), akıl yürütme ve insanlarla etkileşim kurma (etkileşim) gibi potansiyel yetenekler bunlara örnek olarak gösterilebilir (Bommasani vd., 2022: 14-15). OpenAI’nin GPT serisinin en büyük modeli olarak piyasaya sürülen, hem dili hem de görüntüleri işleyebilen ve üretebilen çok modlu GPT-4, simüle edilmiş bir baro sınavını sınava girenlerin ilk %10’u civarında bir puan ile geçmiştir (OpenAI, 2023). Temel modeller, model mimarileri ve eğitim prosedürleri, veri sistemleri, evalüasyon ve teori gibi altta yatan teknolojileri ile YZ sistemlerinin nasıl inşa edildiğini kökten dönüştürerek bir paradigma değişimini tetikleme potansiyeline sahiptir; bu modellerin biyotıp, sağlık, hukuk ve eğitim gibi yüksek etkili uygulamaları etkileyerek ekonomik büyümede Sanayi Devrimi’ne benzer şekilde artışlara yol açabileceği ekonomistler tarafından ileri sürülmektedir (Bommasani vd., 2022: 15-16-151).

1.3.5. Robotik Sistemler

Çoğu durumda robotik, YZ gerektirmese de YZ alanındaki gelişmeler vesilesiyle ev eşyalarında, tarım, sağlık ve hizmet sektörlerinde ve imalat sanayiinde, uzay araçlarında ve askeri sahada bu teknolojiye uygulanan YZ örnekleri artık sıklıkla karşımıza çıkmaktadır. Çevreye fiziksel kuvvet uygulamalarını sağlayan bacaklar, tekerlekler, mafsallar ve kısıkaçlar gibi efektörler, çevrelerini algılamalarını sağlayan kameralar ve lazerler, kendi hareketlerini

ölçmelerini sağlayan jiroskoplar ve ivmeölçerler dahil olmak üzere çeşitli sensörler ile donatılan robotlar, fiziksel dünyayı manipüle ederek belirli görevleri yerine getiren fiziksel ajanlardır ve üç ana kategori altında incelenmeleri mümkündür (Russell ve Norvig, 2009: 971-972).

Uluslararası Uzay İstasyonu'nda (ISS) ziyaret araçlarının yanaşmasında ve ABD ve Japon modüllerindeki harici yüklerin yakalanmasında önemli bir rol oynayan “Canadarm2”, Japon Deney Modülü Uzaktan Manipülatör Sistemi ve Avrupa Uzay Ajansı tarafından tasarlanan, ISS Rus segmentine hizmet veren “Avrupa Robotik Kolu” (European Robotic Arm: ERA) gibi “**manipülatörler**” veya “**robotik kollar**”, en yaygın endüstriyel robot türüdür; görevlerin yalnızca robotun çalışma alanının sınırları içinde gerçekleştirilebildiği fabrika zeminlerine cıvatalanırlar (ESA, 2010; Pedersen, 2021).

Gözetim (sürveyans), ekin ilaçlama ve askeri operasyonlar için yaygın olarak kullanılan dronelar ya da insansız hava araçları (UAV), genellikle UAV'leri desteklemek için kullanılan İnsansız Kara Araçları (Unmanned Ground Vehicle: UGV), derin deniz keşiflerinde kullanılan otonom su altı araçları (Autonomous Underwater Vehicle: AUV) ve İnsansız Deniz Araçlarını (Unmanned Marine Vehicle: UMV) da içeren, mayın karşı tedbirleri, deniz güvenliği ve denizde engelleme operasyonlarına destek gibi görevleri yerine getirmek için kullanılan İnsansız Suüstü Araçları (Unmanned Surface Vehicle: USV) gibi araçlar “**mobil robotlar**” olarak adlandırılmaktadır, ikinci kategoride yer alırlar (Stolfi vd., 2021: 1).

Son robot türü ise mobilite ve manipülasyonu bir araya getirmektedir; efektörlerini sabitlemiş manipülatörlerden daha uzağa uygulayabiliyor olsalar da sabitlemenin sağladığı rijitliğe sahip olmadıkları için görevleri daha da zorlaşmaktadır. Bu robotlar, patlayıcı maddelerin imhası (Explosive Ordnance Disposal: EOD), muharebe mühendisliği, havacılık, enerji, petrol ve gaz dahil olmak üzere çok sayıda sektördeki tehlikeli görevlerde kullanılmaktadır (Pedersen, 2021). İnsana benzeyen ve tipik olarak gerçek insan ifadelerini, etkileşimlerini ve hareketlerini taklit etmek üzere tasarlanan Ameca (Engineered Arts), Atlas (Boston Dynamics), Pepper (Softbank Robotics), Robonaut 2 (NASA ve General Motors) ve Sophia (Hanson Robotics) gibi humanoid robotlar ve androidler de “**mobil manipülatörler**” kategorisi altında ele alınır. Burada, etik YZ alanına dair farkındalığın gereğiyle yakından ilgili olduğu için, robotikçi Masahiro Mori'nin 1970 yılında “Energy” adlı bir Japon dergisinde yayınlanan “*Bukimi No Tani*” başlıklı makalesinde ortaya attığı “**Tekinsiz Vadi**” hipotezine

ilişkin bir parantez açmak, ilerleyen tartışma için faydalı olabilir. Bu hipotez, bir robot giderek bir insana benzedikçe bireyin robota aşinalığının da ürkütücülük ve şüphe sebebiyle nispeten ani ve dramatik bir düşüşün söz konusu olduğu bir noktaya kadar yükselmeye devam ettiğini (sıkmayı tercih ettiğiniz bir robotik elde bir insan elinin kemiksi özelliklerini hissedemediğinizde yaşayabileceğiniz gibi); robotun insana benzerliği arttıkça da sonunda gerçek bir insan formuyla neredeyse aynı olan daha gelişmiş bir versiyonla karşılaşılana kadar aşinalığın tekrar yükselişe geçtiğini ileri sürmektedir (Mori, 2012; Paulsen, 2017: 64; Wang vd., 2015: 393). Mevcut araştırmalar, hipotezdeki bağımsız ve bağımlı değişkenlerin, tekinsizlik fenomeninin bilişsel temellerinin daha derinlemesine anlaşılması, sadece görünüşe dayalı robotik yönlerin değil, kaçınılmaz olarak iç içe geçmiş pek çok boyutun dikkate alınması ve tekinsiz vadiyi tespit etme ihtiyacı ile sınırlı kalınmadan, bilhassa tekinsizlik fenomeninin incelenmesi gibi gereksinimlere işaret eder (Wang vd., 2015: 403). Yine de Mori'nin “ikinci zirveye ulaşmak için **insana benzerlik derecelerini artırmaya çalışmanın doğasında var olan risk nedeniyle**, tasarımcıların, orta derecede bir insan benzerliği ve önemli bir yakınlık duygusu ile sonuçlanan ilk zirveyi hedef almaları” ve “*kasıtlı olarak* insan dışı bir tasarım peşinde koşarak güvenli bir yakınlık düzeyi yaratmaları” önerisi vurgulanmalıdır (Mori, 2012).

Robotik alanı ayrıca, insanlar için yapay uzuvlar, kulaklar ve gözler gibi protez cihazlarını, akıllı ortamları (sensörler ve efektörlerle donatılmış bir evin tamamı gibi) ve iş birliği yapan küçük robot sürüleri aracılığıyla robotik eylemin gerçekleştirildiği çok gövdeli sistemleri de içermektedir.

1.4. Sektörel Uygulamalar

YZ sistemleri, her gün daha fazla ağa bağlı gelen toplumumuzun siber fiziksel çerçevesini ve dokusunu dikkate değer bir ölçüde şekillendirmektedir (Leslie, 2019: 3; Stahl, 2023: 2). Karar vermeye destek olabildiği, tavsiyelerde bulunabildiği ve belirli görevlerin otomatikleştirilmesine yardımcı olabildiği için tarım ve çiftçilik, otomotiv, bankacılık ve finansal teknoloji, savunma, eğitim, enerji, sağlık, hukuk, üretim, medya ve eğlence, karma gerçeklik, kamu sektörü, perakende ve pazarlama, güvenlik, uzay teknolojisi ve telekomünikasyon gibi pek çok sektörde kullanılır (ISO/IEC 22989: 2022(E): 54-55). E-öğrenme, izleme, ölçme ve değerlendirme ve öneri sistemleri (T.C. Cumhurbaşkanlığı Bilim, Teknoloji ve Yenilik Politikaları Kurulu, 2021: 203), akıllı ev teknolojisi ve elektroniği, alışveriş ve e-ticaret, giyilebilir cihazlar, çevrimiçi navigasyon, Amazon Echo ve Alexa gibi sesli asistan hizmetleri, sosyal ağ hizmetleri, video ve müzik akışı hizmetleri, dil çeviri

hizmetleri, flört uygulamaları, çevrimiçi uçuş rezervasyonu uygulamaları, pasaport kontrolü, akıllı telefonların donatıldığı yüz tanıma, AR ve VR benzeri teknolojiler YZ'nin entegre edildiği, günlük hayatta sıklıkla etkileşimde olduğumuz kullanım alanlarındandır (Ko ve Leem, 2021: 218). Bilgisayar oyunları da daha iyi donanım tasarlamak veya yasal sorunlarla uğraşmak gibi temel YZ araştırmalarının kapsamı dışında kalabilecek ayrıntıları soyutlayarak ya da ortadan kaldırarak YZ teknikleri için sağlam bir test ortamı sağlar (Bringsjord ve Govindarajulu, 2022). Oyuncuların sınırlı kaynaklara sahip olduğu bir orduyu yönettiği oyunlar olan “gerçek zamanlı strateji oyunları”, YZ'nin ticari kullanımı için oldukça verimli görülen bir oyun alt sınıfıdır (Robertson ve Watson, 2014: 76).

Önemli bir strateji kaynağı olarak da uluslararası rekabetin odak noktası haline gelen YZ, jeopolitik rakipler olarak tasvir edilseler de askeri bütçelerinde açık ara fark söz konusu olan ABD, Çin ve Rusya başta olmak üzere pek çok ülke tarafından bir dizi askeri işlev için savaş sistemlerine de entegre edilmektedir (Sayler, 2020: 20-26). Geliştirilmiş otonom navigasyon kontrolü ve hedef tanıma, siber saldırı yetenekleri, karar desteği ile güçlendirilmiş insan gücü, lojistik ve stok yönetimi, otonom anti-denizaltı ve karşı mayın sistemleri gibi uygulamalar YZ sistemlerinin askeri uygulamalarına örnek olarak sıralanabilir (Boulain vd., 2020b: 4).

YZ sistemlerinin, vahşi yaşam popülasyonlarını tahmin etmek (Gonzalez vd., 2016: 15; Norouzzadeh vd., 2018: 5717; Li, 2020: 1; Sharma vd., 2023: 2) gibi daha pek çok kullanım alanı mevcut iken bu **kullanım alanlarına göre bireylerin bu sistemlere güven ve desteğinin de değişebiliyor olması**, söz konusu sistemlerin geliştirilmesi ve konuşlandırılması için bağlamsal bir yaklaşım benimsenmesinin önemine işaret etmektedir (Gillespie vd., 2021, 55). YZ'nin hızla uygulandığı ve birçok kişi tarafından kullanılması veya birçok kişinin etkilenmesi muhtemel alanları temsil ettikleri için sağlık hizmetleri, güvenlik, insan kaynakları ve tavsiye sistemleri hakkında daha detaylı bir izlenim sağlamak gerekir (Gillespie vd., 2023: 10).

1.4.1. Sağlık Hizmetleri

Teşhis ve tedavi süreçlerinin daha etkin hale getirilmesi (1), kanser, kalp hastalıkları, Alzheimer ve Parkinson (Lane, 2023) başta olmak üzere teşhisi zor olan ya da erken teşhisi kritik öneme sahip hastalıkların tespitini iyileştirme (2), yardımcı sağlık sistemleri aracılığıyla doktorların hasta başına ayırdığı toplam zamanın azaltılması (3), semptomlar, test ve tarama sonuçları, tıbbi geçmiş, aile geçmişi, yaş, kilo ve cinsiyet vb. verilerin birçok hastaya dayanan büyük veri kümeleriyle karşılaştırılarak kişiye özel potansiyel hastalık belirleme (4), durum

tahmini ve takip sistemlerinin geliştirilmesi, bu sayede kişisel sağlık takibi, kişiye özel tedavi yöntemleri, asistan hizmeti ile önleyici hekimliğin iyileştirilmesi (5), ilaç geliştirme (6), bulaşıcı hastalık tespiti, yayılım tahmini ve önlenmesi (6), YZ'nin sağlık hizmetlerinde halihazırda kullanılmakta olduğu ve hedeflendiği çalışma alanları olarak sayılabilir (T.C. Cumhurbaşkalığı Bilim, Teknoloji ve Yenilik Politikaları Kurulu, 2021: 156-184).

1.4.2. İnsan Kaynakları

YZ sistemlerinden, bir iş için en uygun başvuru sahiplerinin seçilmesine yardımcı olmak, iyi performans gösterme olasılığı en yüksek olan çalışanları belirlemek ve kimin işten ayrılma olasılığının en yüksek olduğunu tahmin etmek için eğitim gibi sektörlerde dahi faydalanılmaktadır (O'Neil, 2017). Bu sistemler, çalışanların özelliklerini ve verilerini, zaman içindeki performanslarını bir araya getirip karşılaştırarak hangi niteliklerin daha iyi iş performansı ve işte kalma ile ilişkili olduğunu analiz ederek çalışır. İşe alım yöneticileri, hem ekonomik gerilemenin hem de uzaktan çalışmanın pandemi sonrası yaygınlaşmasının bir sonucu olarak, artan başvuru havuzlarıyla ve sınırlı insan kaynakları bütçeleriyle başa çıkmak amacıyla özellikle büyük çaplı işe alım ve terfi ile ilgili kararlar için bu sistemleri kullanmaktadır. Daha büyük şirketlerde otomasyon, külfetli idari görevleri hafifletme ve maliyetleri düşürürken verimliliği artırma amacıyla da tercih edilebilmektedir. “Aday Takip Sistemi” (Applicant Tracking System: ATS) gibi otomatik istihdam karar araçları (Automated Employment Decision Tools: AEDT), çevrimiçi bir asistan gibi davranarak işe alım yöneticilerinin iş tanımlarını yazmasına, aday özgeçmişlerini taramasına ve görüşmeleri planlamasına yardımcı olmaktadır ve Fortune 500 şirketlerinin yaklaşık %99'u adayları Workday, Taleo, Jobvite, Greenhouse veya Lever gibi büyük bir ATS aracılığıyla filtrelemektedir (Davis, 2023). Buna karşılık, iş ilanlarına erişim sağlayan algoritma tabanlı kariyer platformları, anahtar kelime eşleştirmeyi artıran özgeçmiş optimizasyon yazılımları ve ön yazı hazırlamaya aracı olan üretken YZ platformları gibi bir dizi çevrimiçi araç da piyasada mevcuttur.

1.4.3. Güvenlik

Şüpheli veya suç teşkil eden davranışları, güvenlik tehditlerini ve polisin ilgisini çeken kişileri tespit etmeye yardımcı olmak için kullanılan YZ sistemleridir. Ülke sınırları gibi yerlerde toplanan yüz ve parmak izi taramaları ve güvenlik kameraları aracılığıyla toplanan halka açık yerlerdeki insanların ve araçların fotoğrafları ve canlı görüntüleri gibi bir dizi bilgiyi

işleyerek ve analiz ederek çalışır. Polis ve güvenlik kurumları, kamu emniyeti ve güvenliği ile ilgili kararlarda bu sistemlerden yararlanmaktadır (Gillespie vd., 2023: 10).

1.4.4. Tavsiye Sistemleri

Kullanıcıya en uygun içerik ve ürünleri sağlayarak haberler, sosyal medya ve alışveriş önerileri benzeri hizmetleri kişiselleştirmek gibi görevler için kullanılan YZ sistemleridir. Kişinin seçimlerini ve tercihlerini yaş, cinsiyet, konum gibi özelliklerine, geçmiş davranışlarına, ilgi alanlarına veya tercihlerine ve benzer kullanıcıların davranışlarına göre tahmin ederek çalışmaktadır. Çevrimiçi reklamlar, müzik ve film önerileri, gazetelerin ve yayın şirketlerinin web siteleri, sosyal ağ hizmetleri gibi birçok çevrimiçi yayıncının ve arama motorlarının sundukları içerikler, bu sistemlerin işlediği alanlara örnek olarak verilebilir.

1.4.5. Bankacılık, Finansal Hizmetler ve Sigortacılık

YZ teknikleri, mevcut verilerin bolluğu ve uygun fiyatlı bilgi işlem kapasitesi sayesinde varlık yönetimi, algoritmik ticaret, finans kuruluşlarının belirli bir kişi ya da şirkete kredi veya kredi kolaylığı sağlamadan önce onun kredi itibarını değerlendirdiği süreç olan kredi yüklenimi ya da blok zinciri tabanlı finans gibi çeşitli alanlarda yaygınlaşmaktadır. Bu uygulamaların finansal ve finansal olmayan riskleri doğurabileceği veya yoğunlaştırabileceği; önyargılı, haksız veya ayrımcı tüketici sonuçları ya da veri yönetimi ve kullanım endişeleri gibi YZ ilişkili potansiyel riskleri ise tartışma konusudur (OECD, 2021: 3).

Sohbet robotları, dolandırıcılık tespiti ve önleme, müşteri ilişkileri yönetimi, tahmine dayalı analitik, özellikle FinTech ve dijital bankacılık pazarında popülerlik kazanmakta olan kredi riski yönetimi, YZ'nin bu sektörde en etkili olduğu bazı kullanım örneklerindendir. Sigortacılıkta da bilgisayarlı görü ve doğal dil işleme gibi teknolojiler sayesinde telematik, uzaktan sensörler, uydu görüntüleri veya dijital sağlık kayıtları gibi verilerin sigortalama için eyleme geçirilebilir içgörülere dönüştürülmesi ve müşterilere özel teminat ve fiyatlandırma sunulması ya da araba kazası benzeri sahtekarlıkların azaltılması gibi değer zincirini etkileyecek çeşitli kullanım alanları bulunmaktadır (Ladva ve Grasso, 2023).

İKİNCİ BÖLÜM

YAPAY ZEKÂ VE ETİK

2.1. Etik

Felsefe ve felsefenin bir alt disiplini olarak *ahlak felsefesi* (ya da bu çalışmada sıklıkla başvurulan ifadesiyle *etik*), araçsallaştırılmaya ve suistimale açık, dolayısıyla “ontolojik olarak kusurlu bir formalist metodoloji” olarak görülmekte ve ahlak felsefesine bağlılık, alternatif ve daha pratik metodolojilerle ele alınması gereken karmaşık sorunların yalnızca entelektüelleştirilmesi olarak tasvir edilse de bu indirgemeciliğe direnerek **ahlak felsefesini, “teknoloji politikası stratejilerinin değerlendirilmesini kolaylaştıran bir bilgi arama ve sorgulama biçimi” olarak görmek** icap eder (Bietti, 2021: 1). Zira, Bietti’ye göre ahlak felsefesi, teknoloji ile diğer değerler arasındaki ilişkiyi açıklayarak ve teknolojiyi insani, sosyal ve siyasi ağına içine yerleştirerek gerekli perspektiflerin kapısını aralar. Bu noktada, özellikle, felsefi terminolojide yer alan soyutlanmış anlam katmanlarını kapsayabilecek eğitime sahip olmayanların dikkate alınması ile ilgili teorilerin popüler tüketim için aşırı basitleştirilmemesi arasındaki dengeyi de korumaya özen göstermek gerekmektedir (IEEE, 2023: 197).

"Ne yapmamız gerektiğinin incelenmesi" veya geniş anlamda hayatlarımızı nasıl yaşamamız gerektiği ile ilgili olan etik, tam da bu nedenle hem kendimizi hem de başkalarını gözeten pratik kaygıları içerir (Merrill, 2011: 3). Bir disiplin olarak kavramsal analiz, düşünce deneyleri ve argümantasyon dahilinde doğru ve yanlış davranış nosyonlarını sistematikleştirmeyi, savunmayı ve tavsiye etmeyi kapsar (Yam, 2018: 76). Geleneksel olarak meta-etik, normatif etik ve uygulamalı etik olmak üzere üç kategoriye ayrılmaktadır. Mevcut çalışma herhangi bir ahlaki felsefe kategorisi üzerine temellendirilmemiş ve çeşitlendirilmiş bir YZ etiği yaklaşımı doğrultusunda potansiyel araştırma istikametlerine yönelik bir çağrıda bulunuyor olduğu için “2.2. Yapay Zekâ Etiği” başlığına geçmeden önce, bu kategorilerin kısa birer açıklamasını sunmak gerekir.

2.1.1. Metaetik

Metaetik, ahlaki düşüncenin, konuşmanın ve uygulamanın “metafizik, epistemolojik, semantik ve psikolojik ön kabullerini ve taahhütlerini anlama girişi” olarak tanımlanabilir; çoğunlukla normatif tümcenin anlamı ve referansı ile muhtemel doğruluk değerlerinin nasıl

belirlenebileceği sorusuna odaklanır ve naturalizm, sezgicilik, duygusalcılık ve kuralcılık gibi teorileri içermektedir (Sayre-McCord, 2023). Aynı zamanda değerler, eylem nedenleri ve insan motivasyonu arasındaki bağlantıyı araştırır, ahlaki standartların bize talep ettikleri şeyi yapmak veya yapmaktan kaçınmak için nasıl nedenler sağlayabileceğini sorar ve genellikle özgürlüğün doğası ve ahlaki sorumluluk için önemi ile bağlantılı olan birçok konuyu ele alır.

2.1.2. Normatif Etik

Normatif ya da preskriptif etik alanındaki pek çok araştırmacı, amacı eylemleri doğru ve yanlış yapan şeyin ne olduğunu sistematik olarak tanımlamak olan etik teorileri araştırmaktadır. Bu teoriler arasında faydacılık, sonuççuluk, Kantçı teoriler, sözleşmecilik, erdem etiği, çeşitli çoğulcu görüşler, doğal hukuk teorisi ve benzerleri yer alır. Normatif etik ayrıca, ahlaki değerlerin doğası ve orantılanabilirliği, ahlaki kısıtlamaların ve seçeneklerin gerekçelendirilmesi, ahlaki ikilemlerin varlığı, iyilik gibi değerlendirici kavramlar ile nedenler, egoizm ve özgecilik, erdemler ve erdemsizlikler, vaatler ve anlaşmalar gibi deontik kavramlar arasındaki bağlantı gibi ahlak felsefesindeki diğer çeşitli konuları da kapsar (Suikkanen, 2023). Platon, Aristoteles ve Konfüçyüs ve Mensiyüs'ün başlıca savunucularından olduğu “*erdem etiği*”, Kant'ın “*deontolojik etiği*” (ödev etiği) ve “*teleolojik (erekbilimsel) etik*” ya da sonuççuluk teorileri altında incelenir (Cointe vd., 2016: 2):

- *Erdem etiği*, bir failin ancak ve ancak ölçülülük, bilgelik, cesaret, sağduyu ve adalet gibi bazı değerlere göre hareket etmesi ve düşünmesi halinde etik olduğunu öne sürer; görevler ya da sonuçlarla ilgilenmez, bunun yerine açık kurallar ve sonuçların yokluğunda insanları etkileyen öznel, kendine özgü ve görünüşte rasyonel olmayan dürtülere odaklanır. Daha “spontane”dir ve “içgüdü ya da ruhsal olarak motive edilmiş bir irade tarafından yönlendirilir” (Merrill, 2011: 12).
- *Deontolojik etiğe* göre bir fail olası durumlarla ilgili yükümlülüklere ve izinlere saygı duyarsa etiktir. Bu yaklaşım, sabit bir dizi görev, kural ve politika eylemleri etik olarak tanımlar ve bu kurallar çiğnendiğinde etik dışı davranmış olunur. Bu ilk ünlü deontolojik teori olan “Kategorik İmperatif” ya da “Kantçılık” teorisi dışında “ilahî buyruk teorisi” gibi farklı deontolojik teoriler de bulunmaktadır.
- *Teleolojik etik* teorileri ise bir eylemin ardındaki kararın, ilgili sonucun değerlendirilmesine göre belirlenmesi gerektiği hipotezine dayanmaktadır. Teleologlar, ahlaki yargıyı bir davranışın sonuçlarına odaklanarak temellendirir. Etik, insanların “aktörün en önemli gördüğü tarafa en fazla iyiliği getirecek eylemi” seçmelerine yardımcı olmalıdır (Merrill, 2011: 11). Jeremy Bentham ve halefi John Stuart Mill'in

felsefeleriyle ilişkilendirilen “faydacılık”, rasyonel karar vermeyi açıklarken ekonomide yaygın olarak uygulanan klasik bir teleolojik teoridir; insan ahlakının temelinde verimliliğin ve daha yüksek bir tatmin düzeyine ulaşma arayışının yattığını varsayar (Baumane-Vitolina vd., 2015: 110).

2.1.3. Uygulamalı Etik

Ahlak felsefesinin son kategorisi olarak uygulamalı etik, genellikle tarihsel olarak belirli bir yeni durumda veya eylem olasılıkları uzayında (çoğunlukla tarihsel olarak benzeri görülmemiş) ne yapmak zorunda olduğumuza (veya ne yapmamıza izin verildiğine) ilişkin etik dalıdır. Uygulamalı etiğin başlıca alanları arasında biyoetik, iş etiği, meslek etiği, sosyal etik, dağıtıcı adalet, kıtlık yardımı ve çevre etiği yer almaktadır (Dittmer, 2023). Ötenazi, kürtaj, çocuk işçiliği gibi tartışmalı konulara eğilen uygulamalı etik; meta-etik ve normatif etik teorileri ile ilgilenmeyi de gerektirir, zira bu üç dal birbirleriyle ilişkilidir.

2.2. Yapay Zekâ Etiği

Etik, doğası gereği nitel bir kavramdır ve kültürel özellikler gibi ölçülmesi zor pek çok özelliğe bağlıdır; bu muğlak doğası, sistematikleştirmeyi de zorlaştırır. Dolayısıyla YZ etiği de değerlendiren kişilere bağlı olarak her zaman bazı öznel unsurlara sahip olacaktır. YZ etiğinin, ek olarak, sentetik biyoloji etiği, nano-teknoloji etiği, genom düzenleme etiği, xenotransplantasyon (tıbbi amaçlarla hayvanlardan organ alınması ve insanlara yerleştirilme süreci) etiği gibi disiplinler ile kesişimleri olsa da bu disiplin, kullanım ölçeği ve çok geniş kapsamlı etik taleplerde bulunuluyor olması nedeniyle diğer teknoloji etiği disiplinlerinden öne çıkmaktadır.

Genellikle uygulamalı etik içinde çok genç bir alan olarak öne çıkan YZ etiği, bu teknolojilerin geliştirilmesinde, dağıtılmasında ve kullanılmasında tutarlı, uygulanabilir ve sürdürülebilir ahlaki davranış ve düzenlemelere kılavuzluk etmek için “yaygın olarak kabul edilen doğru ve yanlış standartlarını kullanan bir dizi değer, ilke ve teknik”tir (AI HLEG, 2018: 9; Leslie, 2019: 3). YZ sistemlerinin suistimali, kötü tasarımı veya istenmeyen sonuçlarının neden olabileceği çeşitli bireysel ve toplumsal zararlara bir yanıt olarak ortaya çıkmıştır. Değer yaratma ve “servet maksimizasyonu ile bireysel gelişmeyi ve kolektif refahı iyileştirme derin potansiyeli” bir yana (Suchman, 2018), gelişmiş kapitalizminki gibi sürekli artan karmaşıklığa sahip, “ideolojik muammayla buğulanmış” bir toplumsal yapı söz konusu iken (Ginzburg, 1980: 27) YZ’nin olası yıkıcı sonuçlarının sıklıkla vurgulanması, haliyle belirsizliklere ilişkin

giderek artan bir farkındalığa katkıda bulunmakta ve araştırmacıları somut etik kaygıları ve ilgili muamele ve koordinasyon gereğini gündeme getirmeye motive etmektedir (Future of Life Institute, 2023; Alaga ve Schuett, 2023; Altman vd., 2023; Hendrycks vd., 2023: 44; Seth, 2023).

YZ teknolojilerinin seçimlerimizi, eylemlerimizi ve kimliklerimizi giderek artan bir biçimde etkileme potansiyeli ve etki alanlarına bağlı olarak ortaya çıkan adalet gibi kavramlara ilişkin etik sorunlar nedeniyle disiplinin amacı, “insanın kendi kaderini tayin hakkının aşınması” gibi bireylerin yaşam kalitesi, özerkliği ve özgürlüğü konularına yönelik endişelerin nasıl gündeme getirilebileceğini belirlemektir (Taddeo ve Floridi, 2018: 751-752). Aslında ilgili etik kaygılar ve bu kaygılara yönelik tartışmalar, sorunların çoğu “yetki devri” ve “sorumluluk” ile ilişkili olmak üzere, dijital bilgi işlemin erken aşamalarına kadar izlenebilir (Samuel, 1960: 741; Taddeo ve Floridi, 2018: 751; Stahl, 2022: 6; Stahl, 2023: 2). YZ etiği, uygulamada genellikle “konu dışı, fazlalık veya teknik kaygılara bir tür eklenti, teknik topluluğun dışındaki kurumlardan dayatılan ve bağlayıcı olmayan bir çerçeve” olarak kabul edilse de (Hagendorff, 2020: 113-114); iç içe geçmiş ve birbirini üzerine inşa edilmiş sosyoteknik sistemlerin bir örneği olan YZ sistemlerine ilişkin etik sorunların nihayetinde insanlardan kaynaklandığı inkâr edilemez (Borenstein ve Howard, 2021: 62). Teknik sistemlerin her zaman etik ve politik mülâhazalar gerektiren tasarım ve uygulamaların sonucu olduğu, en başından itibaren de belirli kültürel, politik ve ekonomik tahayyüller, yatırımlar ve çıkarlar tarafından oluşturulmaları ve bunları pekiştirmeleri sebebiyle hiçbir zaman tarafsız olmadıkları ifade edilebilir; teknolojinin amaçları ve uygulamaları, parçası oldukları daha geniş sosyoteknik sistemler bağlamında değerlendirilmelidir (Suchman, 2018). **Bu sebeptir ki YZ etiği de mevcut ekosistemin tüm damarlarına sirayet edecek tartışmalara açık olmalıdır.** Örneğin, siyaset bilimi, sosyoloji, antropoloji gibi disiplinler kimin güce sahip olduğuna dair sorulara daha net bir şekilde odaklanıyor olsa da uygulamalı etik de kimin hangi güce sahip olması gerektiğini ve bu gücün hangi kullanımlarının ahlâki açıdan meşru olduğunu araştırır, haklar ve ödevler üzerine bireylerin ve devletlerin birbirleri karşısındaki güçlerini inceler; dolayısıyla ilerleyen bölümlerde detaylandırılacağı üzere adalet meselelerini kapsadığında güç dengesizliklerini de ele almak durumundadır (Raicu ve Kwan, 2023). Yapısal kaygıları büyük ölçüde göz ardı eden standart etik söylemin aksine, **feminist etik** veya **bakım etiği** de güç ilişkileri hakkında doğrudan konuşur (Adams, 2018; Miller, 2017). Erdem etiği ise tüm bunlara karakterin ve rol modellerin gücüne odaklanmayı dahil eder. Dolayısıyla, teknoloji etiği ve YZ etiği merceğinden bakıldığında, meselelerin etik ile ilgili değil güç, adalet ya da insan hakları

ile ilgili olduğuna dair ifadeler ancak yanlış yorumlamak olacaktır. Buradaki zorluk, birkaç cümle önce ifade edildiği gibi, **teknik sistemlerin soyağaçlarını kabul etmek ve bunların sonuçlarıyla eleştirel bir şekilde ilgilenmek üzere sürekli sorumluluk almaktır**. Bunu yaparken Alasdair MacIntyre'ın erdem etiğine de dayanarak YZ ve insan ilişkisini toplum ve tarihin öncülleriyle birlikte düşünmek, YZ'nin etik sistemi insanla “ne derece paylaşabileceği”ne, “insan tarihinin anlatısına ve insanın sosyal iş birliği sistemine entegrasyon derecesine” ne kadar paralel olduğuna eğilmek gerekir (Feng, 2018: 472). Kısacası, ancak etik gelenekleri tarihinin özünü yeterince tercüme edebilen bir yaklaşım, etik otonom/akıllı sistemler üzerine yapıcı bir diyalogu destekleyeceği için, önem kazanmaktadır. Bu sebeple, geleneksel etiğin, mevcut söylemi ağırlıklı olarak liberal demokratik değerlere dayanan "sorumlu inovasyon" (Responsible Innovation: RI) ve “sorumlu YZ” söz konusu olduğunda, erdem etiği, deontoloji ve faydacılık gibi Batılı etik temellerin kapsamının ötesine genişletilmesi ve farklı kültürel perspektiflerin karmaşıklığı sorunlu varsayımların ileri sürülmesi ihtimalini barındırıyor olsa dahi diğer etik gelenekleri içermesi icap eder (IEEE, 2023: 203-204).

Kolombiyalı-Amerikalı antropolog Arturo Escobar (2018: 83), Batı rasyonalist geleneğin getirdiği inançların, bildiğimiz tasarım teorisinin ve pratiğinin “varsayılan ayarının bir parçası” olduğunu, başka bir deyişle, tasarımın kaçınılmaz olarak böyle bir “ontolojik arka planın” içinde yer aldığını dile getirir. Bir toplumun ve yönetişiminin tasarımını ve buna bağlı olarak teknoloji tasarımını “yapıbozuma” uğrattırken, bizi, yerel dünyaları yeniden hayal etmek ve inşa etmek için bir araç sunan, birçok dünyanın birbirine uyduğu bir “çoklu-evren”i referans noktası olarak almaya davet etmektedir (Escobar, 2018: 4; Milan, 2020: 5). “Tek tipçilik”, “tek yönlü çözümcülük” ve “daha geniş sosyal ilişkiler, başkalık ve karşılıklı bağımlılık açısından düşünememe” söz ettiği çoklu-evrenin karşıtıdır (Milan, 2020: 5). Dolayısıyla, *kültürlerarası bilgi etiğinin* (Intercultural Information Ethics: IIE) amacıyla olduğu gibi bir yaklaşım benimseyerek “hem yerel hem de küresel sorunları; hem yerel gelenekleri, değerleri ve tercihleri sürdürecektir hem de merkezi etik sorunlara ortak yanıtlar sağlayacak şekilde ele almak” (Ess 2007: 102), “doğası gereği değer yüklü bir proje olan” ve “küresel ağa dahil olmak için normatif kriterleri belirleyen” *standardizasyon* söz konusu olduğunda da yerleşik Batılı etik değer sistemlerinin tekeline ya da bu sistemlere yönelik bir önyargıya dikkat etmek (Wong, 2016: 162) ve “kabul edilmiş değer tekellerinin parametreleri dahilinde çalışmayanlar için” farklı değer sistemlerini barındıracak alternatif yöntemler bulmak gerekli hale gelmektedir (IEEE, 2023: 204).

Antrparantez, Wong (2009: 9), kültürlerarası bilgi etiğinde “**ortak normlar yaklaşımı**”nın “farklı yorumlar ve çeşitli gerekçeler arasında hakemlik yapacak kapsayıcı bir pozisyon olmadan, normatif temelini yanlış bir şekilde asgari ahlâkî paydaya dayandırıldığını” savunmaktadır. Değerlerle ilgilenmek ise “normlara dayalı tartışmalarda marjinal kalan” refah, dijital uçurum ve toplumsal cinsiyet gibi konulara alan açmaktadır; ona göre bu sebeple ahlâkî anlaşmazlıklar konusunda normatif ve değerlendirici yargıları meşru bir şekilde formüle etme görevi için “**paylaşılan normlar**” kullanımı yerine “**paylaşılan değerler**” daha kapsayıcı olacaktır; kültürlerarası bilgi etiği için nesnel ahlaki temeli, paylaşılan normlar değil, bir dizi paylaşılan değer sağlayabilir (2009: 10). Zira, bilgi ve iletişim teknolojileri ile ilgili **etik meselelerin genellikle farklı değerlere sahip kültürlerden kaynaklandığını** iddia edildiğini öne sürer; örneğin Doğu-Batı bağlamında mahremiyet meseleleri yaygın bir şekilde “topluluk temelli değerler ile bireyci değerler arasında bir karşıtlık” olarak yorumlanmaktadır. Dolayısıyla, Wong’a göre “**paylaşılan değerler yaklaşımı**”, söz konusu senaryo ve ilgili değerlerin daha kapsayıcı bir yaklaşımla mercek altına alınması ile “normatif ve değerlendirici yargıları daha meşru bir şekilde formüle etmeye” vesile olabilir.

Öte yandan, mevzubahis gerekliliklerin aciliyeti ortada iken geniş kapsamlı bir araştırma alanı olarak YZ etiği, algoritmik sistemlerin geliştirilme ve mümkün olan her mecra ile bütünleşme hızıyla eşleşmemektedir (Mittelstadt, 2019: 503). Metin tabanlı görevlerde güçlü performansından yararlanmak üzere bir LLM'yi görsel yerleştirmelerle birleştiren DeepMind'ın Flamingo VLM ailesi, temel modellerden kabul edilebilir ve bu duruma bir örnektir (Alayrac vd., 2022: 4). Temel modellerin yükselişi, YZ sistemlerinin tasarlanma ve dağıtılma biçimindeki bir değişimi yansıtmaktadır. “1.3.4. Üretken Yapay Zekâ ve Temel Modeller” alt başlığında belirtildiği üzere, birden fazla “downstream” uygulamasına uyarlanabilen bu modeller, genellikle tek bir aktör tarafından eğitilip piyasaya sürülür ve ardından diğer aktörler tarafından çok sayıda uygulamaya uyarlanır. Durum denetim açısından değerlendirildiğinde ise temel modellerin yetenekleri ve eğitim süreçleri, bu sistemlerin etik, yasal ve teknik olarak güvenilir olmasını temin etmek için geliştirilen araç ve prosedürleri geride bırakmıştır; dolayısıyla piyasaya sürülmeden önce model denetimlerinin, uygulama denetimlerinin ve yönetim denetimlerinin birbirini tamamladığı yeni denetim ve kontrol biçimleriyle desteklenmeleri gerekmektedir (Mökander vd., 2023: 2).

2.2.1. Yapay Zekâ Etiği Çerçevesinin Gündemi

2021’de önerilen ve bu araştırmanın devam ettiği süreç içerisinde hala tam olarak onaylanmayan AB YZ Yasası ilk akla gelen olmak üzere son yıllarda uluslararası örgütler, sivil toplum kuruluşları, sivil toplum meclisleri, meslek birlikleri, farklı ölçeklerden işletmeler ve sendikalar, hükümetler ve uluslararası örgütler gibi çok sayıda aktör tarafından YZ teknolojisinin kullanımına yönelik “normatif” kılavuzlar geliştirilmekte ve düzenleme çalışmaları yürütülmektedir. Ancak çoğu YZ etiği girişimi, özellikle endüstri tarafından desteklenenler, YZ tartışmalarını öncelikli olarak soyut sorunlara ve teknik çözümlere odaklamayı amaçlayan salt erdem sinyali olarak nitelendirilmiş, hukuk tartışmalarını ve düzenlemeleri etkili bir şekilde geciktirme çabası olarak görülür (Nemitz, 2018: 7). Teknoloji şirketlerinin “etik” terimini, düzenlemeden kaçınmak ve akademisyenlerden ve sivil toplumdan gelen eleştirileri savuşturmak; öz-düzenlemeyi veya müdahalesiz yönetişimi desteklemek için kullandıkları öne sürülmektedir (Bietti, 2021: 1). Şimdiye kadar büyük ölçüde belirsiz, üst düzey ilkeler ve eylem kılavuzu olmayı vaat eden değer beyanları üreten bu girişimlerin, “adalet ve mahremiyet gibi kilit kavramlara gömülü temel normatif ve politik gerilimleri ele almakta başarısız” olarak nitelendirilir (Mittelstadt, 2019: 501). Özellikle, MAMAA olarak bilinen Meta, Amazon, Microsoft, Alphabet ve Apple ile diğer önde gelen YZ şirketlerinin elinde olağanüstü bir güç yoğunlaşması söz konusu iken şirketlerin “halkla ilişkiler çalışmalarına düzenli olarak etik hususları dahil etmeleri” ve “olası teknolojik riskleri azaltmak ve kötüye kullanım senaryolarını ortadan kaldırmak için bağlayıcı yasal çerçevelere gerek olmadığı” doğrultusunda yaklaşımları, hem eleştirilen uygulamalar sürdürülürken kamuoyundan yükselen sesleri yatıştırma amacına hem de hesap verebilirliğin demokratik kurumlardan ilgili sektörlerle devredilebileceğini iddia etmeye hizmet edebilir (Hagendorff, 2020:100).

Çizilen etik çerçevelerin ve ilkelerin, çözülmesi gereken YZ sorunları için uluslararası gündemi ve hiyerarşileri belirliyor olmasından dolayı ise “YZ'nin politik ekonomisi, bu sistemlerin çoğunluk dünyası için etik bir çerçevede tasarımının, geliştirilmesinin, kullanımının ve konuşlandırılmasının etkilerinden ayrı tutulamaz” (Ricaurte, 2022: 729). YZ endüstrisi, iş dünyası ve bilim sistemlerinin “kendi kurallarına ait, münhasır, işlevselci bir yönelim olarak anlaşılması gereken özerklikleri” bulunmaktadır; bu sistemlerin kendi hedef değerleri, yapılandırıldıkları ve kararların alındığı kendi ekonomik veya sembolik sermaye türleri vardır (Bourdieu, 1984’ten akt. Hagendorff, 2020: 109). Bu da uzun vadeli ya da daha geniş toplumsal sonuçlar hakkında bilgi eksikliği ile bağlantılı olarak dağıtılmış sorumluluk ve ekonomik teşviklerin, etik ilke ve değerlere bağlılığın önüne geçebileceği ve YZ sistemleri geliştirme ve

uygulama amaçlarının temel haklar veya toplumsal değerler ile örtüşmeyebileceği anlamına gelmektedir (Manders-Huits ve Zimmer, 2009: 42; Hagendorff, 2020: 114). **İtibar riskleri, şirketleri etikle ilgilenmeye itebilecek olsa da bu riskler yalnızca kamu bilincinde kaldıkları sürece ağırlık taşır** ve söz konusu sistemlere etik müdahale ancak sınırlı ölçüde mümkün olduğu için bu müdahalelerin yasal düzenlemelerle desteklenmeleri de gerekir (Boddington, 2017: 8; Hagendorff, 2020: 112; Mittelstadt, 2019: 502).

Diğer taraftan, söz konusu çerçevelerin çokluğu, dahil edilen ilkelerin aralarındaki farklılıklar ve tanımlanma şekillerinin çeşitliliği, YZ paydaşları için epistemik ve operasyonel zorluklar da teşkil etmektedir. Yukarıdaki paragraflarda da ele alındığı üzere “yerel bağlamların, epistemolojilerin ve farklı kültürlerdeki ahlaki sistemlerin dikkate alınmaması” sebebiyle bu kılavuzların kullanışlılığı da sorgulanmaktadır (Abdilla vd., 2021’den akt. Ricaurte, 2022: 729; Hongladarom ve Bandasak, 2023: 2). Bu aşamada, “YZ etik değer ve ilkelerin kimin tarafından belirlendiği, bu değer ve ilkelerin kime yönelik ve amacının ne olduğu, neden ve nasıl takip edilmesi gerektiği, temel kavramlara ilişkin çelişkili yorumların nasıl çözümlenebileceği” gibi çok sayıda soru üzerinde durmak gerekir (Mittelstadt, 2019: 504).

Asimov’un (1950) üç robotik yasasına uzanan ve Murphy ve Woods’un (2009) üç “Sorumlu Robotik” yasası, EPSRC Robotik ilkeleri (2010), Faydalı YZ için Asilomar İlkeleri (Ocak 2017), ACM ABD Kamu Politikaları Konseyi “Algoritmik Şeffaflık ve Hesap Verebilirlik İlkeleri” (Ocak 2017), Japon YZ Topluluğu Etik Kuralları (Şubat 2017), The Future Society Bilim, Hukuk ve Toplum Girişimi’nin Taslak İlkeleri (Ekim 2017), “Sorumlu YZ için Montréal Deklarasyonu” taslak ilkeleri (Kasım 2017), IEEE Etik Otonom ve Akıllı Sistemlerin Genel İlkeleri (Aralık 2017), UNI Küresel Sendika Etik YZ için En iyi 10 İlke (Aralık 2017) şeklinde sıralanabilecek çalışmalar küresel ölçekte etik tavsiyelerde bulunmaya yönelik yaygın olarak bilinen ilk girişimlerdendir (Winfield, 2018). AlgorithmWatch tarafından yürütülen, algoritmik karar verme sistemlerinin etik olarak nasıl geliştirilebileceği ve uygulanabileceğine dair ilkeleri ortaya koymayı amaçlayan etik kılavuzların ve çerçevelerin analizi, çoğunun genellikle “araçsal-ekonomik” ve “etik perspektifler” arasında konumlandırılmış olduğunu ve daha ziyade uyumluluk, suistimale başa çıkma, sosyal açıdan bilinçli girişimcilik, ürünlerin üretildiği koşullar ve olası uzun vadeli sonuçlarına kadar uzanan tavsiyeler yoluyla aslında iş uygulamalarının şekillendirilmesi amacı güden “iş etiği”ne işaret ettiğini ortaya koymuştur (EPRS, 2020: 50). Bu çerçeveler farklı ölçeklerde birbirine bağlı sistemleri ele almaya yönelik

eleştirel bir etik yaklaşımından çok; siyaset, ekonomi ve toplum üzerinde etki yaratmayı, YZ yöntemlerinin çeşitli alanlarda sosyal olarak kabul edilen kullanımını amaçlayan belgelerdir.

2.2.1.1. Hükümetler ve Uluslararası Girişimler

YZ'ye ilişkin ulusal politika ve strateji belgeleri 2016 yılından itibaren yayımlanmaya başlanmıştır ve ABD, Güney Kore ve Çin (2016), Japonya ve Singapur (2017), Almanya, Fransa, Hindistan ve Birleşik Krallık (2018) ve Rusya (2019) başlıca ülkeler olmak üzere 60'tan fazla ülkenin ulusal YZ stratejilerini yayımladığı söylenebilir (CBDDO, 2021: 20). Örneğin, Alman hükümetinin 2018'de yayınlanan YZ stratejisi, "Almanya'nın bir araştırma merkezi olarak seçkin konumunu korumayı, Alman endüstrisinin rekabet gücünü artırmayı ve toplumda vatandaşların yararına somut bir ilerleme sağlamak adına YZ'yi toplumun tüm kesimlerinde kullanmanın birçok yolunu teşvik etmeyi" amaçlamakta ve 12 eylem alanıyla, araştırmanın güçlendirilmesi, küçük ve orta ölçekli işletmelerin desteklenmesi, iş ve işgücü piyasasının korunması, kamu yönetiminde bu teknolojinin kullanılması, verilerin kullanılabilir hale getirilmesi ve kullanımının kolaylaştırılması, düzenleyici bir çerçeve geliştirilmesi, standartların belirlenmesi, ulusal ve uluslararası ağlara katılım ve toplumsal diyalogun teşvik edilmesi gibi geniş bir konu yelpazesini kapsamaktadır (Höne, 2022). Odak alanları dikkate alındığında, bu stratejinin diğer ulusal YZ stratejileri ile de birçok özelliği paylaştığı görülmektedir. Burada, Alman hükümetinin stratejisine ve diğer ulusal YZ stratejileri ile yakınlıklarına değinilmesinin maksadı, YZ etik kuralları altında yatan “sosyal katılım, bireysel eylem özgürlüğü ve toplumun refahından YZ'nin ticari fırsatlarına kadar uzanan karmaşıklıkları ve gerilimleri” göstermektir.

Aslında Türkiye dahil olmak üzere 50'den fazla ulusal hükümetin politikalarını ve uygulamalarını önemli ölçüde şekillendiren, resmi olarak da onayladıkları “OECD/G20 İlkeleri” olmuştur (Center for AI and Digital Policy, 2023: 3). OECD, YZ tarafından karar alma süreçlerinin şeffaflığı ve izlenmesi, ayrımcılığın önlenmesi ve adalet, hesap verebilirlik ve güvenlik gibi etik standartları ortaya koyan, bu teknolojiler için normatif yanıtlar ve politika yönergeleri sağlayan “OECD Dijital Ekonomi Görünümü” raporunu 2017 yılında, “YZ Tavsiye Kararı”nı ise 2019'da yayınlamıştır (Ko ve Leem, 2021: 219). “Tavsiye Kararı” 22 Mayıs 2019 tarihinde OECD Konseyi tarafından Bakanlar düzeyinde onaylanarak yukarıda bahsedilen ilk hükümetler arası standart olmuştur ve G20, OECD YZ İlkeleri'nden yararlanan insan merkezli YZ İlkeleri'ni ise 9 Haziran 2019'da kabul etmiştir.

“Yapay Zeka Küresel Ortaklığı” (Global Partnership on Artificial Intelligence: GPAI) da 2020’de başlatılan, OECD’nin YZ’nin insan-merkezli değerlere dayanan ve güvenilir, sorumlu ve sürdürülebilir kullanımını teşvik eden “Tavsiye Kararı” doğrultusunda, uluslararası iş birliğine bir dayanak olmak amacıyla, ilgili öncelikler konusunda en ileri araştırmaları ve uygulamalı faaliyetleri destekleyerek teori ve uygulama arasındaki boşluğu doldurmak için endüstri, sivil toplum, hükümetler ve akademiden önde gelen uzmanları bir araya getiren oluşumdur (Center for AI and Digital Policy, 2023: 34). “Ulusal Yapay Zekâ Stratejisi (UYZS) 2021-2025” ile YZ stratejisi ana hatları çizilen Türkiye’nin bu ortaklığa üyelik başvurusu ise 22 Kasım 2022’de onaylanmıştır (T.C. Dışişleri Bakanlığı, 2022). 2021/18 sayılı Cumhurbaşkanlığı Genelgesi ile Ağustos 2021’de hayata geçirilen ve OECD, G20, AB ve UNESCO tarafından belirlenen *insan odaklı* YZ ilkelerinin paydaşı olan UYZS’de belirtilen değerler “insan hakları, demokrasi ve hukukun üstünlüğüne saygı, çevreyi ve biyolojik ekosistemi geliştirmek, çeşitliliğin ve kapsayıcılığın sağlanması, barışçıl, adil ve birbirine bağlı toplumlarda yaşamak”; ilkeler ise “ölçülülük, emniyet ve güvenlik, tarafsızlık, mahremiyet, şeffaflık ve açıklanabilirlik, sorumluluk ve hesap verebilirlik, veri egemenliği ve çok paydaşlı yönetim” olarak sıralanmaktadır (CBDDO, 2021: 57-61).

Son olarak, “Güvenilir Yapay Zekâ” Yönergeleri” 2019’da yayınlanarak şu anda nihai aşamasında olan AB YZ düzenlemesine (“YZ Yasası”) doğru bir basamak görevi görmüştür (AI HLEG, 2019). Söz konusu YZ Yasası’nın esası, bir YZ teknolojisinin bir kişinin sağlığı ve güvenliği veya temel hakları için oluşturabileceği risk düzeyini belirleyen bir sınıflandırma sistemidir (Feingold, 2023). Bu sınıflandırma, “**kabul edilemez**”, “**yüksek**”, “**sınırlı**” ve “**asgari**” olmak üzere dört risk kademesini içermektedir. Spam filtreleri veya video oyunları gibi sınırlı ve minimum riske sahip YZ sistemlerinin, şeffaflık yükümlülükleri dışında çok az gereklilikle kullanılmasına izin verilmektedir. Son zamanlarda davranışı kontrol etmek için kullanılan, Çin’deki “Sosyal Kredi Sistemi” (Roberts vd., 2021; Kshetri, 2020) gibi sosyal puanlama sistemleri ve kamusal alanlarda gerçek zamanlı biyometrik tanımlama sistemleri gibi kabul edilemez bir risk oluşturduğu düşünülen sistemler çok az istisna dışında yasaklanmıştır. Otonom araçlar, tıbbi cihazlar ve kritik altyapı makineleri gibi yüksek riskli YZ sistemlerine ise izin verilmektedir, ancak geliştiriciler ve kullanıcılar titiz testler, veri kalitesinin uygun şekilde belgelendirilmesi ve insan gözetimini detaylandıran bir hesap verebilirlik çerçevesi gerektiren düzenlemelere uymalıdır. Önerilen mevzuat ayrıca, farklı risk derecelerine sahip farklı amaçlar için kullanılabilen YZ sistemleri olan, ChatGPT gibi büyük dil modeli üreten sistemleri kapsayan “genel amaçlı YZ” ile ilgili düzenlemeleri de ana hatlarıyla belirtmektedir. İlaveten,

uyumsuzluk ve düzenleyicilere yanlış veya yanıltıcı belge sunmak halinde ağır cezalar da öngören YZ Yasası, yönetmeliğin uygulanmasını denetleyecek ve AB genelinde yeknesak bir uygulama sağlayacak, ortaya çıkan sorunlarla ilgili görüş ve tavsiyelerde bulunmanın yanı sıra ulusal makamlara rehberlik etmekle de görevli olacak bir “Avrupa Yapay Zekâ Kurulu” kurmayı amaçlamaktadır.

2.2.1.2. Sivil Toplum

Avrupa'da sivil toplum kuruluşları, YZ politikası hakkında önemli raporlar yayınlamış, ihlalleri belgelemiş ve reform çağrısında bulunmuştur (CAIDP, 2023: 3). *Latin Amerika*'da bulunan sivil toplum kuruluşları, özellikle yüz tanıma teknolojisinin kullanımı ile bağlantılı olarak aktiftir. Tanınan bir Arjantin insan hakları örgütü olan “Sivil Haklar Derneği (Asociación por los Derechos Civiles)”, Mozfest tarafından düzenlenen Güvenilir Yapay Zekâ (TAI) programına ve dünya çapında farklı sivil toplum kuruluşlarından otuzdan fazla üyenin yer aldığı "Yapay Zekâ Etki Değerlendirmesinde Sivil Sesten Yararlanma" çalışma grubuna dahildir (Asociación por los Derechos Civiles, 2023). Yüz tanıma teknolojisinin artan ve hesaplanamayan kullanımını eleştirmektedir ve özellikle verileri şeffaf olmayan ve hesap vermeyen sistemlere dahil edildiğinde, bu teknolojinin tehlikeleri hakkında farkındalık yaratmayı amaçlamaktadır. “Yüzümle Hayır” (“Con mi Cara No”) sloganını kullanan ulusal bir kampanyanın oluşturulmasına vesile olmuştur. Bağımsız bir Brezilya düşünce kuruluşu olan Igarape Institute de (2019) suçla mücadelede yeni teknolojilerin sunduğu fırsatlara ve tuzaklara genel bir bakış sunan ve şeffaflık ve hesap verebilirliği sağlamak için tavsiyelerde bulunan ve suç tahmini sürecinde insanların güvenliğinin, onurunun ve haklarının güvence altına alınması gerektiğini savunan “Geleceğin Suçu” çalışmasını yayınlamıştır. **Tahmine dayalı araçların kolluk kuvvetlerinin sezgi ve deneyimlerinin yerini almaması**, aksine onları denetlenebilir bir şekilde tamamlaması ve kolluk kuvvetlerinin de bu yeni suç tahmin platformlarını uygulamanın olası etkileri hakkında bilgilendirilmesi gerektiği üzerinde durulmaktadır. “Fundación Karisma” ise Kolombiya'daki gözetim teknolojisi ve biyometrik tanımlama sistemlerinin değerlendirilirken sadece teknik ve etki hususlarının dikkate alındığına; gereklilik, orantılılık ya da teknolojinin insan hakları üzerindeki olası etkileri üzerinden bir analiz yapılmadığına dikkat çekmektedir. Teknolojinin tüm sosyal sorunları kolaylıkla ve düşük maliyetle çözebileceği varsayımı altında, çözülmesi gereken sorunların ciddi bir teşhisi yapılmadan veya daha verimli ve haklarımıza saygılı olabilecek çözüm yolları bulunmadan çözümler uygulanmaktadır (Saenz ve Spanger, 2018: 33). Dikkate değer bir diğer tespit, Uruguay bir yüz tanımlama veri tabanı geliştirmeye başladığında Fundación Karisma dahil

olmak üzere Arjantin, Şili gibi diğer Latin Amerika ülkelerinden de bazı sivil toplum kuruluşlarının, bu sistemin Ulusal Bütçe Yasası kullanılarak bir '**torba yasa**' olarak onaylandığıdır, “bu tür yasaların onaylanmasına ilişkin sürelerin kısıtlı olması nedeniyle konunun düzgün bir şekilde tartışılmasının engellendiği” uyarısında bulunulmuştur (Datysoc, 2020).

Latin Amerika'da dijital ortamda insan haklarını güçlendirmeyi amaçlayan "Al Sur" çatısı altındaki Peru, Panama, Meksika, Paraguay gibi ülkelerdeki çeşitli sivil toplum kuruluşları da Uluslararası Veri Koruma ve Mahremiyet Komiserleri Konferansı (ICDPPC) tarafından desteklenen "Yapay Zekada Etik ve Veri Koruma: Tartışmayı Sürdürmek" konulu çalışmalarıyla kamu istişaresine katkıda bulunmuşlardır (AlSur, 2019).

Son olarak, *Afrika*'da, araştırmalar YZ ile ilgili sorunlara daha sınırlı katılım göstermekle beraber Kamerun, Kenya, Nijerya, Senegal, Zambiya ve Zimbabve'de bölgesel ofisler işleten “Paradigm Initiative”, Nijerya'nın “adli gözetim ve kapsamlı bir veri koruma çerçevesi olmaksızın” gözetim faaliyetleri yürüttüğünü gözlemlemiş ve veri koruma ve mahremiyet için kapsamlı bir çerçevenin ve gözetim faaliyetleri üzerinde adli gözetimin yürürlüğe girmesini tavsiye etmiştir (Center for AI and Digital Policy, 2023: 50). Bu girişimin, **“yerel YZ'ye öncelik verilmesi ve yurt dışından YZ sistemleri için şeffaf bir tedarik sürecinin sağlanması”** çağrısı (Effoduh, 2021: 3), kendi değerlerine uymayan yaklaşımları bastırırken belirli değerleri ve çıkarları dayatan iş modellerinin yani “gelişmekte olan dünya için teknolojik çözümler giydirilmiş algoritmik sömürgeciliğin” reddi olarak anlamlıdır. Bu ürünler tasarlanırken, politika yapıcılarının ve uygulama geliştiricilerin yalnızca belirli bir kullanıcı tipini - dijital okuryazar ve son teknoloji bir akıllı telefona sahip olacak kadar varlıklı olanları göz önünde bulunduruyor olması da çoğu potansiyel kullanıcı için hemen fark edilemeyecek bir dezavantaj teşkil eder (Milan, 2020: 2). Dolayısıyla, girişimin, YZ becerilerinin geliştirilmesi ve yeniden becerilerin kazandırılması vurgusu da bir o kadar önemlidir (Center for AI and Digital Policy, 2023: 50). Raporlarında, ayrıca, yalnızca sistemlerin neden olduğu önyargılara değil, ifade ve örgütlenme özgürlüğünü zayıflatabilecek bu sistemlerin hükümet tarafından silahlandırılmasına, yüz tanıma teknolojilerinin kullanımı yoluyla sürveyansa ve içeriklerin denetlenmesi yoluyla hakların ihlal edilmesine de odaklanılmaktadır.

Center for AI and Digital Policy'nin (2023: 3) “YZ ve Demokratik Değerler 2022” çalışmasının bulgularına göre, ilgili etik çerçeveler ve girişimler ana hatlarıyla değerlendirildiğinde, çok az YZ uygulamasının, **“YZ'nin kamusal alanlarda gözetim için kullanılması”**ndan daha tartışmalı olduğu dikkat çekmektedir, yüz tanımanın genel nüfus üzerinde kullanılması, birçok sivil toplum kuruluşunun bu teknolojilerin yasaklanması gerektiğini öne sürmesine ve yaygın tartışmalara yol açmıştır. Diğer tartışmalı YZ uygulamaları arasında **“vatandaşların puanlanması”**, **“cezai hükümler”**, **“idari hizmet kararları”** ve **“işe alım değerlendirmeleri”** yer almaktadır. Son birkaç yılda başka birçok YZ politika sorunu ortaya çıkmış olsa da küresel politika yapıcılarının dikkatini çeken ilk YZ uygulamalarından biri YZ'nin savaş için kullanılması olmuştur; **“ölümcül otonom silah sistemleri”** riski ve ilgili düzenlenmelerin geleceğine ilişkin endişeler mevcuttur (Schmitt ve Thurnher, 2013; Center for AI and Digital Policy, 2023: 43).

2.3. Yapay Zekânın Etik Boyutları Üzerine

YZ ile ilgili toplumsal ve etik kaygılar “yakın vadeli” ve “uzun vadeli” sorunlar olarak kategorize edilebilir ve bu sorunlar arasında pek çok bağlantı bulunmaktadır (Baum, 2018: 565; Cave ve ÓhÉigeartaigh, 2019: 5). Uzun vadeye odaklanılırken yakın vade göz önünde bulundurulmalıdır, zira “mevcut araştırma öncelikleri, yürürlüğe giren politika önlemleri ve işbirlikleri” YZ sistemlerinin gidişatını ve etkisini daha uzun vadeli kaygılarla alakalı şekilde etkileyebilir. Benzer şekilde, yakın vade sorunlarına odaklanılırken de YZ sistemlerinin yıkıcı potansiyelini irdeleyen uzun vadeli tahmin ve acil durum planlaması üzerinde çalışılması faydalı olacaktır (Cave ve ÓhÉigeartaigh, 2019: 5).

Varsayımsal yapay genel zekâ (AGI) ve yetenekleri hakkında spekülasyonlar bir kenara, YZ'nin geliştirilmesi ve kullanılmasıyla ilgili mevcut sorunları ele almak önceliğiyle, bu ana başlık altında, Ulusal Yapay Zeka Stratejisi'nde (UYZS, 2021) de benimsendiği ifade edilen “adil ve birbirine bağlı toplumlarda yaşamak, ölçülülük, emniyet ve güvenlik, tarafsızlık, mahremiyet, şeffaflık ve açıklanabilirlik, sorumluluk ve hesap verebilirlik” değer ve ilkelerinin irdelenmesi maksadı da güdülerek, YZ etik belgelerine yönelik araştırmalara göre de kaynakların yarısından fazlasında atıfta bulunulan **adalet, şeffaflık (açıklanabilirlik), kötü niyetli olmama (zarar vermeme), gizlilik (mahremiyet) ve sorumluluk (hesap verebilirlik)** ilkelerine ve ilgili etik ve toplumsal çıktılara genel bir bakış sağlanmaktadır (Jobin vd., 2019: 395; Jang vd., 2022: 11638).

2.3.1. Adalet

Verma ve Rubin (2018: 1), NIPS, Büyük Veri, AAAI, FATML, ICML ve KDD gibi makine öğrenimi ve adaletle ilgili önemli konferans ve dergilerdeki yayınları inceledikleri çalışmalarında, “Grup adaleti veya istatistiksel eşitlik”, “Koşullu istatistiksel eşitlik”, “Yanlış pozitif hata oranı dengesi”, “Yanlış negatif hata oranı dengesi”, “Genel doğruluk eşitliği”, “Muamele eşitliği”, “Test doğruluğu veya kalibrasyon” ve “Nedensel ayrımcılık” dahil olmak üzere Ocak 2018 itibarıyla Google Akademik’te en çok öne çıkan yirmi adalet tanımı olduğunu ortaya koymuştur. Aslında YZ’de adalet, hakkaniyet ve kapsayıcılığı ele almak, kritik ve çeşitli bilgileri bünyesinde barındıran kapsayıcı bir işgücünü teşvik etmekten potansiyel haksız önyargı kaynakları için eğitim veri setlerini değerlendirmeye, sorunlu önyargıları ortadan kaldırmak veya düzeltmek için eğitim modellerine, performanstaki eşitsizlikler için makine öğrenimi modellerini değerlendirmeye ve adil olmayan sonuçlara karşı nihai sistemleri test etmeye devam etmeye kadar aktif bir araştırma alanının konusudur. Bu araştırmalar, insan önyargılarının kaynaklarının ve kapsayıcılığın önündeki engellerin bazılarının belirlenmesi hususunda da kullanılabilir, dolayısıyla bir fırsat da teşkil etmektedir.

Adaletin maddi boyutu hem faydaların hem de maliyetlerin eşit ve adil bir şekilde dağıtılmasını sağlamak ve bireylerin ve grupların haksız önyargı, ayrımcılık ve damgalamadan arınmış olmasını sağlamak için bir taahhüdü ima eder (AI HLEG, 2018: 12). Usule ilişkin boyut ise YZ sistemleri ve onları işleten insanlar tarafından alınan kararlara itiraz etme ve bunlara karşı etkili bir telafi arama becerisini gerektirir ki bunu yapabilmek için karardan sorumlu olan kuruluşun tanımlanabilir, karar verme süreçlerinin de açıklanabilir olması gerekir (AI HLEG, 2018: 13). Bu da adaletin sorumluluk ve hesap verebilirlik ile daimî ilişkisini ortaya koyar.

Değerlerin dünyaya dair çok sayıda gerçeğe iç içe geçmiş olması, makine öğrenimi modelleri oluştururken gerçekleri değerlerden temiz bir şekilde ayırma karmaşıklığının bir kaynağıdır (Ramani ve Wang, 2023). Örneğin, "cinsiyet" ile ilgili bir kural, bu kavramı doğru bir şekilde tanıyan bir ML modeli gerektirir. Buolamwini ve Gebru (2018), fenotipik alt gruplara göre otomatik yüz analizi algoritmalarında ve veri setlerinde bulunan yanlışlığı değerlendirmek için bir yaklaşım sundukları çalışmalarında, veri setlerini kullanarak üç ticari cinsiyet sınıflandırma sistemini değerlendirmişlerdir. Bulgularına göre, daha koyu tenli kadınlar, %34.7’ye varan hata oranlarıyla en yanlış sınıflandırılan grup iken, daha açık tenli erkekler için ise maksimum hata oranı %0.8’dir (2018: 77). Bu noktada ‘önyargı’ terimini de açıklığa kavuşturmak gerekir. YZ’de “önyargı” terimi, farklı vakaların farklı muamele

gerektirdiği fikrini ifade eder. Bu anlamda, ML sistemlerinin bir durumun diğerinden farklı olduğuna karar vermesini ve buna göre farklı davranmasını sağlayan şeydir; makine öğrenimi süreci ve davranışın mevcut duruma uyarlanması için esastır (ISO/IEC 22989:2022(E), 2022: 31). Diğer taraftan, **bu çalışmada irdelenen “önyargı” terimi**, sosyal bağlamda genellikle muameledeki belirli farklılıkların adil olmadığı düşüncesine karşılık gelmektedir. YZ sistemlerindeki istenmeyen önyargı kaynakları birbiriyle ilişkilidir ve insan bilişsel önyargısını, eğitim verilerindeki önyargıyı ve mühendislik kararları tarafından önyargıyı içerebilir. YZ sistemlerinin geliştirilme sürecinde kullanılan veri kümeleri eksik olabilir, istenmeyen tarihsel önyargılara veya zayıf yönetim mekanizmalarına sahip olabilir, bu durum insanlar ve algoritmalar için de geçerlidir (Stinson, 2022: 763-764). İnsan bilişsel önyargıları, veri toplama ve işleme, sistem tasarımı, model eğitimi ve diğer geliştirme kararlarını etkileyebilir; ahlaki yargılarımız, muhakemenin "kısıtlı ama önemli bir rol" oynadığı birçok farklı beyin alanının dağınık bir kombinasyonu yoluyla gerçekleştirilir (Greene ve Haidt, 2002: 522). Ancak insanlar, kısa bir süre içinde makul bir şekilde değerlendirilebilmek için çok fazla hesaplama gerektirebilen sorunlara yönelik sınırlı bilişsel kapasiteye ve temsil kısayollarına güvenmekten kaynaklanan sınırlı muhakeme yeteneğine sahiptir ve bu muhakemeye müdahale eden örtük veya bilinçsiz, çeşitli bilişsel, grup içi ve etik önyargılar da sergilerler (Tversky ve Kahneman, 1974: 1130; Hewstone vd., 2002: 576). Aynı zamanda, “karmaşık süreçlerin veya sistemlerin bir fonksiyonu olan” sınırlı bir bilgiye sahiptirler ve bireyselleştirici ayrımlar yapmaya dayalı “bilgi” de her zaman insan merkezli, etnosentriktir ve diğer belirli önyargılara açıktır (Ginzburg ve Davin, 1980, 20).

Makine öğrenimi modelleri gerçek dünyadan toplanan mevcut verilerden öğrendiği için, cinsiyet, ırk, din veya diğer özelliklere dayalı verilerde önceden var olan sorunlu önyargıları öğrenebilir ve hatta artırabilir. Risk, bu tür sistemlerdeki adaletsizliklerin insan yargılarına dayalı karar verme süreçlerine göre daha geniş çaplı bir etkiye sahip olabilmesidir. Dolayısıyla, YZ sistemleri kimin hayat kurtaran ilaçlara erişebileceği, kimin kredi alacağı, kimin iş bulacağı, kimin hapse gönderileceği ya da şartlı tahliye edileceği gibi sayısız sosyal sürecin düzenlenmesinde etkin hale getirilirken, söz konusu özelliklerde farklılıkları tespit etmek ve potansiyel önyargıları YZ uygulanan ürünlerin ve hizmetlerin “**yaşam döngüsü boyunca**” dikkate almak, adil ve kapsayıcı sistemler oluşturmak için kritik önem taşımaktadır (Andrada vd., 2023: 1321).

Algoritma ise sosyal süreçlerin dönüştürülmesinin yanı sıra “bilgi üretimi için yetkili sistemler” içeren karmaşık bir altyapı, insan, veri ve kurumlar kümesini ifade etmeyi gerektirir (Yeung, 2018: 506). Sosyal bilimciler genellikle bu terimi; içinde çalıştıkları hesaplama ağlarını, bunları tasarlayan ve işleten kişileri, üzerinde hareket ettikleri verileri (ve kullanıcıları) ve bu hizmetleri sağlayan kurumları içeren, daha geniş bir sosyal çabaya bağlı olan ve “bilgi üretimi için yetkili sistemler ailesinin bir parçasını oluşturan sosyoteknik topluluğu tanımlamak için” bir sıfat olarak kullanmaktadırlar. Buna göre, bir şeyi "algoritmik" olarak tanımlarken, aslında “tasarımcıları tarafından işlevsel ve ideolojik olarak bilginin hesaplamalı üretimine adanmış olması amaçlanan sosyoteknik bir bilgi sistemi tarafından üretilen veya bu sistemle ilişkili olan prosedürün eklenmesiyle” ilgilenmekteyizdir (Yeung, 2018: 506). Algoritmalar “yaşam temsili ve gerçeklik kavramlarını yeniden yapılandıran sosyal yapılar, epistemik ve kültürel dolayimler”dir; yani özneler, özneler ve nesneler ve nesnelerin kendileri arasındaki ilişkileri dönüştüren ve yeniden yapılandıran bir dizi sosyal, epistemik ve kültürel süreç olan aracılık teknolojileri olarak hizmet ederler (Ricaurte, 2022: 732). Bu noktada “**algoritmik önyargı**” ise, tahsis ve temsil zararları açısından ölçülmelidir. “Tahsis zararı”, bir sistem bir fırsatı veya kaynağı belirli bir gruba adil olmayan bir şekilde tahsis ettiğinde ortaya çıkar; “temsil zararı” da bir sistem bir grubun boyun eğmesini güçlendirecek şekilde klişeleri ve güç dinamiklerini sürdürdüğünde meydana gelir. Algoritmalar, ancak yasal veya etik nedenlerle (cinsiyet, ırk, sınıf) karar vermede kullanılamayan, korunan niteliklere dayalı olarak bireyleri veya grupları ne lehte ne de onlara karşı ayrımcılık sergilemeyen tahminler gerçekleştirdiklerinde adil kabul edilir (Maslej vd., 2023: 130).

Dahası, örneğin bir ikili sınıflandırma modelinin performansını değerlendirmek için doğruluk, duyarlılık, belirginlik gibi metriklere odaklanılırken temel zorluklardan biri, hangilerine öncelik verileceğinin bilinmesidir. Bu tür ölçümler, model tahminlerinin eğitim veri setinden test veri setine ne kadar iyi genelleştirildiğini ifade etmektedir ve bu nedenle bu veri setlerinin elde edildiği koşullara ve dolayısıyla bağlamlara bağlıdır. Bu tür bir genellenebilirlik, perakende tüketici satın alma tavsiyeleri gibi bazı düşük riskli kararlar için yeterli olabilir, ancak daha yüksek bir zarar önleme standardına tabi olan ve “karıştırıcı unsurların sistematik olarak dışlanmasını gerektiren” tıbbi karar verme gibi bağlamlar söz konusu olduğunda yeterli bulunmayabilir (Tal, 2023: 16). Bu hususta da örneğin hedefin cezalandırma mı yoksa yardım mı olduğu gibi değişkenler ve hatalarla ilişkili maliyetlerin dikkate alınması gerekir. Bir yaptırım (para cezası) uygulanabilmesi için dolandırıcılık faaliyetini tespit edecek bir modelden faydalandığında, bireylere yanlışlıkla müdahale edilerek zarar verme riskinin ayırdında olmak

gerekir. Ya da bir mali yardımın yapılacağı adayların belirlenmesi gibi yardımcı müdahalelerde amaç ihtiyacı olan bireylere yardım etmek iken, bireylere ihtiyaç duyduklarında müdahale edilmeyerek de zarar verilebilir. Ancak titizlikle yönetilmiş bir metrolojik yani ölçümsel değerlendirme, hedef belirleme yanlışlığını azaltma, kullanıcılara daha net ve eyleme geçirilebilir bilgiler sağlama, adaleti ve kamu güvenini artırma gibi avantajlara sahip olacaktır (Tal, 2023: 17).

Aslında bir sistem için uygun adillik kriterlerinin belirlenmesi ve temin edilmesi kullanıcı deneyimi, kültürel, sosyal, tarihsel, politik, yasal ve etik gibi birçok hususun hesaba katılmasını gerektirir ve bunların çoğu bir diğerinden ödün vermeye de neden olabilir. Bu yüzden daima bir dengenin gözetilmesi şarttır, zira basit görünen durumlarda bile neyin adil olduğu konusu ekseriyetle fikir ayrılığına düşülmesinin sebebidir.

Örneğin, ChatGPT gibi büyük dil modelleri, en iyi, metin ve ses verilerinin çevrimiçi olarak bol olduğu İngilizce gibi dillerde çalışmaktadır. Milyonlarca insan tarafından konuşuluyor olsa da internette çok az bulunan Kannada gibi “düşük kaynaklı” dillerdeki YZ sistemleri, sıklıkla doktorların erkek ve hemşirelerin kadın olduğunu varsayarak önyargılı olabilir veya yerel lehçeleri anlamakta zorlanabilir (Perrigo, 2023). 100 milyon insan ve altı dil barındıran Etiyopya örneğinde ise Facebook’un -şu anki adıyla Meta’nın- bütünlük sistemleri için bu dillerden yalnızca ikisini desteklediği öne sürülmüştür (Hao, 2021). Myanmar’da Rohingya halkına yönelik “nefret söylemi”nin Facebook’un algoritmik sistemleri aracılığıyla viral olarak yayılmış olması sebebiyle de İngilizce dışındaki dillerde NLP teknolojilerinin geliştirilmesine yatırım yapılması gereği ortadır. Buna karşılık şirket, “sistemlerinin Birmanca ya da Burmaca metinleri yorumlayamaması nedeniyle, nefret söylemini bildiren kullanıcılara büyük ölçüde güvenmeye devam etmiştir” (Stecklow, 2018; Ricaurte, 2022: 732).

Bağlamla ilgili farkındalık eksikliğinin zararlı sonuçları sağlık sektöründe de kendini göstermektedir. Black ve Richmond (2019: 1), 9 Sahra Altı Afrika ülkesi ve yüksek gelirli ülkeler arasındaki erken meme kanseri tespit uygulamalarını inceleyen karşılaştırmalı çalışmalarında, Batı’da ‘başarılı’ olan mamografilerin Sahra Altı Afrika’da uygulanmasının meme kanserinden kaynaklanan ölümleri azaltmada etkili olmadığını ortaya koymuştur.

2.3.2. Şeffaflık (Açıklanabilirlik)

Bir sistemin kararlarının, nedenlerinin ve alt süreçlerinin izlenebilir ve bağlama uygun olarak açıklanabilir olması (bu alandaki YZ araştırmaları, “açıklanabilir YZ - Explainable AI: XAI” olarak adlandırılır) ve tüm sürecin doğrulanabilir ve yorumlanabilir olması şeklinde tanımlanabilecek olan şeffaflık, YZ etiği kılavuzlarında en sık atıfta bulunulan ilkedir; bununla birlikte yorumlanması, gerekçelendirilmesi ve uygulanması da büyük ölçüde değişmektedir (Jobin vd., 2019: 395; Ko ve Leem, 2021: 219).

Bireyler, öncelikle güvenliklerini veya insan haklarını etkilediği zamanlar dahil olmak üzere, bir karar YZ algoritmalarına dayalı olarak alındığında tam olarak bilgilendirilmeli ve bu durumlarda ilgili YZ aktöründen veya kamu sektör kurumlarından açıklayıcı bilgi talep etme fırsatına sahip olmalıdır, ayrıca bu tür bir açıklama zamanında yapılmalı, ilgili paydaşın (örneğin meslekten olmayan kişi, düzenleyici veya araştırmacı) uzmanlığına göre uyarlanmalıdır (CBDDO, 2021: 61). Bununla beraber, hak ve özgürlüklerini etkileyen bir kararın gerekçelerine erişebilmekten öte, kararı gözden geçirip düzeltebilecek özel sektör şirketi veya kamu sektörü kuruluşunun görevlendirilmiş bir personeline bildirimde bulunabilme seçeneğine de sahip olmalıdırlar. Bu yolla şeffaflık, yolsuzluğu ve ayrımcılığı azaltabilen ve ayrıca insan hakları üzerindeki olumsuz etkilerin tespit edilip önlenmesine yardımcı olabilecek kamu denetimine izin verir (UNESCO, 2021: 9). Aynı zamanda, insanların bir YZ sisteminin her aşamasının, YZ sisteminin bağlamına ve hassasiyetine uygun olarak nasıl devreye alındığını anlamasını sağlayabilir; belirli bir öngörüü veya kararı etkileyen faktörlere ve güvenlik veya adalet önlemleri gibi uygun güvencelerin yürürlükte olup olmadığına ilişkin içgörü de içerebilir. Şeffaflığın olmaması ise YZ sistemlerinin ürettiği sonuçlara dayalı olarak kararlara etkili bir şekilde itiraz etme olasılığını baltalar; bu nedenle adil yargılanma ve etkili başvuru hakkını da ihlal edebilir, dolayısıyla bu sistemlerin yasal olarak kullanılabileceği alanları sınırlar (UNESCO, 2021: 9). Zira insanlar karara nasıl ulaşıldığını anlayamadıkları sürece, özellikle de kararın kişisel düzeyde kendileri için bir şekilde olumsuz olduğu durumlarda (örneğin, bir kredi başvurusunun reddedilmesi) o karara güvenmeme eğilimindedirler (ISO/IEC 22989:2022 (E), 2022: 209).

Etik açıdan daha uygun, teknolojik şeffaflık ile ve insan edimi arasındaki ilişkiyi daha yararlı bir hale getirebilen YZ sistemleri tasarlanırken, “insan-teknoloji etkileşimlerinin çeşitli boyutlarıyla ilişkilendirilebilecek farklı şeffaflık kavramları”nın dikkate alınması gerekir (Andrada vd., 2023: 1321). Kullanılan algoritmaların doğası ve bir eğitim sürecinde kullanılan

veri çeşitleri gibi bilgi türlerini hedefleyen “**bilgi şeffaflığı**”, belirli bir enformasyonel sistemi gerçekleştiren donanımın lüzumunu, üretim ve bakım gereksinimlerini ve buna eşlik eden emek ilişkilerini içeren “**maddi şeffaflık**” ve YZ teknolojisinin yaygın kullanımının ortaya çıkardığı nöral ve bedensel dönüşümlerle ilgili olan “**dönüşümsel şeffaflık**” bunlardan bazılarıdır (2023: 1324-1325).

Bir YZ sistemi hakkında şeffaflık sağlamak; sistem hakkında hedefler, bilinen sınırlamalar, tasarım seçimleri, varsayımlar, algoritmalar, eğitim yöntemleri, modeller ve kalite güvence süreçleri gibi uygun bilgilerin paydaşlara iletilmesini içerebilir (ISO/IEC 22989:2022(E), 2022: 30). Ek olarak, bir YZ sisteminin şeffaflığı, paydaşların, sistemi üretmek için kullanılan verilerin ayrıntıları (ne, nerede, ne zaman, neden veri toplandığı ve nasıl kullanıldığı gibi) ve sistemin amacı, nasıl inşa edildiği ve konuşlandırıldığı ile kişisel verilerin korunması hakkında ya da ilgili kararları almak için kullanılan işlem ve otomasyon düzeyi hakkında bilgilendirilmesini de içerebilir. YZ sisteminin örgütsel karar verme sürecini, sistemin tasarım seçimlerini ve onu dağıtma gerekçesini etkileme ve şekillendirme derecesine ilişkin açıklamalar mevcut olmalıdır; kısacası “iş modeli” şeffaflığını da sağlamak gerekir.

YZ sistemlerinde "kod yolu" milyonlarca parametre ve matematiksel işlem içerebilir ve hatalı bir karara yol açan belirli bir sorunu tam olarak belirlemek bu sebeple çoğunlukla zordur. Özellikle derin öğrenme sinir ağları sistemlerinin karmaşıklığı, sistemin bir karara nasıl ulaştığına dair anlamlı bir açıklama yapmayı zorlaştırabilir, dolayısıyla genellikle “**kara kutu**” olarak nitelendirilirler; iç işleyişleri gözlemlenemez ve anlaşılamayabilir ve aynı zamanda bu sistemlerin kurumsal bağlamları içinde uygulamaları ve konuşlandırılmaları da şeffaf olmaktan uzak olabilir (Valentino-DeVries, 2020). Bu “**opaklık**” farklı sebeplerden dolayı ortaya çıkabilmektedir; risk endişeleri nedeniyle bazen sistemler kasıtlı olarak kara kutu olarak tasarlanmaktadır, zira şeffaflık amacıyla bazı bilgilerin ifşa edilmesi güvenlik, gizlilik veya mahremiyet gerekliliklerine ters düşebilir (ISO/IEC 22989:2022(E), 2022: 30). Yaşam döngüleri boyunca YZ sistemlerinin şeffaflığını ve açıklanabilirliğini artırmak için çaba gösterilmesi gerekse de mahremiyet, emniyet ve güvenlik gibi diğer ilkeler arasında bir denge kurma gereği nedeniyle şeffaflık ve açıklanabilirlik düzeyi her zaman **bağlama ve etkiye uygun** olmalıdır (UNESCO, 2022: 22). Örneğin, bir algoritma hakkındaki bilgiler, sahibinin fikri mülkiyet haklarını korumak için kısıtlanabilir (Mittelstadt vd., 2016: 9). Oldukça şeffaf bir sistemde, bir saldırganın sistemin arızalanmasına neden olacak bir girdiyi dikkatlice tasarlaması halinde birçok YZ algoritmasının kandırılabilir olması ihtimali de söz konusudur. Adalet,

ayrımcılık ve güven hususlarına yönelik endişelerin hafifletilmesi ve etik sorunların önüne geçilebilmesi amacıyla daha şeffaf uygulamalar geliştirmeye ihtiyaç duyulurken aynı zamanda sistem ile ilgili açıklamaların etik açıdan kendi risklerini ve gözden kaçabilecek dezavantajları da beraberinde getiriyor olması, kasıtlı manipülasyonların varlığı olasılığında zararın artması ve sisteme ilişkin güven kaybına yol açabilme potansiyeli YZ'nin “**şeffaflık paradoksu**” olarak adlandırılabilir duruma da işaret etmektedir (Burt, 2019; Shokri vd., 2021: 1). Misal, kamu sektörü verilerinin her zaman “**açık kaynak**” olmaması ve halkın hükümet politikalarına güveninin sağlanması gibi dikkate alınması gereken detaylar da bulunmaktadır.

Yorumlanabilirlik ve tahmin gücü arasında süregelen bir değiş tokuş da mevcuttur. Yüksek düzeyde yorumlanabilir modellerin karmaşık ilişkileri yakalamada sınırlamalarla karşılaşabilmesi ve bunun da doğruluğun azalmasına neden olabilmesi söz konusudur. Karmaşık modeller ise genellikle yorumlanabilirlik pahasına üstün doğruluk elde etmektedir. Dolayısıyla bu hususların, **uygulama alanının özel gereksinimlerinin de göz önünde bulundurularak** dengelenmesi önem arz eder; sağlık hizmetleri veya finansal sektör gibi bağlamlarda yorumlanabilirlik ve şeffaflık, mevzuata uygunluğun sağlanması ve etik hususların ele alınmasında çok önemli roller oynar (Albahri vd., 2023; Kurshan vd., 2021). Doğruluğun çok önemli olduğu tıbbi görüntü veya sinyal tanıma gibi diğer alanlarda ise odak noktası yorumlanabilirlikten ziyade tahmin gücü olabilir (Komorowski vd., 2023; Yang vd., 2023: 163).

Son olarak, Hongladarom'un (2020), teknolojinin dinamik doğasının yanı sıra sosyo-tarihsel bağlamlara gömülü olduğunu vurgulayan postfenomenoloji analiz biçimini öneren Ihde'nin çalışmalarını takip ederek, yüz tanıma teknolojisinin fenomenolojik ve hermeneutik bir ön analizini sunduğu çalışmasına değinmek yerinde olacaktır. Hongladarom (2020: 2), yüz tanıma fenomeninin tamamını ve bunun altında yatan makine makine öğrenimi ve YZ'yi bir çifte aracılık olgusu olan ‘makine hermeneutiği’ aracılığıyla anlayabileceğimizi ileri sürer. Bu, hermeneutiğin failinin günümüzde sadece biz insanlar olmadığı, giderek artan bir şekilde makinelerin de kendi hermeneutik işlerini yaptığı gerçeğini yansıtmak içindir ve insanların iki yorumsama katmanına tabi olduğu anlamına gelmektedir. İlk katman, bilimsel araçlar yoluyla insan algısı düzeyinde gerçekleşir, ikinci katman ise algoritmik makinenin kendisi veri girdisini yorumladığı ve insanınkinden farklı olmayan bir şekilde zaten işlenmiş ve yorumlanmış bir şey olarak çıktı verdiği zaman mevcuttur. Hongladarom (2020: 7), makineler böyle bir yeteneğe sahip olduklarında “işlerini etik bir şekilde yaptıklarından ve ürettiklerinin doğru olduğundan nasıl emin olabileceğimiz epistemolojik felsefi sorunu”nun ortaya çıktığını savunur; gerçeklik

ve yorumlanan şeyin birbirine karıştığı ve belki de ikisini birbirinden ayırmanın hiçbir yolu olmadığı anlamına gelir, algıladığımız dışsal, nesnel gerçeklik olarak kabul ettiğimiz şey giderek daha fazla, ‘makine hermeneutiği’nin bir sonucu olmaktadır. Bunun fenomenolojik açıdan anlamını, kendimizi giderek daha fazla makine dünyasına entegre olmuş halde bulmamız olarak açıklar. Bu, **bilhassa yüz tanıma gibi kullanım alanları YZ etiği gündeminin en tartışmalı hususları iken** (Big Brother Watch, 2018), algoritmaların en başından itibaren etik olacak şekilde programlanmasının önemine işaret etmektedir; Hongladarom’a göre böylelikle değerler ve hedefler algoritmaların kendi hedeflerinde mevcut olacaktır. Dolayısıyla, şeffaflık ve bununla birlikte açıklanabilirlik, izlenebilirlik, denetlenebilirlik ve hesap verebilirliğin temininin daha da elzem hale geldiği söylenebilir.

2.3.3. Kötü Niyetli Olmama (Zarar Vermeme)

Bu ilke, YZ’nin insanlara zarar vermekten kaçınacak şekilde geliştirilmesini, konuşlandırılmasını ve kullanılmasını savunur. Katil robotlar, otonom araçlar ve drone teknolojisi, bu hususta irdelenen en belirgin başlıklar olmuştur (Jang vd., 2022: 11640). Ancak, fiziksel zarar ve siber güvenlik risklerinin ötesinde YZ’nin uzun vadede sosyal refah, altyapı ve psikolojik, duygusal ve ekonomik yönler üzerinde olumsuz etkileri olabileceği önermesini de kapsamaktadır.

Yankı odası etkilerinden propaganda botlarının kullanımına veya sahte haberlerin yayılmasına kadar YZ’nin her zaman sosyal uyumu azaltma ve radikalleşmeyi teşvik etme, sosyal bölünme üzerinde kilit bir rol oynadığı görülür (Vosoughi vd., 2018: 1146; Hagendorff, 2020: 110). YZ’nin politika ve demokratik süreçler üzerindeki etkisine ilk örneklerden biri Facebook Cambridge Analytica skandalıdır (Nemitz, 2018: 2). “2.3.1. Adalet” başlığı altında da değinilen, Facebook’ta Myanmar Silahlı Kuvvetleri’ne bağlı hesaplar üzerinden Rohingya halkına yönelik nefret söyleminin viral bir şekilde yayılmasına karşı şirketin, siyasi ve ekonomik çıkarlarıyla ilişkilendirilebilecek siyasi ‘tarafsızlığını’ beyan ederek herhangi bir işlem uygulamamış olması da bir başka örnektir (Ricaurte, 2022: 732-733).

Büyük dil modellerinin yanlış bilgi ve zararlı içerik oluşturmak için olası kötüye kullanımı konusuna da dikkat çekilmektedir (Coeckelbergh ve Gunkel, 2023: 1). Zira, propagandacılar genellikle dezenformasyon üretmek için önemli bir zamana ihtiyaç duyarlar ve bu modeller toplu metin üreterek endüstriyel ölçekte dezenformasyon üretmeyi kolaylaştıracaktır. Yakın tarihli araştırmalar da GPT serisi gibi modellerin, insan kaynaklı

olanlardan ayırt edilemeyecek haber makaleleri ve çok az insan müdahalesi ile büyük miktarlarda dezenformasyon üretebildiğini, dezenformasyonun yayılmasını şiddetlendirdiğini, ayrımcılığı ilerlettiğini ve devletlerin ve şirketlerin casusluk yapmasını da kolaylaştırdığını ortaya koymaktadır (Kreps vd., 2022; Jakesch vd., 2023; Buchanan vd., 2021; Spitale vd., 2023). Yeni modeller çevrimiçi hale geldikçe ve genellikle “sosyal etkileşim gibi insana özgü olduğuna inanılan alanlara yayıldıkça”, insanları tasarlanan planları gerçekleştirmeleri için manipüle etme konusunda daha yetenekli sistemler ve bu yeteneklerden fayda gözeten karmaşık sosyal mühendislik saldırıları söz konusu olabilir (Murphy ve Criddle, 2023). Medya içerikleri ve analizler, YZ’nin, halk arasında derin sahte (deep fake) olarak bilinen, giderek daha gerçekçi hale gelen fotoğraf, ses ve video dijital sahteciliğindeki rolünü de vurgulamaktadır (Chesney ve Citron, 2019: 1753). Yeterli veri ve bilgi işlem gücü ile derin öğrenme modelleri, gerçekçi görüntüler, ses ve metin oluşturmak da dahil olmak üzere son derece güçlü olabilir; hatta o kadar etkilidirler ki derin öğrenmedeki “derin” kelimesi “derin sahte” terimini doğurmuştur (Byman vd., 2023: 3). Derin sahtelerin ve “yeni dezenformasyon savaşı”nın seçimler üzerindeki etkisinden ve “kamuoyunun gerçeği algılamasından” endişe duyulan bir “hakikat sonrası jeopolitik çağ” tanımlanmaktadır (Farid, 2023; Solon, 2023). Medyanın doğası gereği büyük ölçüde partizan olduğu, sosyal medya platformlarının yeni bilgi iletişim ortamının büyük bir bölümünü oluşturduğu, haber ve bilginin hem üretilme hem de tüketilme biçimini değiştirdiği ülkelerde dil modelleri gibi “üretken YZ modelleri” (Goldstein vd., 2023: 63) ve daha inandırıcı ve erişilebilir derin sahte sentez teknikleri ile fotoğraf ve video materyallerinin daha kolay manipüle edilebilmesi; dezenformasyon kampanyalarını geleneksel medya dönemindeki kapsamı ve etkinliğinin ötesine taşıyarak tehdit boyutunu daha da artırmaktadır (Helmus, 2022: 6). Dolayısıyla yanlış haberler üretmek, ünlü şahsiyetleri karalamak, hükümet yetkililerine şantaj girişiminde bulunmak, kamuoyu söylemini ve seçimleri etkilemek (Westling, 2019: 4), vatandaşların kurumlara ve kamu otoritelerine güvenini aşındırmak, kutuplaşmayı artırmak (Waldemarsson, 2020: 10) gibi çok sayıda kötü niyetli faaliyet için kullanılabilecek derin sahte teknolojisinin tespit edilmesi ve önlenmesi pratik bir güvenlik sorunu teşkil etmektedir (Hussain vd., 2022: 2; Sayler ve Harris, 2022: 1).

Konuşma tanıma teknolojisinin yaygın bir şekilde telefon dinlemelerine ve böylelikle sivil özgürlüklerin kaybına yol açabileceğine de dikkat çekilmektedir. Bilhassa İngiltere gibi “liberal demokrasilerde” geniş gözetleme kamerası ağlarının yaygınlaşması ile sivil toplum örgütleri, mahremiyet kaybı endişeleriyle beraber devletin birey üzerindeki güç asimetrisiyle

mücadele etme gereğini gündeme getirmiştir (Feldstein, 2019: 2; Russell ve Norvig, 2009: 1036).

Diğer taraftan en tartışmalı YZ uygulamalarından otonom YZ sistemleri, muharabe ortamında yaygın olarak kullanılmaktadır, robotik silahlar da ek riskler doğurmaktadır. “İnsanların karar verme mekanizmasının ateşleme döngüsünden çıkarılması durumunda” masum sivillerin öldürülmesine yol açan kararlar verilebilir. Füze savunma sistemleri gibi otonom sistemlerde durum tahmininin yanlış olabilme ihtimali ve bunun yanlış girişimlere neden olabilme riskini azaltmanın yolu, tek bir durum tahmin hatasının sistem boyunca kontrolsüz bir şekilde yayılmaması için kontrol ve dengelemelere sahip bir sistem tasarlamaktır (Russell ve Norvig, 2009: 1037). Daha büyük ölçekte, güçlü robotlara sahip olmak, “sağlam kasklara sahip olmak gibi” bir ulusa aşırı güven verebilir ve pervasızca savaşa girmesine neden olabilir, halihazırda çoğu savaşta, taraflardan en az biri askeri yeteneklerine aşırı güvenmektedir (2009: 1035).

YZ modellerinin biyolojik silah saldırılarına olanak sağlamasının nasıl önlenebileceği de bir başka endişe kaynağıdır. Bu doğrultuda politika yapıcılarının, DNA sentez şirketlerinin tehlikeli patojenlerle (veya potansiyel patojenlerle) ilgili DNA dizilerini yetkisiz müşterilere göndermesini engelleyen düzenlemeler oluşturmaları, hangi genetik dizilerin tehlikeli olabileceğini belirlemeye çalışan DNA sentez şirketlerini desteklemeleri, yeni patojenlerin takibi için sürekli olarak kanalizasyon ve havaalanlarını gözetlemeleri gerekmekte; siber güvenlik firmalarının da bilgisayar sistemlerinin tehlikeli YZ modelleri aracılığıyla hacklenmesini önlemek maksadıyla güvenlik açıklarını bulmaları ve bu açıkları yamamak için başka YZ sistemleri geliştirmeleri gibi çeşitli faaliyetler yürütmeleri gerekmektedir (Anderljung ve Scharre, 2023). İstihbarat sistemlerinin bir dizi hedef için konuşlandırılabilmesi (Bostrom, 2014: 107) göz önüne alındığında, geliştirmek veya konuşlandırmak için çok az uzmanlık gerektiren yüksek yetenekli sistemlere, onları hackleyerek veya baştan geliştirerek yeni, tehlikeli hedefler verilebilecek olması da bir başka kaygı unsurudur (Brundage vd., 2018: 65). Biyolojik ve kimyasal silahlar şu anda sadece eğitimli profesyoneller tarafından üretebiliyor olsa da bilgi işlem donanımının daha uygun maliyetli hale gelmesi, daha güçlü YZ modellerini daha düşük bir fiyata eğitmeyi mümkün kılması ve YZ modellerinde ilerlemeler neticesinde gelecekte düşman devletler ya da teknolojik bir avantaj arayan kötü aktörler, bilimsel uzmanlık gereksinimi duymadan ölümcül patojenler üretebilir, açık kaynaklı yazılım araçları ve kamu veritabanlarını dolduran birçok veri kümesi de hiçbir gözetim olmaksızın

kullanılabilir (Urbina vd., 2022: 189). Nisan 2023'te Carnegie Mellon Üniversitesi'nden bir grup akademisyen, bilimsel deneylerin otonom tasarımı, planlanması ve yürütülmesi için birden fazla büyük dil modelini birleştiren bir “Akıllı Ajan” sistemi sundukları çalışmalarında, Ibuprofenin tarifini ve nasıl üretileceğini bilen araçlarının, Advil'den çok daha tehlikeli kimyasalları sentezleyebileceğini tespit ettiklerini belirtmişlerdir (Boiko vd., 2023: 3). Dahası, ölümcül sinir gazı olarak bilinen “sarin”i sentezleme aşamasına geçmeden Google'da bu bileşiğin karanlık geçmişini araştıran sistem, kimyagerlerin, "arama işlevi"nin "terminoloji değiştirilerek kolayca manipüle edilebileceğini düşünmelerine sebep olmuştur (Anderljung ve Scharre, 2023). Bulgular, YZ sistemlerinin “sentetik patojenler yaratabilmesi” veya “enerji santralleri gibi kritik altyapı sistemlerini hackleyebilmesi” ihtimallerine işaret etmektedir. Hassas deneylerin, özellikle de potansiyel olarak zararlı maddeler veya metodolojiler içerenlerin gözden geçirilmesi ve onaylanması için;

- i. döngü içinde bir insan bileşeninin varlığı,
- ii. sistem internetten ve operasyonel dokümantasyondan topladığı verilerin kalitesine dayandığı için sistemin güvenilirliğini korumak amacıyla veri kaynaklarının sürekli olarak iyileştirilmesi ve güncellenmesi, sistemin karar verme sürecini belirlemek için en güncel ve doğru bilgilerin kullanılmasının sağlanması ve
- iii. büyük dil modelleri ve otomatik deneyler de dahil olmak üzere birden fazla bileşenin entegrasyonu güvenlik riskleri oluşturduğu için sistemi yetkisiz erişim, kurcalama veya kötüye kullanımdan korumak maksadıyla şifreleme ve erişim kontrolü gibi sağlam güvenlik önlemlerinin uygulanması önerilmektedir (Boiko vd., 2023: 13).

Son olarak, kötü niyetli olmama vurgusunun etiği teknik sorunlar için teknik çözümler bulma meselesi haline getirmemesi gereğinin de altını çizmek gerekir. Teknik sistemlerin içine gömülü olduğu daha geniş etik ve toplumsal bağlam unutulmakta, sorunlar teknik "düzeltmelerle" ya da iyi tasarım yoluyla çözülebilir olarak görülmektedir. Böyle bir yaklaşım ile YZ'nin kontrol, yönetim ve toplumsal boyutlarını yönlendiren birçok önemli konu göz ardı edilecektir. Evgeny Morozov'un "teknolojik çözümcülük" olarak adlandırdığı bu hal, teknolojinin neden olduğu sorunların her zaman daha fazla teknoloji ile çözülebileceği inancıdır. Bir sorun icat ederek, bu kurguyu gerçek ve acil bir ikilem olarak yanlış tanıtarak ve sorunu çözmek için teknolojinin kullanılmasını savunarak kötü sorunların asal hususlarını göz

ardı etmekten kaynaklanır, nihayetinde derin ve zor etik sorunlar aşırı basitleştirilmekte ve cevapsız kalmaktadır (Sancho-Gil, 2020: 198).

2.3.4. Gizlilik (Mahremiyet)

Dijital dünyada, kimin hangi veriye erişebildiğini kontrol etmenin, analog dünya ile kıyaslandığında çok daha zor olduğu söylenebilir. Fotoğraf ve videolarda yüz tanıma ve yaygın olarak kullanılan cihaz parmak izi gibi tanımlama teknikleri, kimlik tespitine ve bireylerin profilinin çıkarılmasına olanak sağlamaktadır; arkamızda bıraktığımız *dijital ayak izleri* de aslında ücretsiz hizmetlerimizin nasıl ödendiğidir (Müller, 2023). Kalluri vd. (2023), özellikle bilgisayarlı görü alanının sadece tarafsız bir bilgi arayışı olmadığını, bir sürveyans (gözetim) paradigması için temel bir katman oluşturduğunu ortaya koymaktadırlar. Bulgular, makale ve patentlerin %90'ının, teknolojilerinin insan verilerini hedefleyebilmesini güçlü bir yön olarak vurguladığına ve sadece insan verilerinin geniş çapta hedef alınmakla da kalınmadığına, aynı zamanda bu makale ve patentlerin çoğunluğunun (%68) açıkça insan vücut parçalarının (örneğin yüzler) ve insan bedenlerinin gözetimine odaklandığına işaret etmektedir (2023: 5).

"Gözetim ekonomisinde" ya da "gözetim kapitalizminde" sosyal medyanın, oyunların ve internetin büyük bir kısmının birincil odak noktası da dikkat çekmek, bu dikkati sürdürmek ve yönlendirmek, dolayısıyla da veri temin etmektir. Şirketlerin "*mikro hedefleme*", "*dürtme*" ve "*kullanıcı deneyimi tasarımı*" benzeri tekniklerle "insan özerkliğini sorgulatacak kullanıcı davranışını manipülasyon yolları" söz konusuysa, insan onurunun ve failliğinin korunması için gerekli bir hak olan mahremiyete YZ sistemlerinin yaşam döngüsü boyunca saygı gösterilmesi, mahremiyetin korunmasının teşvik edilmesi gerekir (Fogg, 2002: 89; Matz vd. 2017: 12717; Hagendorff, 2020, 110). Öte yandan bu durum, bazı şirketler için yine "iyi bir fiyat karşılığında satılabilecek bir rekabet avantajı" teşkil etmektedir. Zira devlet kuruluşları, Google veya Meta gibi özel şirketler ve perakendeciler tarafından toplanan, güçlü bilgi işlem yetenekleri ve veri analitik araçlarıyla desteklenmiş veri tabanları artık "nesnelerin interneti", "akıllı ev", "akıllı şehir" ve "akıllı yönetim" gibi pek çok kavram ile farklı türlerde, gerçek zamanlı ve daha ayrıntılı veriler sunan bir veri toplama mekanizmasının parçası haline gelmektedir.

Feldstein'in (2019: 16) makalesinde sunduğu, dünya çapında YZ gözetim kullanımına ilişkin ampirik verileri derleyen YZ Küresel Gözetim Endeksi'ne göre, 176 ülkenin en az 75'i YZ teknolojilerini gözetim amacıyla aktif olarak kullanmaktadır;

- hizmet sunumunu, şehir yönetimini ve kamu güvenliğini kolaylaştırmak için gerçek zamanlı veri ileten sensörlere sahip “**akıllı/güvenli şehirler**” (56 ülke),
- bireylerin depolanmış veya canlı görüntülerini veri tabanlarındaki görüntülerle eşleştirmek için kameraları (hareketsiz görüntüler veya video) kullanan ya da toplu demografik eğilimleri değerlendiren veya yüz tanıma kalabalık taraması yoluyla daha geniş “duygu analizi” yapan biyometrik teknoloji olarak tanımlanabilecek “**yüz tanıma sistemleri**” (64 ülke),
- gelecekteki suçlar hakkında tahminlerde bulunmak için de algoritmik analiz içerebilen, soruşturmaları ve polis müdahalesini kolaylaştırmak amacıyla kullanılan veriye dayalı analitik teknoloji olan “**akıllı polislik**” (54 ülke).

Gözetim programlarını desteklemekten doğrudan sorumlu olmayan, ancak uygulamaların hayata geçirilmesi için gerekli olan kritik yetenekleri sağlayan ikincil ya da “etkinleştirici” teknolojiler ise,

- **otomatik sınır kontrol sistemleri:** Havaalanı veya sınır erişimini otomatik olarak kontrol etmek için yüz tanıma ile çalışan biyometrik sistemler (biyometrik pasaport ve kimlikleri de içerebilir),
- **bulut bilişim:** YZ işleme ve operasyonlarını mümkün kılan altyapı bileşenleri, ağlar ve sensörler (bulut sunucuları, veri merkezleri, IOT ağları gibi),
- **nesnelerin interneti (IoT):** Bulutta analitik işleme için verilerin paylaşılmasına izin veren internet üzerinden bağlanan cihazlar,
- dijital devlet, YZ eğitim merkezleri ve YZ araştırma enstitüleri gibi **diğer ilgili teknolojiler** olarak sıralanabilir (Feldstein, 2019: 21).

Ayrıca başlıcaları Huawei, Hikvision, Dahua ve ZTE olarak sayılabilecek Çinli şirketlerle bağlantılı teknolojilerin, 36’sı Çin’in Kuşak ve Yol Girişimi’ni imzalamış olan, Türkiye’nin de dahil olduğu toplam 63 ülkeye YZ gözetim teknolojisi sağladığı dikkat çekmektedir; 14 ülke ile sırayı takip eden en büyük tedarikçi de Japon NEC Corporation’dır (Feldstein, 2019: 25-28). IBM, Palantir ve Cisco en önemlileri olmak üzere ABD firmaları tarafından tedarik edilen YZ gözetim teknolojisi ise toplam 32 ülkede mevcuttur.

Bahsedilen çalışmaya dair değinilmesi gereken son bir husus da söz konusu endeksin, bir ülkenin askeri harcamaları ile hükümetin YZ gözetim sistemlerini kullanması arasında güçlü bir ilişki mevcudiyetini ortaya koymuş olmasıdır; kümülatif askeri harcamalara göre dünyanın

en çok askeri harcama yapan ilk 50 ülkesinden 40'ı aynı zamanda YZ gözetim teknolojisini de kullanmaktadır (Feldstein, 2019: 2).

Diğer taraftan, genel olarak “YZ sistemlerine ilişkin verilerin ilgili ulusal, bölgesel ve uluslararası yasal çerçevelere saygı gösterilerek, uluslararası hukuka uygun ve belirtilen değer ve ilkeler doğrultusunda toplanması, kullanılması, paylaşılması, arşivlenmesi ve silinmesi” önemlidir (UNESCO, 2022: 21). “Kişisel verilerin kimden, nasıl sağlandığı ve veriye dayalı olarak alınan kararların insanları nasıl etkileyeceği her zaman denetime açık bir şekilde izlenebilir olmalıdır.” (CBDDO, 2021: 60). “YZ sistemleri, bir sistemin tüm yaşam döngüsü boyunca gizliliği ve veri korumasını garanti etmelidir”; bu, başlangıçta kullanıcının sağladığı bilgilere ek olarak, “YZ sisteminin belirli kullanıcılar için ürettiği çıktılar” ya da “kullanıcıların belirli önerilere nasıl yanıt verdiği” gibi kullanıcının sistemle etkileşimi boyunca kullanıcı hakkında oluşturulan bilgileri içerir ve bu dijital kayıtlar, YZ sistemlerinin, bireylerin tercihleriyle beraber yaşlarını, cinsiyetlerini, cinsel yönelimlerini, dini veya siyasi görüşlerini de çıkarsamasına izin verebilir (AI HLEG, 2019:17). Verilere erişim konusunda ise bireylerin verilerini işleyen herhangi bir organizasyonda, veri erişimini yöneten ve verilere kimlerin hangi koşullar altında erişebileceğini ana hatlarıyla belirten veri protokolleri uygulamaya konmalıdır; yalnızca yetkinliği olan ve usulüne uygun olarak kişinin verilerine erişme ihtiyacı bulunan nitelikli personelin bunu yapmasına izin verilmelidir (AI HLEG, 2019: 17).

Aslında dünya çapında özel hayatın gizliliği hukukunun gelişimini, Avrupa İnsan Hakları Sözleşmesi'nin özel hayatın gizliliğine ilişkin 8. Maddesi etkilemiştir (CAIDP, 2023: 6). Kişisel verilerin toplanması ve işlenmesine eşlik edebilecek suistimallere karşı bireyi koruyan ve kişisel verilerin sınır ötesi akışını düzenleyen ilk bağlayıcı uluslararası belge olan 108 sayılı Avrupa Konseyi Sözleşmesi (1981) ise 2018 yılında Avrupa Konseyi tarafından değiştirilmiş ve “erişim haklarını kullandıklarında veri sahiplerine iletilecek bilgi kataloğunu genişletmekte” olan 108+ sayılı Avrupa Konseyi Modernize Sözleşmesi imzaya ve onaya açılmıştır. Madde 9(1)(c), özellikle YZ karar verme sürecini ele almaktadır (CAIDP, 2023: 6). “Ayrıca, veri sahipleri, sonuçları kendilerine dayatılan veri işlemenin altında yatan gerekçeler hakkında bilgi edinme hakkına sahiptir. Bu yeni hak, bireylerin profillerinin çıkarılması açısından özellikle önemlidir” (CAIDP, 2023: 7). Avrupa Konseyi üyelerinden olan Türkiye, Gizlilik Sözleşmesini tadil eden Protokolü (No 223) henüz imzalamamış ve onaylamamış olan devletlerdendir, ancak kırk iki devlet bu Protokolü imzalamıştır (Council of Europe Treaty Office, 2023).

2.3.5. Sorumluluk (Hesap Verebilirlik)

İlişkisel bir kavram olan sorumluluk akıllı sistemler bağlamında ele alındığında, “bu tür sistemlerin tasarımından ve kullanımından ve bunların sonuçlarından kimin veya neyin sorumlu olduğu”, “bu sorumluluğun nasıl dağıtıldığı” ve “sorumluluğun atfedilmesinin sonuçlarının kim tarafından veya ne olduğu” gibi takip eden sorulara işaret etmektedir (Stahl, 2023: 3). Diğer taraftan, YZ teknolojilerinin sorumlu bir şekilde geliştirilmesi ve kullanılmasına yönelik referanslar tüm kılavuzlarda yer alırken sorumluluğun tanımı bir bütün olmaktan daha ziyade, farklı yönlerine yapılan atıflarla ifade edilmektedir.

Akıllı sistemlerden kimin ya da neyin sorumlu olduğu sorusunun cevabı için öne çıkan adaylar arasında geliştiriciler, çeşitli bağlamlarda kullanıcılar (örneğin tüketiciler ve aynı zamanda ortak tasarımcılar), bireysel ya da kurumsal sahipler, düzenleyiciler ve yasa koyucular dahil olmak üzere toplumsal aktörler ve hatta teknik sistemin kendisi yer almaktadır (Stahl, 2023: 1). Ayrıca “sorumlu özne” ve “öznenin sorumlu olduğu nesne ve sorumluluk ilişkisinin pratik sonuçlarını belirleyen ve uygulayan otorite” gibi sorumluluk ilişkisinin temel bileşenleri bulunmakla birlikte, “sorumluluğun türü” (ahlaki sorumluluk, yasal sorumluluk veya rol sorumluluğu gibi) gibi bu sorumluluk ilişkisinin sosyal gerçekliğini etkileyen kavramlar da mevcuttur (Stahl, 2023: 3). Bir sorumluluk ilişkisinin pratikte nasıl görüneceğini belirleyen diğer yönler ise “yaptırımlar” (cezalar ve ödüller), “otorite türü veya geçişliliği” (bir başkasına atfedilir olması), “zamansal boyut” (ileriye veya geriye dönük), “öznenin türü” (bireysel, organizasyonel veya teknik yapı gibi) veya “bir öznenin yerine getirmesi gereken koşullar” (örneğin; rasyonellik, yaptırımlara tepki verme yeteneği) olarak sıralanabilir. Bununla beraber bu kavramlar bir dereceye kadar birbirine bağımlıdır ve çok sayıda permütasyona da yer açmaktadır (Stahl, 2023: 3).

Stahl’ın “*meta-sorumluluk*” argümanında, “felsefe ve sosyal bilimler perspektifinden YZ etiği sorularını takip eden akademisyenler”den “bu sistemleri tasarlayan ve geliştiren uzmanlar”a kadar çeşitli paydaş grupları muhatap alınır. Stahl (2023: 2) bilhassa bu uzmanların, çalışmalarına ilişkin etikle ilgili tartışmaların farkında olsalar da genellikle çabalarının sonuçları için ahlaki sorumluluk sorusuna pratik yanıtlar vermekte zorlandıklarını öne sürer ve sorumluluğu akıllı sistemlerin ekosistemlerine yerleştirmenin, onlara etik sorular üzerinde daha bütüncül bir şekilde düşünme ve böylece sosyal olarak kabul edilebilir olma olasılığı daha yüksek seçenekler bulma yolu sağlayacağını savunur. Bazı araştırmacılar ise, “*dağıtılmış sorumluluğu*” savunmaktadır (Floridi, 2013: 727; Taddeo ve Floridi, 2018: 751). Bu yaklaşıma

göre, YZ tabanlı kararların veya eylemlerin sonuçları genellikle tasarımcılar, geliştiriciler ve kullanıcılar gibi çok sayıda aktör arasındaki çoklu etkileşimlerin ürünüdür; YZ ile ilgili konulara birden çok kurum dahil olduğu için sorumluluk da dağıtılmalıdır.

YZ sistemlerinin yaşam döngüsünün herhangi bir aşaması için ve YZ sistemleriyle ilgili telafi durumlarında da gerçek kişilere veya mevcut tüzel kişilere etik ve yasal sorumluluk atfetmenin her zaman mümkün olmasını sağlanmalıdır (UNESCO, 2021: 9). Hesap verebilirliği sağlamak için ise “muhbirlerin korunması” dahil olmak üzere uygun gözetim, etki değerlendirmesi, denetim ve durum tespiti mekanizmaları geliştirilmelidir ve YZ sistemlerinin denetlenebilirliği ve izlenebilirliği özellikle insan hakları normları ve standartları ile çatışmalar ve çevresel ve ekosistem refahına yönelik tehditlerin ele alınması açısından hem teknik hem de kurumsal tasarımlar tarafından sağlanmalıdır -ki bu beraberinde özellikle insan yaşamını etkileyen kritik görev uygulamalarında şeffaflığın teminini de zorunlu tutar (UNESCO, 2021: 10). Çünkü, önceki başlıklar altında da ele alındığı üzere, YZ sistemlerinin genellikle daha büyük bir teknolojik sistemin ve veri geçmişinin parçası olması sebebiyle karşılaşılan soruna YZ modelinin mi yoksa sistemin bir başka bölümünün mü sebep olduğunu ayırt etmek bile çoğunlukla başlı başına bir problem teşkil etmektedir.

Bu noktada, YZ sistemlerine ne ölçüde özerlik verilmesi gerektiği veya beklenmedik sorunlara karşı koruma sağlamak için bu sistemlerin ne dereceye kadar “**insan gözetimi ve tespiti**”ne tabi kalması, yani insanların ne dereceye kadar “döngüde” kalması gerektiği tartışmalarına da dikkat çekmek icap eder (Mazzolin, 2020). İnsan gözetimi, yalnızca bireysel insan gözetimini değil, “**kapsayıcı kamu gözetimi**”ni de ifade etmektedir; insanların karar verme ve hareket etme konusunda yararlı oldukları düşüncesiyle YZ sistemlerine güvenmeyi tercih edebileceği durumlar söz konusu olabilir, yine de sınırlı bağlamlarda kontrolü bırakma kararı insanlara ait olmaya devam eder, bir YZ sistemi nihai insan sorumluluğun ve hesap verebilirliğinin yerini alamaz.

2.4. Toplumsal ve Etik Çıktılar

2019'da özel bir proje olarak başlatılan ve YZ teknolojisiyle ilişkili etik sorunları izleyen kapsamlı bir girişime dönüşen, bağımsız ve herkese açık YZ, Algoritmik ve Otomasyon Olayları ve Araştırmaları (AI, Algorithmic, and Automation Incidents and Controversies: AIAAIC) veri tabanı kayıtları, yeni bildirilen YZ olaylarının ve tartışmalarının sayısının 2021'de 2012'ye göre 26 kat daha fazla olduğunu göstermektedir. Bildirilen olaylardaki bu artış,

hem YZ'nin gerçek dünyayla ne kadar iç içe geçtiğinin hem de etik olarak kötüye kullanılabileceği yollara dair artan bir farkındalığın muhtemel kanıtıdır (Maslej vd., 2023: 133). Raporlar, bazı Amerikan cezaevlerinin mahkumların telefon görüşmelerini taramak için YZ tabanlı sistemler kullandığını göstermektedir ve **sürveyans, mahremiyet** ve ayrımcılıkla ilgili endişelere yol açmaktadır. Örneğin, *Verus ABD Hapishanesi Mahkûm Çağrı İzleme (Konuşma Tanıma, Şubat 2022)* vakasına ilişkin olarak, sestene-metne sistemlerin siyahî bireyler için transkripsiyonda daha az doğru olduğuna ve buna karşılık ABD'de hapsedilen nüfusun büyük bir bölümünün siyahi olduğuna dair kanıtlar bulunmaktadır (Maslej vd., 2023: 135). Şubat 2022 tarihli bir başka vaka raporu, Londra Büyükşehir Polis Teşkilatı'nın "Çete Şiddeti Matrisi" adı verilen binden fazla sokak çetesi üyesinden oluşan bir veri seti tuttuğunu ve her çete üyesinin ortaya koyduğu risk potansiyelini sıralamak için YZ araçlarını kullandığını öne sürmektedir. Bu veri setinin doğru olmadığı ve belirli **etnik ve ırksal azınlıklara karşı ayrımcılık** yapma eğiliminde olduğu sonucuna varan farklı araştırmalar mevcuttur (Maslej vd., 2023: 136). Yine 2022 Mart ayında, sosyal medyada ve bir Ukrayna haber sitesinde dolaşan bir videoda, Ukrayna Devlet Başkanı Volodymyr Zelenskyy'nin orduyu Rusya'ya karşı mücadeleyi teslim etmeye yönlendirdiğini gösterdiği iddia edilmiş ve videonun bir **derin sahte** olduğu ortaya çıkmıştır (Maslej vd., 2023: 134). Bir başka yakın tarihli vaka raporu (Nisan 2022), örüntü tanıma teknolojisi üzerine, Intel'in Zoom'da öğrencilerin duygusal durumunu belirleyecek YZ tabanlı bir teknoloji oluşturmak için "Classroom Technologies" adlı bir eğitim girişimiyle çalışmasına ilişkindir. Bu teknolojinin kullanımı da mahremiyet ve ayrımcılık endişelerini beraberinde getirmiş, **öğrencilerin gereksiz yere izleneceği ve sistemlerin duygularını yanlış tanımlayabileceği korkusuna** yol açmıştır (Maslej vd., 2023: 135). Eylül 2022'de ise, metin açıklamalarından görüntüler üreten şirketlerden yalnızca biri olan Midjourney'e karşı, **telif hakkı** (sistemin kaynaklarını kabul etmeksizin insan yapımı görüntülerden oluşan bir külliyat üzerinde eğitilmiş olması), **istihdam** (Midjourney gibi sistemlerin insan sanatçıların işlerinin yerini alacağı korkusu) ve mahremiyet (Midjourney'nin şirketin kullanma iznine sahip olmayabileceği milyonlarca görüntü üzerinde eğitilmiş olması) dahil olmak üzere çeşitli etik eleştiriler gündeme getirilmiştir (Maslej vd., 2023: 136). Aslında GPT-4 ve Stable Diffusion gibi çok sayıda üretken YZ modeli, genellikle telif hakkıyla korunan metinler ve resimler içeren çok büyük eğitim veri kümelerine dayanmaktadır (Reisner, 2023). Bu, AB'nin YZ Yasası'nda "model oluşturucuların telif hakkı yasası kapsamında korunan eğitim verilerinin kullanımına ilişkin bilgileri yayınlamasını" gerektiren, sayısız davaya ve hükümlere yol açmıştır (European Parliament, 2023).

Büyük YZ şirketlerinin modellerini oluştururken telif hakkı ihlalleri nedeniyle davalarla karşı karşıya kalması gibi, bazı davacılar da bu modelleri eğitmek için gereken terabaytlarca veriyi toplamak için lazım olan “**toplu web kazıma**”nın (mass web scraping) mahremiyete saldırı teşkil ettiğini savunmaktadır (De Vynck, 2023). ChatGPT kullanıcılarının diğer kullanıcıların sohbet geçmişlerine ve hatta ödeme bilgilerine erişmesine olanak tanıyan, yama uygulanmış bir veri güvenlik açığının anlaşılması daha fazla endişeye yol açmıştır; hatta İtalya eğitim verileriyle ilgili söz konusu gizlilik endişeleri sebebiyle bu hizmeti bir süreliğine yasaklamıştır (Burgess, 2023; GPDP, 2023). Sosyal medya ve çevrimiçi reklamcılıkta da benzer konulara odaklanılmaktadır; **reklamlarda hedefleme için kişisel verilerin kullanılmasının** tamamen yasaklanması ve web sitelerinin yalnızca sitenin dar bir işleviyle ilgili verileri toplayabildiği “**veri minimizasyonu**” gerektirecek ABD Federal Hükûmeti’ne bağlı Federal Ticaret Komisyonu eylemi dahil olmak üzere ortak öneriler mevcuttur (Bernstein, 2023).

YZ sistemleri yazılarda, fotoğraflarda, yasal kararlarda ve daha pek çok bağlamda klişeleri ve önyargıları yansıtan veri kümelerinden yararlandıkları için bu algoritmik sistemlerin genellikle kadınlara, etnik ve ırksal azınlıklara ve diğer marjinal gruplara zarar verme potansiyeline sahip önyargılar sergilediğine yönelik endişeler, çok sayıda çalışma ve girişim tarafından, öncelik başlıklar altında da değinildiği üzere, ileri sürülmektedir. Haber makaleleri üzerinde “eleştirel olmayan bir şekilde” düzenli olarak eğitilen makine öğrenimi algoritmaları, mevcut dillere gömülü olan toplumsal önyargıları öğrenecek ve yeniden üreterek şiddetlendirecektir. Doğal dil işlemede popüler bir teknik olan “**kelime temsili**”nin (word embedding), mevcut toplumsal cinsiyet klişelerini şiddetlendirdiği ortaya konulmuştur (Bolukbasi vd., 2016: 4350). Dolayısıyla bu sistemlerin, önyargı, etkinlik ve diğer faktörler açısından değerlendirilmelerinin ve ilgili düzenlemelerin gerekliliği kaçınılmazdır. Örneğin, bir önceki paragrafta da adı geçen Federal Ticaret Komisyonu, bu hususta “ırksal olarak önyargılı algoritmaların satışını veya kullanımını” önlemek için mevcut yetkilerini kullanarak sert önlemler alacağını açıkça ifade etmiştir (Jillson, 2021).

Yukarıda sıralanan endişelerin ötesinde, **ulusal güvenlik veya aşırı risk perspektifinden** çalışmalar ise, insan kontrolünden kaçabilecek daha güçlü haydut YZ'leri (rogue AI) veya YZ sistemlerinin, “2.3.3. Kötü Niyetli Olmama – Zarar Vermeme” başlığı altında da irdelendiği üzere, terörist gruplar veya düşman devletler tarafından silah ya da diğer zararlı ürünler geliştirmek için kötüye kullanılmasının nasıl önlenebileceğini ele almaktadır (Schuett vd., 2023). Yeterince güçlü YZ modelleri bilgisayar korsanlığı konusunda insanüstü

yetenekler kazanarak kötü niyetli kullanıcıların hassas verilere ve hatta askeri teçhizata erişmesini sağlayabilir ya da ölümcül patojenleri tasarlamak ve yerleştirmek için kullanılabilirler (Sandbrink, 2023: 8). Varoluşsal riskin önemli bir unsuru olan güçlü YZ sistemlerinin tehlikeli kontrol kaybı potansiyeli ya da “her yönüyle insanlardan daha zeki” süper-zeki YZ’lerin geleceği hususlarında anlaşmazlıklar olsa da ve bu tür kaygılardan ziyade daha somut sorunların önceliklendirilmesi gereği öne sürülse de insanların sonlandıramayacağı veya güvenli bir hedefi takip etmeye zorlayamayacağı, yüksek yeteneklere sahip gelecekteki bir haydut YZ ihtimali toplumlar ve hatta biyosfer için ciddi tehdit faktörüdür (Bengio, 2023). Burada ilgi çekici bir hikâyeye temas etmek gerekir. Eylül 1983’te, görevi tüm radarları izleyerek bir ABD nükleer füze saldırısı durumunda Sovyetler’i uyarmak olan SSCB Yarbayı Stanislav Petrov’un küresel bir termonükleer savaşı önlediği anlatılagelmiştir. Bu emsal, YZ’nin daha akıllı hale gelmesi ve bir noktada “Petrov’un döngüden çıkarılması” ihtimalinde, füzeler fırlatılmadan önce döngünün bir yerinde sezgileri olan bir “insanın” bulunmayacak olmasını hatırlatması açısından önemlidir (Smith, 2018).

Bu çalışmanın belirli başlıklar altında vurgulandığı üzere, YZ tabanlı gelişmelerine dair tekrar belirtilmesi gereken önemli bir nokta, bu sistemlerin geleneksel olarak geliştirildiği ulusların ihtiyaç ve değerlerine dayanmasıdır. Etik inceleme, şeffaflık ve demokratik kontrolün olmadığı bölgelerde YZ teknolojisi ve büyük veri kullanılırsa, **milliyetçiliği, azınlıklara karşı nefreti** körükleyebilir ve **seçim sonuçlarını** etkileyebilir (Vinueza vd. 2020: 3). Yazılım platformları, sofistike manipülasyon tekniklerini kullanarak “ikna edici bilişim” hedefine ilerlerken, kararları yönlendirmek için psikolojik zayıflıklardan yararlanmak üzere büyük veri ve YZ’nin kullanılmasını temsil eden, propaganda ve davranışsal kontrol yoluyla vatandaşın hedefli olarak etkisizleştirilmesi anlamına gelen “**büyük dürtme**” (big nudging), toplum gibi karmaşık bir sistemde hem sosyal uyuma, demokratik ilkelere ve insan haklarına zarar vermekte hem de suçluların, teröristlerin ve aşırılık yanlılarının eline geçme potansiyelini barındırmaktadır (Helbing vd., 2019: 75). Yeterli şeffaflık ve demokratik kontrol olmadığında sistemin içeriden erozyona uğraması, **bireylerin seçim özgürlüğünü** kısıtlayan bu manipülatif teknolojileri kim kontrol ederse etsin lehine sonuçlar elde etmek için kullanmasına imkân tanıyacaktır. Ülkelerde çoğunlukla yalnızca bir arama motoru ya da sosyal medya platformunun pazar payı üstünlüğü de bu mecralar aracılığı ile kamuoyu gündeminin belirleyici bir biçimde etkilenebiliyor olması açısından bu sorunları derinleştirmektedir (StatCounter, 2024). Her bir birey için yeterince özelleştirilmiş öneriler ile manipülasyonun fark edilmeden kalabilmesini mümkün kılan “**rezonans etkisi**”, yerel eğilimlerin tekrarlanarak giderek güçlenmesine ve

“filtre balonu” ya da “yankı odası etkisi”ne yol açar ve nihayetinde **toplumsal kutuplaşmaya**, birbirini anlamayan ve aralarındaki çatışmanın giderek şiddetlendiği ayrı grupların oluşmasına sebep olur (Helbing vd., 2019: 76). Dahası, psikolojik zayıflıklardan yararlanan ve eleştirel düşüncüyü ortadan kaldırmak amacıyla mahremiyeti de ihlal eden bu manipülatif yöntemler, bireylerin karar verme şeklini değiştirebilmektedir. Dolayısıyla geniş ölçekli kullanımları, dijital dünyadaki davranışların acımasızlaştırılması dahil olmak üzere ciddi sosyal hasara neden olabilir, bilgisel özerklik ve en önemli anayasal haklarımızdan bireysel kendini geliştirme hakkını zedeleyerek "kolektif zekanın" temelini zayıflatma tehlikesini dahi beraberinde getirir (2019: 76-77).

Son olarak, II. Dünya Savaşı sırasında Nazi Almanyası’nın silahlanmadan sorumlu bakanı Albert Speer’in (1970: 614-616) işaret ettiği gibi, “teknolojiyle güçlendirilmiş bir diktatörlük tuzağına düşme riskiyle karşı karşıya kalınabileceği”ni de ifade etmek gerekir; zira dünya ne kadar teknolojik hale gelirse, teknolojiye karşı bir denge unsuru olarak insanın bireysel özgürlük ve öz farkındalık talebi de o kadar elzem olacaktır.

Bir sonraki bölümde, yürütülen karma araştırma metodolojisi detaylandırılmakta ve bulgular sunulmaktadır.

ÜÇÜNCÜ BÖLÜM

ÜNİVERSİTE ÖĞRENCİLERİNİN YAPAY ZEKÂ ETİĞİNE İLİŞKİN TUTUMLARI ÜZERİNE BİR KARMA YÖNTEM ARAŞTIRMASI

YZ teknolojisi daha sofistike ve yaygın hale geldikçe öngörülemeyen etik konular ve sosyal sorunların göz ardı edilme potansiyeli hakkında endişeler de şiddetlenmektedir (Ko ve Leem, 2021: 218). Kapitalizmin artı değerinin genetikten astronomik verilere kadar birçok yaşamsal alana ulaşarak üretilmesiyle sosyal ilişkilerin ve emek ilişkilerinin yeniden formüle edilmesine olanak sağlayan; yeni bir aracılık biçimi, “bir sosyokültürel artefakt” olarak “varoluşların otomasyonunu mümkün kılan” (Stiegler, 2016: 19) bu sistemlerin, “çoğunluk dünyası için orantısız etkilerini” değerlendirebileceğimiz parametreleri ise YZ etiği sağlayacaktır (Ricaurte, 2022: 727).

Öte yandan, Ko ve Leem (2021: 224), artan kolaylık ve verimlilik açısından faydaları hususunda genel bir kabul olsa da YZ sistemlerinin olumsuz yönleri söz konusu olduğunda hala bir farkındalık eksikliği olduğunu da ortaya koymuştur. Bunun nedenini ise, söz konusu teknolojinin uygulanmasının erken aşamalarında olması ve günlük yaşamda karşılaşılabilecek ilgili sistemler ile desteklenen ürün ve hizmetlerin sayısı giderek artsa da kullanıcıların genellikle bu teknolojiye maruz kaldıklarını dahi bilmemeleri olarak açıklamışlardır. Dolayısıyla yalnızca “yukarıdan aşağıya” bir yaklaşım benimseyerek kılavuzlar geliştirmenin ötesinde, bireylerin muhtemel sorunlara ilişkin endişelerini anlama, ifade edilmemiş ihtiyaçlarını ortaya çıkarabilme ve farkındalıklarını artırmalarına dair çabalar anlamlı olacaktır.

Bu noktada öncelikle “mikro, mezo ve makropolitik ilişkilerin tarihsel, politik, kültürel ve duygusal-özel boyutlarıyla iç içe geçtiğinin” göz önünde bulundurularak bu seviyelerin karmaşık eklemlenmesinin dikkate alınması gerekir (Ricaurte, 2022: 727-728). Ancak FAccT ve AIES gibi önde gelen iki multidisipliner YZ etiği konferansının analizi, bu tür konuların yapısal ve sosyal faktörlerden ve belirli potansiyel zararlardan yoksun, soyut bir mahiyette tartışılmasına yönelik büyük bir eğilim olduğunu ortaya koymuştur. Bu bağlamlara çekimser ya da tarafsız yaklaşan veya ancak zayıf bir şekilde değinen bildiriler çoğunluğu oluşturmaya da devam etmektedir (Birhane vd., 2022: 13). Toplumu şekillendirmeye yönelik eleştiriye bağlı potansiyel ise ancak mezo ve mikro düzeylerde bulunabilir (Hagendorff, 2016: 3). Buna karşılık birçok devlet kurumu, sivil toplum örgütü, araştırma merkezi ve özel şirket YZ etiğine dair

alışmalar yrtyor olsa da bireylerin bu sistemlerin sosyal inřasını ve kabuln řekillendirebileceėi mikro analiz dzeyine literatrde ok az yer verildiėi gzlemlenmektedir. Bilhassa ulusal literatrde -son yıllarda YZ farkındalıėı (Ferikoėlu ve Akgn, 2022), YZ kaygısı (Terzi, 2020) ve YZ'ye ynelik genel tutum (Kaya vd., 2024) zerine lekler kazandırılmıř olsa da- bireylerin YZ etiėine ynelik yaklařımlarını irdeleyen alışmalar dikkat ekmemektedir.

Dahası, YZ yayınlarının sayısı artmaya devam ederken ABD ve in'in son on yılda lkeler arası en fazla iř birliėine sahip lkeler olduėu grlmekte (Zhang vd., 2022: 22); bu teknolojiden elde edilen en byk ekonomik kazanımların yine kresel ekonomik etkinin neredeyse %70'ini oluřturan Kuzey Amerika ve in'de yoėunlařacağı ngrlmektedir (PwC, 2017). Daha nce de vurgulandıėı zere, bu sistemlerin etik ve sorumlu bir řekilde geliřtirilmesi ve uygulanması hususunda yayınlanan kılavuzlar da menře lkelerine ve fon saėlayan kuruluřlara gre incelendiėinde, ABD-Batı eksenindeki zel řirketlerin tartıřmalara hâkim olduėu da sylenebilir (Hickok, 2021: 41-43). Ancak YZ'yi evreleyen etik konuların karmařıklıėı uluslararası, blgesel ve ulusal toplulukların eřitli dzeylerinde ve sektrlerinde ok sayıda paydařın iř birliėini ve kresel ve kltrlerarası bir diyalėa dayalı olarak sorumluluk almalarını gerektirir (UNESCO, 2022).

Bu kritik doėrultusunda, 2021-25 UYZS'de yer alan "eřitliliėin ve kapsayıcılıėın saėlanması" bařlıėı altında "YZ sistemlerinden yararlanmaya ve kapsayıcılıėı geniřletmeye ynelik farkındalık seviyesi ykseltilmelidir" ve "Toplumun tm kesimlerinin YZ sistemlerinin kullanımı hakkında bilinli kararlar alabilmesi ve YZ sistemlerinin istenmeyen etkilerinden korunabilmesi iin etkin halk katılımı saėlanmalıdır" ifadelerinden de (CBDDO, 2021: 59) yola ıkarak, YZ sistem yařam dngs boyunca farklı paydařları temsilen gnmz kullanıcısı, geleceėin potansiyel geliřtiricisi ve politika yapıcısı bireyleri olarak farklı akademik disiplinlerden niversite ėrencilerinin, ilgili sistemlerin geliřtirilmesi ve konuřlandırılmasına dair ncelikleri ile muhtemel etik ve toplumsal ıktılar konusunda zel yaklařımlarını ortaya ıkarmak amalanmıřtır. Zira, niversiteler her yıl binlerce mhendis yetiřtirmekte iken, onların mfredatı dahi bu konulara aėırılık vermemektedir, zaten etik sorunlar ve ynetiřim tamamen YZ geliřtiricilerine de bırakılmaz (Reese, 2022). Bununla beraber, bir YZ etiėinden sz ediyorsak ve en olumsuz etkilenen paydařların refahını ve esenliėini gzetmek hedefleniyorsa, etik, bu paydařların deneyimleri ve bakıř aıları da dikkate alınarak ynlendirilmelidir (Birhane vd., 2022: 13). Tm endstrilerin etrafında yn alacaėı bu

sistemlerin ardında yatan değerleri anlamak, muazzam işlem gücü sağlayacak yeni çipler geliştiren ve topladığı veriler ile kâr elde eden şirketlere ve teknokratlara da bırakılmamalıdır. Üretim, dağıtım ve mübadele süreçlerinin yanı sıra kamuoyu ve siyaseti organize etme yollarına girdiğini halihazırda kanıtlamış olan ve doğası gereği disiplinlerarası bir şekilde ele alınması gereken YZ ile toplum arasındaki etkileşimin karmaşıklığıyla yüzleşmek kolektif bir çaba gerektirir. Bu araştırma ile YZ'ye biçmek isteyeceğimiz rol hakkında farkındalığın ana akım tartışmalara girmesinde etkili olmak ve uluslararası literatürde bu hususta Türkiye'de yürütülen araştırmaların temsilini artırmaya katkı sağlamak da umut edilmektedir.

Bu motivasyon ile mevcut araştırma, Jobin vd.'nin (2019)'nin bulgularına dayanarak Kore'de öğrenim gören lisans öğrencilerinin YZ etiğine yönelik tutumlarını ölçmek için AT-EAI ölçeğini geliştiren ve tutumlar arasındaki farklılaşmayı cinsiyet ve YZ eğitimlerine dair deneyimlerine göre de inceleyen Jang vd.'nin (2022) çalışmaları ile temellendirilmiştir. Ayrıca Jang ve ekibi, ulusal ve kültürel faktörlerin YZ etiğinde önemli bir rol oynaması sebebiyle (Awad vd., 2018) sonuçlarının genellenebilirliğinin sınırlı olabileceğine ve YZ etiğine yönelik tutumlarda cinsiyet temelli farklılıklara katkıda bulunan faktörleri belirlemek için çalışmalar yapılabileceğine de işaret etmiştir (Jang vd., 2022: 11654). Söz konusu istikamette ölçeğin orijinaline bağlı kalınarak Türkçeye çevrilmesi ve Türkiye'de öğrenim gören, temel teknoloji bilgisine sahip eğitim düzeyi yüksek örneklemelerde araştırma amacıyla kullanılması hedeflenmiştir. Akademik disipline dayalı tutum farklılıkları da bu aşamada irdelenmiştir. Nicel veri çözümlemesi neticesinde elde edilen bulgular doğrultusunda, tutumları ve farklılıkları besleyen faktörleri açığa çıkarmak maksadıyla yarı yapılandırılmış görüşmeler de gerçekleştirilmiştir.

Bu bölümde araştırma metodolojisi detaylandırılmakta ve bulgular aşamalı olarak sunulmaktadır.

3.1. Araştırma Tasarımı ve Metodoloji

Nitel ve nicel metodolojiler belli “karşıt epistemolojik varsayımları” temel alan farklı araştırma anlayışlarını temsil etmekte iken “temelleri pragmatik ve transformatif paradigmalara dayanan” **karma yöntem deseni** ile alandaki bu “metodolojik hizipleşmenin önüne geçerek” potansiyel araştırma bulgularını ortaya çıkarmak mümkün hale gelebilmektedir (Johnson ve Onwuegbuzie, 2004: 15; Aydın-Çakır ve Türkeş-Kılıç, 2021: 2-4). Bu sebeple, belirtilen motivasyon çerçevesinde, salt nicel ya da salt nitel yaklaşımı kullanmaya kıyasla mevcut

araştırma problemlerine daha bütüncül bir anlayış sağlamak maksadıyla nicel araştırmanın, “araştırmacının kendisinin veri oluşturma aracı olmasına, araştırma kapsamında katılımcılar ile bire bir ilişki kurmasına.. imkan tanıyan” nitel araştırma ile desteklenmesi hedeflenmiştir (Tekindal ve Uğuz Arsu, 2020: 158). Dolayısıyla karma yöntem deseni, bulguların detaylandırılması, artırılması ve açıklığa kavuşturulması maksadıyla “tamamlayıcılık” ve nitel verilerin nicel safhanın gelişimine yardımcı olmak amacıyla kullanılması olan “gelişim” gerekçelerine dayandırılmış (Baki ve Gökçek, 2012: 4), tartışma aşamasında bulguların sentezine gidilmiştir. Araştırma amacı ve sorularına hitap eden bir tipoloji ortaya koyabilmek (Johnson ve Onwuegbuzie, 2004: 20) adına, araştırma safhalarının sırayla yürütüldüğü ve karma yöntem araştırmacılarının en sık kullandığı ana tasarımlardan olan “Sıralı açıklayıcı” tasarımdan faydalanılmıştır. Nicel verilerin toplanıp analiz edilmesini takiben nitel veriler, çalışılan olguyu ve saptanan tutum farklılıklarını daha detaylı irdeleyebilmek, daha derin içgörülere ulaşabilmek için toplanmıştır (Creswell, 2003: 213 Baki ve Gökçek, 2012: 9-10).

Yürütülecek karma araştırma stratejisinin Etik Kurul izni, T.C. Akdeniz Üniversitesi Rektörlüğü Sosyal ve Beşerî Bilimler Bilimsel Araştırma ve Yayın Etiği Kurulu 13.04.2023 tarihli ve 207 sayılı kararı ile sağlanmıştır.

Nicel ve nitel araştırma ayrıntılarına ve bulgularına ayrı ana başlıklar altında yer verilmekte, bulgular tartışma safhasında bir arada yorumlanmaktadır.

3.2. Nicel Araştırma

3.2.1. Veri Toplama Aracı ve Türkçeye Çevrilmesi

Bireylerin YZ etiğine yönelik tutumlarını veya tercihlerini analiz eden, ancak çoğunlukla başka amaçlar için geliştirilen ve daha sonra YZ etiğine uyarlanan araçları kullanan önceki çalışmaların aksine Jang vd. (2022: 11651) çalışmalarında, orijinal olarak lisans öğrencilerinin "YZ etiğine yönelik tutumlarını" ölçmek için beş boyutlu, 17 maddeden oluşan “**Yapay Zeka Etiğine Yönelik Tutum Ölçeği**”ni (AT-EAI) geliştirmiş ve bu aracı çeşitli yöntemler kullanarak kapsamlı bir şekilde bir doğrulama sürecine tabi tutmuşlardır. Açımlayıcı Faktör Analizi (AFA) ve Doğrulamalı Faktör Analizi (DFA) ile aracın geçerliliği onaylanmış, aracın güvenilirliğine ilişkin olarak ise her boyutunda (Cronbach's alpha = 0.753–0.865) ve genel olarak (Cronbach's alpha = 0.843) tatmin edici psikometrik özelliklere ulaşılmıştır. Dolayısıyla, aracın, üniversite öğrencilerinin YZ etiğine yönelik tutumlarını ölçmek için kullanılabilecek geçerli ve güvenilir bir araç olduğu söylenebilir.

Bu ölçeğin Türkçeye çevrilerek kullanılmasının tercih edilmesinin sebepleri ise;

- YZ geçmişi olmayan öğrencilerin tutumlarını doğru bir şekilde ölçebilmek amacıyla YZ hakkında önceden bilgisi olmayan öğrenciler tarafından kullanılmak üzere, gerçek dünya deneyimleriyle ilgili olması hedeflenerek tasarlanmış ve
- Shin (2021)'in sadece şeffaflık, adalet ve sorumluluğu ölçmek için mevcut bir aracı değiştirerek YZ etiği bağlamında doğruladığı çalışmanın aksine, önceki araştırmaların ve yayınlanan YZ etik yönergelerinin gözden geçirilerek ve Jobin vd.'nin (2019) YZ için etik ilkeler ve kılavuzlar içeren halka açık 84 belgeyi incelediği araştırmalarının bulgularına dayanılarak literatürde en öne çıkan ilkeler olarak ileri sürdüğü “**adalet, şeffaflık, gizlilik (mahremiyet), sorumluluk ve kötü niyetli olmama (zarar vermeme)**” ilkelerinin tümünü kapsayacak şekilde, “kötü niyetli olmama” ve “mahremiyet” boyutlarının da dahil edilmiş olmasıdır (Jang vd., 2022: 11652).

Adalet boyutu, YZ geliştirme için veri toplarken çeşitliliğin dikkate alınması ve geliştirilen YZ'nin ayrımcılık yapılmaksızın herkese açıklanması gibi maddeleri kapsamaktadır. **Şeffaflık** boyutu, YZ'nin açıklanabilirliğinin gerekliliğine yönelik tutumları ölçen maddelerden oluşmuştur. **Kötü niyetli olmama** boyutu, YZ ile ilişkili çeşitli kurumların kötüye kullanımını önlemenin önemine ilişkin yanıtları ortaya çıkaran maddeleri kapsamaktadır. **Gizlilik** boyutu, YZ oluşturmak için veri toplarken ve kullanırken gizliliğin korunmasına yönelik tutumlarla ilgili soruları içermektedir. Son olarak **sorumluluk** boyutu ise, YZ ile ilgili sorunlar ortaya çıktığında sorumlulukların toplumsal uzlaşmaya göre mi paylaşılması gerektiği yoksa geliştiriciler ve kullanıcılar gibi belirli grupların mı tüm sorumluluğu üstlenmesi gerektiği konusundaki tutumları ortaya çıkaran maddeleri içermektedir.

Maddeler, (1) “Kesinlikle katılmıyorum” ile (5) “Kesinlikle katılıyorum” arasında değişen 5’li Likert ölçeğinde sunulmuştur. Ek olarak, anket sürecinde YZ konusunda yetersiz bilgiye sahip öğrencilerin soruları yanıtlamakta zorlanacağı olasılığı da göz önünde bulundurularak her boyutun kısa bir açıklaması eklenmiştir. Boyut seçimi ve her boyut için açıklayıcı materyallerin oluşturulması da iki YZ uzmanının yardımıyla gerçekleştirilmiştir (Jang vd., 2022: 11651).

Yapay Zekâ Etiğine Yönelik Tutumlar Ölçeği’nin (AT-EAI) geliştiricilerinden gerekli izin sağlanarak (ilgili izin dokümanı EK 1’de mevcuttur) ve İngilizce orijinale bağlı kalınarak, öncelikle, birbirinden bağımsız iki uzman tarafından madde çevirileri yapılmıştır. Çeviri, üçüncü bir uzman tarafından da incelendikten sonra maddelerin dilsel ve anlamsal bakımdan büyük ölçüde örtüştüğü görülerek aracın çevirisinin son hali elde edilmiştir. İngilizce ifadelerin

Türkçe uygunluğunu araştırmak için Fen-Edebiyat Fakültesi Batı Dilleri ve Edebiyatları Bölümü mezunu iki dil uzmanına EK 2’de mevcut “Çevirileri Değerlendirme Formu” doldurtulmuştur. Bu aşamada ölçek maddelerinin çevirilerinin değerlendirilmesi için Seçer (2015: 70) tarafından geliştirilen formdan faydalanılmıştır. “Orijinal metin”, “Değerlendirme”, “Türkçe çeviri” ve “Yorum” kolonlarından oluşan formda “Değerlendirme” kısmında yer alan seçenekler; “1- Hiç uygun değil (Maddenin uygun şekle getirilmesi gerekir)”, “2- Biraz uygun (Uygun, ancak küçük değişiklik gerekir)”, “3- Uygun”, “4- Çok uygun” ve “5- Tamamen uygun” şeklindedir. Zıt ifadelerden oluşan iki madde dışında iki dil uzmanı da tüm ifadeleri “Tamamen uygun” olarak puanlandırılmıştır; söz konusu iki madde ise “Çok uygun” olarak değerlendirilmiştir. Ölçek bu doğrultuda, nihai formuna ulaştırılmıştır. EK 3’te Türkçeye çevrilmiş Yapay Zekâ Etiğine Yönelik Tutumlar Ölçeği’nin “Anket Formu” yer almaktadır.

3.2.2. Veri Toplama Süreci

Karma yöntemlerde kullanılan en yaygın örnekleme kombinasyonu benimsenmiş (Onwuegbuzie ve Collins, 2007: 284) ve birincil nicel aşama, Jang vd.’nin (2022: 11644) çalışmalarında tercih edildiği gibi gönüllülük esasına dayalı olarak gerçekleştirilmiştir (Tiit, 2021: 9). Google Forms tabanlı çevrimiçi bir anket bağlantısı potansiyel katılımcılara, hem ülke genelinde devlet ve vakıf üniversitelerinin dekanlıklarına bildiri olarak hem de sosyal medya aracılığıyla ulaştırılmıştır. Katılımcılar ankete eriştiklerinde ilk olarak çalışmanın amacı, katılımlarının gönüllülük esasına dayandığı ve toplanan tüm verilerin anonim ve gizli olacağı konusunda bilgilendirilmişlerdir. Veri toplama süreci, 17.05.2023 – 21.08.2023 tarihleri arasında gerçekleşmiş, Türkiye’de öğrenim gören 395 üniversite öğrencisinin katılımı sağlanmıştır.

3.2.3. Geçerlilik ve Güvenilirlik

İç tutarlılığı değerlendirmek üzere 17 ifade orijinal ölçekte kullanıldığı haliyle güvenilirlik analizine tâbi tutulduğunda, tüm ölçek için Cronbach’s alpha değeri $\alpha = 0.628$ olmakla beraber, bazı alt boyutlar için alpha değerlerinin güvenilir değerlerin altında olduğu gözlemlenmiştir. Zıt ifadelerden oluşan ve ters kodlama gerektiren söz konusu beş ifade (ADL1, ŞEF2 ve 3, KNO2 ve GİZ3) ilgili boyutlardan çıkarıldığında ise alt boyutlar için Cronbach’s alpha güvenilirlik katsayılarının 0.407 ile 0.782 arasında değerler aldığı belirlenmiştir. Farklılık analizlerinde alt boyutlardaki maddelerin ortalamasının alınması gereği göz önünde bulundurularak söz konusu ifadeler bu analizlerde kapsam dışı bırakılmıştır. Aynı zamanda, bu ifadelerden ŞEF2 ve KNO2, “Çevirileri Değerlendirme Formu”nda dil

uzmanlarından birinin “4” ile puanlandıkları maddelerdir. Bu zıt ifadelerin katılımcıların sadece olumlu yanıt verme eğilimlerinin önüne geçmek amacıyla oluşturulduğu ve kapsam dışı bırakıldıklarında boyut içeriğini etkilemeyecekleri göz önünde bulundurularak, ilgili ifadeler farklılık analizlerinin güvenilirliğini artırmak maksadıyla analiz sürecine dahil edilmemişlerdir. 12 maddeye indirgenen ölçeğin güvenilirlik düzeyi $\alpha = 0.701$ olarak elde edilmiştir. Cronbach's alpha'yı yorumlamak için belirli kriterler üzerinde net bir anlaşma bulunmamakla beraber kat sayının yaygın bir yorumu, “düşük güvenilirlik” için $\alpha < 0.50$, “orta (kabul edilebilir) güvenilirlik” için $0.50 < \alpha < 0.80$ ve “yüksek (iyi) güvenilirlik” için $\alpha > 0.80$ şeklindedir (Ekolu ve Quainoo, 2019: 25; Ursachi vd., 2015: 681). Kalaycı'ya göre de $0.40 \leq \alpha < 0.60$ aralığındaki değerler düşük düzeyde olmakla beraber “güvenilir” kabul edilmektedir (2008: 405). Dolayısıyla ölçeğin genel iç tutarlılık katsayısının 0.701 değeri ile $0.60 \leq \alpha < 0.80$ aralığında yer alarak “oldukça güvenilir” olduğu, faktörlerin de sırasıyla “Adalet” boyutu için $\alpha = 0.407$, “Şeffaflık” boyutu için $\alpha = 0.545$, “Kötü niyetli olmama” boyutu için $\alpha = 0.782$, “Gizlilik” boyutu için $\alpha = 0.609$ ve “Sorumluluk” boyutu için $\alpha = 0.413$ değerleri ile güvenilirlik sağladığı söylenebilmektedir (Kalaycı, 2008: 405).

Kültürlerarası ölçek uyarlama çalışmalarında da önerildiği üzere aracın faktör deseni için doğrudan doğrulayıcı faktör analizi (DFA) ile başlanmıştır (Çokluk vd., 2014: 283). Mevcut modelin yapısal geçerliliğini ortaya koymak maksadıyla kullanılan DFA, uygunluk istatistiklerinin örneklem büyüklüğüne, serbestlik derecesine veya model karmaşıklığına duyarlılık farklılıkları göz önünde bulundurularak çeşitli “mutlak uygunluk ölçütleri” ve “aşamalı uygunluk ölçütleri” bir arada incelenmiştir (Byrne, 2011: 664).

IBM SPSS Amos 26 kullanılarak gerçekleştirilen DFA sonucunda uyum değerleri, modelin istatistiksel olarak uygun ve anlamlı olduğu ortaya koymuştur. Ki-kare testi (X^2/sd) oranı 2,577 ile iyi uyuma, Ortalama Hata Karekök Değeri (RMSEA) 0,063 ile kabul edilebilir uyuma, Kök Artık Kareler Ortalaması (RMR) ,109 ile kabul edilebilir uyuma, Jöreskog ve Sorbom Uyum İyiliği İndeksi (GFI) 0,956 ile iyi uyuma, Uyarlanmış Uyum İyiliği İndeksi (AGFI) 0,923 ile iyi uyuma, Karşılaştırmalı Uyum İndeksi (CFI) 0,940 ile kabul edilebilir uyuma, Normlaştırılmış Uyum İndeksi (NFI) ise 0,907 ile kabul edilebilir uyuma işaret etmektedir. Dolayısıyla faktörlerin teorik yapıyı açıklamada yeterli olduğu, gerçekleştirilen uyum istatistikleri ile desteklenmektedir (Özdamar, 2013: 236). Uyum indeksleri ve ölçütler Tablo 3.1’de yer almaktadır.

Tablo 3.1. DFA Sonucunda Ulaşılan Uygunluk İstatistikleri Değerleri ve Referans Aralıkları

Uyum İstatistiği	Araştırma Modeli Değeri	İyi Uyum	Kabul Edilebilir Uyum
χ^2/sd	2,577	$2 < \chi^2/sd \leq 3$	$3 < \chi^2/sd \leq 5$
RMSEA	,063	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$
RMR	,109	$0 \leq RMR \leq .05$	$.00 < RMR < .10$
GFI	,956	$.95 \leq GFI \leq 1$	$.90 \leq GFI < .95$
AGFI	,923	$.90 \leq AGFI \leq 1$	$.90 \leq AGFI < .95$
CFI	,940	$.95 \leq CFI \leq 1$	$.90 \leq CFI < .95$
NFI	,907	$.95 \leq NFI \leq 1$	$.90 \leq NFI < .95$

Kaynak: (Doll vd., 1994: 456; Baumgartner ve Homburg, 1996: 153; Hu ve Bentler, 199: 6; Karadağ, 2013: 1416; Sümer, 2000: 59; Schermelleh-Engel vd., 2003: 52; Hooper vd., 2008: 54; Çokluk vd., 2012: 307, 400).

3.2.4. Verilerin Analizi

Analizler IBM SPSS Statistics 23 aracılığıyla gerçekleştirilmiştir. İstatistiksel olarak anlamlı sonuçları değerlendirmek için %95 güven düzeyi kullanılmıştır. Çevrimiçi anketi tamamlayan toplam 395 katılımcının verileri standart veri kalitesi kontrollerinden geçirilmiştir. 217 erkek (%54.9) ve 176 kadın (%44.6) katılımcı tanımlanmış, cinsiyetini belirtmek istemeyen ve cinsiyetini ifade etmeyen birer katılımcının cinsiyeti kaydedilmemiştir (%.6). Katılımcıların çoğunluğu 267 (%67.6) lisans öğrencisidir; 52 (%13.2) yüksek lisans, 38'er doktora (%9.6) ve ön lisans (%9.6) öğrencisi bulunmaktadır. Araştırma amacı doğrultusunda 188 (%47.6) üniversite öğrencisi İnsan ve Toplum Bilimleri, 207 (%51.4) öğrenci ise Fen Bilimleri Teknoloji ve Mühendislik grubu altında değerlendirilmiştir. En fazla katılımın gözlemlendiği altı bölüm sırasıyla; Yönetim Bilişim Sistemleri (45), Bilgisayar Mühendisliği (28), Hukuk (22), Sigortacılık (ve Aktüerya) (15), İşletme (11) ve Mimarlık (10) bölümleridir. Katılımcılara ilişkin betimsel bilgiler Tablo 3.2'de sunulmuştur.

Tablo 3.2. Katılımcıların Betimsel Bilgileri

Değişkenler	Kategoriler	N	%
Cinsiyet	Erkek	217	54.9
	Kadın	176	44.6
	Belirtmek istemeyen	1	0.3
	Bilinmeyen	1	0.3
	Toplam	395	100
Eğitim düzeyi	Ön lisans	38	9.6
	Lisans	267	67.6
	Yüksek lisans	52	13.2
	Doktora	38	9.6
	Toplam	395	100

Akademik disiplin	Fen Bilimleri, Teknoloji ve Mühendislik	207	51.4
	İnsan ve Toplum Bilimleri	188	47.6
	Toplam	395	100

3.2.5. Bulgular

Farklılık analizleri neticesinde elde edilen bulgulara, cinsiyet ve akademik disiplin değişkenlerine yönelik olmak üzere ayrı birer başlık altında yer verilmektedir.

3.2.5.1. Cinsiyet Değişkenine Yönelik Farklılık Analizleri

Cinsiyetini belirtmek istemeyen ve cinsiyetini ifade etmeyen birer katılımcı, sınırlı temsilleri nedeniyle cinsiyet temelli farklılık analizleri kapsamı dışında tutulmuş, analizler 393 katılımcının verileriyle gerçekleştirilmiştir. 393 kişiden elde edilen veriler ile YZ etliğinin her bir boyutu için cinsiyete ilişkin anlamlı bir farklılık olup olmadığı incelenmiştir. Bu aşamada ilgili boyuta ait maddelerin ortalama değeri kullanılmıştır. Alt boyutlar için Kolmogorov-Smirnov ve Shapiro-Wilk testlerinin kadın ve erkek gruplarına ilişkin önemlilik değerleri $p > 0.05$ ön şartını sağlamamaktadır; dolayısıyla bu testler ile normal dağılım varsayımı ileri sürülemez. Diğer taraftan, Jang vd.’nin (2022) cinsiyete ilişkin ön analiz sonuçları da aykırı değer bulunmamakla beraber verilerin Kolmogorov-Smirnov ve Shapiro-Wilk testlerine göre normal dağılmadığını göstermiştir. Ancak çarpıklığın mutlak değeri üçü ve basıklığın mutlak değeri sekizi aşmıyorsa dağılımın normal olarak kabul edilebileceği argümanına dayanarak normal dağılım varsayımını kabul etmişlerdir (Kline, 2005’ten akt. Jang vd., 2022: 11648). Hatta Kline (2011: 63), basıklık mutlak değerinin 10’u aşmasının problem arz ettiğini öne sürer. Bu aşamada, Klein’in argümanlarını esas alarak “Kötü niyetli olmama” boyutunda kadınlarda göze çarpan “10.675” basıklık değeri haricinde tüm boyutlar için hem kadın hem erkek gruplarında asimetri ve basıklık değerlerinin kabul edilebilir olduğu söylenebilmektedir. Bu sebeple bu boyut dışında tüm boyutlar için kadın ve erkeklerde normal dağılım varsayımı geçerli kabul edilmiş ve “Bağımsız Örneklem T-Testi” gerçekleştirilmiş, “Kötü niyetli olmama” boyutu için ise parametrik olmayan “Mann-Whitney U” testi analiz sonuçları dikkate alınmıştır.

Tablo 3.3 ve Tablo 3.4’te gösterildiği gibi, tüm boyut ortalamalarında kadınların ortalaması erkeklerin ortalamasından yüksektir. Jang vd.’nin (2022: 11648) çalışmasında ise kadınların “Şeffaflık” boyutunda erkeklerden ortalama olarak daha yüksek puan almadığı dikkat çekmektedir.

Öncelikle, “Adalet”, “Şeffaflık”, “Gizlilik” ve “Sorumluluk” boyutlarında kadın ve erkek gruplarında varyansların homojen dağılım gösterip göstermediğini görmek için “Bağımsız Örneklem T-Testi” gerçekleştirilmiştir. Levene’s test sonuçları, “Adalet”, “Şeffaflık” ve “Gizlilik” boyutları için homojenlik varsayımının geçerliliğini desteklememiştir (Adalet: $F = 18.485, p < 0.05$; Şeffaflık: $F = 12.157, p < 0.05$; Gizlilik: $F = 19.581, p < 0.05$). Sadece “Sorumluluk” boyutu için varyansların homojen dağılımı ön şartı gerçekleşmiştir ($F = 2.101, p > 0.05$). Bununla beraber, “Adalet” ($t_{389.775} = 4.015, p < 0.05$) ve “Gizlilik” boyutlarında ($t_{385.241} = 3.084, p < 0.05$) cinsiyet değişkenine göre anlamlı farklılıklar saptanmıştır. “Şeffaflık” ($t_{386.918} = 1.279, p > 0.05$) ve “Sorumluluk” ($t_{391} = 1.249, p > 0.05$) boyutlarında ise gruplar arasında istatistiksel olarak anlamlı bir fark tespit edilmemiştir. Bu bulgu, Jang vd.’nin (2022) bu boyutlara ilişkin cinsiyet değişkenine göre farklılık bulgularını destekler niteliktedir.

Tablo 3.3. Cinsiyet Değişkenine İlişkin T-Testi Sonuçları

Boyut	Cinsiyet	N	$\bar{x}$	S	sd	t	p
Adalet	Erkek	217	3.74	1.03089	389.775	4.015	0.000*
	Kadın	176	4.11	0.78954			
Şeffaflık	Erkek	217	3.96	0.96515	386.918	1.279	0.202
	Kadın	176	4.07	0.70453			
Gizlilik	Erkek	217	4.17	0.92328	385.241	3.084	0.002*
	Kadın	176	4.56	0.66036			
Sorumluluk	Erkek	217	3.05	0.79630	391	1.249	0.213
	Kadın	176	3.14	0.69804			

Normallik varsayımının sağlanamadığı “Kötü niyetli olmama” boyutu için ise Mann-Whitney U testi sonuçları incelendiğinde, Tablo 4’te görüleceği üzere, gruplar arasında kadın öğrenciler lehine istatistiksel açıdan anlamlı bir farklılık saptanmıştır ($U = 15973.500, p < 0.05$).

Tablo 3.4. Cinsiyet Değişkenine İlişkin U-Testi Sonuçları

Boyut	Cinsiyet	N	Sıra Ortalaması	Sıra Toplamı	U	p
Kötü niyetli olmama	Erkek	217	214.74	37794.50	15973.500	0.004*
	Kadın	176	182.61	39626.50		

Ek olarak, “Sorumluluk” boyutu haricinde tüm boyutlarda hem erkek katılımcılar hem de kadın katılımcılar “yüksek” düzeyde tutum belirtmişlerdir; “Sorumluluk” boyutunda ise hem erkek katılımcıların ($\bar{x} = 3.05$) hem de kadın katılımcıların ($\bar{x} = 3.14$) tutumu “orta” düzeydedir (Bkz. Tablo 3.3).

Jang vd.’nin (2022) bulgularıyla tutarlı olarak “Şeffaflık” ve “Sorumluluk” boyutlarında cinsiyete ilişkin anlamlı bir farklılık tespit edilememekle beraber, “Adalet”, “Kötü niyetli olmama” ve “Gizlilik” boyutlarında gruplar arasında kadınlar lehine istatistiksel açıdan anlamlı bir farklılık gözlemlenmiştir.

3.2.5.2. Akademik Disiplin Değişkenine Yönelik Farklılık Analizleri

395 katılımcıdan elde edilen veriler doğrultusunda, “İnsan ve Toplum Bilimleri” ve “Fen Bilimleri, Teknoloji ve Mühendislik” akademik disiplinleri altında bir araya getirilen bölümler dahilinde, akademik disiplin değişkeninin söz konusu boyutlara yönelik anlamlı farklılığa işaret edip etmediğini belirlemek amacıyla “Bağımsız Örneklem T-Testi” gerçekleştirilmiştir. Kolmogorov-Smirnov ve Shapiro-Wilk testlerine göre yine normallik varsayımı ileri sürülememekle beraber, çarpıklığın mutlak değeri üçü ve basıklığın mutlak değeri sekizi aşmıyorsa dağılımın normal olarak kabul edilebileceği argümanına dayanarak tüm boyutlar için normallik varsayımı geçerli kabul edilmiştir (Kline, 2005’ten akt. Jang vd., 2022: 11648). Levene’s test sonucuna göre tüm boyutlar için varyansların homojenliği varsayımı da sağlanmıştır (Adalet: $F = .006, p > 0.05$; Şeffaflık: $F = .122, p > 0.05$; Kötü niyetli olmama: $F = 1.710, p > 0.05$; Gizlilik: $F = 2.191, p > 0.05$; Sorumluluk: $F = 1.147, p > 0.05$).

Tablo 3.5’te yer verilen sonuçlara göre sadece “Şeffaflık” boyutunda akademik disiplin değişkenine göre istatistiksel olarak anlamlı farklılık tespit edilmiştir ($t_{(393)} = 2.009, p < 0.05$). Bu boyutta, İnsan ve Toplum Bilimleri (İTB) öğrencilerinin ortalamaları ($\bar{x} = 4.09$), Fen Bilimleri, Teknoloji ve Mühendislik (FTM) grubu öğrencilerinin ortalamalarından ($\bar{x} = 3.92$) yüksek bir farkla olmamakla birlikte anlamlı bir şekilde farklılaşmaktadır. Diğer boyutlar için ise anlamlı bir farklılıktan söz edilememektedir (Adalet: $t_{(393)} = 0.846, p > 0.05$; Kötü niyetli olmama: $t_{(393)} = 1.329, p > 0.05$, Gizlilik: $t_{(393)} = 0.939, p > 0.05$, Sorumluluk: $t_{(393)} = -0.149, p > 0.05$).

Aynı tabloda görüleceği üzere, “Sorumluluk” boyutu haricinde tüm boyutlarda hem FTM katılımcıları hem de İTB katılımcıları “yüksek” düzeyde tutum belirtmişlerdir.

“Sorumluluk” boyutunda ise FTM ($\bar{x} = 3.09$) ve İTB ($\bar{x} = 3.08$) katılımcılarının “orta” düzeyde tutum sergiledikleri dikkat çekmektedir.

Tablo 3.5. Akademik Disiplin Değişkenine İlişkin T-Testi Sonuçları

Boyut	Akademik Disiplin	N	$\bar{x}$	S	sd	t	p
Adalet	FTM	207	3.86	0.95896	393	0.846	0.398
	İTB	188	3.94	0.95446			
Şeffaflık	FTM	207	3.92	0.89479	393	2.009	0.045*
	İTB	188	4.09	0.83467			
Kötü niyetli olmama	FTM	207	4.25	0.91240	393	1.329	0.185
	İTB	188	4.36	0.78453			
Gizlilik	FTM	207	4.38	0.88602	393	0.939	0.348
	İTB	188	4.46	0.76780			
Sorumluluk	FTM	207	3.09	0.70864	393	-0.149	0.882
	İTB	188	3.08	0.80780			

*FTM: Fen Bilimleri, Teknoloji ve Mühendislik, İTB: İnsan ve Toplum Bilimleri

Nicel veri analizleri neticesinde elde edilen bulguları özetlemek gerekirse;

- ⇒ “Adalet”, “Kötü niyetli olmama” ve “Gizlilik” boyutlarına yönelik olarak, kadın katılımcıların ortalaması erkek katılımcıların ortalamasından yüksektir ve ortalamalar arası bu fark istatistiksel olarak anlamlıdır.
- ⇒ Akademik disiplin değişkeni temelli anlamlı farklılık sadece “Şeffaflık” boyutunda mevcuttur ve İnsan ve Toplum Bilimleri bölümlerinden katılımcılar, Fen Bilimleri, Teknoloji ve Mühendislik bölümlerinden katılımcılara göre bu hususta daha yüksek ortalamalara sahiptir.
- ⇒ Sorumluluk boyutunda ise ne cinsiyet ne de akademik disiplin değişkeni temelli anlamlı bir farklılığa rastlanmıştır.

3.3. Nitel Araştırma

3.3.1. Araştırma Tasarımı

“Bütüncül yaklaşım” niteliğiyle araştırma değişkenlerinin alâkadar diğer değişkenler ile etkileşimini göz önünde bulunduran kalitatif yöntemler çatısı altında, katılımcıların bir fenomene ya da bir kavrama dair perspektiflerini ve hislerini ortaya çıkarabilmek için derinlemesine keşfe odaklanırken felsefi tartışmalara da olanak sağlayan açık uçlu bir araştırma metodolojisi olarak **fenomenolojik yaklaşım** benimsenmiştir (Rose vd., 1995: 1124; Tekindal

ve Uğuz Arsu, 2020: 158-160). Bu aşamada, araştırma soruları dahilinde cinsiyet ve akademik disiplin odaklı gruplandırılmış katılımcıların, araştırmanın temellendirildiği “Adalet”, “Şeffaflık”, “Kötü niyetli olmama”, “Gizlilik” ve “Sorumluluk” YZ etiği ilkelerine hassasiyetlerinin yapısını ve özünü yakalamaya (Patton, 2018: 133; Tekindal ve Uğuz Arsu, 2020: 163) ve bu ilkeleri YZ fenomeni bağlamında nasıl deneyimlediklerine ve anlamlandırdıklarına yoğunlaşmış, veri toplama aracı olarak ise bu araştırmalarda öne çıkan “**yarı yapılandırılmış görüşme**” tercih edilmiştir.

Çalışmanın nicel safhasında elde edilen bulgular doğrultusunda nitel araştırma süreci, aşağıdaki temel araştırma soruları üzerine inşa edilmiştir:

- 1- Kadın üniversite öğrencilerinin YZ etiği bağlamında adalet, kötü niyetli olmama ve gizlilik ilkelerine yönelik tutumları erkek öğrencilere göre **nasıl** farklılaşmaktadır?
- 2- Kadın öğrencilerin adalet, kötü niyetli olmama ve gizlilik YZ etiği ilkelerine yönelik tutumlarında erkek öğrencilere göre daha duyarlı olmalarının altında yatan **temel faktörler** nelerdir?
- 3- “İnsan ve Toplum Bilimleri” ile “Fen Bilimleri, Teknoloji ve Mühendislik” öğrencileri arasında YZ etiği ilkelerinden şeffaflığa ilişkin, **akademik disiplinlerine özgü öne çıkan bakış açıları ve öncelikler** nelerdir?

Temel araştırma sorularına dayanarak oluşturulan temel mülakat ve sondaj sorularının yer aldığı “Yarı Yapılandırılmış Mülakat Formu” EK 4’te sunulmuştur. Her bir görüşmeye, katılımcıların genel olarak YZ’yi nasıl tanımladıkları üzerine açık uçlu bir soru ile başlanmış; böylelikle zihinlerinde YZ sistemlerinin ya da YZ destekli uygulamaların canlandırılmasına zaman tanınmış olması amaçlanmıştır. Ayrıca bu örneklerin, geleneksel ve sosyal medya içeriklerine yönelik deneyimlerine yönelik sondaj soruları aracılığıyla çeşitlendirilmesi ve ilerleyen mülakat soruları için de bir zemin oluşturulması planlanmıştır. Öncesinde takip sorularının bir ana listesi hazırlanmış (Leech, 2002: 668), temel görüşme ve sondaj sorularının sayısı sınırlı tutularak (Weiss, 1994: 65) doğal etkileşimi sekteye uğratmamaya özen gösteren esnek bir görüşme akışı sağlamaya çalışılmıştır. Bu aşamada, betimlemeyi destekleyecek detayları ortaya çıkarmaya yarayacak soruların türetilmesi hedeflenmiştir (Polat, 2022: 169). Aynı zamanda tek katmanlı, “kolay anlaşılır” ve “tanındık” (Patton, 2018: 358-361); akademik özgülüğünden arındırılmış bir dil (Merriam ve Tisdell, 2015: 117; Rubin ve Rubin, 2012: 132; Polat, 2022: 170); sorular arası geçişte ise katılımcıların “pratik, yorumsal, deneyimsel anlatıları” kadar bunların “çağrışım ve keşifleri”ne yer açabilecek ve sessizliklerinin analizine

imkan tanıyacak boşluklar oluşturma esaslarına dayanılmıştır (Polat, 2022: 168-170). Görüşme soruları, araştırma sorularının soyut, tümdengelsel boyutu yerine katılımcıların algı ve deneyimlerine odaklı spesifik sorular olarak şekillendirilmeye çalışılmıştır (Josselson, 2013: 136). Empatik akışı sürdürebilmek adına temel mülakat sorularında yönlendirici tepki ve ifadelerden kaçınılarak (Weiss, 1994: 74), görüşmecinin “kendi varsayımlarından arındırılmış, girift yanıtlar almaya uygun” açık uçlu sorular (Roberts, 2020: 3190) ve beklenmedik yanıtlara açık bir üslup benimsenmiş (Patton, 2018: 357), görüşmeci “katılımcıya kendi sözcüklerini geri yansıtmaya” tekniğinden de faydalanmıştır (Edwards ve Holland, 2013: 73). Görüşme akışındaki bazı sondaj soruları ise “takip, teyit, açıklık getirmek için” daha kapalı uçlu olarak sorulmuş, “büyük resmi ortaya koyan, kapsamlı sorular ile genelden özele inen, küçük ölçekli soruların dengesi” gözetilmeye gayret edilmiştir (Roberts, 2020: 3194-3196).

3.3.2. Araştırma Alanı ve Çalışma Grubu

Küçük bir gönüllü katılımcı havuzundan faydalanılmış, **çalışma amacına uygun** diğer katılımcılara erişebilmek maksadıyla **kartopu stratejisi** kullanılmıştır (Tekindal ve Uğuz Arsu, 2020: 167). Mülakatlar, derinlemesine veri elde etmek maksadıyla “YZ” fenomenine ilgi duyan, Akdeniz Üniversitesi bünyesinde öğrenimine devam eden 16 katılımcı ile yürütülmüştür. 9 erkek ve 7 kadından oluşan katılımcılar ağırlıklı lisans öğrencisidir. “İnsan ve Toplum Bilimleri” ve “Fen Bilimleri, Teknoloji ve Mühendislik” grupları arasında nicel veri analizleri neticesinde tespit edilen akademik disiplin temelli tutum farklılıklarına ilişkin daha ayrıntılı bulgular ortaya koyabilmek amacıyla, nicel verilerin toplanması sürecinde ankete en fazla katılım gösteren bölümlerin temsil edilmesini sağlamak önceliklendirilmiş, örneklem sayısının değişik perspektifleri kapsayacak düzeyde tutulmasına özen gösterilmiştir (Arksey ve Knight, 1999: 58). 9 katılımcı Fen Bilimleri, Teknoloji ve Mühendislik (FTM), 7 katılımcı ise İnsan ve Toplum Bilimleri (İTB) altında gruplandırılarak mülakata tabi tutulmuştur. Katılımcılara ait demografik bilgiler Tablo 3.6’da detaylandırılmaktadır.

Tablo 3.6. Katılımcılara Ait Demografik Bilgiler

Katılımcı Kodu	Bölüm	Eğitim Düzeyi
E1-FTM	Yönetim Bilişim Sistemleri	Lisans
E2-FTM	Yönetim Bilişim Sistemleri	Lisans
K1-FTM	Yönetim Bilişim Sistemleri	Lisans
K2-İTB	Hukuk	Lisans
E3-İTB	Hukuk	Lisans
E4-İTB	Radio Televizyon ve Sinema	Lisans
K3-FTM	Tıp	Lisans

E5-FTM	Bilgisayar Mühendisliği	Doktora
E6-FTM	Bilgisayar Mühendisliği	Lisans
E7-FTM	Bilgisayar Mühendisliği	Lisans
K4-FTM	Bilgisayar Mühendisliği	Lisans
K5-FTM	Bilgisayar Mühendisliği	Lisans
K6-İTB	İşletme	Lisans
E8-İTB	İşletme	Lisans
K7-İTB	İşletme	Lisans
E9-İTB	İşletme	Lisans

*E: Erkek katılımcı / K: Kadın katılımcı

**FTM: Fen Bilimleri, Teknoloji ve Mühendislik, İTB: İnsan ve Toplum Bilimleri.

3.3.3. Veri Toplama Süreci

Nitel araştırmanın örneklem seçimi, bir önceki başlık altında belirtildiği üzere nicel araştırma bulguları neticesinde belirlenmiştir (Aydın-Çakır ve Türkeş-Kılıç, 2021: 8). Araştırmanın örneklem ölçütüne uygun özelliklere sahip, daha sonra ana araştırma grubuna dahil edilmesi de görüşmeler neticesinde uygun görülen 3 katılımcı ile pilot çalışma gerçekleştirilmiştir. Bu katılımcıların geribildirimleri doğrultusunda temel ve sondaj mülakat soruları iteratif bir şekilde mümkün olduğunca anlaşılır kılınmaya çalışılmıştır; mülakatlar seyrinde sorular katılımcı tarafından anlaşılmadığı durumlarda tekrar düzenlenip düzeltilerek katılımcıya yöneltilmiştir (Polat, 2022: 172). 18.10.2023 – 10.11.2023 tarihleri arasında gerçekleştirilen mülakatlar 12 saatte tamamlanmış, katılımcı başına ortalama 45 dakika süren görüşmeler gerçekleştirilmiştir. Mülakata başlamadan, araştırmanın amacı ve yarı yapılandırılmış mülakat tekniği ile ilgili usuller katılımcılara açıklanmıştır. EK 5’te mevcut “Gönüllü Katılım ve Bilgilendirilmiş Onam Formu” ve “Yarı Yapılandırılmış Mülakat Formu” yazılı olarak katılımcılarla paylaşılmıştır. Katılımcıların onayı dahilinde hem ses kaydı alınmış hem de mülakata dair gözlemlere dair not tutulmuştur. Görüşmeler, katılımcıların konuya ilişkin eklemek isteyebileceklerine de alan açmak adına bir kapanış sorusuyla sona erdirilmiştir.

Araştırma sorularına verilen yanıtlar tekrar etmeye başladığında örneklem büyüklüğünün belirleyicisi olan doyum noktasına ulaşıldığı için görüşmeler sonlandırılmıştır, dolayısıyla toplanan verinin araştırmanın geçerliği açısından yeterli düzeyde olduğu söylenebilir (Yıldırım ve Şimşek, 2011: 115).

3.3.4. Verilerin Analizi

Planlanan görüşmelerin tamamlanmasının ardından, arşivlenmiş ses kayıtları bilgisayar ortamında manuel olarak bire bir yazılı hale getirilmiş ve yaklaşık 320 sayfalık veri elde edilmiştir. Elde edilen görüşme metnlerinin tamamı detaylıca çözümlenerek “**İçerik Analizi**”ne tâbi tutulmuştur, her bir görüşme metninde katılımcıların yanıtları esasında tematik kodlama gerçekleştirilmiştir. Metinler yinelemeli olarak gözden geçirilerek oluşturulan kodlar rehberliğinde mülakatlar arasında örüntüler arama süreci gerçekleşmiş ve nihayetinde ana ve alt temalara ulaşılmıştır. Bu aşamada nicel araştırma bulguları doğrultusunda temellendirilen tutum farklılıklarının alt katmanları irdelenerek, öne çıkan ifade alıntılarının da üzerinde durulmuştur.

3.3.5. Geçerlilik ve Güvenilirlik

Nicel perspektifler ile karşılaştırıldığında zengin veriler üreterek teorileri mümkün kılmak için insan sistemlerini araştıran nitel araştırmaların, güvenilirlik ve geçerlilik kriterlerinin kavramsallaştırılması hakkında süregelen bir tartışma vardır (Cypress, 2017: 253). Bu çalışmada öncelikle, doğası gereği bütünlük ve geçerliliğe ulaşmak için dahili olarak gerekliliği arz eden güvenilirlik kavramı gereği, araştırma titizliğini sağlamaya yönelik stratejilerin bilhassa araştırmacı tarafından tüm nitel araştırma sürecine dahil edilmesine özen gösterilmiştir (Cypress, 2017: 260). Geçerliliğin değerlendirilmesi safhasında da birincil (inandırıcılık, özgünlük, eleştirelilik ve bütünlük) ve ikincil (açıklık, canlılık, yaratıcılık, kapsamlılık, uyumluluk ve duyarlılık) kriterleri de esas alınmıştır (Whitemore vd., 2001: 522). Öncelikle, toplanan katılımcı ifadelerinin özgünlüğüne olabildiğince sadık kalınması ve gerektiğinde bu ifadelerin doğrudan alıntılanarak sunulması ile özgünlüğün geçerliğini artırılması amaçlanmıştır (Baltacı, 2019: 377).

İnandırıcılığı ifade eden iç geçerlik kapsamında veri toplama evresinde araştırmacı, mevcut “Gönüllü Katılım ve Bilgilendirilmiş Onam Formu” aracılığıyla etik ilkeler ve gizlilik taahhüdü esasında katılımcılar ile güven odaklı bir etkileşim temin etmeye itina etmiş, görüşmeler boyunca yönlendirici sorular sormaktan da imtina edilmiştir (Creswell, 2007: 207-208). Veri çözümleme ve ölçümleme evresinde ise ana ve alt temalar arasında korelasyonlar detaylıca ele alınarak bütünlüğün sağlanmasına gayret edilmiştir (Miles ve Huberman, 1994: 279).

Araştırmanın aktarılabilişliğini ifade eden dış geçerlik kapsamında, akademik disiplin temelli tutum farklılıklarının da derinlemesine irdenebilmesi maksadıyla zaman kısıtı da göz önünde bulundurularak çalışma grubu, erişilebilirlik ve gönüllülük esasları temelinde oluşturulmuş (Yıldırım ve Şimşek, 2011: 112), amaca yönelik örnekleme yöntemlerinden **maksimum çeşitlilik (heterojen) örnekleme** ile en fazla bilgi alınabilecek katılımcıların stratejik olarak seçilmesine özen gösterilmiştir. Ayrıca nicel ve nitel yöntemlerin bir arada kullanıldığı bu araştırmanın tasarımı ve bulgularına ilişkin değerlendirmeler detaylıca aktarılmıştır (Yıldırım ve Şimşek, 2011: 258-259).

Bulguların tutarlılığının yorumlanmasında kullanılan iç güvenilirliğin temin edilebilmesi açısından, tematik kodlama yöntemi tercih edilerek 16 mülakat verisi de yinelemeli bir şekilde titizlikle bire bir çözümlenmiş (Creswell, 2007: 209-210) ve bulgular yorum katılmaksızın ortaya konulmuştur (Yıldırım ve Şimşek, 2011: 262).

Araştırma sonuçlarının teyit edilebilirliğini yansıtan dış güvenilirliğini artırmak amacıyla da araştırmacı, varsayımlarından etkilenmeksizin yine toplanan verilere sadık kalınmasını sağlamıştır (Miles ve Huberman, 1994: 278). Bununla beraber, çalışmanın tüm süreçleri ayrıntılarıyla dışarıdan bir uzman tarafından teyit incelemesine tâbi tutulmuştur (Yıldırım ve Şimşek, 2011: 272).

3.3.6. Bulgular

Mülakat verilerinin tematik analizi doğrultusunda, nicel veri çözümlemelerinde katılımcıların tutumlarında anlamlı farklılık tespit edilen “Adalet”, “Şeffaflık”, “Kötü niyetli olmama” ve “Gizlilik” boyutlarına ilişkin dokuz ana tema saptanmıştır. Cinsiyet temelli tutum farklılıkların söz konusu olduğu “Adalet” boyutu “Adalet algısı” ve “Adalet etkisi” olarak iki ana tema, “Kötü niyetli olmama” boyutu “Kötü niyet algısı”, “Kötü niyet motivasyonu” ve “Kötü niyet çıktıları” olmak üzere üç ana tema ve “Gizlilik” boyutu “Gizlilik algısı” ve “Gizlilik toleransı” ile iki ana tema altında toplanmıştır. Öncelikle bu ana temalara ilişkin bulgular ayrı birer başlık altında sunulacaktır. Akademik temelli tutum farklılıklarının tespit edildiği “Şeffaflık” boyutu için ise “Şeffaflık algısı” ve “Şeffaflık toleransı” ana temaları belirlenmiştir. Cinsiyet temelli tutum farklılıklarının tespit edildiği boyutlara ilişkin bulgulardan sonra “Şeffaflık” boyutu bulguları ele alınmaktadır. En son ise “Sorumluluk” boyutuna ilişkin bulgulara yer verilecektir.

3.3.6.1. Adalet

Cinsiyet temelli tutum farklılıklarının tespit edildiği “Adalet” boyutu “Adalet algısı” ve “Adalet etkisi” ana temalarını ortaya koymuştur. “Adalet algısı” ana teması altında “objektiflik”, “rasyonellik”, “geliştirici perspektifi”, “affektif ve kognitif insani nitelikler”, “teknik sağlamlık ve güvenilirlik”, “yasa merkezlilik”, geniş sprektrumda veri altyapısına işaret eden “çeşitlilik ve esneklik” ve “gizlilik” olmak üzere sekiz alt tema belirlenmiştir. “Adalet etkisi” ana teması ise “savaş ve çatışma koşulları - anomi”, “verilerin kötüye kullanımı”, “karar süreçleri”, “YZ’ye güven kaybı” ve “sorgulama ve inisiyatif yitimi” alt temalarını kapsamaktadır. Ana ve bağlantılı alt temalar, görüş bildiren katılımcı kodlarının ve ilgili katılımcı ifadelerinin görülme sıklığının da dahil edildiği Tablo 3.7’de sunulmaktadır.

Tablo 3.7. Adalet Boyutuna İlişkin Ana ve Alt Temalar

Ana Temalar	Alt Temalar	Görüş Bildiren Katılımcılar	Frekans
Adalet algısı	Affektif ve kognitif insani nitelikler	E3, E4, E7, E8, E9, K2, K4, K5, K7	9
	Objektiflik	E1, E2, E6, E7, E9, K1, K3, K4	8
	Geliştirici perspektifi	E3, E6, E7, E9, K1, K3, K5	7
	Çeşitlik ve esneklik	E3, E4, E5, E8, K3, K5, K6	7
	Rasyonellik	E1, E2, E7, K3, K4	5
	Yasa merkezlilik	E1, E3, E5, E7, K1	5
	Teknik sağlamlık ve güvenilirlik	E5, E9	2
	Gizlilik	K7	1
Adalet etkisi	Karar süreçleri	E1, E2, E3, E4, E5, E6, E7, E8, E9, K2, K3, K5, K6	13
	Savaş ve çatışma koşulları – Anomi	E1, E3, E7, K1, K2, K3, K6	7
	YZ’ye güven kaybı	E2, E5, E7, K2	4
	Verilerin kötüye kullanılması	E1, K7	2
	Sorgulama ve inisiyatif yitimi	E1, E4	2

*E: Erkek katılımcı / K: Kadın katılımcı

3.3.6.1.1. Adalet Algısı

Tablo 3.7’de “Adalet algısı” ana temasında gözlemlenebileceği üzere, öncelikle, erkek katılımcıların kadın katılımcılardan farklı olarak “teknik sağlamlık ve güvenilirlik” ve “yasa merkezlilik” alt temaları üzerinde durduğu tespit edilmiştir. Erkek katılımcılar, yasa merkezlilik ile paralel olarak bir teftiş mekanizmasının ve itiraz mercii gereğinin de altını çizmektedir. Bu

gerekliliğe dair ilgili ifadeleri, adalet boyutunun hesap verebilirlik ve sorumluluk boyutuyla kuvvetli ilişkisini de desteklemektedir.

Diğer yandan, adaletin tesis edilmesi hususunda erkek katılımcıların ağırlık olarak, özellikle yasal karar süreçlerinde YZ sistemlerinin objektiflik ve rasyonellik potansiyel niteliklerini göz önünde bulundurarak kazanımlara eğildiği, kadın katılımcıların ise “vicdan” ve “empati” gibi affektif ve kognitif yani duygusal ve bilişsel insani niteliklerin noksanlığı sebebini öne sürerek ekseriyetle ceza hukuku gibi kritik pratiklere yönelik endişelerini ve karşıtlığını dile getirdiği dikkat çekmektedir. “Affektif ve kognitif insani nitelikler” alt temasına ait katılımcı ifadelerinden alıntılara aşağıda yer verilmiştir:

“..vicdan mesela adalette önemli bir unsurdur, yapay zekanın vicdanlı davranıp davranmaması yani.. adaletli olurken vicdan var mı yok mu mesela burada çok ayrıma gidilmesi gerekiyor bence. çünkü vicdan biraz daha duyguya bağlı. ..bir yapay zekanın.. çok daha keskin ve sistematik konularda vereceği kararlarda bence adaletli olması düşünülebilir..” (E3)

“..günümüzün adalet sisteminde insan faktörü için çok fazla içerisinde. ..örnek vermek gerekirse.. hakimın ruh hali.. bilgi birikimi veyahut bildiğiniz gibi rüşvet alma durumları vs.. stresli olup olmaması.. davanın belki sabah erken saatte görüşülmesiyle öğleden sonra görüşülmesi arasında bile bir fark oluşabilir.. çok mikro faktörler var, makro faktörler de var.. günümüzde de görüyoruz, bazı emsal kararlar olmasına rağmen hala aksi kararlar verilebiliyor mahkemelerde. ..dolayısıyla çok adil veya işin tabiri caizse kitabına uygun olarak yapıldığını düşünmüyorum insan bazında. Yapay zekanın da tabii ki bu anlamda çok katı çizgileri var.. onu çok manipüle edip değiştiremiyorsunuz veya kendi keyfine göre bir karar vermesi çok söz konusu olmuyor. ..işin teorik bilgisinin ya da geçmişe yönelik incelenmesi açısından yapay zeka çok daha başarılı, çok daha hızlı, çok daha kesin sonuçlar veriyor..”

“..belki duyguları burada söyleyebiliriz. sonuçta insan duygusal bir varlık ve duygularımızla hareket ediyoruz ama yapay zekanın şu aşamada en azından duyguları yok; tamamen semantik, mantıklı bir şekilde, olması gerektiği gibi, objektif bir yol izleyerek.. aslında kavramları birbirine bağlayarak bir yol haritasında bir sonuca ulaşıyor.. orada çok yani YZ’ye ne bileyim rüşvet veremezsiniz ya da bir makinenin yorgun olması gibi bir durumdan söz edilemez, hani düzgün çalıştığı takdirde her şey kontrol mekanizmalarıyla beraber.. insan duygularından ekarte ettiğini söyleyebilirim..” (E7)

“..yapay zeka aslında bir nevi.. vicdandan yoksun. ..vicdan olmadığı zaman da ben adaletin olacağını düşünmüyorum. Çünkü adalet doğru ve yanlışlardan değil de hak ve hukuktan doğan şeyler. Yani ben hakkı, insan hakkını, yapay zekanın yani bir bilgisayarın algılayabileceğini düşünmüyorum. Adaletli olması o yüzden çok zor, sadece ona verilen bilgileri uygulayabilir, adaletli olamaz, diye düşünüyorum..”

..empati yapabiliyor olması gerekirdi. ..insanları, insani duygularla düşünebiliyor olması gerekirdi. ..bir konu hakkında karar verirken meşru müdafanın ne olduğunu bilmesi gerekirdi mesela. Onu bilmediği ya da onu anlamadığı için, anlayamayacağı, anlamlandıramayacağı için aslında adaletli kararlar da veremez. ..çok uç noktalarda şeyler bunlar, meşru müdafaa için empati kurabilmek gerekiyor ve bir yapay zekanın empati yeteneğine sahip olabileceğini sanmıyorum ben. Yani öyle bir şeyi bir insanın bir bilgisayara aktarabilmesi çok zor.

..kurallardan kanunlardan çok insani duygular derim. ..bir bilgisayara aktaramayacağımız sonuçta.. sevgi, üzüntü ya da bağlılık.. bağlılığın ne demek olduğunu bilmesi gerekir mesela..

..yapay belli bir noktaya kadar kullanılabilir adaletli olmak için. ..daha ileri boyutlarda, mahkeme noktasına geldiği zaman işler biraz daha değişir diye düşünüyorum; günlük hayattaki adaleti sağlar ama sonuçları büyük olan şeylerde kullanılmasının doğru olduğunu düşünmüyorum.. bir insanı hapse atmaya geldiği zaman konu, bunu kullanmamalıyız çünkü o noktada işin içine vicdan ve empati giriyor.. herkesin eşit durumda olduğu noktalarda kullanılabilir ama birinin bir konu hakkında suçlandığı zaman kullanılmamalıdır.” (K2)

Literatür ile uyumlu olarak, insan bilişsel önyargısı, veri önyargısı ve mühendislik kararları gibi etmenlerin “Geliştirici perspektifi” alt teması dahilinde hem erkek hem de kadın katılımcılar tarafından dolaylı olarak ifade edildiğinin ve adaletsizliğe yol açacağının sıklıkla vurguladığının da belirtilmesi gerekir. Bu alt temaya ait katılımcı ifadelerinden alıntılara aşağıda yer verilmiştir:

“..uygulamayı geliştirenlerde aslında, biraz daha onlarda.. biraz daha geri planına gidecek olursak aslında onları da biraz incelemek lazım bu konuda. ..onların vereceği karar da biraz daha şey yapıyor.. çünkü YZ uygulamasını geliştiren onlar yani, öncelikle onları yorumlamak gerekiyor.” (E3)

“..o uygulamalar da sonuçta bir insan tarafından yazılıyor ve bu yüzden adaletsizliklere yol açılabilir.. nasıl yazıldığını bilmiyoruz, arkasındaki mekanizmayı.. sonuçta bizler hep kendi çıkarımızla düşünürüz.. o yüzden onu yapan kişilere tam olarak güvenmem.” (E6)

“..programlayan kişinin adaletine kalmış bir şey bence.. gerçekten adaletli ise.. oluşturduğu model de ona göre tepkiler veriyor.. yani kişinin kendi adaleti.. yapay zekanın kendinin bir adaletinin olduğunu, kendisinin bir rasyonel ya da radikal bir karar verebileceğini düşünmüyorum. Henüz o düzeyde değil.. onu modelleyen araştırmacının kendi adaleti.” (K1)

“..yapay zekayı da yapan sonuçta bir insan olduğu için bunun direkt yapay zekanın verdiği bir karar olduğunu düşünmüyorum ve o adaleti sağlayan bence insan oluyor bu durumda.” (K3)

“..belirli etik kurallar var; ‘adaletli olması için şöyle, şöyle olunmalıdır; şu konuda şu şekilde bakılmalıdır.’ bu bilgiler yapay zekaya yükleniyor ve yapay zeka ona verilen bilgiler doğrultusunda bize sonuç döndürdüğü için de adalet, adil olmak kavramları nasıl tanımlandıysa ona göre bir dönüş yapıyor. ..biz eğer eksik veya yanlış bir bilgi yönlendirirsek ona, verilen sonuç adil veya adaletli olmayabilir.”
(K5)

3.3.6.1.2. Adalet Etkisi

Tablo 3.7’de işaret edildiği üzere “Adalet etkisi” ana teması bünyesinde incelenen “savaş ve çatışma koşulları - anomi”, “verilerin kötüye kullanımı”, “karar süreçleri”, “YZ’ye güven kaybı” ve “sorgulama ve inisiyatif yitimi” alt temalarına yönelik ise, “karar süreçleri” ve bu süreçler ile ilintili olarak “savaş ve çatışma koşulları” alt temalarının hem erkek hem kadın katılımcıların ifadelerinde yoğun olarak tespit edildiği söylenebilir. YZ aracılı adaletsizlik ve haksızlık neticesinde toplumsal kabul ve kullanımın etkilenmesi ile “YZ’ye güven kaybı” kadın katılımcılardan ziyade erkek katılımcılar tarafından dile getirilmektedir. “Sorgulama ve inisiyatif yitimi” alt teması ise adalet boyutuyla bağlantılı olarak sadece iki erkek katılımcının, insanın düşünsel olarak pasifleşmesine ve sorgulama kabiliyetinin körelmesine yönelik ifadelerinde gözlemlenmektedir:

“..global olarak tamamen.. işi uygun olarak yaparlarsa, kodlayan insanlar da veri setleri de, tamamen objektif olarak karar verebilir. ..diğer türlü çok tehlikeli olur ve tüm insanlar, yapay zeka olduğu için onu kabul edecektir.. herkes onun kararının doğru olduğunu düşünür çünkü.. ben bin tane bilgiyi aklımda tutabiliyorsam yapay zeka bir milyar tane veriyi aklında tutabiliyor.. bu da insanların ona saygı duymasını gerektirir; ama o veri yanlış olursa, biz yanlış kişinin arkasından gitmiş gibi oluruz bir nevi..” (E1)

“..adil sonuçlar veriyorsa tabii ki de güvenebiliriz. ..zamanla yüksek ihtimalle kendimizi teslim de ederiz. ..mesela artık bir insanın elinden belki geçmeyebilir; tam anlamıyla bir yapay zekanın işleyip sonuç verdiği bir sisteme çıkabilir.. bu ne kadar gelişmiş ve ne kadar gelecekte olduğuyla bağlantılı..

..sürekli olarak olumlu, insan mantığına yatan, etiğimize ve duygularımıza yatkın sonuçlar oluşturunca ve bunda hata payı çok azsa hatta yoksa.. şöyle bir sonuç olabilir; insanı her anlamda tembelleştireceği için.. bir şey etik mi değil mi, diye düşünme gereği duymayabiliriz belki.. doğru mu, yanlış mı diye düşünme gereği duymayabiliriz.. bahsettiğim şey çok ütöpik ama çok gelecekte ve yapay zekanın çok hâkim olduğu bir ortamdan bahsediyorum..” (E4)

Görüş sıklığının yüksek olduğu (13/16) “Karar süreçleri” alt temasına ait katılımcı ifadeleri incelendiğinde, kullanım bağlamına göre katılımcı tutumlarında değişkenlik söz konusu iken hukuki, tıbbi ve eğitime yönelik karar alma süreçlerinde insan yaşamını

etkileyebilecek “üstünkörü” kararlar konusunda katılımcıların çoğunluğunun hassasiyet gösterdiği tespit edilmiştir. Söz konusu alt temaya ilişkin ifade alıntıları aşağıda yer almaktadır:

“..gelecekte diyelim ki artık yapay zeka ile sağlanıyor mahkeme sistemleri atıyorum, ama doğru düzgün bir sonuca varamadı ve haklı olan birisini çok da iyi savunamadı ya da haksız buldu.. bu kişi mesela haksız yere yargılanabilir..” (E4)

“..doğal dil işleme ile aslında öğrencilerin cevaplarını değerlendirdiğimiz bir sistemimiz olsun.. puanlandırma.. öğrencisinin anlattığı bir şey ya da o kelime dağarcığı eğitim setimizde yoksa ve öğrenci aslında bunu açıkladıysa ona düşük puan vermek hiç adil bir sonuç değil..” (E5)

“..yanlış bir karar verdiği durumda.. otuz yıl, otuz beş yıl suçsuz yere cezaevinde bulunan insanlar var.. bir insanın bir saatini, bir gününü bile telafi edemeyiz.. zaman telafi edilebilen bir şey değil.. elbette ki burada doğruluğu ve güvenilir olması çok çok önemli bir nokta..

..sağlık sektörü üzerinde düşünecek olursak doğru teşhisin konulması, doğru tedavinin öneriliyor olması.. sonuçta doğru tedavi önerilmediği takdirde insanın hayatını kaybetmesine veya çok daha ağır hastalıklara mal olabilecek bir durum.. yapay zekanın kullanıldığı hemen hemen her alanda güvenilir ve objektif olmasını istiyor olunuz, diye düşünüyorum..” (E7)

“..küçük olaylardan, küçük ipuçlarından insanları biz suçlu ya da suçsuz duruma getirebiliyoruz. yaptığı ufak bir davranış, o anda düşündüğünü, olduğu durumu değerlendirdiğimizde o insan görünürde suçlu olsa bile aslında o küçük davranışla hiç ceza almadan kurtulabiliyor mesela.. yapay zeka bu durumu düşünmeyecek çünkü insani bir duygu var o işin içinde.. o insan suçsuz olması gerekirken.. hiç ceza almaması gerekirken.. yapay zeka değerlendirdiği için, sadece kurallarla değerlendirdiğimiz için hapse gidecek.. burada yapay zekanın kullanılması adaletsizlik oluyor, bana göre.. ama öbür türlü, daha herkesin eşit durumda olduğu.. mesela miras paylaştırılırken ya da insanların arasında küçük sorunlar çıktığı zaman, daha günlük, daha herkesin çok daha stabil şekilde çözebileceği olaylarda yapay zeka her zaman daha adaletli olur çünkü o zaman da insani duyguların olmaması gereken noktalar olur.” (K2)

3.3.6.2. Kötü Niyetli Olmama

Bu boyut, üç ana tema altında incelenmektedir. “Kötü niyet algısı” ana teması altında “askeri YZ ve siber tehditler”, “derin sahte tehlikesi”, “süper zekâ”, “veri ve gizlilik ihlali”, “davranışsal manipülasyon” ve “otomasyon” alt temaları belirlenmiştir. “Kötü niyet motivasyonu” ana teması üç alt temayı kapsamaktadır: “maddi kazanç arzusu”, “bireysel duygusal ve çıkar temelli dinamikler”, “devletler arası rekabet ve jeopolitik çekişme”. Son ana tema ise “Kötü niyet çıktıları”dır ve “bilgi manipülasyonu ve rıza imalatı”, “YZ’ye güven

kaybı”, “lkeler arası politik gerilim”, “artan yapısal eşitsizlikler ve güç asimetrisi”, “veri ve gizlilik ihlali” ve “toplumsal bağlarda erozyon ve güven sorunları” olmak üzere altı alt temadan oluşmaktadır. Ana ve bağlantılı alt temalar, görüş bildiren katılımcı kodlarının ve ilgili katılımcı ifadelerinin görülme sıklığının da dahil edildiği Tablo 3.8’de sunulmaktadır.

Tablo 3.8. Kötü Niyetli Olmama Boyutuna İlişkin Ana ve Alt Temalar

Ana Temalar	Alt Temalar	Görüş Bildiren Katılımcılar	Frekans
Kötü niyet algısı	Veri ve gizlilik ihlali	E1, E3, E4, E5, E7, E8, E9, K1, K4	9
	Askeri YZ ve siber tehditler	E1, E5, E6, K1, K2, K4, K6	7
	Derin sahte tehlikesi	E2, E5, K5	3
	Süper zeka	E4, E6, K7	3
	Davranışsal manipölasyon	E7, E8, K3	3
	Otomasyon	K2, K7	2
Kötü niyet motivasyonu	Maddi kazanç arzusu	E2, E3, E5, E7, E8, E9, K1, K2, K3, K4, K5	11
	Bireysel duygusal ve çıkar temelli dinamikler	E1, E3, E5, E6, E7, E9, K1, K2, K3, K6	10
	Devletler arası rekabet ve jeopolitik çekişme	E6, K2, K4, K5, K6, K7	6
Kötü niyet çıktıları	Bilgi manipölasyonu ve rıza imalatı	E1, E4, E6, E7, E9, K3, K4, K5, K6	9
	lkeler arası politik gerilim	E3, E6, E9, K1, K3, K4, K5, K6, K7	9
	Toplumsal bağlarda erozyon ve güven sorunları	E1, E2, E4, E9, K1, K2, K3, K6	8
	Artan yapısal eşitsizlikler ve güç asimetrisi	E5, E7, K2, K3, K4, K6, K7	7
	Veri ve gizlilik ihlali	E4, E8, K1, K2, K4, K5, K7	7
	YZ’ye güven kaybı	E2, E5	2

*E: Erkek katılımcı / K: Kadın katılımcı

3.3.6.2.1. Kötü Niyet Algısı

Tablo 3.8 incelendiğinde, “Kötü niyet algısı” ana temasında ilk olarak “otomasyon” alt temasına yönelik katılımcı ifadelerinin sadece iki kadın katılımcıya ait olduğu dikkat çekmektedir:

“..fabrikada kullanılan yapay zekâ bile bazen kötü niyetli olabiliyor çünkü istihdamı azaltıyor, insanların çalışmasını azaltıyor.. yapay zekâ tıp alanında arttıkça mesela.. sadece tıp değil, bu her alanda geçerli; fabrikalarda, tıpta, hukukta.. insanların artık çalışmak gibi bir amacı kalmıyor, çalışmıyorlar.. çalışmayan bir toplum ne kadar yükselebilir.. insanlar sadece makinelerden bir şeyler yapılmasını bekliyorlar artık, eskiden araba

fabrikalarında insanlar çalışıyordu, artık araba fabrikalarında hiç insan çalışmıyor mesela, tamamen yapay zeka ile üretilmiş makineler çalışıyor..” (K2)

“..yapay zekânın insanları işsiz bırakacağını da düşünüyorum bir noktadan sonra.. kasiyersiz kasalar.. artık bir kasiyere ihtiyaç duymayacağız o zaman ileride diye düşünüyorum çünkü gayet rahat işimizi halledebiliyoruz onlarsız da..” (K7)

“Deepfake” benzeri manipülasyon tekniklerinin dolandırıcılık ve toplumsal düzeni bozma amaçlarıyla kullanımı endişesi “derin sahte tehlikesi” alt teması altında iki erkek, bir kadın katılımcı ifadelerinde saptanmıştır. Literatürde YZ etiğine dair uzun vadeli sorunlar başlığı altında işlenen ve hararetli bir tartışmalara yol açan “süper zekâ” kavramı da iki erkek ve bir kadın olmak üzere üç katılımcının yanıtlarında kaygı unsurudur. “Davranışsal manipülasyon” alt teması, dijital ayak izlerinin tüketim alışkanlıklarını, siyasi tutumları ve psikolojik durumları manipüle etmek maksadıyla kötü niyetli kullanımına yönelik endişeleri özetlemekte iken yine iki erkek ve bir kadın katılımcının ifadelerinde tespit edilmiştir. “Veri ve gizlilik ihlali” (9) “askeri YZ ve siber tehditler” (7) alt temalarıyla ilintili ifadeler ise “Kötü niyet algısı” teması altında en sık gözlemlenenlerdir. Diğer taraftan bu temayla bağlantılı olarak “veri ve gizlilik ihlali”ne yönelik endişeler yedi erkek katılımcı ve sadece iki kadın katılımcı tarafından dile getirilmiştir ki bu değinilmesi gereken bir farka işaret etmektedir. “Veri ve gizlilik ihlali” alt temasına ait dikkat çeken katılımcı ifadelerinden birkaçı aşağıdaki yer almaktadır:

“..izinsiz veri transferi sağlıyor olabilir.. çünkü yapay zeka.. uygulama olarak da aslında çoğunu kullanabiliyoruz ve bunlar telefonadır, bilgisayaradır indirdiğimizde erişim izni vererek haberimiz bile olmayan şeylerin karşı tarafa aktarılmasını sağlıyoruz.. ..bu verilerin alınması olarak diye düşünüyorum..

..atıyorum fotoğraf galerisi tarzında bir uygulama yüklersiniz, yapay zeka ile yüze göre filtreleme koyar; yüz seçeneklerinden Ahmet diye bir arkadaşınız vardır, Ahmet’le.. Mehmet’le fotoğrafınız vardır.. yüz taramasından sınıflayabiliyordur, stoklayabiliyordur ama bu veriyi aslında stoklarken kendi cloud sistemine de yüklüyor olabilir ve izin olmadan da yapabilir..” (E8)

“..her şey internetle gerçekleşiyor.. en ufak bir sızıntıda her şey, bütün bilgiler ele geçirebilir bence.. ..kişilerden veriler alıp satan bir yapay zeka yapmış olabilirler.. insanların konumunu öğrenip onlarla alakalı veri elde edip.. çok saçma gelebilir ama belki hırsızlık bile yapabilirler onların evlerinde.. insan kaçakçılığı olabilir.. bankalarda artık canlı destek yapay zekâ tarafından desteklenmeye başlandı.. canlı desteğe ulaşamasanız bile yapay zekâ destekli banka uygulamalarında oradan yardım alınabiliyor.. onlarda açık olup.. kullanabilirler..” (E9)

Buna karşılık, “askeri YZ ve siber tehditler” alt teması kadın katılımcı ifadelerinin ağır bastığı görülmektedir (4/7):

“..savaş robotu.. şeyi gördüğümünden beri, Boston Dynamics’in robotlarını.. onların savaşta çok rahat kullanabileceği aklıma geliyor.. karşı koymak çok mümkün değil gibi duruyor insan gücüyle..” (E5)

“..silahlı olan robotlar var; “insanlığa ileride saldırabilir” gibi haberler çıkıyor mesela.. Sonuçta biz onları üretiyoruz belki ama bir yerden sonra onun arkasındaki etmenleri işte bilmiyoruz hani kime saldıracak.. kendini korumak için diyor ama belki de spesifik bir yere saldıracak, bunu öngöremiyoruz, bilemiyoruz...” (E6)

“..kötü niyet dediğim zaman aklıma ilk gelen.. ülkelerin birbiriyle savaşları oluyor.. Filistin’e karşı kullanılan savaş uçakları, yapay zeka kullanılan insansız hava araçları mesela.. bunların kullanılması bana direkt kötü niyetli olarak geliyor çünkü insanlar birbirlerini öldürmeyi amaçlıyorlar burada ve bir insanın öldürülmesi, sebebi ne olursa olsun, kötü niyettir benim için..” (K2)

3.3.6.2.2. Kötü Niyet Motivasyonu

“Kötü niyet motivasyonu” ana temasının alt temalarından “maddi kazanç arzusu”na ve “bireysel duygusal ve çıkar temelli dinamikler”e yönelik yanıtlar hem erkek katılımcılar hem de kadın katılımcılar tarafından verilmiştir. “Maddi kazanç arzusu” alt temasında daha fazla, daha kolay ve daha hızlı para kazanma hevesleri sıklıkla dile getirilmektedir:

“..aslında kötü amaçlı olan birçok şeyin temelinde maddi kaygı yatıyor.. yani bu işten para kazanma, bir şekilde gelir elde etme.. en temelde maddi kaygı taşıdığını düşünüyorum kötü amaçlı yazılım geliştirmenin..” (E7)

“..kısa sürede yüksek hacimli paralar kazanmak.. çünkü şu an günümüzde en değerli şey veri.. artık bilgi de değil.. çok veriler transfer ediyor ülkeler, bir sürü anketler yapıyor, veri analizi yapıyor, bunları topluyor ve çok yüksek paralar değerlendiriliyor bu konuda.. bunlar hani çalınıyor da.. kötü niyetli yapay zekalar geliştirerek bu verileri de hızlı bir şekilde aslında alabilirler..” (E8)

“..uygulamayı yapan kişiler daha fazla para kazanmak veya kâr etmek amacıyla.. hani sadece uygulamadan kazandıkları değil.. insanların bilgilerini satarak vs..” (K4)

“Bireysel ve çıkar temelli dinamikler” alt teması ise bireysel hırs, güç ve statü arayışı ile hasımlık ya da geçmiş romantik ilişkilere dayanan motivasyonları içermektedir. Son olarak bu ana temada öne çıkan bir diğer bulgu, “Devletler arası rekabet ve jeopolitik çekişme” alt

temasında, kadın katılımcıların sayıca üstünlüğüdür; ilgili ifade alıntılarına aşağıda yer verilmiştir:

“..siber saldırı olarak gerçekleşiyor neredeyse çoğu saldırı ve bu yüzden.. kendini göstermesi lazım artık herkesin, ülke üzerinden düşünürsek her ülkenin..” (E6)

“..her konuda güç, ekonomik güç.. yeraltı gücü, deniz gücü, ülkeler geliyor aklıma direkt aslında, ülkelerin birbirine karşı kurmak istediği üstünlük çabası gibi geliyor yapay zekanın kötüye kullanılması..” (K2)

“..kaynakları ele geçirmek!?!.. ileride su savaşlarının olacağı söyleniyor çünkü.. devlet güçleri tarafından yapılmış bir şey olabilir.. insanlardan daha güçlü varlıklar olabilir, ileride robotlar işlerine yarayabilir savaş için.. Amerika direkt.. Çin de gayet rahat yapar, yani dev ülkeler aslında aklıma geliyor bunu yapabilecek.. Japonya..” (K7)

3.3.6.2.3. Kötü Niyet Çıktıları

Bu boyutun son ana teması olan “Kötü niyet çıktıları”nda en fazla katılımcı görüşü sıklığı gözlemlenen alt temalar “bilgi manipülasyonu ve rıza imalatı” ve “ülkeler arası politik gerilim” alt temaları olmakla beraber “YZ’ye güven kaybı” haricinde tüm alt temalarda katılımcı ifade sıklıklarının yakın seyrettiği belirtilebilir. “Adalet etkisi” ana temasında da tespit edilen ve daha çok erkek katılımcı ifadelerinde rastlanan “YZ’ye güven kaybı” alt teması, bu ana tema kapsamında da sadece iki erkek katılımcının yanıtlarına yöneliktir.

“Bilgi manipülasyonu ve rıza imalatı” alt teması; uydurma, manipüle edilmiş, sahte, yanıltıcı, sponsorlu içerik türlerini ve propagandayı içeren dezenformasyonun ve sentetik medyanın bireysel ve toplumsal etkilerine yönelik endişeleri ifade etmektedir. Bireysel karar verme süreçlerinin (alışveriş tercihleri, siyasi tutumlar gibi), seçim süreçlerinin manipüle edilmesine ve toplumsal etik değerlerin etkilenmesine yönelik endişeler sıklıkla yanıtlarda gözlemlenmektedir. Burada, “Kötü niyet algısı” ana temasında ele alınan “derin sahte”nin finansal dolandırıcılık, sahte haberler ve aldatmacalar maksadıyla kullanımının olası olumsuz sonuçlarından da söz edilmektedir:

“..ileriki dönemlerde şöyle bir şey olmasından korkuyorum; diyelim ki bir pantolon alacağım, ama yapay zeka beni öyle bir manipüle eder ki ben pantolon almaktan vazgeçerim, şort alırım, bu en küçük örnek.. bunu başka şeylerde.. siyasi görüşte yaptığını.. veya bir olaya bakış açımı değiştirirse.. mesela bir savaşa ben hümanist şekilde bakarken bir anda beni o savaşın içine çekmeye, “evet sen de saldırmalısın” gibi şeyler göndererek önüme.. beni manipüle ederse bu ciddi sorun olur, bunu da verilerle yapabilir bence

“çünkü benim de neyden etkileneceğimi, neremden vurulabileceğimi, hangi konularda hassas olduğumu..”

(E1)

“..geçtiğimiz yıllarda Amerika başkanı seçilirken mesela, Rusya'nın seçimleri etkilediği ile alakalı bir olay var, aklıma ilk olarak bu geliyor.. aslında çok uç bir örnek ama önemli.. lafının geçmesi bile bence kötü niyetli bir yapay zeka sonuçlarının nerelere varabileceğini gösteriyor.. belki gerçekten gerçekleşmedi ama ileride gerçekleşmesi çok da muhtemel..” (E3)

“..hepimiz çok fazla internette yaşıyoruz ve sosyal medya kullanıyoruz.. işte tarayıcılar kullanıyoruz.. buralarda aslında bir nevi manipüle edilebiliyorsunuz, yani ihtiyacınız olmayan bir şeyi ihtiyacınız varmış gibi satın almak zorunda hissedebiliyorsunuz..

..en büyük örneği Cambridge Analytica.. Amerikan seçimlerinde insanlar çok fazla manipüle edildi, aslında kararsız olarak kategorize edilen insanlara belli bir tarafa eğilim göstermeleri adına özelleştirilmiş içerikler sunuldu ve bu gerçekten işe yaradı...

..alışveriş tercihlerinize, siyasi tutumlarınıza.. belki bağımlılıklarınıza.. canınıza kast etmenize dair ya da başka birine kötü bir şey yapmanıza kadar birçok alanda etkisi olabilecek bir sistem olabilir yani.. satın alma veya seçim kısmından ziyade sizin psikolojik olarak ruh sağlığınızı bile etkileyebilecek bir şey.. belki tükenmişlik sendromuna ya da işte gördüğümüz şeyleri “alamıyorum, almak için daha çok borçlanıyorum”.. bir nevi kapitalist sistem içerisine de belki sizi daha fazla dahil edebilecek bir şey, sonuçta bu sistemler sizi sizden çok daha iyi tanıyor, size ne zaman neyi önereceğini çok iyi biliyor..”

(E7)

“..Türkiye'de üretilmeyen bir yapay zekânın Türkiye'nin iyiliği adına çok bir şey de yapacağını düşünmüyorum ülkeler birbirini bu kadar bitirme şeyindeyken yani.. Türkiye'de üretilen bir yapay zeka olup da Türkiye'nin kötülüğü adı altında olmaz diye düşünüyorum ama işte kaynakları da bilinmediği için.. birçok görüş çatışması.. propaganda tarzı.. komplo düzeneği tarzı şeyler de olabilir.. şimdi teknoloji artık çoluk çocuk herkesin elinde, küçücük çocukların bile ellerinde telefon var, eskaza o yapay zekayı bir çocuk kullansa çok yanlış bilgiler edinip çok farklı düşüncelere de girebilir..

..başka bir ülkenin kötü niyetiyle geldiğini düşünürsek en temel vurabileceği kesinlikle siyasi konular.. birçok konu var aslında, mesela değerlerimiz.. zaten biraz Batılılaşma şeyinde olduğumuz için Türkiye olarak, bu da yapay zekaya aktarılarak çok kolay bence bilinçaltına yerleştirilebilir insanların diye düşünüyorum..

..atıyorum toplum değerleri konusunda kötü niyetli bir yapay zeka var ortada ve ne araştırılırsa, ne sorulursa, ne karar verilirse verilsin hep toplum değerlerinin tam tersi bir şekilde bir sonuca götürüyorsa.. bu nesle kadar doğru bildiğimiz ve bir şekilde bir yere taşıdığımız değerlerimizi bir sonraki nesle aktarmada zorluk yaşarız.. şu ana kadar inandığımız, bildiğimiz değerlerimiz hiçbir şekilde

kalmayacak.. belki biz o kötü niyetle bize verilen değerlere inanarak gideceğiz bu sefer.. yeni nesli ele alalım.. zaten çoğu şeyi telefonlarından, bilgisayarlarından, internetten öğreniyorlar.. böyle bir yapay zekâ ile hayatlarına devam ederlerse zaten onlar o şekilde o görüşe inanarak gitmiş olacak.. bir de.. ben kötü niyetli gördüğümün ya da güvenmediğim arkasından gitmem, dedim, evet.. ama Türkiye'de çoğu insanın bu şekilde olduğunu düşünmüyorum, altını araştırıp yaptıklarını düşünmüyorum hiçbir şeyi.. o yüzden büyük oranda etkileyeceğini düşünüyorum, ama kısa vadede değil, uzun vadede..” (K6)

“Ülkeler arası politik gerilim” ve “artan yapısal eşitsizlikler ve güç asimetrisi” alt temalarında kadın katılımcıların baskın ifade sıklığı dikkat çekmektedir. “Artan yapısal eşitsizlikler ve güç asimetrisi” alt temasında görüş bildiren katılımcılar arasında, YZ sistemlerine erişim imkânı olan bireylerin iş gücü piyasasında ve uluslararası arenada dijital rekabet gücü yüksek ülkelerin de askeri teknolojilere dayanan avantajlarını güçlendireceği kaygısı yaygındır:

“..ülkeler arası sistemlerde aslında.. İsrail.. mesela onların Gök kubbesi (Burada, İsrail’in taşınabilir hava savunma sistemi “Iron Dome - Demir Kubbe”den bahsedilmektedir.) bir yapay zeka sistemi.. sahip olan taraf aslında nükleer güç gibi, karşı tarafı ezme gücüne sahip oluyor şu anda..

..borsa sistemleri ve ekonomi.. yani burada da dehşet bir şekilde kullanılıyor; ekonomiyi kalkındırabilirsiniz sahipseniz.. manipülasyondan da bahsedebiliriz.. bir ülkenin aslında banka altyapısını çökerttiğinizde hani, topla tüfekle girmenize gerek yok..” (E5)

“..güç dengesi yok olurdu tamamen kötü niyetli yapay zekayı birbirimize karşı kullanmaya başladığımızda, ülkeler birbirine karşı kullanmaya başladığında.. artık güçlüler zayıfları çok daha rahat ezebilecek, ezmek meşru olacak.. savaşlar çıkacak.. siyasi alanda insanlar birbirlerine artık saygı duymayacak çünkü ortada “Ben zaten senden güçlüyüm” duygusu oluşacak ve saygı ortadan kalkacak..” (K2)

“..mesela bir şeyde hızlı olmak önemlidir, atıyorum bir soru sorulmuştur; hızlı olmak önemlidir, ChatGPT size çok hızlı cevap verebilir, yapay zeka.. bir insan belki o problemi çözene kadar vs.. bunu gerçekten hak eden birinin önüne geçilebilir.. belki derslerimizden söyleyebilirim, bir problem veriliyor işte.. “Bunu çözen ilk on kişiye herhangi bir puan vs. verilecek.. ChatGPT bunu çözebiliyorsa eğer çok hızlı çözer ve o kişi onu kullanır ama gerçekten uğraşan bir kişi ona göre yavaş kalır tabii ki..” (K4)

Kötü niyetli olmama boyutuna dair yöneltilen ilk temel mülakat sorularında kadın katılımcıların “veri ve gizlilik ihlali”ne değinen yanıtlarına az rastlanmış olsa da ilerleyen temel mülakat ve sondaj soruları aracılığıyla kötü niyetli YZ geliştirme ve kullanımının toplumsal

çıktıları bağlamında “veri ve gizlilik ihlali” endişelerine erişildiği de belirtilmelidir; bu alt temada görüş bildiren yedi katılımcının beşi de kadındır.

3.3.6.3. Gizlilik

Gizlilik boyutunda “Gizlilik algısı” ve “Gizlilik toleransı” ana temaları belirlenmiştir. “Gizlilik algısı” ana teması altında “vatandaşlık verileri”, “finansal veriler”, “siber risk yönetimi”, “şeffaflık – rızaya dayalı veri işleme ve erişimi/kullanımı”, “toplumsal uyum baskısı” ve “veri yönetimi ve paylaşım sınırlamaları” alt temaları oluşturulmuştur. “Gizlilik toleransı” ise “kullanım azaltma ve alternatif yaklaşımlar”, “kullanıcı odaklı veri kullanımı ve acil durumlar” ve “kamu yararı (ve suçla mücadele)” alt temalarına ayrılmaktadır. Ana ve bağlantılı alt temalar, görüş bildiren katılımcı kodlarının ve ilgili katılımcı ifadelerinin görülme sıklığının da dahil edildiği Tablo 3.9’da yer almaktadır.

Tablo 3.9. Gizlilik Boyutuna İlişkin Ana ve Alt Temalar

Ana Temalar	Alt Temalar	Görüş Bildiren Katılımcılar	Frekans
Gizlilik algısı	Veri yönetimi ve paylaşım sınırlamaları	E1, E2, E3, E4, E6, E8, K3, K4, K5, K6, K7	11
	Şeffaflık – Rızaya dayalı veri işleme, erişimi ve kullanımı	E2, E3, E4, E5, E6, E7, K1, K2, K5, K6, K7	11
	Toplumsal uyum baskısı	E1, E6, K1, K2, K4, K5, K7	7
	Vatandaşlık verileri	E2, E6, E9, K5, K6	5
	Siber risk yönetimi	E1, E2, E5, K1, K2	5
	Finansal veriler	E3, E5, E7, K5	4
Gizlilik toleransı	Kullanım azaltma ve alternatif yaklaşımlar	E2, E4, E6, E7, E8, E9, K3, K4, K5, K6	10
	Kullanıcı odaklı veri kullanımı ve acil durumlar	E1, E3, E4, E5, E6, E8, E9, K1, K2, K6	10
	Kamu yararı (ve suçla mücadele)	E5, E7, E9, K3, K5, K7	6

*E: Erkek katılımcı / K: Kadın katılımcı

3.3.6.3.1. Gizlilik Algısı

“Gizlilik algısı” ana teması incelendiğinde görüş sıklığı en yüksek “şeffaflık – rızaya dayalı veri işleme, erişimi ve kullanımı” ve “veri yönetimi ve paylaşım sınırlamaları” alt temalarının erkek ve kadın katılımcıların ifadelerini dengeli bir şekilde içerdiği gözlemlenmiştir.

“Veri yönetimi ve paylaşım sınırlamaları” alt teması, katılımcıların, kişisel verilerinin, veri sorumluları (gerçek ya da tüzel kişi) tarafından saklanması, imhası (silinmesi, yok edilmesi, anonim hale getirilmesi) ve sürelerine ilişkin T.C. Kişisel Verileri Koruma Kurumunca (2019: 4) hazırlanan “Kişisel Veri Saklama ve İmha Politikası”na dair “6698 Sayılı Kişisel Verilerin Korunması Kanunu ve KVKK’nin ikincil düzenlemesini teşkil eden 28 Ekim 2017 tarihli ve 30224 sayılı Resmi Gazete’de yayımlanan Kişisel Verilerin Silinmesi, Yok Edilmesi veya Anonim Hale Getirilmesi Yönetmeliği uyarınca” yükümlülüklerini yerine getirmesi hakkında beklentilerini içeren dolaylı ifadelerini kapsamaktadır.

“Şeffaflık – rızaya dayalı veri işleme, erişimi ve kullanımı” alt temasında katılımcıların, bilhassa kamu görevlilerinin ve özel sektör üst kademe yöneticilerinin kişisel verilere kullanıcı rızası dışında erişimi ihtimaline ilişkin rahatsızlıklarını ifade ettiği belirtilmelidir. Aynı alt tema kapsamında, fotoğraf, konum bilgisi, kişisel yazışmalar – konuşmalar ve özellikle bulut tabanlı yönetimsel araçların işleyişi hususunda yine kullanıcı rızası vurgulanmaktadır ve kişisel verilerin gizliliği ve güvenliğine yönelik şeffaflık talebi dile getirilmektedir. İlgili katılımcı ifadelerinden alıntılara aşağıda yer verilmiştir:

“..veriyi sistem tutmalı ve aşırı güvenli olmalı, hiçbir yönetici.. yönetici dahi olsa girip kişisel veri aratamamalı. ..örnek olarak polislerin elindeki bir veritabanı sisteminde T.C. kimlik numarasıyla herkesi yazabilirsiniz. Polis olmuş bir insan sırf eğlencesine veya öbür türlüştüne çevresindeki insanların kimlik numaralarını öğrenip o sisteme girerse bu.. yanlış bir kullanım olmuş olur.. daha çok sistem tutmalı, sistem güvenli olmalı, istediğinde yalnızca kullanıcıya getirmeli.. mesela bir yönetici de açıkçası kullanıcıların e-postalarını ve telefon numaralarını görmek istememeli, işi olmamalı..” (E2)

“..kişisel bilgilerimi veriyorsam tabii buna fotoğraf dahil, aile bağlarım dahil.. konum bilgisi.. bunda da bilgi gizliliği gerekir mesela.. ..bir izin alması lazım kişiden, kullanıcıdan.. “Bilgileriniz şu şekilde kullanılacaktır” diye.. aslında bunu biraz daha sık yapması lazım. ..mesela erişim izni istiyor bir uygulama.. “Fotoğraflarınıza erişim lazımdır” falan diye.. bunu aslında biraz daha.. özel bilgileri kullanırken mesela.. biraz daha sık sorabilir yani, bu özenin olması gerekebilir bence yapay zekada, bir uygulamada..” (E3)

“..örnek vermek gerekirse, şu an tıptan kimsenin benim verime ulaşamıyor olması lazım. Eğer ben bir giriş yaptırmadıysam, hastane kaydı yaptırmadıysam.. şu an bir doktorun benim adımla girip verilerime bakamıyor olması lazım.. burada “usual ve unusual behaviorlar” var.. aile hekiminin belki arada kontrol etmesi gerekiyor olabilir işte ilacı bitti mi, şusu oldu mu busu oldu mu.. ama atıyorum, tıpta bir doçentin, benimle hiç alakası olmayanın bakamaması gerekiyor; bu “unusual” behavior yani..” (E5)

“..tabii ki bilmek istiyorum.. bugün gerçekten şunu söyleyebilirim ki gizliliğinize gerçekten önem vererek yaşamak isteyen bir insan olursanız muhtemelen hiçbir şeyi kullanamazsınız.. yani bir iş yerine gittiğinizde, bir okula gittiğinizde, bir alışveriş yapmak istediğinizde muhakkak ki bir şeylerinizi feda etmek durumunda kalıyorsunuz çünkü sistem tamamen bu şekilde kurulmuş.. bahsettiğim örnekler dolayısıyla, siz farkında olmadan dahi farklı bir yere çekiliyor olabilirsiniz.. hastayken hasta olduğunuzu fark edememek gibi bir şey.. aslında kendiniz karar veriyor olduğunuzu düşünüyorsunuz, öyle zannediyorsunuz ama aslında yönetiliyorsunuz bir taraftan.. dolayısıyla veri güvenliği çok çok önemli bir husus bu noktada..

..özellikle muhatap olduğum uygulamanın bu konuda bilgi alabileceğim bir birimi varsa veya bir iletişim kanalı varsa veya zaten kendileri halihazırda üye olduğum veya kullanıyor olduğum esnada bir bilgilendirme metni, formu, sözleşmesi sunuyorsa da bunları incelemeye çalışıyorum olabildiğince.. içerisinde anlamadığım, beni muallakta bırakabilecek ifadeler olduğunda o şirketle iletişime geçip öğrenmeye çalışıyorum ve bunun neticesinde bir karar vermeye çalışıyorum..” (E7)

“..mesela bireysel yapılan konuşmalar; sosyal medyadan ya da herhangi bir platformda iletişime geçebiliyoruz, bunlar yazışma şeklinde oluyor.. ..birisi bir şey sorduğunda benim o konuşmam üzerinden bir örnek verebiliyorsa, o konuşmamı, orada edindiği bilgileri kullanıp bir başkasına cevap oluşturabiliyorsa aslında benim gizliliğimi ihlal etmiş oluyor, çünkü bire bir yapılan bir konuşma o.. benim özel, sadece bir kişinin ya da belirli bir grubun bildiğini düşündüğüm herhangi bir bilgi.. bu çok önemli, kritik bir bilgi olmak zorunda değil, bu bilgi kullanılıyorsa, bir başkasına sunuluyorsa o noktada benim bilgilerim ihlal edilmiş olur..” (K5)

“Toplumsal uyum baskısı” alt temasının ise frekans değeri ikinci sıradadır ve irdelenmesi gereken bir farka da işaret etmektedir; bu alt temada görüş bildiren yedi katılımcının beşi kadındır.

“Siber risk yönetimi” alt temasında, “akıllı sözleşme” (smart contract) ve “uçtan uca şifreleme” (End-to-end encryption: E2EE) gibi teknolojiler aracılığıyla gizliliğin ve güvenliğin teminatının gereği katılımcılar tarafından vurgulanmaktadır:

“..akıllı sözleşmeler teknolojisi.. kriptoloji teknolojisi.. bunlar kullanıldığı sürece, gayet güvenli ve gizli.. bir de uçtan uca şifreleme diye biliyorum.. dışarıdan gelen saldırılara karşı.. inanılmaz derecede bir önlem alıyor..” (E1)

“..asla çalınmaması gibi bir şeyden bahsetmiyorum.. hani bu işler bu şekilde yürüyor; birileri çalacak ki ona karşı önlem geliştireceğiz, sistemleri daha da iyi hale getireceğiz.. ama bunun önlemini alması lazım, telafi etmesi lazım..” (E5)

Ek olarak, katılımcıların başta kimlik numaraları olmak üzere “vatandaşlık verileri” ve banka ve kart bilgilerini ifade eden “finansal veriler” konusunda daha tutucu olduğu saptanmıştır. Sosyal medya hesabının çalınması, iş verilerine ve kişisel fotoğraflarına erişilmesi ihtimaline yönelik endişelerini dile getiren katılımcılar da bulunmaktadır. Ruhsal durumunu, harcama alışkanlıklarını ve siyasi görüşünü belirtmekten kaçındığını ifade eden bir katılımcı da mevcuttur.

3.3.6.3.2. Gizlilik Toleransı

“Gizlilik toleransı” ana teması, “kullanım azaltma ve alternatif yaklaşımlar”, “kullanıcı odaklı veri kullanımı ve acil durumlar”, “kamu yararı (ve suçla mücadele)” alt temalarından oluşmaktadır.

“Kullanım azaltma ve alternatif yaklaşımlar” alt temasında, hem erkek katılımcıların hem de kadın katılımcıların ve toplam katılımcı sayısının çoğunluğunun, genel itibarıyla veri gizliliği ve güvenliğinin karşılanmadığı durumlarda söz konusu uygulama kullanımına son verme, uygulamayı kullanırken bıraktığı veri izini ya da dijital ayak izini en aza indirme, muadil rakip ürünlere yönelme veya sanal özel ağ (VPN), anonim hesap kullanma olanaklarını göz önünde bulundurabilecekleri tespit edilmiştir. Aynı zamanda katılımcılar, kullanım bağlamına ve uygulama türüne göre tutumlarının değişebileceğini de ifade etmektedirler. Ek olarak, katılımcıların yanıtlarında ekseriyetle kişisel verilerinin gizliliği ve güvenliğine itimat etmediklerine dair ifadeler yer almaktadır.

Tablo 3.9, “Kullanıcı odaklı veri kullanımı ve acil durumlar” alt temasında görüş bildiren on katılımcıdan yedisinin erkek olduğunu göstermektedir. Katılımcılar, alışveriş süreçlerini ve finansal süreçleri kolaylaştırmak amacıyla kullanıcı lehine durumlarda, kullanıcının herhangi bir zarara uğramasını önleme veya kullanıcıya yönelik muhtemel bir riski erken saptayarak önleme benzeri amaçlar doğrultusunda ve kullanıcının sağlık durumunun şüphe uyandırıcı veyahut kritik olduğu şartlarda kişisel veri ihlalinin makul karşılanabileceği yönünde görüş bildirmişlerdir. Katılımcı ifadelerinden alıntılar aşağıda sunulmuştur:

“..işime yararsa gizliliği ihlal edebilir..” (E1)

“..örnek olarak bir sağlık uygulamasının aslında benim tüm verime erişebiliyor olması bir sıkıntı yaratmaz, çünkü atıyorum bir trendi izliyor olabilir baştan sona kadar vs..” (E5)

“..spesifik bir durum yok da eğer benim lehime bir şey ise.. mesela telefon numaramı girdim bir yere ve lehime o anda ihtiyacım olan bir şey de gördü o verilerden herhangi bir uygulamada, bir satın alma.. herhangi bir şeyi olabilir ve onun için bana işte başka bir yerden bildirim geliyorsa.. ve bu olay benim lehimeyse çok sorunlu karşılamam.. ” (E6)

“..sağlık olabilir. Kan değerlerimin paylaşılması ya da röntgen görüntüleri.. kullandığım ilaçlar.. gittiğim hastane, muayene olduğum doktor..” (K1)

“..şöyle bir durum olursa, artık bankalarda vs. güvenilir bir yapay zekâyı kullanırsak.. kolaylık olabilir.. mesela benim bankanın ihtiyaç duyduğu bilgilerimi paylaşabilir; böyle bir durumda yani kabul edebilirdim.. atıyorum bir mükellef düşünün, yani bir iş yeri, işletme sahibi, mizanlarını falan.. bir yapay zekanın içinde her dönemin de bulunmasını.. o bankalar çünkü hep mizanlarını falan istiyor mükelleflerin kredi vermek için, o anki ay güncellemesi, yıl güncellemesi.. yapabilmek için, ciro bilgisi istemek için.. o yapay zekadan bulup, işte “şu mükellefin şu bilmem.. üçüncü dönem mizanını” falan, “geçici vergisi” falan dediğinde o bilgileri direkt alabilmesini isterdim mesela, güzel olabilirdi..” (K6)

Son olarak, “Kamu yararı (ve suçla mücadele)” alt temasında katılımcıların, özellikle eğitimde fırsat eşitliği odaklı girişimlere ve sağlık hizmetleri sektörüne yönelik toplumsal fayda gözetilerek geliştirilen uygulamalar söz konusu olduğunda ve adaletin tesisi temelli çabalar ile suçun önlenmesi, asayişin temin edilmesi kapsamında, veri mahremiyeti hususunda tolerans ve ilgili faaliyetler esasında gönüllülük de gösterme eğilimine sahip olabilecekleri görülmektedir:

“..çok daha fazla insanın hayatını iyileştirmek için, belki bulunduğumuz topluma, kültüre.. ya da uluslararası anlamda.. çok büyük bir popülasyona, örnekleme hitap eden bir yapay zekânın geliştirilmesi için ve iyi amaçlı olarak şeffaf bir şekilde.. gizliliğine önem verdiğim veriler işlenecek ise ve ben bundan %100 emin olabiliyorsam buna tamam diyebilirim.. kişisel verilerim olabilir, sağlık verilerim olabilir.. yaş bilğim, hangi ilaçları kullanıyor olduğum, hangi alerjik reaksiyonlara.. nasıl tepki verdiğim, hangi bölüme, hangi doktora, ne kadar muayene olduğum gibi verileri belki paylaşıyor olabilirim.. ..belki de işte hangi öğrencilere burs verileceğini tespit eden bir yapay zekâ sistemi yapılyordur ve insanların nasıl harcama yaptığının bu modele eğitilmesi gerekiyordur, o noktada belki ben de gönüllü olarak verilerimin eğitilmesi için verebilirim gibi..” (E7)

“..suçlu birinin takibinde, yani bir katil vardır ve onun verilerine erişerek, konumunu, yerini.. tüm bilgilerine erişmek bence kural ihlali sayılmaz, o kişi için sayılır ama toplum için sayılmaz bence, zorunlu bir durum..” (E9)

“..verdiğim bilgi.. başka birinin hayatını kötü etkileyecek, benim yaptığım bir suçu içeriyorsa belki de.. kabul ederdim.. psikiyatrist-psikolog üzerinden düşünüyorum, onlara bir suçlu suçunu itiraf ediyorsa bu bence mahkemeye taşınmalı, yüzde yüz psikiyatri bunu, bu suçu görmezden gelmemeli.. bir noktada suçu itiraf ettiyse ve bu suçtan eğer daha önce ceza almadıysa bu kişi.. o suç açığa çıkmalı.. bence adaletli olan o.. başka insanların hayatlarını etkilediysem ben o suç üzerinden.. bence bunu, bu bilgiyi gerekli merciye ulaştırabilir..” (K3)

“..çok ciddi durumlarda, bir adli karar verilecek mesela ve örnekler gerekiyor, o dava için birden fazla örneğe ihtiyacı var.. benim yaşadığım herhangi bir şey.. o bilgileri yapay zekâ biliyor ve o olay üzerinde bir benzerlik var, onu kullanması gerekiyor ama bazı kararlarda mesela gizlilik durumu olur, paylaşılmaz medyada.. yani orada o bir kişi kim, o olay nerede nasıl gerçekleşti.. tam böyle spesifik bir durumu, olayı, kişiyi göstermeyecek şekilde bir muhakeme yapılabilir belki.. çünkü belki de oradan alacağı örnek bir davanın sonucunu değiştirecek.. ama orada da işte dediğim gibi ben kimim, bu açıklanmamalı ya da olay belki Antalya’da gerçekleşti ama İstanbul’da gerçekleşti şeklinde medyaya yansıtılabilir.. koruyucu bir şey eklenerek bu şekilde olabilir.. yani kişi, zaman, mekan.. çok da olayı, nedenini, nasılımı etkilemeyecek etmenler ve kişiyi tam olarak göstermeyecek durumlar değiştirilerek kullanılabilir..” (K5)

3.3.6.4. Şeffaflık

Nicel veri analizleri aşamasında İnsan ve Toplum Bilimleri (İTB) grubu dahilinde değerlendirilen Hukuk, İktisadi ve İdari Bilimler ve İletişim fakültelerinden katılımcılar ile Fen Bilimleri, Teknoloji ve Mühendislik (FTM) grubuna dahil edilen Uygulamalı Bilimler, Tıp ve Mühendislik fakültelerinden öğrenciler arasında “Şeffaflık” boyutuna yönelik istatistiksel olarak anlamlı bir farklılık saptanmıştır (İTB, $\bar{x} = 4.09$; FTM, $\bar{x} = 3.92$). Dolayısıyla bu başlık altında mülakat verilerinin tematik analizine ait bulguların akademik disiplin temelli farklılıklar ekseninde incelenmesi önceliklendirilmiştir.

İlk olarak, literatürde şeffaflığın komşu kavramları olarak karşımıza çıkan “açıklanabilirlik”, “yorumlanabilirlik”, “anlaşılabilirlik” ve “kara kutu” kavramlarının katılımcı ifadelerinde karşılık bulduğu görülmektedir.

Bu boyuta dair “Şeffaflık algısı” ve “Şeffaflık toleransı” ana temaları belirlenmiş olmakla beraber “Şeffaflık algısı” ana teması “adalet odaklılık”, “kötü niyetli olmama teminatı”, “kişisel verilerin gizliliği” ve “kaynak toplama ve işleme süreci” alt temalarını; “Şeffaflık toleransı” ise “ulusal güvenlik - askeri YZ ve stratejik istihbarat”, “hasta mahremiyet hakkı”, “teknik opaklık”, “ticari sır (know-how)” ve “kullanım azaltma ve alternatif yaklaşımlar” alt temalarını içermektedir.

Ana ve bağlantılı alt temalar, görüş bildiren katılımcı kodlarının ve ilgili katılımcı ifadelerinin görülme sıklığının da dahil edildiği Tablo 3.10’da yer almaktadır.

Tablo 3.10. Şeffaflık Boyutuna İlişkin Ana ve Alt Temalar

Ana Temalar	Alt Temalar	Görüş Bildiren Katılımcılar	Frekans
Şeffaflık algısı	Kaynak toplama ve işleme süreci	E6-FTM, E7-FTM, K4-FTM, K5-FTM, E4-İTB, K6-İTB, E8-İTB, E9-İTB	8
	Adalet odaklılık	E1-FTM, E2-FTM, E3-İTB, E7-FTM, K2-İTB	5
	Kişisel verilerin gizliliği	E3-İTB, E8-İTB, E9-İTB, K3-FTM, K4-FTM	5
	Kötü niyetli olmama teminatı	E9-İTB, K2-İTB, K6-İTB	3
Şeffaflık toleransı	Kullanım azaltma ve alternatif yaklaşımlar	E1-FTM, E2-FTM, E6-FTM, E7-FTM, E9-İTB, K1-FTM, K2-İTB, K3-FTM, K6-İTB	9
	Ulusal güvenlik – askeri YZ ve stratejik istihbarat	E4-İTB, E9-İTB	2
	Hasta mahremiyeti hakkı	E9-İTB, K3-FTM	2
	Ticari sır (know-how)	E7-FTM, K4-FTM	2
	Teknik opaklık	E5-FTM	1

*E: Erkek katılımcı / K: Kadın katılımcı

**FTM: Fen Bilimleri, Teknoloji ve Mühendislik disiplinlerinden katılımcılar, İTB: İnsan ve Toplum Bilimleri disiplinlerinden katılımcılar

3.3.6.4.1. Şeffaflık Algısı

“Şeffaflık algısı” ana teması, şeffaflık boyutunun adalet, kötü niyetli olmama ve mahremiyet (gizlilik) boyutlarıyla birbirlerini tamamlayıcı nitelikte olduğunu ortaya koymaktadır. Şeffaflığı bir “kötü niyetli olmama teminatı” olarak gören katılımcılar İTB bölümleri öğrencileridir.

Hem İTB disiplinlerinden katılımcılar hem de FTM bölümlerinden öğrenciler, “kaynak toplama ve işleme süreci” üzerinde durmaktadır. FTM disiplininden bazı katılımcıların “açık kaynak kodlu yazılımlar”ı tercih ettiklerini belirtmeleri dikkat çekmektedir:

“..bunu şöyle bir analogiyle açıklayabilirim; bir restorana gittiniz, yemek yiyeceksiniz ve yemeğin nasıl yapıldığını bilmiyorsunuz arka planda.. arkada yapılıyor ve getiriliyor.. hangi verileriniz toplanıyor, hangi amaçla işleniyor, arkasında ne gibi bir durum var.. bunu bilmek ben isterim açıkçası.. bu noktada da şeffaflık benim için önemli bir nokta.. yani arkadaki algoritmanın nasıl çalıştığı, benim verilerimi nasıl işleyeceği veya bunun ne gibi şeylere yol açacağını bilmek tabii ki isterim..

..en nihayetinde açık kaynak kodlu olmayan bir projeyse o projeyi yöneten kişilerin kişisel çıkarları veya müşterilerine aslında her şey yolundaymış gibi gösterip aslında arka kapılardan farklı şekillerde veri ihlalleri ya da çeşitli illegal olmayan durumlar oluşabilir, diye düşünüyorum.. algoritmalar vs. dışında insan faktörü de var içinde, sonuçta o bir ekip işi, yani bu çok iyi denetlenmediği zaman, çok iyi gözetlenmediği zaman, en basitinden açık kaynak kodlu olmadığı zaman.” (E7-FTM)

“..direkt aldığı veritabanları aklıma geliyor, nereden kaynağı aldığı geliyor. Eğer o kaynaklar güvenilir, bilindik, uluslararası yerlerdense zaten daha rahat.. daha güvenilir olur.. ama daha küçük çaplı kaynaklardan geliyorsa ve ben bunu biliyorsam güvenmezdim açıkçası..” (E8-İTB)

“..verdiği bilgilerin kaynağını göstermeli bence, nereden aldığını, neye göre o bilgiyi verdiğini..” (E9-İTB)

“..o verinin kaynağı benim için önemlidir tabii ki.. elime nereden ulaştığını bilmem gerekiyor.. ..ya da bana o bilgiyi kim veriyor, kim tarafından ben o bilgiyi alıyorum.. belki o bilgiyi aldığı başka yerler de var, başka sitelerle belki bağlantıda olabilir.. atıyorum, ‘Wikipedia’den aldım, şuradan aldım, buradan aldım’ ya da ‘şu sonuçları birleştirdim..’ gibi bir kaynakça vermesi daha uygun olabilir diye düşünüyorum..” (K6-İTB)

“Kişisel verilerin gizliliği” alt teması, katılımcıların, hem kişisel verilerinin gizliliği ve güvenliğine riayet edilmesi konusunda hassas hem de veri sorumluları tarafından kişisel verilerinin gizliliği ve güvenliğinin sağlanmasına yönelik detaylar hakkında şeffaf bir tutum sergilemesi beklentisinde olduklarını göstermektedir:

“..ben bir şey araştırıyorum ya da bir konu hakkında çalışma yapacağımdır.. onların kaydedilip başka insanlara, başka şirketlere satılıp satılmadığından emin olamam.. onların güvencesinin verilmesi gerekiyor bence..” (E9-İTB)

“..özel hayat ve kişi haklarının korunması gibi konularda tabii ki de yüzde yüz şeffaflık gerekemeyebilir..” (K3-FTM)

“..yapay zekâda herhangi bir kişisel veri kullanılıyorsa.. atıyorum ben ona işte ismimi.. giriyorsam.. bunların tabii.. kişisel verilerin paylaşılmaması iyi olur, yani mantıklıdır..” (K4-FTM)

3.3.6.4.2. Şeffaflık Toleransı

YZ sistemlerinin kullanımı neticesinde verilen kararların açıklanmamasını gerektiren bağlamlara yönelik katılımcı ifadeleri “Şeffaflık toleransı” ana teması altında bir araya getirilmiştir. Katılımcı ifadeleri, şeffaflığın kullanıcıların adalet ve güvene dair endişelerinin hafifletilmesine yardımcı olabileceği gibi, açıklamalar ve yayınlanan ek bilgiler aracılığıyla sistemin saldırılara karşı daha savunmasız hale getirilebilmesine de sebep olabileceğini göstermekte; literatürle uyumlu olarak “şeffaflık paradoksu”na işaret etmektedir.

“Kullanım azaltma ve alternatif yaklaşımlar” incelendiğinde, görüş bildiren dokuz katılımcıdan çoğunluğunun FTM disiplinlerinden öğrenciler olduğu görülmektedir (6/9). Bu alt temada da katılımcılar, “gizlilik toleransı” alt temasına benzer şekilde, kullanımı sonlandırmaya, asgariye düşürmeye ya da muadil rakip uygulamalara yönelmeye ilişkin ifadeler kullanmışlardır. Diğer taraftan, sadece FTM disiplinlerinden öğrencilerin, sisteme dair açıklamaları yetersiz ya da mantıksız bulmaları durumunda önyargıyla yaklaşmama, ilgili nedenleri tekrar gözden geçirme, tartma veya daha ayrıntılı bir açıklama talebinde bulunma eğiliminde olacakları gözlemlenmiştir. Son olarak, kullanılan YZ uygulamasına ve bağlama göre tutumunun farklılaşabileceğini ifade eden katılımcıların çoğunluğunun İTB disiplinlerinden katılımcılar olduğunun da belirtilmesi gerekir.

“Ulusal güvenlik – askeri YZ ve stratejik istihbarat” alt teması sadece İTB bölümlerinden iki öğrencinin ifadelerine yöneliktir, tüm katılımcıların ilgili ifade alıntıları aşağıda sunulmaktadır:

“..belki çok daha büyük mevzulara karar veren bir robotumuz vardır ve “derin devlet” bir robot bu.. verdiği kararlar.. ulusal değer, anlam taşıyordur..”(E4-İTB)

“..askeri alanlarda da kullanılan YZ’ler var.. savunma sanayiinde de yapay zekâ kullanılıyor mesela.. yani operasyona çıkan bir birliğin kullandığı ortak bir sistem vardır yapay zekâ destekli.. bunlara düşman birliklerinin ya da düşmanların erişmesi soruna yol açar.. çünkü yapay zekânın destekçisinin kim olduğu bilinmiyor.. ve verilerin gizliliği konusunda şüphe varsa ya da sistemsel açık varsa karşı taraf bunu öğrenebilir..”(E9-İTB)

“Ticari sır (know-how)” alt teması da sadece iki FTM bölümlerinden katılımcıların ifadelerine yönelik oluşturulmuştur. Hasta mahremiyetinin sağlanmasına yönelik ise, iki gruptan da birer katılımcının ifadeleri tespit edilmiştir. Bu alt temalardan öne çıkan ifade alıntılarında bazılarının aşağıda yer verilmiştir.

“Hasta mahremiyeti hakkı” alt teması üzerine;

“..tıpta olabilir.. hasta ilişkilerinin gizliliği.. bunlar gizli tutulabilir bence yapay zeka kullanıyorsa eğer.. psikoloji anlamında olabilir.. şu an var mı bilmiyorum ama ilerleyen zamanlarda gerçekleşebilir bence psikoloji alanında yapay zekanın desteği.. ” (E9-İTB)

“..hastayla ilgili bilgilerin bana verilmemesi olabilir.. çünkü onun özel hayatı..” (K3-FTM)

“Ticari sır (know-how)” alt teması üzerine;

“..belki işte patentle ilgili konular olabilir.. mesela Apple’da bildiğim kadarıyla çok fazla yeni teknoloji geliştiriyorlar ve tabii ki bunun sızdırılmasını, açıklanmasını ya da rakip firmalar tarafından çalınmasını istemedikleri için bu konu hakkında patentleyip bilgi vermeme haklarını kullanabiliyorlar.. tabii ki burada sonuçta henüz kullanmıyor olduğum bir ürün olduğu için bu konuda bana bilgi verilmemesi herhangi bir sorun teşkil etmiyor benim için.. ama.. direkt olarak benim hayatıma, verilerime dokunan bir kısım olduğunda çok bir gerekçe göremiyorum açıkçası..” (E7-FTM)

“..kullanılan yapay zekanın mekanizmasıyla ilgili bir şey saklı tutulabilir.. teknik bir şey belki.. hani başkaları tarafından aynı mekanizma, algoritma kullanılmasın diye..” (K4-FTM)

Son alt tema olan “Teknik opaklık” alt teması, sadece bir bilgisayar mühendisliği öğrencisinin, özellikle Evrişimsel Sinir Ağları (Convolutional Neural Networks: CNN) gibi derin öğrenme algoritmalarının kara kutu niteliğine yönelik ifadesi dikkate alınarak oluşturulmuştur, söz konusu katılımcı yine de şeffaf olunmamasını hiçbir bağlamda kabul edilebilir bulmamakla birlikte “algoritmanın, YZ algoritması da olsa, net bir şekilde açıklanabilir olması gerektiği”ni dile getirmiştir.

3.3.6.5. Sorumluluk

Nicel veri çözümlemesi neticesinde “Sorumluluk” boyutuna ilişkin gruplar arası anlamlı bir farklılık tespit edilmemiş olsa da diğer boyutlara ve bu boyutlar ile sorumluluk boyutu arasında potansiyel doğrusal ilişkilere dair daha detaylı bulgular ortaya çıkarabilmek amacıyla katılımcılara bu boyut kapsamında da temel mülakat ve sondaj soruları yöneltilmiştir.

YZ geliştiricileri, yasa koyucular (düzenleyiciler), teknoloji sağlayıcıları (tedarikçiler), bireysel ve tüzel (kuruluşlar) kullanıcılar ve sistemin kendisi olmak üzere farklı YZ ekosistemi

aktörlerinin (Minkkinen vd., 2023: 104) sorumluluklarına ve sorumlulukların belirlenmesinin kritik olduğu kullanım bağlamlarına yönelik katılımcı yanıtları içerik analizine tâbi tutulmuştur.

Bir YZ sisteminin uygulanması sonucunda istenmeyen bir durum söz konusu olduğunda YZ ekosistemi aktörlerinden hangisini en sorumlu bulduğu sorulduğunda, tüm katılımcıların yedisinin (7/16) yasa koyucuların yani düzenleyicilerin sorumluluğunu ifade ettiği, ancak burada akademik disiplinlerin anlamlı bir tutum farklılığına işaret etmediği görülmüştür. Diğer taraftan, kadın katılımcıların çoğunluğu (5/7) ilk sırada düzenleyicilerin sorumlu olduğu yönünde görüş bildirmişlerdir. Erkek katılımcıların yanıtlarında ise düzenleyiciler, geliştiriciler ve veri sağlayıcılar arasında dengeli bir dağılım görülmektedir. YZ'nin özerklik seviyesine (1 katılımcı), sorunun kaynağına (6 katılımcı) göre sorumluluğun kime ait olduğunun değişkenlik gösterebileceğine ilişkin katılımcı ifadeleri de bulunmaktadır. Katılımcıların çoğunluğunun (9/16), kullanıcıları en az sorumlu aktör olarak değerlendirdiği de belirtilmelidir.

“Dağıtılmış sorumluluk beklentisi” olarak nitelendirilebilecek bir katılımcının ifadesi göze çarpmaktadır:

“..sağlık alanında da kullanılıyor bazı hastalıkların hücrelerini tespit etmek için.. eğer yapay zekâ yanlış buluyorsa sonuçları, belki yapay zekâyı geliştirenler o alanda sorumlu olabilir.. Başka alanlarda.. mesela kişisel verilerin paylaşılmasından konuştuk, bunda da yasa koyucuların sorumlu olabileceğini düşünüyorum.. hatta hani şey de çok önemli, tek bir sorumlu bulmanın da zor olduğunu düşünüyorum.. hastalıklı hücre bulma örneğinde de belki geliştiricilere verilen imkanlarla da ilgili olabilir, o imkanları da atıyorum onların şirketleri sağlıyordur, bu sefer onlar sorumlu olabilir..” (K4-FTM)

Sorumluluğun belirlenmesinin önem arz ettiği kritik kullanım bağlamlarına ilişkin ise, akademik disiplin tutum farklılığı teşkil etmemekle birlikte, 6'sı erkek olmak üzere 7 katılımcı yargı sistemlerinde, 3 erkek ve 3 kadın katılımcı sağlık endüstrisinde, 2 erkek katılımcı ulusal güvenlik ile ilişkilendirilebilecek hususlarda, 1 erkek ve 1 kadın katılımcı finansal hizmetler sektöründe, 1 kadın katılımcı askeri teknolojiye, 1 erkek katılımcı da eğitim sisteminde ve ulaşımda YZ kullanımı söz konusu olduğunda sorumluluğun belirlenmesinin önemini vurgulamıştır. İlgili ifade alıntılarına aşağıda yer verilmektedir.

“Yargı sistemleri” üzerine;

“..hukuk olabilir yine, adalet olabilir.. orada da yine en nihayetinde çok telafisi olmayan bir karara bağlanabilir konu yanlış bir şekilde.. örneğin suçsuz olduğunuz halde belki de bir şekilde deliller sizi suçlu gösteriyordur ve bu konunun bir şekilde incelenip, detaylı bir şekilde araştırılıp karara bağlanması

gereken yerde yapay zeka direkt, net bir şekilde olaya baktığı için “Tamam bu delil var, bu delil var.. olay yerinde bu parmak izi var, bu suçludur..” diyebilir ve o bağlamda siz cezaevine girebilirsiniz.

..en azından şu aşamada.. sadece insan da olmamalı, sadece yapay zeka da olmamalı, ikisinin birlikte birbirlerini destekleyici bir şekilde beraber.. ve aslında burada da şöyle önemli bir nokta var; aralarında bir konsensus yani bir uzlaşma olmalı, o hukuk çalışanı, ilgili kişi de yapay zeka da aynı noktada buluşuyorsa bu bence çok iyi bir şey, bir dava görüldü atıyorum.. gerçekten iki taraf da aynı sonuca varıyorsa çok rahat hissettirebilir.. ” (E7-FTM)

“..insan yaşantısını ele alan durumlar aslında.. sağlığı.. özgürlüğü.. yaşam tipi.. bu konularda karar veren noktalarda.. bu şimdi hukuk ve tıp oluyor sadece, o konularda kimin geliştirdiği ya da sorumluluğun kimde olduğu önemlidir..” (K2-İTB)

“Sağlık endüstrisi” üzerine;

“..sağlık uygulamaları, insan hayatı çünkü.. hukuk.. çünkü hukuk çökerse sistem çöker..” (E5-FTM)

“..sağlık sektörü geliyor aklıma.. özellikle mesela ilaç sektörü olabilir, yani o konuda şeffaf davranılmadıysa ya da herhangi bir şekilde yapay zekânın yol açtığı bir hata varsa kesinlikle o şirketin sorumlu tutulması gerekiyor çünkü bu telafisi olan ya da belki karşılayabileceğiniz bir durum olmayabilir.. bir hastaya örneğin bir ilaç tedavisi uygulanacak ve çok hayati bir önem taşıyor ya da bir ameliyat yapılacak örneğin uzaktan bağlanarak farklı ülkelerden doktorların ameliyat yaptığı sistemler var ya da işte çeşitli tanı, teşhis koyma sistemleri var.. bunların insan sağlığına dokunduğu noktalarda.. her ne kadar yapay zekâ işin içinde olsa dahi kesinlikle insan faktörünün de sorumlu tutulması gerektiğini düşünüyorum, çünkü böyle bir konuda, yani kaza yaptınız, arabanızın belki yenisini alabilirsiniz ya da onu yaptırabilirsiniz veya bir şekilde uzlaşmaya varabilirsiniz, ama sağlığını kaybettiğinizde ya da yakınlarınızın, sevdiğinizin sağlığına herhangi bir şekilde dokunduğunda tabii ki de bir muhatap, bir sorumlu aramak zorundasınız..” (E7-FTM)

“Ulusal güvenlik” üzerine;

“..ülkeler örneğini verdik ya mesela, büyük bir şey etkiliyorsa bir toplumu, bir ülkeyi.. onun sorumlusunun kim olduğunun belli olması gerekiyor.. bir saldırı gibi dedik, olmayan belki de.. kötü amaçlı bir bilgi ama büyük bir soruna yol açıyor ülke adına.. onun nereden kaynaklandığının bilinmesi gerekiyor, kim tarafından..

..sorumluluk bence aranmalı yayınlandığı mecrada da, sonuçta onu yayınlıyorsan sen de kontrol etmelisin, hani kötü bir bilgi yayılıyor ve sen bunun farkındaysan o zaman kendi mecranda tutmamalısın.. bunu aslında gazeteler gibi düşünebiliriz, yalan yanlış yazan bir gazete yazarı varsa onu kendi gazetende tutmazsın yani.. ” (E6-FTM)

“..Savunma Bakanlığı olabilir, herhangi bir devletin..” (E9-İTB)

“Finansal hizmetler sektörü” üzerine;

“..hukuk boyutuna giderse karar almalarda %100 uygun olması lazım.. sonra.. yüksek piyasalı işlemlerde %100 olması lazım finans sektöründe, çünkü günlük hacimlerde çok yüksek paralar dönüyor.. sağlık alanında %100 olması lazım, çünkü eğer sağlık alanındaki çoğu cihaz yapay zekaya geçerse ölümlü sonuçlara bile gidebilir yanlış bir kullanımda..” (E8-İTB)

“..mesela bir muhasebeci de yani.. maddi bir durum söz konusu olduğu için eğer tedarikçide bir problem varsa ya da geliştiricide bir problem varsa bu kişilerin mağdurların mağduriyetini gidermeleri gerekiyor.. bu noktada bence çok önemli ya da kullanıcı suçluysa bir şeyleri talep edemiyor olması gerekir, bu çünkü geliştiriciyi de bu sefer zora sokmak olur..” (K5-FTM)

“Askeri teknoloji” üzerine;

“..savaşlar için geliştirilen yapay zekalar, yani günümüzün en kritik şeyleri o şu an..” (K1-FTM)

“Eğitim sistemi ve ulaşım” üzerine;

“..adalet, çünkü kararlar bir insanın canına bile mal olabilir, ulaşım da olabilir, eğitim de olabilir..” (E1-FTM)

Tüm sorunlardan arınmış bir YZ sisteminin olanaklılığı hakkında katılımcı görüşleri değişmekte iken en fazla fikir ortaklığı “tüm boyutların bir arada karşılanmasının mümkün olmadığı” üzerinedir, 6 katılımcı bu doğrultuda görüş bildirmiştir ve bu katılımcılardan 5’inin kadın olması dikkat çekmektedir. İlgili ifadelerden bazıları aşağıda mevcuttur:

“..şeffaflığa gelince aslında hepsinde de var diyemem, tamamen şeffaf olan bir yapay zeka şu anda olduğunu düşünmüyorum.. Gizliliğin de tamamen korunduğu bir uygulama şu an olduğunu sanmıyorum, yoktur yani.. çünkü her türlü.. bu yapay zekayı üreten insanlar bir şekilde bunlara ulaşıyorlar.. yani gizliliği ele alırsam, kullanan kişinin kişisel verilerine bir şekilde ulaşıyorlardır.. ya da benim bilmediğim ama onların bildiği durumlar vardır ve bu şeffaflığa aykırıdır mesela.. üretilebileceğini sanmıyorum..” (K2-İTB)

“..bunu da geliştiren bir insan olduğu için bunların hiçbirinin %100 uygulanabilir özellikler olduğunu düşünmüyorum açıkçası, bir noktada insan bakış açısı, insanın kendi hayatından çıkardığı tecrübeler ve işte deneyimleri.. merhameti belki.. bir sürü unsur olduğunu düşünüyorum etkileyecek.. %100 olmadığına eminim.. bu beş özelliği de sağlayacak bir yazılım bence geliştirilemez.. on üzerinden puanlasam yedi..” (K3-FTM)

“..böyle bir şeyin mümkün olduğunu düşünmüyorum.. yapay zeka da insanların elinden çıkıyor sonuçta.. insanlar da bu özelliklerin hepsine sahip olamadığı gibi bir yapay zekanın da sahip olamayacağını düşünüyorum.. hani birkaçı birlikte olabilir ama hepsine birden olabileceğini düşünmüyorum..” (K4-İTB)

Gözden çıkarılabilecek ilk boyutun hangisi olduğu sorusuna ilişkin görüş bildiren katılımcı kodlarının ve ilgili katılımcı ifadelerinin görülme sıklığının yer aldığı Tablo 3.11 aşağıda bulunmaktadır. K4-FTM, kullanım amacına göre gözden çıkarılabileceği ilkelerin “Adalet” ve “Şeffaflık” olarak değişeceğini ifade etmiştir, bu nedenle Tablo 3.11’de, katılımcı koduna iki boyut altında da yer verilmiştir.

Tablo 3.11. Gözden Çıkarılan İlk Boyuta İlişkin Görüş Bildiren Katılımcılar

Gözden Çıkarılan İlk Boyut	Görüş Bildiren Katılımcı	Frekans
Şeffaflık	E3-İTB, E4-İTB, K1-FTM, K2-İTB, K3-FTM, K4-FTM, K5-FTM, K7-İTB	8
Gizlilik	E1-FTM, E5-FTM, E6-FTM, K5-FTM	4
Sorumluluk	E2-FTM, E8-İTB, K6-İTB	3
Adalet	E9-İTB, K4-FTM	2
Kötü niyetli olmama	-	-

*E: Erkek katılımcı / K: Kadın katılımcı

**FTM: Fen Bilimleri, Teknoloji ve Mühendislik disiplinlerinden katılımcılar, İTB: İnsan ve Toplum Bilimleri disiplinlerinden katılımcılar

Katılımcılarının yarısının ilk sırada “Şeffaflık” boyutundan feragat edeceği ve bu katılımcılarının çoğunluğunu kadınların oluşturduğu göze çarpmaktadır (6/8). Dahası iki katılımcı da (E2-FTM ve E5-FTM) ikinci sırada bu boyutu belirtmişlerdir ki bu, diğer boyutlar ile karşılaştırıldığında yüksek bir farka işaret etmektedir.

“Şeffaflık” boyutunu, “Gizlilik” boyutu takip etmektedir, bu boyutun ihmal edilebileceğini düşünen katılımcıların erkek ağırlıklı olmak üzere tümünün FTM disiplinlerinden öğrenciler olduğu da görülmektedir.

Kadın katılımcılar, uygulama şeffaf olmasa da kişisel verilerinin gizli tutuluyor olmasını tercih edeceklerini ya da sorumluluk boyutuyla bunun telafi edilebileceğini açıkça belirtmişlerdir. Bu bulgu, aynı zamanda, kantitatif araştırma aşamasında saptanan, kadın katılımcıların “Gizlilik” boyutunda erkek katılımcılara göre yüksek hassasiyetini de destekler

nitelikte bir bulgu olarak yorumlanmaya da açıktır. Öte yandan, “Şeffaflık” ilkesine riayet edilmediği takdirde diğer ilkelerin sağlandığından nasıl emin olabilecekleri sorulduğunda ise katılımcılar “*umut etmek dışında elden bir şey gelmeyeceği*”, “*yine de diğer ilkelerin sağlanmasının öncelik arz ettiği*” benzeri ifadelerle başvurmuşlardır:

“Açıkçası emin olmam sanırım, özellikle gizliliğinden.. bence bir noktada kesişiyorlar, o da onu etkiliyor.. yine de.. şeffaf olmadığında verilen kararı en azından bir noktada kendim sorgulayıp kendi bakış açımla belki değerlendirebilirim..” (K3-FTM)

“Sanırım ilk şeffaflığı çıkarabilirim gözden.. Kötü niyetli olmamalı, çok ağır sonuçları olabilir.. adil olmalı, çünkü bunun nerede kullanıldığı çok fazla şeyi değiştirir.. yani bir mahkeme salonunda kararı yapay zeka veriyorsa orada çok ciddi bir adaletten bahsetmemiz gerekir, çok vazgeçilebilecek bir etken olduğunu düşünmüyorum. Gizliliğimin de mümkün olduğunca korunmasını tercih ederim, her ne kadar çok fazla güvenmesem de bu konuda, en iyi şekilde korunulabiliyor olduğunu bilmek isterim.. sorumluluk.. ben bir problem yaşıyorsam bana birileri bu konuyla ilgili hesap vermeli, “Bu bundan kaynaklandı” gibisinden.. şeffaflıkta.. hani arkasındaki nedeni bilmesem de belki o sorumluluk kısmında.. biraz daha tolere edilebilir gibi düşündüm.. tamam en başında bana “Şu, şu şu..” diye her şeyi vermiyor ama sonrasında, o sorumluluk kısmında ben problemime bir dönüt alabiliyorsam o da bazı şeyleri çözmeme yardımcı olabilir. Hani yerini biraz da olsa kapatabileceğim bir şey olduğunu gördüğüm için şeffaflıktan vazgeçtim..” (K6-İTB)

Bire bir görüşmelerde ayrıca, genel itibariyle, akademik disiplin fark etmeksizin kadın ve erkek öğrencilerin, ChatGPT başta olmak üzere Bard gibi büyük dil işleme modellerini günlük hayatlarına dahil ettikleri ve akademik çalışmalarında faydalandıkları tespit edilmiştir. Aynı zamanda, yeni nesil dijital içerik ve medya platformlarında sıklıkla gündeme getirilen Robot Sophia’nın geleneksel medyaya kıyasla zihinlerindeki YZ kavramını şekillendirme hususunda baskın geldiği de söylenebilir. Katılımcı ifadelerinden alıntılar aşağıda sunulmuştur:

“yani aslında çok başarılı bir şey ama bazı yerlerde tehlikeli olduğunu düşünüyorum açıkçası.. hani bazı cevapları.. gerçekten bir anda çok fazla arama motoruna erişebildiği için bilgi odaklı olduğu zaman gayet yararlı olduğunu düşünüyorum.. çok hızlı bir şekilde cevap verip birçok şeyi de açıklığa kavuşturuyor ama farklı derin mevzulara girdiğimiz takdirde.. biraz yani.. tehlikeli oluyor açıklamaları ya da insanlara görüşleri hakkında verilen cevaplarında biraz sıkıntı çıkacak gibi duruyor bana göre - gülüyor- ..ilerleyen zamanlarda yapay zekâyı tamamen hayatımıza koyarsak biraz hani görüş çakışması olacağı için insanlar buna tepki koyabilir ve çatışmaya yol açabilir diye düşünüyorum..” (K6-İTB)

“Gerçek bir insan gibi ya.. gerçekten.. korkutucu yani.. konuşmaları, tavırları bayağı ukala -gülüyor- baktığınız zaman, ürkütücü duruyor.. duyguları yok çünkü en başında, bu da tehlike arz edebilir..” (K7-İTB)

3.4. Tartışma

Salt nicel bir metodolojinin YZ etiği anlayışını ölçmek için en uygun yol olmayacağı varsayımıyla sıralı açıklayıcı karma araştırma stratejisinin benimsendiği bu çalışmada, çevrimiçi anket aracılığıyla toplanan nicel veriler ile üniversite öğrencilerinin, baskın olarak Z ve Y kuşaklarından katılımcıların, YZ etiğinin “Adalet”, “Şeffaflık”, “Kötü niyetli olmama”, “Gizlilik” ve “Sorumluluk” boyutlarına ilişkin olası cinsiyet ve akademik disiplin temelli tutum farklılıklarının ölçümlenmesi üzerine, bu farklılıkların kaynaklarına inebilmek ve boyutlar arası bağıntılara dair daha ayrıntılı bulgulara ulaşabilmek maksadıyla nitel veriler yarı yapılandırılmış görüşme yoluyla toplanmış ve içerik analizine tabi tutulmuştur.

Awad vd. (2018: 64), otonom araçlar özelinde makinelerin ahlaki kararları nasıl alacağına dair toplumsal endişeler ve makine davranışına rehberlik etmesi gereken etik ilkeler hakkındaki beklentileri ölçmek; bireysel etik tercihlerin, yaş, eğitim, cinsiyet, gelir, siyasi ve dini görüşler tarafından modüle edilip edilmediğini değerlendirmek üzere yürüttükleri çalışmalarında en kayda değer etkilerin cinsiyet ve dindarlık değişkenleri tarafından kaynaklandığını halihazırda ortaya koymuşlardır. Mevcut karma araştırma modelinde ise katılımcıların dini tercihlerine yönelik herhangi bir veri toplanmamıştır; odak, cinsiyet ve akademik temelli potansiyel tutum farklılıklarıdır ve bulgular bu doğrultuda kapsamlı çıkarımlara işaret etmektedir.

Öncelikle, bu çalışmada, çevrimiçi ankete gönüllü katılan üniversite öğrencilerinin yanıtları ile elde edilen nicel veri çözümlemeleri neticesinde, “Adalet”, “Kötü niyetli olmama” ve “Gizlilik” etik boyutlarına yönelik tutumların cinsiyete göre istatistiksel olarak anlamlı bir farklılık gösterdiği saptanmıştır. Kadın katılımcıların tutum düzeyleri bu boyutlarda erkek katılımcılara göre daha yüksektir. Bu bulgu, Jang vd.’nin (2022: 11652) Kore’de lisans öğrencilerine uyguladıkları anket sonuçları ile de örtüşmektedir. Bu başlık altında söz konusu üç boyuta dair bulguların içerik analizi neticeleriyle bir arada değerlendirilmesini takiben, yine nicel veri analizi sonucunda akademik disiplin temelli tutum farklılığının saptandığı “Şeffaflık” boyutuna ve farklılık analizlerinin anlamlı bir sonuç vermediği “Sorumluluk” boyutuna ilişkin bulgular da tematik analiz çıktıları kapsamında aşamalı olarak ele alınmaktadır.

3.4.1. Adalet

Önceki araştırmalara göre cinsiyetin, insanların adalet algıları üzerinde önemli bir etkisinin bulunduğu söylenebilir (Grgić Hlača vd., 2022: 3). YZ etiği yörüngesinde çalışmalar incelendiğinde de benzer bulgular söz konusudur. Örneğin kadınlar, makine öğrenimi tahmincilerine karşı erkeklerden daha hassas bir adalet algısına sahiptir (van Berkel vd., 2021: 11). Pierson da (2017: 2) çalışmasında, bir YZ algoritmasının doğruluğu azalsa dahi kadınların, adaletin sağlanmasının daha önemli olduğuna inandığını ortaya koymuştur. Yakın zamanda Pew Research Center (2022: 28) tarafından yürütülen bir araştırma, erkeklerin kadınlara kıyasla, tutarlı bir şekilde adil kararlar verebilen YZ programları tasarlanmasının mümkün olduğuna inanma olasılığının daha yüksek (%38'e karşı %22) ve kadınların ise bunun mümkünatından emin olmadıklarını söyleme olasılığının daha yüksek (%46'ya karşı %35) olduğunu açığa çıkarmıştır. Aynı araştırmaya göre kadınların farklı cinsiyetlerden insanları dahil etmenin “son derece” veya “çok önemli” olduğunu belirtme olasılığı da erkeklerden daha yüksektir (%67'ye karşı %58) (2022: 51). Dolayısıyla YZ programları tasarlanırken kapsayıcılığa erkeklerden daha fazla vurgu yapan ve adil kararlar verebilen YZ programlarının tasarlanmasının olanaklılığına erkeklerden daha az yanaşan kadınlar, belirli grupları orantısız bir şekilde etkileyebilecek YZ sistemlerinde ayrımcılık, eşitsizlik ve haksızlık algılayarak adaletsizlik ihtimaline karşı daha fazla hassasiyet sergiliyor olabilir.

Dahası, 16 katılımcı ile yürütülen bire bir yarı yapılandırılmış mülakatların çözümlenmesi sürecinde oluşturulan “**Adalet etkisi**” **ana teması** altında görüş sıklığının, diğer alt temalar ile karşılaştırıldığında açık ara en yüksek olduğu ve hem erkek katılımcıların hem kadın katılımcıların ifadelerinin sıklıkla tespit edildiği “Karar süreçleri” alt temasına ait ifadeler incelendiğinde de, kullanım bağlamına göre katılımcı tutumlarında değişkenlik görülse de bilhassa hukuki, tıbbi ve eğitime yönelik karar alma süreçlerinde insan yaşamını etkileyebilecek hakkaniyete aykırı “üstünkörü” kararlar konusunda çoğunluğun duyarlı olduğu saptanmıştır. Öte yandan cinsiyet temelli tutum farklılıklarının kaynağına yönelik emareler irdelendiğinde “**Adalet algısı**” **ana teması** kapsamında, adaletin tesis edilmesi hususunda erkek katılımcıların ağırlıklı olarak, özellikle yasal karar süreçlerinde YZ sistemlerinin “objektiflik” ve “rasyonellik” potansiyel niteliklerini göz önünde bulundurarak kazanımlara eğildiği görülmekte iken, kadın katılımcıların, algoritmik sistemlerin “vicdan” ve “empati” gibi “affektif ve kognitif insani nitelikler” açısından noksanlığını ileri sürerek bu sistemlerin ekseriyetle ceza hukuku gibi kritik pratiklere yönelik kullanımlarına ilişkin kaygı ve karşıtlık dile getirdiği açığa çıkarılmıştır. Bu, özellikle Amerikan kadınların, erkeklerden daha yüksek

bir oranla (%43'e karşı %27) YZ programlarının insanlar için önemli yaşam kararları alabilmesi durumunda çok veya biraz endişeli olacaklarını saptadıkları araştırma bulgularıyla da tutarlıdır (Pew Research Center, 2022: 24). Bilhassa kadın katılımcı ifadelerinde vurgulanan "empati" kavramı, "Bir başkasının içsel referans çerçevesini doğru bir şekilde algılamak" olarak (Gold ve Rogers, 1995: 79) tam anlamıyla insan olmanın merkezinde konumlandırılır (Zahn-Waxler ve Radke-Yarrow, 1990: 108) ve ampirik araştırmacılar, kadınların genellikle erkeklerden daha yüksek empati seviyelerine sahip olduğunu da ifade etmektedir (Toussaint ve Webb, 2005: 675-682). Dolayısıyla, kadın katılımcıların mevcut eğiliminin, ahlaki gelişim ve adalet için kritik öneme sahip olan ve toplumsal uyum ve birlik için bir katalizör görevi gören empati (Toussaint ve Webb, 2005: 674) gibi duygusal ve bilişsel bileşenlere sahip mefhumları gözetmelerinden kaynaklandığı ileri sürülebilir. Bu argümanı destekler bir başka bulgu olarak, "Adalet etkisi" ana temasına dönecek olursak, "Savaş ve çatışma koşulları – anomi" alt temasında da kadın katılımcıların endişelerine yönelik ifadelerin sayıca üstün olduğu görülmektedir. Burada "anomi", toplumda mevcut kuralların tesirini yitirmesi, sosyal entegrasyonda belirsizlik ve bu neticede anormalliklerin tetiklenmesi olarak açıklanabilir.

Buna karşılık, nicel çözümlemeler kadın katılımcıların erkek katılımcılara göre bu boyutta daha fazla hassasiyet gösterdiğine işaret ediyor olmakla beraber "Adalet algısı" ana teması dahilinde "Yasa merkezlik" alt teması ile paralel olarak bir teftiş mekanizması ve itiraz mercii gereğinin altını çizen katılımcıların çoğunlukla erkek katılımcılar olduğunu da belirtmek gerekir. Kadın katılımcılar "Sorumluluk" boyutuna yönelik mülakat sorularına yanıtlarında regülatörlerin sorumluluğunu birincil gördüklerini belirtmiş olsalar da bu katılımcıların "Adalet" boyutu üzerine konuşurken bir teftiş mekanizması ve itiraz mercii ya da düzenleme, politika ve yönetim gereğine pek atıfta bulunmamış olmaları dikkat çekmektedir. Bu noktada "Gizlilik algısı" ana teması kapsamında frekans değeri ikinci sırada bulunan "Toplumsal uyum baskısı" alt temasında irdelenen ifadelerin çoğunluğunun kadın katılımcılara ait olduğunu hatırlatmak gerekir (5/7). "Toplumsal uyum baskısı" faktörünün, YZ aracılı adaletsizlik ve haksızlık neticesinde kadınların YZ sistemlerini kabul ve kullanımında ve gerektiğinde bir itiraz merciiine başvurup başvurmamasında potansiyel bir etkiye sahip olup olmadığı üzerine araştırmalar faydalı olabilir. Yine de elde edilen "Yasa merkezlik" alt teması ile adalet boyutunun hesap verebilirlik ve sorumluluk boyutlarıyla geçirgenliği de belirginlik kazanmaktadır. Gillespie vd. (2023: 26) de araştırmalarında, YZ kullanımını güvenli ve etik hale getirmek için düzenleme, politika ve yönetim eksikliğine ilişkin kaygıları ortaya koymuştur. Söz konusu çalışma Güney Kore, Finlandiya, Güney Afrika, Singapur, Brezilya,

Estonya, Avustralya, Fransa, İsrail, Kanada, ABD, İngiltere, Hollanda, Japonya, Çin, Almanya ve Hindistan üzerine karşılaştırmalı bir analiz sunmakta, ancak Türkiye’ye yönelik herhangi bir veri içermemektedir. Bu sebeple yarı yapılandırılmış görüşmeler sonucu katılımcıların bu husus doğrultusunda dışavurumları genelleme kısıtı teşkil etse de Türkiye’de öğrenim gören üniversite öğrencilerinin görüşlerini yansıttığı olması açısından önem arz etmektedir.

“Adalet etkisi” kapsamında oluşturulan “Sorgulama ve inisiyatif yitimi” alt teması da genel itibariyle Gillespie vd.’nin (2023: 26) bulgularını doğrular niteliktedir; zira araştırmaları dahilinde bir üst paragrafta belirtilen ülkelerdeki katılımcıların “İnsan kontrolünün ve eylemliliğinin kaybı ve karar vermede insan muhakemesinin kaybı, istenmeyen sonuçlara yol açmaktadır (YZ’nin 'yönetimi devralması' dahil).” yönünde endişe duyduklarını açığa çıkarmışlardır. Ancak Tablo 3.7’de sunulduğu üzere, mevcut çalışmada sadece iki erkek katılımcı, insanın düşünsel olarak pasifleşmesine ve sorgulama kabiliyetinin körelmesine yönelik tedirginliğini doğrudan dile getirmiştir. “Algoritma takdiri” ya da aşırı güven olarak nitelendirilen durumda kullanıcının sistem çıktılarını aktif olarak izlemeyi bıraktığı veya her tavsiyeyi sorgusuz sualsiz uyguladığı “aylaklık” halini (Zerilli vd., 2022: 2) işaret eden dolaylı bir tedirginliğin karşılık bulduğu katılımcı ifadelerine rastlanmıştır, ilgili alıntılara “Bulgular” bölümünde “3.3.6.1.2. Adalet etkisi” başlığı altında yer verilmektedir. Aslında bu saptama, insan faktörleri mühendisliği ve insan-bilgisayar etkileşiminde uzun yıllardır kapsamlı bir şekilde araştırılan “otomasyon rehaveti” ve “otomasyon önyargısı” olarak karşımıza çıkan olgular istikametinde yorumlanmaya da açıktır (Wiener, 1981: 119; Mosier ve Skitka, 1996: 205; Parasuraman ve Manzey, 2010: 382-391).

Bu boyuta yönelik değinilmesi gereken son bulgu da YZ aracılı adaletsizlik ve haksızlık neticesinde “YZ’ye güven kaybı”nın, kadın katılımcılardan ziyade erkek katılımcılar tarafından dile getirilmiş olmasıdır. Burada bir güven kaybı olasılığının, “Adalet algısı” ana teması altında “Geliştirici perspektifi” alt temasını oluşturan ifadeler ile ilişkisinden de bahsetmek gerekir. Zira, güven kavramı ilk etapta insanlar ve makineler arasında değil, kişiler arası bir tutum olarak örgütsel ve sosyal psikolojide de eğinilen araştırma konularındandır ve bu alanlarda güvenin, vekilin yetkinliği ve vekilin iyi niyet sergileme derecesi ya da güvenilen kişide kötü veya art niyet olmaması faktörlerinden etkilendiği anlaşılmaktadır (Zerilli vd., 2022: 2). Mülakatlarda da yapay ajanları geliştirenlerin ve piyasaya sürenlerin adil bir tutum sergileyeceğine inanma gereksinimini belli eden katılımcı ifadeleri öne çıkmıştır. Katılımcılar sıklıkla, geliştirenin adalet anlayışına ve niyetine vurgu yapmaktadırlar; zira sistem her zaman geliştiricisinin etik

tutumunu yansıtacaktır. Antrparantez, bu noktada, teknoloji uzmanlarının kendi değer ve inançlarına odaklanan, “siyasi niteliklere” sahip eserlerin (Winner 1986: 20) belirli kişilere, fikirlere ve olaylara diğerlerinden daha fazla görünürlük ve güç kazandırdığı iddiasına dayanan, "tasarımcıların ve üreticilerin, eserlerinin mükemmelliğinin değerlendirildiği kriterler dizisine değerleri bilinçli olarak nasıl dahil ettiklerini" sorgulayan erdem etiği yaklaşımını benimseyen eleştirel çalışmaların lüzumu da kendini göstermektedir (Flanagan vd., 2008: 322; Ananny, 2016: 95). Hagendorff (2020: 112) da etiğin, “sadece deontolojik olarak esinlenilen bir onay kutusu alıştırması” olarak değil, “kişilikleri geliştirme, tutumları değiştirme, sorumlulukları güçlendirme ve etik olmadığı konusunda uzlaşılan belirli eylemlerden kaçınmak için bir cesaret kazanma projesi” olarak anlaşılması gerektiğini savunur.

3.4.2. Kötü Niyetli Olmama

Öncelikle, tıp alanında yapılan araştırmalarda kötülük yapmamaya yönelik tutumlarda cinsiyet farkı olmadığını öne süren sayılı çalışma aksine, Kore örneğinde tutum farklılığı ortaya koyan Jang vd.’nin (2022) araştırmaları haricinde YZ veya otomatik sistemlerin kötüye kullanılmamasına yönelik tutumlarda cinsiyet farkı olup olmadığını analiz eden bir çalışmaya rastlanmamış olması sebebiyle, Jang vd.’nin (2022: 11652) tespitlerini destekleyen nicel bulgulara ulaşan bu araştırma önem kazanmaktadır. İzleyen nitel aşamada yarı yapılandırılmış görüşmelerin içerik çözümlemesi de bu boyuta ait “Kötü niyet algısı”, “Kötü niyet motivasyonu” ve “Kötü niyet çıktıları” olarak belirlenen ana temalar ile cinsiyet temelli tutum farklılıklarına yönelik daha detaylı keşifler ortaya çıkarmıştır.

İlk olarak göze çarpan “**Kötü niyet algısı**” ana temasında “otomasyon” alt temasına yönelik katılımcı ifadelerinin sadece iki kadın katılımcıya ait olduğudur. İş kaybı ve teknolojik işsizlik, Gillespie vd.’nin (2023: 26) on yedi ülkenin karşılaştırmalı analizini sundukları araştırmanın, YZ kullanımının bu ülkeler arasında algılanan risklerine ilişkin bulgularına göre de önde gelen kaygı unsurlarındandır, bilhassa Hindistan ve Güney Afrika’daki insanlar otomasyon nedeniyle iş kaybından endişe duymaktadır. Mevcut çalışmada ise kadınlar, Kaya vd.’nin (2024: 506) bulgularıyla da paralel olarak otomasyona yönelik kuşkularını dile getirirken erkek katılımcılar için otomasyonun işsizliğe dair bir endişe kaynağı olmaktan çok, insanın ağır işlerden arınıp insan gücünü ikâme eden makinelerin kontrolörü pozisyonuna geçme ihtimaline meyleden kanaatler sebebiyle daha olumlu karşılandığı söylenebilir:

“..çok büyük katkılarının da olacağını düşünüyorum.. şu an yapay zekânın nereye gidebileceğinin ucu bucağı yok.. eğitim sektörü, tıp sektörü.. belki de çoğu insanın gerçekten fiziki gücünü de makinelerle beraber artırarak çoğu insanın mesleki hayatını bile değiştirecek, belki de tamamen çalışan kesimden,

işçi kesiminden kontrol kesimine geçeceğiz; makinelerin kontrolünü, devriyesini, talimini yapan kesime geçeceğiz.. Ben o yüzden aslında sıcak bakıyorum çünkü ilerlemeyi hızla arttıracak, birkaç insanın yapamayacağı çoğu şeyi bilgi kapasitesi sayesinde çok hızlı.. en iyi film senaristlerinden daha iyi filmler yazabiliyorlar şu anda.. oran olarak hiçbir zaman %100 olmayacak ama ben başarı yönünden çok yükseklerle gideceğine adım kadar eminim.” (E8-İTB)

Aslında bulgular, erkeklerin genel olarak YZ teknolojilerine kadınlardan halihazırda daha olumlu yaklaştığını gösteren çeşitli çalışmalarla tutarlıdır (Ikkatai vd., 2022: 1589). ABD’de erkekler, insan gücünü artırma potansiyeli sebebiyle geliştirilmekte olan ve bazen sensör verilerini kullanan yerleşik bir YZ sistemi içeren “giyilebilir robot” olarak da adlandırılan “robotik dış iskeletlerin” el emeği için yaygın olarak kullanılmasının toplum için iyi bir fikir olduğunu (%46) belirtirken, %19’u kötü bir fikir olduğunu ifade etmiştir (Pew Research Center, 2022: 109). Buna karşılık, kadınların sadece %21’i bunun toplum için iyi bir fikir olacağını belirtirken, %29’u kötü bir fikir olduğunu düşünmektedir. Ayrıca aynı araştırmaya göre, tartışmalı ve alabildiğine kompleks bir mevzu olan beyin çipi implantlarının yaygın kullanımına yönelik heves sınırlı ve bunun toplum için iyi bir fikir olduğunu söyleyenlerin oranı düşük iken dahi bu fikre daha olumlu yaklaşanlar erkeklerdir; erkeklerin %20’si bunu iyi bir fikir olarak değerlendirirken kadınların sadece %6’sı bunu iyi bir fikir olarak gördüğünü beyan etmiştir (2022: 86). Ya da polemiğe açık bir başka başlık olarak bebeklerde hastalık riskini azaltmak için gen düzenlemenin kullanılmasını erkeklerin kötü bir fikirden ziyade (%29) iyi bir fikir (%36) olarak nitelendirme eğiliminde olduğu, kadınların ise tam tersi yönde tutum göstererek gen düzenlemenin bu amaçla yaygın olarak kullanılmasının iyi bir fikirden ziyade (%24) kötü bir fikir (%32) olacağını ifade ettiği görülmektedir (2022: 97). Almanya’da şirket temsilcileri ve öğrenciler arasında bir çevrimiçi anket çalışması da kadınların aksine erkeklerin YZ’de daha fazla fırsat görmekte olduğunu ortaya koymuştur (Franken ve Mauritz, 2020). Grassini ve Sævidl Ree’nin (2023: 5), Birleşik Krallık ve ABD ülkelerinden katılımcılar ile gerçekleştirdiği araştırmaları da erkeklerin YZ’yi kadınlardan daha yararlı ve genellikle daha olumlu olarak algıladıklarını destekler niteliktedir; erkek katılımcıların kadın katılımcılara göre YZ sistemleri için daha yüksek umutlar sergilediğini göstermektedir.

Yine “Kötü niyet algısı” teması altında “veri ve gizlilik ihlali”ne yönelik endişeler ise ağırlıklı olarak erkek katılımcılar tarafından dile getirilmişken, sadece iki kadın katılımcının bu doğrultuda tedirginliğini yansıtan doğrudan ifadesi bulunmaktadır. Diğer taraftan, bu boyuta dair yöneltilen ilk temel mülakat sorularında kadın katılımcıların “veri ve gizlilik ihlali”ne

değinen yanıtlarına az rastlanmış olsa da ilerleyen temel mülakat ve sondaj soruları aracılığıyla kötü niyetli YZ geliştirme ve kullanımının toplumsal çıktıları bağlamında oluşturulan “Kötü niyet çıktıları” ana teması bünyesinde “veri ve gizlilik ihlali” endişelerine erişilmiştir; bu alt temada görüş bildiren yedi katılımcının beşi kadındır. “Askeri YZ ve siber tehditler” alt temasında ise yine kadın katılımcı ifadeleri ağır basmaktadır. Bu alt tema ile paralellik izleyen **“Kötü niyet motivasyonu” ana teması** altında beliren “devletler arası rekabet ve jeopolitik çekişme” alt temasında da görüş bildiren kadın katılımcıların sayıca üstünlüğü göze çarpmaktadır. Öte yandan, hem erkek katılımcılar hem de kadın katılımcılar kötü niyetle YZ geliştirme, dağıtma ve kullanma motivasyonları olarak “maddi kazanç arzusu” ve “bireysel duygusal ve çıkar temelli dinamikler”i sıklıkla dile getirmişlerdir.

“Kötü niyet çıktıları” ana teması incelendiğinde hem erkek katılımcıların hem de kadın katılımcıların kaygılarını ifade ettiği “Bilgi manipülasyonu ve rıza imalatı”nın görüş frekansı en yüksek alt tema olarak sıvrıldığı görülmektedir. Birçok şirket algoritmalarının ayrıntılarını açıklamak istemese de belirli mesajları oluşturmak ve yaymak için karmaşık bir motivasyon ağı; içeriği güçlendirmek için sayısız içerik türü ve tekniğinden istifade etmektedir. Bu içeriği barındıran ve yeniden üreten sayısız platform ve ekranlar arasındaki iletişimin hızı sebebiyle, yanlış bilginin zarar vermek amacıyla bilerek paylaşılmasını (dezenformasyon) ve gerçek bilginin zarar vermek amacıyla, genellikle özel kalması gereken bilgilerin kamusal alana taşınmasını (mal-enformasyon) kapsayan enformasyon düzensizliği ve bununla bağlantılı filtre balonları, yankı odaları ve botların kullanılması olan dijital 'astroturfing' ve yeni propaganda biçimleri gibi potansiyel zorlukların doğrudan ve dolaylı etkilerine yönelik farkındalığın ve önüne geçme maksatlı proaktif çabaların artırılmasının önem kazandığı söylenebilir (Wardle ve Derakhshan, 2017: 5; Keller vd., 2019). Gillespie vd.’nin (2023: 26) araştırmaları da ele aldıkları ülkelerin neredeyse tümünde bireylerin en çok siber güvenlik riskleri, gizlilik ihlalleri ve bilgisayar korsanlığı başta olmak üzere, hizmet sağlayıcılar ve hükümetler tarafından kötüye kullanım dahil, manipülasyon, süveyans ve kontrol amacıyla zararlı kullanım konusunda kaygılandığını ortaya koymuştur.

Aynı ana tema bünyesinde, “ülkeler arası politik gerilim” ve “artan yapısal eşitsizlikler ve güç asimetrisi” alt temalarında da kadın katılımcıların baskın ifade sıklığı mevcuttur; dolayısıyla bu alt temalar da kadın katılımcıların bu boyuta yönelik hassasiyetinin kaynağına işaret eden diğer unsurlar olarak karşımıza çıkmaktadır. Burada cinsiyet farklılıklarının toplumlar arasında önemli ölçüde değişiklik göstermesi sebebiyle sabit farklılıkları öngören

herhangi bir teoriye dayanmaya şüpheli yaklaşma gereği hususunda bir parantez açmak da gerekir; zira kadınların ve erkeklerin görüş farklılıkları ekonomik büyüme, siyasi koşullar ve tutumsal temelli diğer koşulların bir fonksiyonu olarak belirmektedir (Eichenberg, 2007: 7). Cinsiyet temelli görüş farklılıkları özellikle daha yüksek refah seviyelerinde artmaktadır da; yine de farklılıkların yönü her zaman ve her yerde kadınların askeri güç kullanımını erkeklerden daha az desteklediği yönündedir (Eichenberg, 2007: 6-7). Mevcut veriler de bu ayrımı devam ettirmektedir. Yakın tarihli bir araştırma makalesinde de 2021 ve 2022 yıllarında toplanan iki odak grup tartışması verisine istinaden Türkiye’de dış politikada güç kullanımına yönelik kamuoyu algılarında cinsiyet farklılıkları olduğu ileri sürülmüştür; kadınlar dış politikada güç kullanımını erkeklere kıyasla daha az desteklemektedir ve bilhassa sınır güvenliği söz konusu olmadıkça güç kullanımını gereksiz görmektedir (Sayan, 2023: 1328). Söz konusu çalışmanın bulguları, güç kullanımına yönelik toplumsal tutumlarda cinsiyet farklılığına ilişkin yabancı yazın tarafından da desteklenmektedir (2023: 1342). “Artan yapısal eşitsizlikler ve güç asimetrisi” alt temasında görüş bildiren katılımcılar arasında, YZ sistemlerine erişim imkânı olan bireylerin işgücü piyasasında kazanımını artıracak ve uluslararası arenada halihazırda dijital rekabet gücü yüksek ülkelerin de askeri teknolojilere dayanan avantajlarını güçlendireceği kaygısı yaygındır. Bu sebeple “Kötü niyet algısı” ana temasında iki kadın katılımcının ifadelerinde karşılaşılan “otomasyon” tedirginliği de bu alt temalarla anlam kazanmaktadır. İşgücüne katılım oranı düşük ve bu piyasada dezavantajlı konumda bulunan kadınların yapısal eşitsizliklerin YZ sistemleri aracılığıyla şiddetlenmesi ihtimalini algılayarak daha duyarlı yanıtlar verdiği çıkarımında bulunulabilir. Bu alt temalar ile bağlantılı olarak değinilmesi gereken bir tespit de “kendi değerlerine uymayan yaklaşımları bastırırken belirli değerleri ve çıkarları dayatan iş modellerini besleyen düzenin reddi” ile toplumsal etik değerlerin ve ulusal güvenliğin gözetilmesi lehine “yerel YZ’ye öncelik verilmesi” yönünde dolaylı katılımcı ifadelerine rastlanılmış olmasıdır. Bu temenni, özellikle Afrika ülkelerinde sivil toplumun, yerel YZ uygulamalarının önceliklendirilmesi ve yurt dışından tedarik edilen sistemler için şeffaf bir tedarik sürecinin sağlanması çağrısı ile de kesişmektedir (Effoduh, 2021: 3). Nitekim benzer sosyal normları paylaştığı iddia edilebilecek Birleşik Krallık ve ABD’den bireylerin dahi YZ’ye yönelik tutumları arasında eşitsizlik söz konusudur. Grassini ve Sævilid Ree (2023: 5), saptanan eşitsizliğin bu ülkelerin eğitim sistemlerindeki farklılıklara, medyalarının YZ tasvirlerine, YZ ve ilgili teknolojilerin benimsenmesiyle ilgili diğer politika ve girişimlerine dayanılarak da açıklanabileceğini belirtse de, YZ sistemleri geliştirmede önde gelen şirketlerin çoğunun ABD merkezli olmasının Amerikan vatandaşlarının bu ürünlere yönelik güvenini ve genel algısını olumlu yönde etkileyebilme potansiyeline de vurgu yapmıştır

ve görüşünü ulusal veya vatansever ürünlere yönelik bu tür tercihlerin diğer ürünlerde de gözlemlendiği bulgusuyla kuvvetlendirmiştir. Nihayetinde bu paragrafta yorumlanan alt temaların varlığı aynı zamanda, bilhassa yerel yazında, güç ilişkilerini inceleyen bir etik yaklaşım ihtiyacına da işaret etmektedir; bu da bakım etiği ve feminist etik gibi güç ilişkileri araştırmalarına eğilen disiplinler için potansiyel araştırma istikametlerini göstermektedir (Raicu ve Kwan, 2023). YZ'ye ilişkin etik tartışmalar bu sistemlerin jeopolitik ve tekno-politik bağlamdaki rolüyle nadiren ilgileniyor olsa da “sosyo-materyal bir eser olarak, güç asimetrisini pekiştiren sembolik ve kültürel bir aracılık olarak” YZ, dekoloniyal ve kesişimsel feminizm türleri, Eleştirel Irk Çalışmaları, Eleştirel Medya Çalışmaları ve yeni materyalizmler gibi farklı disiplinlerin sunduğu çeşitli analitik ve ampirik araçlar aracılığıyla daha teferruatlı irdelenebilir (Ricaurte, 2022: 729).

Son olarak, “Adalet etkisi” ana temasında da tespit edilmiş ve ilgili paragrafta izah edilmiş olan, daha çok erkek katılımcı ifadelerinde rastlanan “YZ’ye güven kaybı” alt teması, “Kötü niyet çıktıları” ana teması kapsamında da elde edilmiştir ve sadece iki erkek katılımcının yanıtlarına yöneliktir. “Belirsizlik ve kırılabilirlik ile karakterize edilen bir durumda bir ajanın bireyin hedeflerine ulaşmasına yardımcı olacağına dair bir tutum” olarak tanımlanan makinelere güven, hususiyetle YZ arayüzleri tasarlayanlar için önemli bir dizi sorun teşkil eder (Zerilli vd., 2022: 1). Güven kaybı olasılığının hem “Adalet” boyutunda hem de bu boyut altında karşılık bulması, özellikle katılımcıların ifade sıklığının yoğunlaştığı alt temaların araştırmacılar, geliştiriciler, regülatörler gibi ilgili paydaşlar tarafından etraflıca dikkate alınmasını gerektirmektedir.

3.4.3. Gizlilik

Çevrimiçi gizlilik, anonimlik ve kendini ifşa etmeye yönelik kullanıcı tutumları literatürde yaygın olarak araştırılan konulardır ve çoğu birey tamamen anonim olmanın mümkün olduğuna inanmamakta olsa da bu araştırmaların büyük kısmı, çoğunluğun çevrimiçi gizliliklerine ve anonimliklerine yönelik tehditler hususunda ‘endişeli olmak’ ile ‘çok endişeli olmak’ arasında tutum sergilediği ve mahremiyetlerini korumak için harekete geçmeye istekli oldukları sonucuna varmaktadır (Weinberger vd., 2017: 5). Belirlenen ana temaların yorumlanmasına geçmeden, anonim olmanın ve gizliliği muhafaza etmenin tamamen mümkün olmadığı görüşünün bu çalışma dahilinde yürütülen yarı yapılandırılmış görüşmelerde de hem erkek katılımcıların hem de kadın katılımcıların yanıtlarında karşılık bulunduğunu vurgulamak gerekir. Öyle ki, çözümlemeler neticesinde, katılımcılara mülakat sonuna yaklaşıırken “ilk

feragat edecekleri araştırma kapsamındaki YZ etiği boyutunun hangisi olacağı”na ilişkin soru yöneltildiğinde “Gizlilik” boyutunun ihmal edilebilirliğini, halihazırda bu ilkeye riayet edilmediğine dair kuvvetli bir varsayımına dayanarak daha makul bulan katılımcıların yanıt frekansının ikinci sırada yer aldığı görülmüştür. Bu boyut ihmal edilebilirlik frekansı en yüksek olan “Şeffaflık” boyutundan hemen sonra gelmektedir. Ayrıca nicel veri analizinde saptanan gizliliğe yönelik cinsiyet temelli tutum farkını da destekleyen bir tespit olarak, “Gizlilik” boyutunun ihmal edilebileceği yanıtını veren bu katılımcılar ağırlıkla erkek öğrencilerdir (3/4). Kadın katılımcıların yüksek bir frekansla “Şeffaflık” boyutunun ilk sırada gözden çıkarılabilirliğini dile getirmiş olmalarının sebebi ise bir uygulamanın şeffaf olmasındansa öncelikli olarak kişisel verilerinin gizli tutuluyor olmasını tercih etmeleridir, bunu net bir biçimde dile getirmişlerdir.

Kadınların bilgi gizliliği konusunda daha endişeli olduğunu, özellikle de sosyal ağlarda gizlilikleri konusunda erkeklerden daha fazla tedirginlik duyduklarını ortaya koyan çalışmalar mevcuttur. McGill ve Thompson’ın (2021: 859) araştırmalarına göre kadınlar ortalama olarak güvenlik tehditlerine karşı kendilerini daha savunmasız hissetmemelerine rağmen, etkilerin kendileri için erkeklerden daha kötü olacağına inanmaktadırlar. Erkekler kâr amacı güderek risk kabul etmeye istekliyken, kadınların kişisel bilgilerinin olası ifşasına yönelik kaygıları nedeniyle elektronik ticarete katılmaktan dahi caydığı ya da kadınların sosyal ağlar ortamında telefon numaraları ve ev adresleri gibi daha hassas kişisel verilerini ifşa etmeme eğiliminde olduğu yönünde bulgular da örnek verilebilir (Weinberger vd., 2017: 6). Hatta bu eğilim, mevcut görüşmelerde kadın katılımcıların yanıtlarında da gözlemlenmiştir; sıklıkla, telefon numaraları ya da adres bilgilerinden ziyade zorunluluk dahilinde e-posta adreslerini paylaşmayı yeğleyeceklerini belirtmişlerdir.

Ancak bazı katılımcı ifadeleri, son yıllarda pek çok alanda YZ destekli uygulamaların yaygınlaşarak günlük hayatın alışılmışlığına karışması ile ve özellikle son aylarda rahatsızlık sebebi olan veri sızıntılarını da takiben, bireylerin kişisel verilerinin gizliliğini koruma çabası konusunda daha umursamaz olmaya yatkın olabileceğini de düşündürmektedir ve bu olası bir çözüme yönelmektense bir nevi öğrenilmiş çaresizlik belirtisi olarak değerlendirilebilir. Bu da Cho’nun (2022: 150-167), internet kullanıcılarına dair ileri sürdüğü “mahremiyet çaresizliği”nin “gizlilik risklerine ilişkin önceki deneyimler” ve “algılanan savunmasızlık” teorik faktörleri tarafından yordanabileceği ve kullanıcıların, kaçınılmaz başarısızlık beklentisiyle mahremiyet yönetiminden vazgeçmeleri halinde söz konusu çaresizliğin herhangi

bir politika veya teknolojik çözümün uygulanmasında önemli bir psikolojik engel teşkil edebileceği tespitlerini destekler bir çıkarım olarak nitelendirilmeye açıktır.

“Verilerimin gizliliği konusunda eskiden çok dikkat etmeye çalışıyordum ama bir noktada artık sızdırılma haberlerini gördükten sonra ben de çok dikkat etmemeye başladım. Mesela önceden banka kartı bilgilerimi verme konusunda falan daha hassas davranırdım, ‘kod gelmeden banka kartımdan çekmeme özelliği’ gibi özelliklerim açıkken.. açıkçası şu an çok da önemsemiyorum, çünkü aslında bir şekilde zaten bilgilerimin birilerinin elinde olduğuna emin gibiyim; o yüzden artık gizlilik boyutunu çok da önemsemiyorum kendi hayatımda öyle.. Yani WhatsApp’ten tanımadığım, artı sayılarının hangi ülkeye ait olduğunu bile bilmediğim numaralardan mesajlar alıyorum.. numaram sızdırıldığına göre numaramı verdiğim, numaramı bilen bir uygulamadan olabilir..”

..çok net bildiğim bir uygulamaya tabii ki de bilgilerimi vermiyorum, ama onun dışında, soru işaretleri olan uygulamalarda çok da hassas davranmıyorum uygulamaya karşı.. Yemeksepeti ile ilgili bir ara banka kredi kartı verilerinin sızdırılmasına yönelik haberler çıkmıştı ve mesela ben hala kullanmaya devam ediyorum aslında...” (K3-FTM)

“..tabii ki böyle söyleyince insan herhangi bir bilgisini çok paylaşmak istemiyor hiçbir yerde.. ama hepimiz uygulamalar kullandığımız için alıştığımız bazı bilgilerimizi vermeye.. ismimiz olsun, telefon numaramız ve mail zaten her uygulamada var; bunları vermekten çok çekinmiyorum artık.. en başlarda daha bu kadar telefon, internet, uygulama kullanmaya alışmamışken bir çekiniyordu insan.. hatta internet alışverişi de çok şey geliyordu başta kredi kartı bilgisi paylaşıldığı için.. ama zamanla çok yaygın bir hale geldiğinden kredi kartı bilgisinin paylaşılması bile çok normalleşti.. çekindiğim şey olmuyor artık..” (K4-FTM)

Sasse vd. (2001: 127) tarafından güvenlik algıları üzerine yapılan araştırma, kullanıcıların bilgilerinin başkaları için değerli olmadığını düşünme eğiliminde olduklarını ve bu nedenle hedef alınacak kadar önemli olmadıklarını varsaydıklarını ortaya koymuştur. Görüşmelerde bu doğrultuda katılımcı ifadelerine de rastlanmıştır:

*“..banka ve bu tarz veriler.. onun dışında hiçbir verimi saklamıyorum, hatta üç-dört cloud ortamında yedeklidir, bilgisayarda yedeklidir.. ona izin vermeyeyim, bunu almasın vs. gibi hiçbir takıntım yok.. **önemli bir insan olduğumu da düşünmüyorum**, o yüzden verilerimin de önemli olduğunu düşünmüyorum..” (E5-FTM)*

*“..aslında tüm bilgilerin gizliliği çok önemli; yani şunu paylaşırsam olmaz, bunu paylaşırsam olur diye bir şey söz konusu olmamalı.. Ama çok, yani benim düşüncem bu, saçma gelir veya gelmez insanlara, **önemli bir kişi olmadığım için, sıradan bir vatandaş olduğum için**.. elimden geldiğince verilerimin gizliliğini sağlamaya çalışırım ama çalınsa çok bir şey olacağını düşünmüyorum çünkü çok önemli bir*

kişi değilim.. şöyle bir kişi değilim, bir devlet yöneticisiyle benim aramdaki önemlilik seviyesi anlamında..” (E9-İTB)

Diğer taraftan kadın kullanıcıların gizliliğe yönelik bir güvenlik ihlalini erkeklerden daha riskli olarak değerlendirme eğiliminde olmaları, bilgilerinin belli hususlarda daha fazla tehlike arz ettiğini düşünmelerinden ya da saldırıya uğrama olasılığını daha yüksek görmelerinden kaynaklanıyor olabilir. Bilhassa kişisel fotoğraflar söz konusu olduğunda kadın katılımcıların ifadeleri bu argümanı desteklemektedir:

“kendi sosyal medyamdan örnek verirsem; arkadaşlarımı paylaşıyorum ve onlarla ilgili aslında spesifik olarak bir veri paylaşmıyorum ama görsellerini sonuçta paylaşıyorum.. o görsellerden art niyetli insanlar tabii ki onlarla ilgili özel çıkarımlar yapabilir.. birisi belki takıntı haline getirip.. sarışın olmasına bile takabilir.. çok uç noktalarda insanlar var, kimin neyi ne kadar önemsediyini bilemiyorum..

..gizliliğe önem veren bir uygulamayı kullanma konusunda daha rahat olurum tabii ki.. ama işte dediğim gibi ben çok da aslında bilgilerimin gizli tutulduğunu düşünen tarafta değilim, biraz pessimist kalıyorum sanırım bu konuda -güliyor-” (K3-FTM)

“Fotoğraflarımın paylaşılmasına çok bakarım. ..mesela Instagram’ın ‘profilin açık veya gizli olması’ seçeneği vardır.. hesabımı herhangi bir durumda açık yapmam. ..gerçi hani herkes şöyle düşünüyor; “Kim benim hesabıma girecek..” ama olabilir, çünkü açık olduğu zaman fotoğraf kullanılabilir hale geliyor artık senin takipçin olmayan insanlar tarafından da.. belki uygunsuz fotoğraflara silüetiniz yerleştirilebilir.. başka bir, bulunmak istemeyeceğiniz bir fotoğrafa sizi koyabilir, silüetinizi, suratınızı..” (K4-FTM)

“..fotoğrafların da paylaşılmaması gerekiyormuş, diye duydum.. Instragram, sosyal medya.. gözlerimizden DNA’larımızı alabildiğini duydum bir yerden yapay zekânın, yani gözlerimizden bir kopyamızı kaydettiğini.. aslında fotoğraflarımı paylaşmak istemiyorum ama çok geçir diye düşünüyorum bunun için, o yüzden kaldırma gereği de duymuyorum..” (K7-İTB)

Mevcut saptamalarla birlikte ilginç olan, nispeten çoğunluk gizliliğe riayet edilmediğine inanma eğiliminde olsa da bilhassa kadın katılımcıların endişelerinin önlem almaya yönelik doğrudan bir davranışa dönüşmeyebilmesidir. Gizlilik tutumları ve davranışları arasındaki bu tür bir eşitsizlik, yazında “mahremiyet paradoksu” olarak adlandırılmıştır (Weinberger vd., 2017: 6). Burada da “**Gizlilik algısı**” **ana teması** altında frekans değeri ikinci sırada bulunan “toplumsal uyum baskısı” alt temasında görüş bildiren katılımcıların kadın ağırlıklı olması (5/7), gizlilik tutumları ve davranışları arasındaki eşitsizlik ile ilişkilendirilebilir. Katılımcılar

ekseriyetle, güvenmeme eğiliminde olsalar dahi bir noktada topluma ayak uydurabilmek ve işlerini yürütebilmek için uygulamaları kullanmaya devam etme eğilimlerini dile getirmektedirler:

“...sosyal medya uygulamalarında bilgilerin güvenilirliği çok fazla tartışma konusu haline geliyor ama yine de bir şekilde kullanıyoruz.. en basitinden okulla ilgili bilgi ağı sosyal medya, WhatsApp, ya da onun dışındaki platformlar.. bunların hepsinden ortak bir şekilde aslında bilgi ediniyoruz; doğrudan bir öğrenci gidip bir bilgi edinmiş olabilir ve bunu bize mail yoluyla bildirmiyor; çok daha basit, çok daha günlük kullanılan uygulamalar üzerinden bildiriyor ve bir noktada aslında kullanmak zorunda kalıyoruz haberdar olmak için..” (K5-FTM)

McGill ve Thompson’ın (2021: 858-860) bulgularıyla benzer bir şekilde, bu araştırma verileri de mahremiyeti koruma davranışının erkeklerde kadınlardan daha sık görüldüğünü ortaya koymaktadır. **“Gizlilik toleransı” ana teması** kapsamında “Kullanım azaltma ve alternatif yaklaşımlar” alt teması, erkek katılımcıların, arama geçmişini takip etmeyen bir arama motoru kullanma veya gizlilik artıran bir eklenti kullanma, sanal özel ağ (VPN) kullanma, kendisi hakkında yanlış veya yanıltıcı bilgi verme, anonim hesap ya da alternatif bir kullanıcı adı / e-posta adresi kullanma, uçtan uca şifreli uygulamaları tercih etme gibi davranışlara ilişkin ifadelerini içermektedir. Dolayısıyla bireylerin, bilhassa kadınların, kişisel bilgi gizliliğini korumaya yönelik harekete geçme eğilimini hangi faktörlerin etkilediğinin ya da artırdığının gelecek araştırmalar ile aydınlatılması yerinde olacaktır. Ek olarak, bu alt temada hem erkek katılımcıların hem de kadın katılımcıların ve toplam katılımcı sayısının çoğunluğunun genel itibarıyla veri gizliliği ve güvenliğinin karşılanmadığı durumlarda söz konusu uygulamanın kullanımına son verme, uygulamayı kullanırken bıraktığı veri izini ya da ‘dijital ayak izi’ni en aza indirme, muadil rakip ürünlere yönelme seçeneklerini göz önünde bulundurabilecekleri de saptanmıştır. Güvensizlik davranışıyla ilintili, “Algoritma isteksizliği” olarak tanımlanan “sistemin hatalarına tanık olmanın aşırı bir eylemi”, en uç durumda sistemle etkileşime girmeyi reddetme veya tavsiyelerini açıkça göz ardı etme ile sonuçlanmaktadır ve “muhalefet” olarak da tarif edilen tutumdur (Zerilli vd., 2022: 2). Ancak katılımcılar, kullanım bağlamına ve uygulama türüne göre bu hususta tutumlarının değişebileceğini de açıkça belirtmişlerdir.

Yine “Gizlilik toleransı” kapsamında “kullanıcı odaklı veri kullanımı ve acil durumlar” alt temasında görüş bildiren katılımcıların ağırlıkla erkek olduğu göze çarpmaktadır (7/10). Katılımcılar, alışveriş süreçlerini ve finansal süreçleri kolaylaştırmak maksadıyla kullanıcı lehine durumlarda, kullanıcının herhangi bir zarara uğramasını önleme veya kullanıcıya yönelik

muhtemel bir riski erken saptayarak önleme benzeri amaçlar doğrultusunda ve kullanıcının sağlık durumunun şüphe uyandırıcı veya kritik olduğu şartlarda kişisel veri ihlalinin makul karşılanabileceği yönünde görüş bildirmişlerdir. Ayrıca, “kamu yararı (ve suçla mücadele)” alt temasında katılımcıların, toplumsal fayda gözetilerek sağlık hizmetlerine yönelik geliştirilen uygulamalar, eğitimde fırsat eşitliği odaklı girişimler ve adaletin tesisi temelli çabalar kapsamında veri mahremiyeti hususunda tolerans ve ilgili kamu yararına faaliyetler esasında gönüllülük de göstermeye yanaşabilecekleri görülmektedir. Dikkat çeken, suçun önlenmesi, asayişin temin edilmesi kapsamında da bazı katılımcıların kişisel veri ihlalinin olağan karşılayabileceklerine yönelik ifadeleridir. Awad vd.’nin (2018: 68), bireyci kültürler ile kolektivist kültürler arasında gözlemlediği sistematik farklılıklar burada etkin olabilir; bireyci ve kolektivist değerlerin unsurlarını içeren (Kaya vd., 2024: 510), “yüksek rölativizm” ve “yüksek idealizm” eğilimleri gösteren, bireysel muhakeme analizini tercih eden ve ahlâki yargıları bu doğrultuda durum ve olay çözümlemelerine dayanan “Sitüasyonistler” ya da “Durumcular” kategorisinde ele alınabilecek bir toplum olmanın tesirleri olarak da yorumlanabilir (Forsyth, 1980: 176; 1992: 462-463; Forsyth vd., 2008: 823).

Örneğin, hala genel olarak kolektivist bir ülke olarak kabul edilen Çin’de, ticari ve yerel yönetim sosyal kredi sistemleri (SCS’ler) *henüz pilot aşamasındayken*, devlet tarafından “gözetim altyapısı” niteliğinde ve sosyal yönetim için kullanıldığı tartışılan bu girişimler üzerine Sinolog Genia Kostka (2019) tarafından kamuoyu görüşünü açığa çıkarmak maksadıyla ülke çapında bir anket çalışması yürütülmüştür. Araştırmada, SCS’lerin katılımcı gruplar arasında yüksek düzeyde onay bulduğu; mahremiyet ihlâli yerine fayda yaratma, toplumda ve ekonomide dürüst ilişkileri teşvik etme çerçevelerinde yorumlandıkları ileri sürülmektedir (Kostka, 2019: 1565). Makalenin metodolojisine ilişkin eleştiriler bulunsa da ve bunun “yaygın bir destek duygusundan ziyade genel bir farkındalık eksikliği” olarak değerlendirilmesinin daha uygun olabileceği belirtilse de, yine de YZ etiğinin, etkilenen toplumun kültürü, ideolojisi ve kamuoyu dinamikleri açısından anlaşılması gereği vurgulanmalıdır; dolayısıyla Kostka’nın ulaştığı kamuoyu bakışının “sosyal sorumluluğa, grup ve topluluk ilişkilerine daha fazla vurgu yapan, bireysel haklara nispeten daha az odaklanan” Çin toplumunun benimsediği prensipleri ve Çin’e özgü düşünce geleneklerini yansıtmaması da beklenebilir (Roberts vd., 2021: 67-68). Diğer taraftan, çoğunlukla bireyci kültürleri çatısı altında toplayan AB’nin, bir YZ teknolojisinin bir kişinin sağlığı, güvenliği veya temel hakları için oluşturabileceği risk düzeyini belirleyen bir sınıflandırma sistemine dayandığı “YZ Yasası” ise benzer nitelikte sosyal puanlama sistemlerini “kabul edilemez” risk kademesi kapsamında değerlendirmektedir.

(Feingold, 2023). Buna karşın, YZ etiğinin toplumsal yansımalarına yönelik argümanları kültürler arasında sistematik farklılıklar doğrultusunda genellemeler ile dayanaklandırmaya yanaşırken temkinli de olunmalıdır. “Ahlaki temeller gibi doğuştan gelen zihinsel yapıların, atalarımızın çok uzun bir süre boyunca karşılaştığı adaptif zorluklara verilen yanıtlar” olması mümkün olduğundan, zamanla değişebilecekleri de göz önünde bulundurulmalıdır ve bu, daima farklı metodolojiler ve birbirini destekleyen araştırmalar ile kuvvetlendirilmelerini gerektirir (Graham vd., 2013: 67).

3.4.4. Şeffaflık

Çevrimiçi anket verilerinin analizleri sonucunda, İnsan ve Toplum Bilimleri (İTB) grubu dahilinde değerlendirilen Hukuk, İktisadi ve İdari Bilimler ve İletişim fakültelerinden katılımcılar ile Fen Bilimleri, Teknoloji ve Mühendislik (FTM) grubuna dahil edilen Uygulamalı Bilimler, Tıp ve Mühendislik fakültelerinden öğrenciler arasında “Şeffaflık” boyutuna yönelik, akademik disiplin temelli anlamlı bir tutum farklılığına ulaşılmıştır. İTB disiplinlerinden katılımcılar, FTM bölümlerinden katılımcılara göre, YZ sistemleri aracılığıyla verilen kararların nedenlerinin açıklanmasını daha önemli olarak değerlendirmektedirler ve kararların nedenleri açıklanabilir değilse söz konusu sistemlerin güvenilirmez olduğunu düşünmeye de daha meyillidirler. Bu farklılığı doğrudan destekleyen ya da bir kontra argüman ileri süren çalışmaya yerli ya da yabancı literatürde rastlanılmamıştır, bu sebeple bulgunun daha özgül ve/veya daha kapsamlı ileri araştırmalarla irdelenmesi gerekir.

Ek olarak, kadın öğrencilerin sayıca üstünlüğü dikkat çekmekle birlikte, önceden de belirtildiği üzere, katılımcılar “ilk feragat edecekleri araştırma kapsamındaki YZ etiği boyutunun hangisi olacağı”na ilişkin soruya en çok “Şeffaflık” boyutunun ihmal edilebilirliği yönünde yanıt vermişlerdir (8/16). Bu doğrultuda görüş bildiren katılımcıların akademik disiplinleri incelendiğinde ise bir disiplinin baskınlığı ileri sürülememektedir; iki disiplinden de eşit sayıda katılımcının bu cevabı verdiği görülmektedir. Ancak, iki katılımcı da (E2-FTM ve E5-FTM) ikinci sırada bu boyutu belirtmiştir ki bu, diğer boyutlar ile karşılaştırıldığında hem yüksek bir farka işaret etmektedir hem de nicel aşamada saptanan akademik disiplin temelli tutum farklılığının görüşme yanıtları ile de İTB lehine desteklenebileceğini düşündürmektedir. Çoğunlukla Bilgisayar Mühendisliği ve Yönetim Bilişim Sistemleri öğrencilerinin oluşturduğu FTM grubunda katılımcılar, etkileşime girdikleri sisteme ilişkin zihinlerinde beliren soru işaretlerine yanıt arama konusunda teknik açıdan yetkin olduklarına

güvenme eğiliminde olabilirler: “Şeffaf olmadan da emin olamam.. ama onu test edebilirim, test edip şeffaf olduğunu..” (E5-FTM).

Öte yandan, tematik analiz sonucunda “Şeffaflık algısı” ve “Şeffaflık toleransı” ana temaları belirlenmiştir. “Şeffaflık algısı” ana teması kapsamında “kaynak toplama ve işleme süreci”, “adalet odaklılık” ve “kişisel verilerin gizliliği” alt temalarında İTB ve FTM disiplinlerinden katılımcıların ifadelerinin dengeli bir şekilde dağılım gösterdiği söylenebilmektedir. Öncelikle katılımcılar, genel itibarıyla YZ sistemlerine yönelik kaynak toplama ve işleme süreci hakkında ve şahsi verilerinin gizliliğinin teminatı hususunda uygulamayı hizmete sunanlardan şeffaflık beklentisindedirler. Bu noktada çoğunluğun daha ziyade, Andrada vd.’nin (2023: 1324) “yansıtıcı şeffaflık” adı altında yorumladığı ve “belirli bir YZ sistemi tarafından kullanılan algoritmaların doğası ve bir eğitim rejiminde kullanılan veri türleri hakkında bilgi” gibi bilgi türlerini hedefleyen “bilgi şeffaflığı”nın sağlanmasını gözettiği çıkarımında bulunulabilir. Bununla birlikte, YZ teknolojilerinin, onlarla etkileşimimizin bilişsel kapasitelerimiz üzerindeki derin dönüştürücü etkisiyle bağlantılı olarak kullanıcılar için opak olduğu da kabul edilmektedir ve bu bağlamda “dönüşümsel opaklık” da söz konusu teknolojinin, failin (kullanıcının) bilişsel, etik ya da epistemik kapasiteleri, hatta benlik veya kişilik duygusu üzerinde önemli etkileri olan gizli değişikliklere sebep olduğu durumlarda ortaya çıkmaktadır (Andrada vd., 2023: 1325). “Adalet etkisi” ana teması altında “sorgulama ve inisiyatif yitimi” temasında ve “Kötü niyet algısı” ana teması altında “davranışsal manipülasyon” temasında dolaylı kaygılarını dile getirmiş birkaç katılımcı bulunsa da yansıtıcı şeffaflığın bir diğer alt türü olan, YZ teknolojisinin yaygın kullanımının ortaya çıkardığı nöral ve bedensel dönüşümlerle ilgili “dönüşümsel şeffaflık” ile ilişkilendirilebilecek doğrudan bir katılımcı yorumuna ise ulaşamamıştır.

Ayrıca katılımcı ifadeleri, şeffaf bir yaklaşımın kullanıcıların ayrımcılık ve haksızlığa dair endişelerinin hafifletilmesine yardımcı olabileceğini göstermektedir; katılımcılar, şeffaflığı, uygulamanın adil olduğuna bir garantör olarak görmektedirler. Diğer taraftan İTB disiplinlerinden katılımcıların şeffaflığa ilişkin algılarında FTM grubu katılımcılarınkinden ayırt edici olan, “kötü niyetli olmama teminatı” alt temasıdır. Bu alt tema, “Şeffaflık algısı” ana temasında en düşük görüş sıklığına sahip alt tema olsa da görüş bildiren katılımcıların hepsi İTB disiplinlerinden öğrencilerdir. Dolayısıyla şeffaflık, doğruluğun yanı sıra iyi niyeti doğrulamanın da bir aracı olarak görülmektedir; nasıl ki samimî ve dürüst bir kimsenin herhangi bir gizli gündemi olmadığı düşünülüyorsa şeffaflık da bu bağlamda yapay araçlar açısından art

niyet barındırmamanın bir ifadesidir (Zerilli vd., 2022: 2). Aslında bir YZ sisteminin nasıl tasarlandığı, geliştirildiği, teste tabi tutulduğu, konuşlandırıldığı ve düzenlendiği gibi farklı süreçleri ele alan ‘şeffaflık’, yarı yapılandırılmış görüşmelerde, YZ sistemi hakkında sağlanan herhangi bir bilgiye yönelik ele alınmıştır. Böylelikle literatürde ve etik ve düzenleyici çerçevelerde şeffaflık ile sıkça anılan, bir YZ sisteminin ya da modelinin nasıl çalıştığını, belirli bir çıktının nasıl üretildiğini ifade eden ‘yorumlanabilirlik’ ya da ‘anlaşılabilirlik’ komşu ilkelerinin de katılımcı ifadelerinde karşılık bulduğu görülmüştür (Mittelstadt, 2021: 3). Kısacası bu alt tema kapsamında görüş bildiren katılımcılar için şeffaflık, YZ sistemleri ve bu sistemlerden sorumlu “insanlar” tarafından zarar görmeyecek olmalarının bir teminatıdır.

“**Şeffaflık toleransı**” ana teması altında ise “kullanım azaltma ve alternatif yaklaşımlar”, “ulusal güvenlik - askeri YZ ve stratejik istihbarat”, “hasta mahremiyet hakkı”, “ticari sır (know-how)” ve “teknik opaklık” alt temaları elde edilmiştir. “Kullanım azaltma ve alternatif yaklaşımlar” alt teması çoğunlukla Yönetim Bilişim Sistemleri, Bilgisayar Mühendisliği ve Tıp öğrencilerinin ifadelerini içermektedir (6/9). İnsanlar karara nasıl ulaşıldığını anlayamadıkları sürece, özellikle de kararın kişisel düzeyde kendileri için bir şekilde olumsuz olduğu durumlarda (örneğin, bir kredi başvurusunun reddedilmesi) o karara güvenmeme eğilimindedirler (ISO/IEC 22989:2022 (E), 2022: 29). Bu alt temada görüş sıklığının yüksek olması, güvensizliğin kullanım azaltma ve alternatif mecralara yönelmeye yol açabileceğini de desteklemektedir. Bunun yanı sıra, sadece FTM disiplinlerinden öğrencilerin, sisteme dair açıklamaları yetersiz ya da mantıksız bulmaları durumunda önyargıyla yaklaşmama, ilgili nedenleri tekrar gözden geçirme, tartma veya daha ayrıntılı bir açıklama talebinde bulunma eğiliminde olacaklarının gözlemlendiği de belirtilmelidir. Bu gruptan katılımcıların, daha önce ifade edildiği gibi, etkileşime girdikleri sisteme ilişkin zihinlerinde beliren soru işaretlerine yanıt arama konusunda teknik açıdan yetkin olduklarına güvenme eğiliminde olmaları ihtimaliyle de bu durum bağlantılandırılabilir. Diğer taraftan, kullanılan YZ uygulamasına ve bağlama göre tutumunun farklılaşabileceğini ifade eden katılımcıların çoğunluğu ise İTB disiplinlerindendir.

“Ticari sır” alt teması kapsamında da Bilgisayar Mühendisliği öğrencilerinin, fikri ve sınai mülkiyet haklarının gözetilmesi sebebiyle ve özellikle kullanılan yöntem ve teknolojiler konusunda stratejik olarak şeffaf bir tutum sergilenmemesi durumunda tolerans gösterilebileceği yönünde değerlendirilebilecek ifadeleri mevcuttur. Bunun sebebi, bu katılımcıların kendilerini hem kullanıcı hem de geliştirici pozisyonunda görmeleri olabilir.

İTB disiplinlerinden katılımcıların yaklaşımını “Şeffaflık toleransı” ana temasında FTM disiplinlerinden katılımcıların yaklaşımından ayıran ise “ulusal güvenlik - askeri YZ ve stratejik istihbarat” alt temasıdır. Burada görüş bildiren katılımcıların ikisi de İTB öğrencisidir, özellikle bu meselelerde şeffaflıktan ziyade gizliliğin gereğini vurgulamışlardır. Yanıtları, açıklamalar ve yayınlanan ek bilgiler aracılığıyla bu sistemlerin saldırılara karşı daha savunmasız hale getirilebileceği yönündedir ve “şeffaflık paradoksu”na işaret etmektedir (Burt, 2019).

3.4.5. Sorumluluk

Mevcut çalışma kapsamında sonuncu YZ etiği boyutu olan “Sorumluluk” boyutuna yönelik nicel çözümlemeler ile ne cinsiyet ne de akademik disiplin temelli gruplar arası anlamlı bir farklılık saptanmıştır. Bu aşamada katılımcıların, bir YZ sisteminin herhangi bir soruna sebep olması durumunda, YZ spektrumunun akla ilk gelen faillerinden geliştirici ve kullanıcının sorumluluğuna ve zararın tazmininin kim tarafından ve nasıl gerçekleştirileceği hakkında toplumsal bir uzlaşma sağlanmasının gerekip gerekmediğine yaklaşımları araştırılmıştır. Farklılık analizinden elde edilen bulguları takiben yürütülen yarı yapılandırılmış görüşmelerde ise bu boyuta dair daha detaylı verilere erişebilmek amacıyla katılımcıların, ölçek ifadelerinde bulunan aktörlere ek olarak, YZ sisteminin, veri sağlayıcıların, tedarikçilerin ve regülatörlerin sorumluluğuna yönelik anlayışları da anlaşılmaya çalışılmıştır.

Katılımcılara, herhangi bir bağlamda ‘bir YZ sisteminin uygulanması sonucunda istenmeyen bir durum söz konusu olduğunda geliştirici ve kullanıcı dahil olmak üzere yukarıda sıralanan aktörlerden hangisini *en sorumlu* bulacakları’ sorulduğunda frekansı en yüksek yanıt “regülatörler” yani yasa koyucular ya da düzenleyiciler olarak öne çıkmıştır (7/16). Görüşlerde akademik disiplin temelli belirgin bir farklılık bulunmamaktadır. Öte yandan, bu yönde yanıt verenlerin çoğunluğu kadındır ki bu, kadın katılımcıların çoğunluğunun regülatörlerin sorumluluğunu öncelikli gördüğü anlamına da gelmektedir (5/7). Sondaj soruları ile erkek katılımcıların yanıtlarında ise regülatörler, geliştiriciler ve veri sağlayıcıların sorumluluğu arasında daha dengeli bir dağılıma ulaşılmıştır. Sorunun kaynağına göre sorumluluğun kime ait olduğunun değişkenlik gösterebileceğine ilişkin katılımcı ifadeleri de bulunmaktadır (6/16). YZ’nin özerklik seviyesinin, sorumluluğun dağıtılmasında rol oynayabileceğini vurgulayan bir katılımcının da mevcudiyeti sorumluluk meselesinin çok yönlülüğünü ortaya koymaktadır. Ek olarak, göze çarpan, katılımcıların çoğunluğunun kullanıcıları en az sorumlu aktör olarak değerlendirmesidir (9/16).

Ayrıca, araştırma kapsamında YZ sistemlerinin belirli sektörler ve uygulamalar ile sınırlandırılmamış olması, görüşmeci tarafından mülakatlar süresince bu tutumun benimsenmesine özen gösterilmesi ve katılımcıların geleneksel ve sosyal medyadan çağrışımlarına alan bırakılması, daha zengin verilere ulaşılmasına yardımcı olmuştur. Öyle ki, 6'sı erkek olmak üzere 7 katılımcının “yargı sistemleri”, 3'er kadın ve erkek katılımcının “sağlık endüstrisi”, 2 erkek katılımcının “ulusal güvenlik” ile ilişkilendirilebilecek hususlar, 1'er kadın ve erkek katılımcının “finansal hizmetler sektörü”, 1 kadın katılımcının “askeri teknoloji” ve 1 erkek katılımcının da “eğitim sistemi” ve “ulaşım” söz konusu olduğunda kullanılan YZ sistemlerine ilişkin sorumluluğunun belirlenmesinin öncelikli olarak titizlikle ele alınması gerektiğini düşündükleri ortaya çıkarılmıştır.

“Sonuç” bölümüne geçmeden, görüşmecinin gözlemi ve yinelemeli içerik analizleri neticesinde değinilmesi gereken bir nokta da “YZ” denildiğinde özellikle İTB katılımcılarının zihinlerinde öncelikle ChatGPT gibi popüler chatbotları, Sophia gibi insan-benzeri robotları ve Siri gibi sanal asistanları canlandırdığıdır. “Kullanıcı tarafından görülemeyen, bir bilgisayarın veya başka bir aracın içine gömülü olan” ve bu nedenle “görsel bir temsil veya ayırt edici bir kimliğe” sahip olmayan YZ biçimleri olan, sıralama algoritmaları, tavsiye sistemleri ya da daha az rastlanan müşteri kredi derecelendirme algoritmaları veya suçluların yeniden suç işleme riski araçları gibi “gömülü YZ” sistemlerini içeren yanıtlara az rastlandığı ya da bu katılımcıların ancak ilerleyen sondaj soruları aracılığıyla ilgili ifadelerine ulaşıldığı belirtilmelidir (Glikson ve Williams Woolley, 2020: 627; Zerilli vd., 2022: 2). Bu saptama, Kaya vd.’nin (2024: 510), bu teknolojilerin yazılım ve uygulamalara gömülü olması ve teknoloji alanında uzmanlaşmamış popülasyonun bu sistemlerin altında yatan mekanizmaları fark etmesini zorlaştırması sebebiyle genel nüfusun YZ destekli birçok uygulamayı benimsemesine rağmen, çeşitli kullanımlarının ayırdında olmamasından kaynaklanabileceği yönünde yorumuyla da örtüşmektedir. YZ’nin nasıl deneyimlendiğini incelemek üzere ulusal düzeyde İngiltere’de 4.000 üzerinde yetişkin ile gerçekleştirilen temsili bir anket çalışmasında (Ada Lovelace Institute ve The Alan Turing Institute, 2023: 18), ortaya çıkan politika önceliklerine ve kamu yaşamında artan kullanıma dayalı olarak seçilmiş olan YZ kullanımları (sürücüsüz araçlar, otonom silahlar, hedefli tüketici reklamcılığı, hedefli siyasi reklamcılık, işe uygunluğun değerlendirilmesi, sanal sağlık asistanları, robotik bakım asistanları, sosyal yardıma uygunluğun değerlendirilmesi, kredi ödeme riskinin değerlendirilmesi, akıllı hoparlörler, polislik için yüz tanıma, telefon kilidini açmak için yüz tanıma, sınır kontrolü için yüz tanıma, kanser riskinin değerlendirilmesi, iklim

araştırma simülasyonları, eğitimde sanal gerçeklik ve robotik elektrikli süpürgeler) dikkate alındığında da, mevcut karma araştırma kapsamında katılımcı ifadelerinde bu kullanımlardan “otonom silahlar”, “hedefli çevrimiçi reklamcılık (hedefli tüketici reklamcılığı ve hedefli siyasi reklamcılık)”, “sanal sağlık asistanları”, “kanser riskinin değerlendirilmesi” kullanım alanlarının öne çıktığı söylenebilir. Gündemde olan AB YZ Yasası kapsamında “kabul edilemez” ve “yüksek risk” kategorilerinde düzenlenen, sırasıyla, tartışmalı sosyal puanlama ve insanları “indirgemecilik, zorunlu şeffaflık ve bedensel bütünlüğün ihlali yoluyla daha önce emsali görülmemiş nesneleştirme biçimlerine de maruz bırakan” biyometrik tanıma sistemlerine (Lagerkvist, 2022: 1) yönelik ise doğrudan olumlu ya da olumsuz katılımcı ifadelerine genel itibariyle rastlanılmamıştır.



SONUÇ

Karma metodolojilerden *sıralı açıklayıcı desen*in takip edildiği iki aşamalı bu araştırmada, kolektif bir yanıt potansiyelini de düşünmenin bir yolunu irdelemek gayesiyle, ahlâki fail sıfatıyla günümüzün kullanıcısı olarak ve YZ ekosisteminin potansiyel paydaşlarını temsilen farklı akademik orijinlerden üniversite öğrencilerinin YZ teknolojilerinin etik boyutlarına ilişkin anlayışları açığa çıkarılmaya çalışılmıştır. Btihaj Ajana'nın, “araştırmacıyı, aciliyet arz eden sosyo-politik ve etik meselelerin özüne doğrudan ve duyuşsal olarak yerleştirebilecek akademik yazım yollarını keşfetme” çağrısına kulak verilmiş, aksi takdirde kolayca kurulamayacak farklı bağlantıların gerçekleşmesine imkân vereceği umut edilerek bir metodoloji özgünlüğüne gidilmiştir (Beer, 2014: 331; Lagerkvist vd., 2022: 11). Jobin ve ekibinin (2019: 391), özel şirketler, devlet kurumları, akademik kurumlar ve araştırma kurumları, hükümetler arası veya uluslararası kuruluşlar, kâr amacı gütmeyen kuruluşlar, meslek birlikleri, bilimsel topluluklar, özel sektör ve araştırma ittifakları, bilim vakıfları, işçi sendikaları federasyonları ve siyasi partiler tarafından, çoğunluğu 2016'dan sonra yayınlanan, coğrafi dağılım açısından çeşitlilik gösteren ve birden fazla paydaş grubuna hitap eden, YZ için etik ilkeler ya da kılavuzlar içeren belgelerin “şeffaflık”, “adalet ve hakkaniyet”, “kötü niyetli olmama”, “sorumluluk” ve “mahremiyet” etik ilkeleri etrafında yakınlaştığı tespitine dayanarak da araştırma, bu ilkeler ile temellendirilmiştir.

Birincil *kantitatif aşamada* veri toplama aracı olarak Jang vd.'nin (2022) Yapay Zekâ Etiğine Yönelik Tutum Ölçeği'nden faydalanılması tercih edilmiş ve araç Türkçeye çevrilerek çevrimiçi anket aracılığıyla 395 katılımcı yanıtına ulaşılmıştır. Geçerlik ve güvenilirlik analizlerinin gerçekleştirilmesini takiben farklılık analizleri neticesinde katılımcıların “Adalet”, “Kötü niyetli olmama” ve “Gizlilik” ilkelerine yönelik *cinsiyet* ve “Şeffaflık” ilkesine yönelik *akademik disiplin* temelli tutum farklılıkları keşfedilmiştir. Kantitatif aşamayı izleyen *kalitatif araştırma* yoluyla ise mutlak cevaplar sunmaktan çok araştırma sorularının kökenine inmeye imkân tanıyacağı umuduyla katılımcılar ile gönüllü bir diyalog kurulması amaçlanmıştır. Bu doğrultuda fenomenolojik bir yaklaşımla veri edinme tekniği olarak tercih edilen *yarı yapılandırılmış görüşmeler* ile ilgili grupların ilk aşamada saptanan hassasiyet seviyelerinin yapısını irdelemek ve anlayışlarının altında yatan nedenlere ulaşabilmek maksadıyla, dengeli bir şekilde gruplandırılan 16 katılımcının bu sistemler ile etkileşimleri, deneyimleri ve çalışma kapsamında ele alınan ilkeler ile ilişkiselliklerine yoğunlaşmıştır. İteratif bir yaklaşımla mülakat metinleri *içerik analizine* tabi tutulmuş, bu süreçte kantitatif bulgulara dayandırılan

geniş çaplı araştırma sorularına dair gözden kaçabilecek anlayışları açığa çıkarabilmek için öncelikle katılımcıların ifadeleri arasında benzer örüntüler yakalanmaya çalışılmış ve *temalandırma* gerçekleştirilmiş, ardından kantitatif safhada elde edilen bulguları dikkate alarak alt temalarda gruplar arası farklılıkların ilişkilendirilmesi yoluna gidilmiştir.

Öncelikle, bu araştırma ile, farklı bir metodoloji vesilesiyle, Jang vd.’nin (2022: 11635) lisans öğrencilerinin YZ etiğine yönelik tutumlarını değerlendirmek için geliştirdikleri ölçek aracılığıyla Kore’de kadın ve erkek lisans öğrencilerinin “*Adalet*”, “*Kötü niyetli olmama*” ve “*Gizlilik*” ilkelerine yönelik hassasiyet düzeylerinde dengesizliğe işaret eden kantitatif saptamalarını destekleyen bulgulara varılmış, kadınların YZ teknolojilerine ve bu teknolojilerin ele alınan etik ilkeler ile ilişkiselliğine dair öznel bakış açıları yakalanmaya çalışılarak, erkeklere göre bu boyutlarda daha yüksek hassasiyet seviyelerinin özünü betimleyen zengin bir anlayış ortaya konulmuştur.

Yarı yapılandırılmış görüşme metinlerinin içerik analizi neticesinde “**Adalet**” boyutu için, “Adalet algısı” ve “Adalet etkisi” ana temaları belirlenmiştir. “Adalet algısı” ana teması kapsamında hem erkek katılımcıların hem de kadın katılımcıların YZ sistemleri ile “adalet” ilkesini nasıl bağlamsallaştırdıkları ve adil bir YZ uygulamasını nasıl nitelendirdikleri irdelendiğinde “objektiflik”, “rasyonellik”, “geliştirici perspektifi”, “affektif ve kognitif insani nitelikler”, “teknik sağlamlık ve güvenilirlik”, “yasa merkezlilik”, geniş spektrumda veri altyapısına işaret eden “çeşitlilik ve esneklik” ve “gizlilik” olmak üzere sekiz alt temaya ulaşılmıştır. “Adalet etkisi” ana teması altında şekillenen, bu sistemlerin olumlu ve olumsuz etkilerini tarif eden temalar ise “savaş ve çatışma koşulları - anomi”, “verilerin kötüye kullanımı”, “karar süreçleri”, “YZ’ye güven kaybı” ve “sorgulama ve inisiyatif yitimi” olmuştur. Kadın katılımcıların hassasiyetlerini yönlendiren anlayışlarının, belirli alt temalar ile yordanabileceği sonucuna varılmıştır ki bu bağlantısallıkların ileri araştırma soruları ile ele alınması yerinde olacaktır. Öncelikli olarak, “Adalet algısı” ana teması kapsamında, erkek katılımcıların çoğunlukla, adaletin tesis edilmesi hususunda, özellikle yasal karar süreçlerinde YZ sistemlerinin “objektiflik” ve “rasyonellik” potansiyel niteliklerini göz önünde bulundurarak potansiyel kazanımlara eğildiği, kadın katılımcıların ise algoritmik sistemlerin “vicdan” ve “empati” gibi “affektif ve kognitif insani nitelikler” açısından noksanlığını ileri sürerek ekseriyetle ceza hukuku gibi kritik pratiklere yönelik kullanımlarına ilişkin kaygı duyduğu açığa çıkarılmıştır. “Adalet etkisi” ana teması altında hem erkek katılımcıların hem de kadın katılımcıların görüş sıklığının, diğer alt temalar ile karşılaştırıldığında açık ara en yüksek

olduğu ilgili “Karar süreçleri” alt temasına ait ifadeler incelendiğinde de, kullanım bağlamına göre katılımcı tutumlarında değişkenlik görülse de, bilhassa hukuki, tıbbi ve eğitime yönelik karar alma süreçlerinde insan yaşamını etkileyebilecek hakkaniyete aykırı “üstünkörü” kararlar konusunda çoğunluğun duyarlı olduğu görülmüştür. Kararların geri döndürülemez veya tersine çevrilmesi zor bir etkiye sahip olacağı bağlamlarda nihaî insan kararlılığının gereği katılımcılar tarafından vurgulanmaktadır. Bulgular, genel itibariyle kadınların belirli grupları orantısız bir şekilde etkileyebilecek YZ sistemlerinde ayrımcılık, eşitsizlik ve haksızlık algılayarak adaletsizlik ihtimaline karşı daha fazla hassasiyet sergiliyor olabileceği şeklinde yorumlanmıştır. Aynı ana tema kapsamında “savaş ve çatışma koşulları – anomi” alt temasında kadın katılımcıların endişelerine yönelik ifadelerin sayıca üstünlüğü saptanarak adalet boyutunda mevcut hassasiyetlerinin, “ahlaki gelişim ve adalet için kritik öneme sahip olan ve toplumsal uyum ve birlik için bir katalizör görevi gören” empati (Toussaint ve Webb, 2005: 674) gibi duygusal ve bilişsel bileşenlere sahip mefhumları gözetmelerinden kaynaklanabileceği öngörülmektedir.

“**Kötü niyetli olmama**” boyutu, “Kötü niyet algısı”, “Kötü niyet motivasyonu” ve “Kötü niyet çıktıları” şeklinde temalandırılmıştır. “Kötü niyet algısı” ana teması altında “askeri YZ ve siber tehditler”, “derin sahte tehlikesi”, “süper zekâ”, “veri ve gizlilik ihlali”, “davranışsal manipülasyon” ve “otomasyon” alt temaları oluşturulmuş, “Kötü niyet motivasyonu” ana teması altında “maddi kazanç arzusu”, “bireysel duygusal ve çıkar temelli dinamikler”, “devletler arası rekabet ve jeopolitik çekişme” alt temaları belirlenmiştir. “Kötü niyet çıktıları” ise “bilgi manipülasyonu ve rıza imalatı”, “YZ’ye güven kaybı”, “ülkeler arası politik gerilim”, “artan yapısal eşitsizlikler ve güç asimetrisi”, “veri ve gizlilik ihlali” ve “toplumsal bağlarda erozyon ve güven sorunları” olmak üzere altı alt tema ile katılımcı ifadelerinde karşılık bulmuştur. Kötü niyetle YZ geliştirilmesi ve art niyetli kullanıma ilişkin “Kötü niyet motivasyonu” ana teması altında sıklıkla dile getirilen “maddi kazanç arzusu” ve “bireysel duygusal ve çıkar temelli dinamikler” alt temaları, kadın katılımcıların erkek katılımcılarla görüş ortaklığına işaret eden temalar olarak öne çıkmıştır. “Devletler arası rekabet ve jeopolitik çekişme” alt teması, kadın katılımcı ifadelerinin sayıca üstünlüğü sebebiyle erkek katılımcılardan ayırt edici olarak hassasiyetlerini kanalize eden bir unsur olarak değerlendirilmiştir. “Kötü niyet çıktıları” ana teması incelendiğinde ise yine hem erkek katılımcıların hem de kadın katılımcıların kaygılarını ifade ettiği ve görüş frekansı en yüksek alt tema olarak beliren “Bilgi manipülasyonu ve rıza imalatı”, YZ sistemlerinin bu doğrultuda kullanımına yönelik endişeler sebebiyle bireylerin hassasiyetini tetikleyebilecek bir

başka faktör olarak yorumlanmıştır. Aynı ana tema dahilinde yüksek frekansa ulaşan “veri ve gizlilik ihlali” endişeleri bir diğer tetikleyici unsur olarak tespit edilmiştir. “Askeri YZ ve siber tehditler”, “ölkeler arası politik gerilim” ve “artan yapısal eşitsizlikler ve güç asimetrisi” alt temalarında ise kadın katılımcıların baskın ifade sıklığı, kadınların, erkek katılımcılardan bu boyuta yönelik daha yüksek hassasiyetinin kaynağına işaret eden unsurlar olarak açıklanmıştır. “Artan yapısal eşitsizlikler ve güç asimetrisi” alt temasında görüş bildiren katılımcılar arasında, YZ sistemlerine erişim imkânı olan bireylerin işgücü piyasasında kazanımını artıracak ve uluslararası arenada halihazırda dijital rekabet gücü yüksek ölkelerin de askeri teknolojilere dayanan avantajlarını güçlendireceği tedirginliğinin yaygın bir şekilde gözlemlenmesi ile, “Kötü niyet algısı” ana temasında sadece kadın katılımcıların ifadelerinde karşılaşılan “otomasyon” kaygısı da bu alt temalarla anlam kazanmıştır. İşgücüne katılım oranı düşük ve teknoloji sektörlerindeki istihdamda dezavantajlı konumda bulunan kadınların, yapısal eşitsizliklerin YZ sistemleri aracılığıyla şiddetlenmesi ihtimalini algılayarak daha duyarlı yanıtlar verdiği, hassasiyetlerinde bu korkuların etkili olabileceği sonucuna ulaşılmıştır.

“Gizlilik” boyutuna yönelik “Gizlilik algısı” ve “Gizlilik toleransı” ana temaları yakalanmıştır. “Gizlilik algısı” ana teması altında “vatandaşlık verileri”, “finansal veriler”, “siber risk yönetimi”, “şeffaflık – rızaya dayalı veri işleme ve erişimi/kullanımı”, “toplumsal uyum baskısı” ve “veri yönetimi ve paylaşım sınırlamaları” alt temaları tanımlanmış, “Gizlilik toleransı” ise “kullanım azaltma ve alternatif yaklaşımlar”, “kullanıcı odaklı veri kullanımı ve acil durumlar” ve “kamu yararı (ve suçla mücadele) alt temalarına ayrılmıştır. Kadın katılımcıların hassasiyetinin, bilgilerinin belli hususlarda daha fazla tehlike arz ettiğini düşünmelerinden ya da saldırıya uğrama olasılığını daha yüksek görmelerinden kaynaklanıyor olabileceği neticesine varılmıştır; bilhassa kişisel fotoğraflar söz konusu olduğunda kadın katılımcıların ifadeleri bu argümanı destekler niteliktedir.

Diğer taraftan, İnsan ve Toplum Bilimleri disiplinlerinden katılımcılar, Fen Bilimleri, Teknoloji ve Mühendislik bölümlerinden katılımcılara göre, YZ sistemleri tarafından verilen kararların nedenlerinin açıklanmasını daha önemli olarak değerlendirmektedirler ve kararların nedenlerini açıklanabilir bulmuyorlar ise bu sistemlerin güvenilirmez olduğunu düşünmeye daha meyillidirler. Akademik disiplin temelli tutum farklılığının tespit edildiği tek boyut olan “Şeffaflık” boyutuna yönelik tematik analiz neticesinde ulaşılan ana temalar; “adalet odaklılık”, “kötü niyetli olmama teminatı”, “kişisel verilerin gizliliği” ve “kaynak toplama ve işleme süreci” alt temalarını içeren “Şeffaflık algısı” ve “ulusal güvenlik - askeri YZ ve stratejik

istihbarat”, “hasta mahremiyet hakkı”, “teknik opaklık”, “ticari sır (know-how)” ve “kullanım azaltma ve alternatif yaklaşımlar” alt temalarını kapsayan “Şeffaflık toleransı” olmuştur. Katılımcıların, genel itibarıyla, YZ sistemlerine yönelik “kaynak toplama ve işleme süreci” hakkında ve “kişisel verilerinin gizliliği”nin teminatı hususunda uygulamayı hizmete sunanlardan şeffaflık beklentisinde olduğu ve akademik disiplin fark etmeksizin “bilgi şeffaflığı”nın karşılanmasını önceliklendirdiği sonucuna varılmıştır. “Kötü niyetli olmama teminatı” alt teması, İnsan ve Toplum Bilimleri disiplinlerinden katılımcıların, Fen Bilimleri, Teknoloji ve Mühendislik grubu katılımcılarından ayırt edici olarak, şeffaflığı, doğruluğun yanı sıra iyi niyeti doğrulamanın da bir aracı olarak görmekte olduklarını açığa çıkarmıştır. Şeffaflık, YZ sistemleri ve bu sistemlerden sorumlu “insanlar” tarafından zarara uğramayacaklarına ilişkin bir güvence göstergesidir, bu ilkeyle ilintili açıklanabilirlik ve anlaşılabilirlik da bu grup için, etkileşime girdikleri bu sistemlere ilişkin içgörü sağlıyor olması açısından önem kazanmaktadır. Bununla beraber, katılımcıların, bilhassa kadınların, “ilk feragat edecekleri” ilkenin “Şeffaflık” olarak tespit edilmiş olması, iki Fen Bilimleri, Teknoloji ve Mühendislik disiplinlerinden katılımcının da ikinci sırada bu boyutun ihmal edilebilirliğini makul olarak değerlendirmesi, Fen Bilimleri, Teknoloji ve Mühendislik disiplinlerinden öğrencilerin bu boyutu gözden çıkarmaya İnsan ve Toplum Bilimleri öğrencilerinden daha yatkın olabileceği ihtimalini ortaya koymuştur. Dolayısıyla kantitatif aşamada saptanan akademik disiplin temelli tutum farklılığı kalitatif aşama ile desteklenmiştir. Bu durum, Fen Bilimleri, Teknoloji ve Mühendislik grubunun, etkileşime girdikleri sisteme ilişkin soru işaretlerine yanıt arama konusunda teknik açıdan yetkin olduklarına güvenme eğiliminde olabilmeleri ihtimaliyle ve kendilerini hem kullanıcı hem de geliştirici pozisyonunda görmelerinden kaynaklanıyor olabilecek fikri ve sınai mülkiyet haklarını gözetebilmeleri sebebiyle özellikle kullanılan yöntem ve teknolojiler konusunda stratejik olarak şeffaf bir tutum sergilenmemesi durumunda tolerans gösterilebileceği yönünde değerlendirilebilecek ifadeleri ile açıklanmıştır. Ayrıca, özellikle İnsan ve Toplum Bilimleri disiplinlerinden katılımcıların, kullanılan YZ uygulamasına ve bağlama göre tutumunun farklılaşabileceği de saptanmıştır, bu da geliştiricilerin ve düzenleyicilerin bağlamsal bir yaklaşımla süreçleri ele alması gerektiğini ve bu doğrultuda spesifik araştırmaların yürütülebileceğini göstermektedir. Bunun yanı sıra, “Şeffaflık toleransı” ana teması kapsamında, Hukuk ve İktisadi ve İdari Bilimler fakültelerinden öğrencilerin “ulusal güvenlik - askeri YZ ve stratejik istihbarat” alt temasında bu hususlarda hassasiyeti de öne çıkmıştır; katılımcılar, ‘şeffaflık paradoksu’ kavramını destekler nitelikte, YZ sistemlerine ilişkin açıklamalar ve yayınlanan ek bilgiler aracılığıyla sistemlerin saldırılara karşı daha savunmasız hale getirilebileceği düşüncesiyle şeffaflıktansa gizliliği önceliklendirmektedir.

Son olarak, kantitatif çözümlemeler ile “**Sorumluluk**” boyutuna yönelik ne cinsiyet ne de akademik disiplin temelli gruplar arası anlamlı bir farklılık saptanmış olsa da takibinde yürütülen kalitatif çalışma, katılımcıların çoğunluğunun, ‘bir YZ sisteminin uygulanması sonucunda istenmeyen bir durum söz konusu olduğunda’ kullanıcıları *en az sorumlu* aktör olarak değerlendirmekte iken, *en sorumlu* aktör olarak “regülatörler”i gördüklerini ortaya çıkarmıştır. Yasa koyucuların sorumluluğunu birincil gören katılımcılar, erkek katılımcılara kıyasla kadın katılımcılar olmuştur; öncelikli olarak YZ geliştirilmesi ve kullanımında etkili düzenleyici ve yasal çerçevelerin oluşturulması beklentisindedirler. Erkek katılımcıların yanıtları ise regülatörler, geliştiriciler ve veri sağlayıcıların öncelikli sorumluluğu arasında daha dengeli bir dağılıma işaret etmektedir, dolayısıyla tüm paydaşların sorumluluk almasının gereği de burada karşılık bulmaktadır. Bulgular ayrıca, sorunun kaynağı dikkate alınarak sorumluluk meselesinin bağlamsal bir yaklaşımla ele alınmasının gereğini ortaya koymaktadır. “Yargı sistemleri” ve “sağlık endüstrisi” söz konusu olduğunda kullanılacak YZ sistemlerine ilişkin sorumluluğunun belirlenmesi ise katılımcılar arasında öncelikli olarak titizlikle ele alınması gerektiği düşünülen kullanım bağlamları olmuştur.

Araştırma, gruplar arası tutum farklılıklarını açıklarken, aynı zamanda, YZ söyleminde sıklıkla kabul edilen, YZ yönergelerini kanalize edeceği öngörülen ilkelerin Türkiye’de yükseköğrenim gören öğrencilere yansımalarının keşfedilmesine, ülkeye özgü nüansların yakalanarak anlayışın derinleştirilmesine ve meşru kaygıların tüm çeşitliliğiyle ortaya konulmasına imkân tanımıştır. Dolayısıyla bulguların, katılımcı ifadelerinde de karşılık bulan isteme yanıt olarak, yerel değerler ve tercihler dikkate alınarak, merkezi etik sorunlara ortak yanıtlar sağlayacak kapsayıcı ve bütüncül YZ geliştirme ve ilgili yönetim ve gözetim çabalarına yönelik içgörüler sağlayacağı düşünülmektedir (Ess 2007: 102).

Aşamalı bulguların ve kantitatif safhada elde edilen verilere dayandırılan kalitatif araştırma sorularına dair yukarıdaki çıkarımların gelecek araştırmalar ile desteklenmesi ihtiyacına ilaveten, tartışma başlığı altında da değinildiği üzere pek çok potansiyel araştırma istikameti de belirmektedir:

- “Kötü niyet çıktıları” ana teması kapsamında elde edilen “ülkeler arası politik gerilim” ve “artan yapısal eşitsizlikler ve güç asimetrisi” alt temaları hususiyetle yerel yazında, güç ilişkilerini inceleyen bir etik yaklaşım ihtiyacına da sinyal vermektedir. Dekoloniyal ve kesişimsel feminizm türleri, Eleştirel Irk Çalışmaları, Eleştirel Medya Çalışmaları

gibi farklı disiplinlerin sunduğu çeşitli analitik ve ampirik araçlar aracılığıyla bu alt temalara dair bulgular daha özgül araştırma sorularıyla irdelenebilir (Ricaurte, 2022: 729).

- “Gizlilik algısı” ana teması kapsamında, görüş bildiren katılımcıların kadın ağırlıklı olduğu “toplumsal uyum baskısı” faktörünün, kadınların YZ sistemlerini kabul ve kullanımında ve YZ aracılı adaletsizlik ve haksızlık neticesinde gerektiğinde bir itiraz merciine başvurup başvurmamasında potansiyel bir etkiye sahip olup olmadığı üzerine araştırmalar faydalı olabilir.
- Aynı alt temanın, yazında “mahremiyet paradoksu” olarak nitelendirilen (Weinberger vd., 2017: 6) ve çalışma dahilinde kadınların gizlilik tutumları ve davranışları arasında bulgulanan bir tür eşitsizlik potansiyeli ile ilişkiselliği de ileri araştırmalarla ortaya konulabilir. Bilhassa, kadınların kişisel bilgi gizliliğini korumaya yönelik harekete geçme eğilimini hangi değişkenlerin etkilediğine ya da artırdığına ilişkin araştırmalar yerinde olacaktır.

Ayrıca bulguların, eğitimcilerin ve eğitim kurumlarının YZ etiği müfredatı ve eğitim materyalleri geliştirmesine rehberlik edeceği de umulmaktadır. İnovasyona ve YZ geliştirmeye yatırım yapma amacı güden hususiyetle yerli işletmeler de, endüstri ve pazar yapısının çok çeşitli rekabet dinamikleri olsa da bağlam odaklı bir anlayışla, “şeffaflık paradoksu”nu dikkate alan bir şeffaflık ve gizlilik dengesi gözetken, bu etik ilkeler ile uyumlu YZ teknolojileri geliştirmeyi ve sorumlu kullanımı benimseyerek avantajlarını artırabilirler. “Şeffaflık” ve “Gizlilik” boyutlarında elde edilen, görüş sıklığı yüksek” kullanım azaltma ve alternatif yaklaşımlar” alt temaları, güvensizliğin kullanım azaltma ve alternatif mecralara yönelmeye yol açabileceğini desteklemektedir; dolayısıyla bir hayli rekabetçi bir pazarda faaliyet göstermeyi hedefleyen ticari YZ girişimci geliştiriciler, bu hususları dikkate almalıdır. İşgücüne katılım oranı düşük ve teknoloji sektörlerindeki istihdamda erkeklerin gerisinde kalan kadınların yapısal eşitsizliklerin YZ sistemleri aracılığıyla şiddetlenmesi ihtimalini algılayarak daha duyarlı yanıtlar verdiği çıkarımına da dayanarak, kadın istihdamının desteklenmesi, meslekî bilişim teknolojileri becerilerini geliştirmeye ve piyasada gözlemlenen endüstriyel eğilimler doğrultusunda farklı kariyer yollarında yeniden beceri kazandırmaya yönelik farkındalığın artırılması ve ilgili çabaların önem kazandığı da söylenebilir.

Araştırmada sıralı açıklayıcı desen takip edilerek kalitatif araştırma soruları kantitatif bulgulara dayandırılmış olsa da bu safhada yarı yapılandırılmış görüşmeler neticesinde elde

edilen bulgular, ankete yüksek katılım gösteren akademik departmanlar önceliklendirildiği için belirlenen çalışma grubundaki katılımcıların görüşleriyle sınırlıdır. Görüşmeler süresince tutulan gözlem notlarına ve yinelemeli olarak ses kayıtlarına başvurularak toplanan verilere mümkün olduğunca bağlı kalınması esas alınsa da iki paradigmanın bir arada çalışıldığı karma metodolojinin zaman alması sebebiyle görüşme sonrasında içeriğin gözden geçirilmesi için katılımcılara görüşme dökümlerinin sağlanamadığı, katılımcılardan geribildirim istenemediği de belirtilmelidir.



KAYNAKÇA

- Ada Lovelace Institute ve The Alan Turing Institute. (2023). “How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain”.
- Adjabi, I., Ouahabi, A., Benzaoui, A. ve Taleb-Ahmed, A. (2020). “Past, Present, and Future of Face Recognition: A Review”. *Electronics*, 9(8): 1188.
- AI HLEG. (2018). “The European Commission’s high-level expert group on artificial intelligence: Draft Ethics guidelines for trustworthy AI. Working Document for stakeholders’ consultation”. European Commission, B-1049. Brussels.
- AI HLEG. (2019). “Ethics Guidelines for Trustworthy AI”. High-Level Expert Group on Artificial Intelligence. European Commission, Brussels.
- Alaga, J. ve Schuett, J. (2023). “Coordinated pausing: An evaluation-based coordination scheme for frontier AI developers”.
- Alayrac, J.-B., Donahue, J., Luc, P. vd. (2022). “Flamingo: a Visual Language Model for Few-Shot Learning”.
- Albahri, A.S., Duham, A.M., Fadhel, M.A. vd. (2023). “A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion”. *Information Fusion*, 96: 156-191.
- Allyn-Feuer, A. ve Sanders, T. (2023). “Transformative AGI by 2043 is <1% likely”.
- Ananny, M. (2016). “Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness”. *Science, Technology, & Human Values*, 41(1): 93-117.
- Anders, G. (2018). *İnsanın Eskimişliği. Üçüncü Endüstri Devrimi Çağında Yaşamın Tahribatı Üzerine*. (Çev. H. Belen ve H. Ertürk), İthaki, İstanbul.
- Andrada, G., Clowes, R.W. ve Smart, P.R. (2023). “Varieties of transparency: exploring agency within AI systems”. *AI & Society*, 38: 1321-1331.

- Arf, C. (1959). “Makine Düşünebilir mi ve Nasıl Düşünebilir?” *Atatürk Üniversitesi, Üniversite Çalışmalarını Muhite Yayma ve Halk Eğitimi Yayınları Konferanslar Serisi*, No. 1, Erzurum: 91-103.
- Arksey, H. ve Knight, P. (1999). *Interviewing for social scientists*. Sage Publications, New York.
- Asimov, I. (1950). *Runaround*, in I, Robot, (The Isaac Asimov Collection ed.) Doubleday, New York City. ISBN 0-385-42304-7.
- Astakhov, S. (2020). “Phenomenology vs Symbolic AI: Hubert Dreyfus’s Philosophy of Skill Acquisition”. *Philosophical Literary Journal Logos*.
- Awad, E., Dsouza, S., Kim, R. vd. (2018). “The moral machine experiment” *Nature*, 563(7729): 59–64.
- Aydın-Çakır, A. ve Türkeş-Kılıç, S. (2021). “Bilimsel Çalışmalarda Karma Yöntem Nasıl Kullanılır?” *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 42(Özel Sayı 1): Ö1-Ö15.
- Baki, A. ve Gökçek, T. (2012). “Karma Yöntem Araştırmalarına Genel Bir Bakış”. *Elektronik Sosyal Bilimler Dergisi*, 11(42): 001-021.
- Baltacı, A. (2019). “Nitel araştırma süreci: Nitel bir araştırma nasıl yapılır?”. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(2): 368-388.
- Barad, K. (1998). “Getting Real: Technoscientific Practices and the Materialization of Reality”, *In Differences: A Journal of Feminist Cultural Studies*, 10(2) (Summer).
- Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press, Durham & London.
- Bareis, J. ve Katzenbach, C. (2022). “Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics”. *Science, Technology, & Human Values*, 47(5): 855-881.
- Baum, S. D. (2018). “Reconciliation between Factions Focused on Near-term and Long-term Artificial Intelligence”, *AI & Society*, 33(4): 565–572.

- Baumane-Vitolina, I., Cals, I. ve Sumilo, E. (2015). "Is Ethics Rational? Teleological, Deontological and Virtue Ethics Theories Reconciled in the Context of Traditional Economic Decision Making". 3rd Global Conference on Business, Economic, Management and Tourism, 26-28 November 2015, Rome, Italy. *Procedia Economics and Finance*, 39(2016): 108-114.
- Baumgartner, H. ve Homburg, C. (1996). "Applications of structural equation modeling in marketing and consumer research: A review.". *International Journal of Research in Marketing*, 13(2): 139-161.
- Beer, D. (2014). "The biopolitics of biometrics. An interview with Btihaj Ajana". *Theory, Culture & Society*, 31(7/8): 330–336.
- van Berkel, N., Goncalves, J., Russo, D., Hosio, S. ve Skov. M. B. (2021). "Effect of Information Presentation on Fairness Perceptions of Machine Learning Predictors". In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, NY, USA, 1-13.
- Bietti, E. (2021). "From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy". 1-15.
- Birhane, A., Ruane, E., Laurent, T. vd. (2022). "The Forgotten Margins of AI Ethics". In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, USA, 1-17.
- Black, E. ve Richmond, R. (2019). "Improving early detection of breast cancer in sub-Saharan Africa: why mammography may not be the way forward". *Globalization and Health*, 15:3.
- Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer, Cham.
- Boiko, D. A., MacKnight, R. ve Gomes, G. (2023). "Emergent autonomous scientific research capabilities of large language models".
- Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. ve Kalai, A. (2016). "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". *Advances in Neural Information Processing Systems*, 29 (NIPS 2016): 4349–4357.

- Bommasani, R. vd. (2022). “On the Opportunities and Risks of Foundation Models”. Center for Research on Foundation Models (CRM), Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University. 1-214.
- Borenstein, J. ve Howard, A. (2021). “Emerging challenges in AI and the need for AI ethics education”. *AI and Ethics*, 1(1): 61–65.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford Univ. Press, London, UK.
- Boulanin, V., Saalman, L., Topychkanov, P., Su, F. ve Carlsson, M. P. (2020a). “Artificial Intelligence, Strategic Stability and Nuclear Risk”. *Stockholm International Peace Research Institute (SIPRI)*. Stockholm.
- Boulanin, V., Brockmann, K. ve Richards, L. (2020b). “Responsible Artificial Intelligence Research and Innovation for International Peace and Security”. *Stockholm International Peace Research Institute (SIPRI)*. Solna, Sweden.
- Brundage, M., Avin, S., Clark, J. vd. (2018). “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation”. *arXiv*.
- Buchanan, B., Lohn, A., Musser, M. ve Sedova, K. (2021). “Truth, lies, and automation: How language models could change disinformation” Center for Security and Emerging Technology.
- Buolamwini, J. ve Gebru, T. (2018). “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification” *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research (PMLR)*, 81:77-91.
- Byman, D. L., Gao, C., Meserole, C. ve Subrahmanian, V. S. (2023). “Deepfakes and International Conflict”. Foreign Policy at Brookings. The Brookings Institution, Washington, D.C.
- Byrne, B. M. (2011). *Structural equation modeling with AMOS Basic concepts, applications, and programming*. (Multivariate Applications Series). Routledge, New York.

- CAIDP. (2023). "Artificial Intelligence and Democratic Values Index". *Center for AI and Digital Policy*. Washington, DC.
- Cave, S. ve ÓhÉigeartaigh, S. S. (2019). "Bridging near- and long-term concerns about AI". *Nature Machine Intelligence*, 1: 5–6.
- CBDDO. (2021). "Ulusal Yapay Zekâ Stratejisi 2021-2025". T.C. Sanayi ve Teknoloji Bakanlığı, Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi.
- Chesney, B. ve Citron, D. (2019). "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security". *California Law Review*, 107(6): 1753–1820.
- Cho, H. (2022). "Privacy helplessness on social media: its constituents, antecedents and consequences". *Internet Research*, 32(1): 150-171.
- Coeckelbergh, M. ve Gunkel, D. J. (2023). "ChatGPT: deconstructing the debate and moving it forward". *AI & Society*.
- Cointe, N., Bonnet, G. ve Boissier, O. (2016). "Ethical Judgment of Agent's Behaviors in Multi-Agent Systems". In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS'16)*, 1106-1114.
- Creswell, J. W. (2007). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications, Thousand Oaks California.
- Cypress, B. S. (2017). "Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations". *Dimensions of Critical Care Nursing*, 36(4): 253-263.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2014). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamalar*. Pegem Akademi.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2012). *Sosyal Bilimler İçin Çok Değişkenli İstatistik: SPSS ve Lisrel Uygulamaları*. Pegem Akademi Yayıncılık, Ankara.
- Dennett, D. (1979). "Artificial Intelligence as Philosophy and as Psychology". *Philosophical Perspectives in Artificial Intelligence*, M. Ringle, ed., Atlantic Highlands, NJ: Humanities Press: 57–80.

- Doll, W.J., Xia, W. ve Torkzadeh, G. (1994). “A confirmatory factor analysis of the end-user computing satisfaction instrument”. *MIS QUARTERLY*, 18(4): 357–369.
- Edwards, R. ve Holland, J. (2013). *What is qualitative interviewing?*. Bloomsbury Publishing, London.
- Effoduh, J. O. (2021). “Policy Brief: Towards A Rights-Respecting Artificial Intelligence Policy for Nigeria”. *Paradigm Initiative*.
- Eichenberg, R. C. (2007). “Gender Differences in Support for the Use of Military Force in Cross-National Perspective: The War System, Modernization, and the Universal Logics of Military Action”. 1-18.
- Ekolu, S. O. ve Quainoo, H. (2019). “Reliability of assessments in engineering education using Cronbach’s alpha, KR and split-half methods”. *Global Journal of Engineering Education*, 21(1): 24-29.
- EPRS. (2020). “European framework on ethical aspects of artificial intelligence, robotics and related technologies”. *European Parliamentary Research Service*. Brussels.
- Escobar, A. (2018). *Designs for the Pluriverse. Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press, Dunham, NC.
- Esposito, M. (2019). “In the beginning was the hand: Ernst Kapp and the relation between machine and organism”. *Revista De Humanidades De Valparaíso*, (14): 117–138.
- European Institute of Innovation and Technology (2021). “Creation of a Taxonomy for the European AI Ecosystem: A report of the Cross-KIC Activity, “Innovation Impact Artificial Intelligence””. European Union.
- European Parliament (2023). “Draft Compromise Amendments”. Proposal for a regulation of the European Parliament and the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. (COM(2021)0206 – C9 0146/2021 – 2021 – 2021/0106(COD)).
- Federal Office for Information Security. (2023). “Large Language Models: Opportunities and risks for industry and authorities”. Bonn.

- Feldstein, S. (2019). "The Global Expansion of AI Surveillance". *Carnegie Endowment for International Peace*. Washington, DC.
- Feng, Z. (2018). "Does AI Share Same Ethic with Human Being?". 2nd International Conference on Intelligence Science (ICIS), Nov. 2018, Beijing, China. 465-472.
- Ferikoğlu, D. ve Akgün, E. (2022). "An investigation of teachers' artificial intelligence awareness: A scale development study". *Malaysian Online Journal of Educational Technology*, 10(3): 215-231.
- Flanagan, M., D. Howe ve H. Nissenbaum. (2008). "Embodying Values in Technology: Theory and Practice". In *Information Technology and Moral Philosophy*, edited by J. van den Hoven and J. Weckert, 322-53. Cambridge University Press, Cambridge, UK.
- Floridi, L. (2013). "Distributed morality in an information society". *Sci Eng Ethics*, 19(3):727-743.
- Floridi, L. (2020). "AI and Its New Winter: from Myths to Realities" *Philosophy & Technology*, 33: 1-3.
- Flowers, J. C. (2019). "Strong and Weak AI: Deweyan Consideration". AAAI Spring Symposium: Towards Conscious AI Systems 2019.
- Fogg, B. J. (2002). "Persuasive technology: using computers to change what we think and do". *Ubiquity*, 5, Computers as Persuasive Social Actors: 89-120.
- Forsyth, D. R. (1980). "A taxonomy of ethical ideologies". *Journal of Personality and Social Psychology*, 39(1): 175-184.
- Forsyth, D. R. (1992). "Judging the morality of business practices: the influence of personal moral philosophies". *Journal of Business Ethics*, 11(5): 461-470.
- Forsyth, D. R., O'Boyle, E. H. ve McDaniel, M. A. (2008). "East Meets West: A Meta-Analytic Investigation of Cultural Variations in Idealism and Relativism". *Journal of Business Ethics*, 83: 813-833.

- Franken, S. ve Mauritz, N. (2020). "Gender and Artificial Intelligence – Differences Regarding the Perception, Competence Self-Assessment and Trust". Gender und KI. FH Bielefeld University of Applied Sciences.
- Gasser, U. ve Almeida, V. A. F. (2017). "A Layered Model for AI Governance". *IEEE Internet Computing*, 21(6): 58-62.
- Gillespie, N., Lockey, S. ve Curtis, C. (2021). "Trust in Artificial Intelligence: A Five Country Study", The University of Queensland and KPMG Australia.
- Gillespie, N., Lockey, S., Curtis, C. Pool, J. ve Akbari, A. (2023). "Trust in Artificial Intelligence: A Global Study". The University of Queensland and KPMG Australia.
- Ginzburg, C. ve Davin, A. (1980). "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method", *History Workshop*, 9: 5–36.
- Glikson, E. ve Williams Woolley, A. (2020). "Human Trust in Artificial Intelligence: Review of Empirical Research". *Academy of Management Annals*, 14(2): 627-660.
- Gold, J. M. ve Rogers, J. D. (1995). "Intimacy and isolation: A validation study of Erikson's theory". *Journal of Humanistic Psychology*, 35: 78-86.
- Goldstein, J. A., Sastry, G., Musser, M. vd. (2023). "Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations".
- Gonzalez, L., Montes, G., Puig, E., Johnson, S. vd. (2016). "Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation". *Sensors*, 16(1): 97. MDPI AG.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P. ve Ditto, P. H. (2013). "Moral foundations theory: The pragmatic validity of moral pluralism". In *Advances in Experimental Social Psychology*. Elsevier, (47): 55–130.
- Grassini, S. ve Sævilid Ree, A. (2023). "Hope or Doom AI-ttitude? Examining the Impact of Gender, Age, and Cultural Differences on the Envisioned Future Impact of Artificial Intelligence on Humankind". *Proceedings of the European Conference on Cognitive Ergonomics (ECCE '23)*, September 2023. 15: 1-7.

- Greene, J. ve Haidt, J. (2002). "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences*, 6: 517-523.
- Grgić-Hlača, N., Lima, G., Weller, A. ve Redmiles, E. M. (2022). "Dimensions of Diversity in Human Perceptions of Algorithmic Fairness". In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, Oct. 6–9, 2022, Arlington, VA, USA. ACM, New York, NY, USA, 1-12.
- Hagendorff, T. (2016). "Wirksamkeitssteigerungen gesellschaftskritischer Diskurse". *Soziale Probleme*, 27: 1-16.
- Hagendorff, T. (2020). "The ethics of AI ethics: an evaluation of guidelines". *Mind. Mach.*, 30: 99-120.
- Helbing, D., Frey, B., Gigerenzer, G. ve Hafen, E. (2019). "Will Democracy Survive Big Data and Artificial Intelligence?: Essays on the Dark and Light Sides of the Digital Revolution". In *Towards Digital Enlightenment*, Springer International Publishing. 73-98.
- Helmus, T. C. (2022). "Artificial Intelligence, Deepfakes, and Disinformation: A Primer" RAND Corporation.
- Hendrycks, D., Mazeika, M. ve Woodside, T. (2023). "An Overview of Catastrophic AI Risks". *arXiv*. 1-54.
- Hewstone, M., Rubin, M. ve Willis, H. (2002). "Intergroup bias". *Annual Review of Psychology*, 53(1): 575-604.
- Hickok, M. (2021). "Lessons learned from AI ethics principles for future actions". *AI and Ethics*, 1(1): 41-47.
- Hongladarom, S. (2020). "Machine hermeneutics, postphenomenology, and facial recognition technology". *AI & Society*, 38(6): 2151-2158.
- Hongladarom, S. ve Bandasak, J. (2023). "Non-western AI ethics guidelines: implications for intercultural ethics of technology". *AI & Society*. 1-14.

- Hooper, D., Coughlan, J. ve Mullen, M. (2008). “Structural Equation Modeling: Guidelines for Determining Model Fit”. *Electronic Journal of Business Research Methods*, 6(1): 53-60.
- Hu, L. T. ve Bentler, P. M. (1999). “Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives”. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1): 1-55.
- Hu, Y., Jing, X., Ko, Y. ve Rayz, J. T. (2021). “Misspelling Correction with Pre-trained Contextual Language Model”.
- Hussain, S., Neekhara, P., Dolhansky, B. vd. (2022). “Exposing Vulnerabilities of Deepfake Detection Systems with Robust Attacks”. *Digital Threats: Research and Practice*, 3(3), 30: 1-23.
- IEEE. (2023). “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems”. Version 2 – For Public Discussion. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.
- Igarape Institute. (2019). “Future Crime: Assessing twenty first century crime prediction”.
- Ikkatai, Y., Hartwig, T., Takanashi, N. ve Yokoyama, H. M. (2022). "Octagon Measurement: Public Attitudes toward AI Ethics". *International Journal of Human-Computer Interaction*, 38(17): 1589-1606.
- ISO/IEC 22989:2022 (E). (2022). “Information technology – Artificial intelligence – Artificial intelligence concepts and terminology”. First ed., 2022-07. 1-70.
- Jakesch, M., Hancock, J. T. ve Naaman, M. (2023). “Human heuristics for ai-generated language are flawed”. *Proceedings of the National Academy of Sciences*, 120(11): e2208839120.
- Jang, Y., Choi, S. ve Kim, H. (2022). “Development and validation of an instrument to measure undergraduate students’ attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education”. *Education and Information Technologies*, 27: 11635-11667.

- Jobin, A., Ienca, M. ve Vayena, E. (2019). "The global landscape of AI ethical guidelines". *Naure Machine Intelligence*, 1: 389-399.
- Johnson, R. B. ve Onwuegbuzie, A. J. (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come". *Educational Researcher*, 33(7): 14–26.
- Josselson, R. (2013). *Interviewing for qualitative inquiry: A relational approach*. Guilford Press, New York.
- Kalaycı, Ş. (Ed.) (2008). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Asil Yayıncılık, Ankara.
- Kalluri, P. R., Agnew, W., Cheng, M. vd. (2023). "The Surveillance AI Pipeline". *arXiv*. 1-23.
- Karadağ, R. (2013). "Developing an attitude scale for cursive handwriting: Validity and reliability study". *Academic Journals, Educational Research and Reviews*, 1409-1420.
- Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetisensoy, O. ve Demir Kaya, M. (2024). "The Roles of Personality Traits, AI Anxiety and Demographic Factors in Attitudes toward Artificial intelligence". *International Journal of Human-Computer Interaction*, 40(2): 497-514.
- Kline, R.B. (2011). *Principles and practice of structural equation modeling*. The Guilford Press, New York.
- Ko, Y-H. ve Leem, C-S. (2021). "The Influence of AI Technology Acceptance and Ethical Awareness towards Intention to Use". *Journal of Digital Convergence*, 19(3): 217-225.
- Komorowski, P., Baniecki, H., Biecek, P. (2023). "Towards evaluating explanations of vision transformers for medical imaging".
- Kornwachs, K. (2021). *Teknoloji Felsefesine Giriş*. (Çev. S. Vural Kara), Runik Kitap No:118, Runik Bilgi Serisi: 58.
- Kostka, G. (2019). "China's social credit systems and public opinion: explaining high levels of approval". *New Media & Society*, 21(7): 1565-1593.

- Kreps, S., McCain, R. M. ve Brundage, M. (2022). "All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation". *Journal of Experimental Political Science*, 9(1): 104–117.
- Kshetri, N. (2020). "China's Social Credit System: Data, Algorithms and Implications". *IEEE IT Professional*, 22(2): 14-18.
- Kurshan, E., Chen, J., Storch, V. ve Shen, H. (2021). "On the current and emerging challenges of developing fair and ethical AI solutions in financial services". In: *Proceedings of the 2nd ACM International Conference on AI in Finance (ICAIF'21)*.
- Kurzweil, R. (2016). *İnsanlık 2.0: Tekillğe Doğru Biyolojisini Aşan İnsan*. (Çev. M. Şengel), Alfa Yayıncılık, İstanbul.
- Lagerkvist, A., Tudor, M., Smolicki, J. vd. (2022). "Body stakes: an existential ethics of care in living with biometrics and AI". *AI & Society*, 1-13.
- Larsson, S. (2020). "On the Governance of Artificial Intelligence through Ethics Guidelines". *Asian Journal of Law and Society*, 7(3): 437-451.
- Latour, B. ve Venn, C. (2002). "Morality and Technology". *Theory, Culture and Society*, 19(5/6): 247-260.
- Leech, B. L. (2002). "Asking questions: techniques for semistructured interviews". *Political Science and Politics*, 35(04): 665-668.
- Legg, S. ve Hutter, M. (2007). "A Collection of Definitions of Intelligence". *Frontiers in Artificial Intelligence and Applications*, 157: 17-24.
- Leslie, D. (2019). "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector". *The Alan Turing Institute*.
- Li, C. (2020). "Biodiversity assessment based on artificial intelligence and neural network algorithms". *Microprocessors and Microsystems*, 79: 103321.

- Manders-Huits, N. ve Zimmer, M. (2009). "Values and pragmatic action: the challenges of introducing ethical intelligence in technical design communities". *The International Review of Information Ethics*, 10: 37–44.
- Maslej, N. Fattorini, L., Brynjolfsson, E. vd. (2023). "The AI Index 2023 Annual Report". AI Index Steering Committee, Institute for Human-Centered AI. Stanford University, Stanford, CA.
- Matz, S. C., Kosinski, M., Nave, G. ve Stillwell, D. (2017). "Psychological targeting as an effective approach to digital mass persuasion". *Proceedings of the National Academy of Sciences of the United States of America*, 114: 12714–12719.
- McCarthy, H., Minsky, M. L., Rochester, N. ve Shannon, C. E. (2006). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955". *AI Magazine*, 27(4): 12.
- McCulloch, W. S. ve Pitts, W. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*, 5: 115-133.
- McGill, T. ve Thompson, N. (2021). "Exploring potential gender differences in information security and privacy". *Information & Computer Security*, 29(5): 850-865.
- McLuhan, M. (2013). *Understanding Media: The Extensions of Man*. Gingko Press, Berkeley, California.
- Merriam, S. B. ve Tisdell, E. J. (2015). *Qualitative Research: A Guide to Design and Implementation*. Jossey-Bass, San Francisco.
- Merrill, J. C. (2011). "Theoretical Foundations for Media Ethics". In *Controversies in Media Ethics*, 3rd ed., ed. A. D. Gordon, J. M. Kittross, J. C. C. Merrill, W. Babcock, and M. Dorsher, 3-32. Routledge, New York.
- Milan, S. (2020). "Techno-solutionism and the standard human in the making of the COVID-19 pandemic". *Big Data & Society*, 7(2): 1-7.
- Miles, M. B. ve Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage Publications, Thousand Oaks, California.

- Miller, S. (2017). "Feminist Ethics". In Carol Hay (Ed.), *Philosophy: Feminism*. Farmington Hills, MI: Macmillan. 189-213.
- Minkkinen, M., Zimmer, M. P. ve Mäntymäki, M. (2023). "Co-Shaping an Ecosystem for Responsible AI: Five Types of Expectation Work in Response to a Technological Frame". *Information Systems Frontiers*, 25, 103-121.
- Mittelstadt, B. D., Allo, P., Taddeo, M. vd. (2016). "The ethics of algorithms: mapping the debate". *Big Data & Society*, 3(2): 1-21.
- Mittelstadt, B. (2019). "Principles alone cannot guarantee ethical AI". *Nature Machine Intelligence*, 1(11): 501–507.
- Mittelstadt, B. (2021). "Interpretability and Transparency in Artificial Intelligence". In Carissa Véliz (ed.), *The Oxford Handbook of Digital Ethics (online ed., Oxford Academic, 10 Nov. 2021)*.
- Mosier, K. L. ve Skitka, L. J. (1996). "Human decision makers and automated decision aids: Made for each other?" In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application*, 201–220. Mahwah, NJ: Erlbaum.
- Mökander, J., Schuett, J., Kirk, H. R. ve Floridi, L. (2023). "Auditing large language models: a three-layered approach". *arXiv*. 1-36.
- Murphy, R. ve Woods, D. D. (2009). "Beyond Asimov: The Three Laws of Responsible Robotics". *IEEE Intelligent Systems, Human-Centered Computing*, 24(4): 14–20.
- Nemitz, P. (2018). "Constitutional democracy and technology in the age of artificial intelligence". *Phil. Trans. R. Soc. A.*, 376: 20180089.
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M. vd. (2018). "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning". *PNAS*, 115(25): E5716-E5725.
- NSTC. (2016). "Preparing for the Future of Artificial Intelligence". Executive Office of the President. National Science and Technology Council Committee on Technology. United States.

- OECD (2021). “Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers”.
- Onwuegbuzie, A. J. ve Collins, K. M. T. (2007). “A Typology of Mixed Methods Sampling Designs in Social Science Research”. *The Qualitative Report*, 12(2): 281-316.
- Özdamar, K. (2013). *Paket programlar ile istatistiksel veri analizi-I: SPSS-MINITAB*. Nisan Kitapevi, Eskişehir.
- Parasuraman, R. ve Manzey, D. H. (2010). “Complacency and bias in human use of automation: an attentional integration”. *Hum Factors*, 52(3): 381-410.
- Patton, M. Q. (2018). *Nitel Araştırma ve Değerlendirme Yöntemleri*. 2. Baskı. (Çev. Ed. M. Bütün ve S. Demir), Pegem Akademi, Ankara.
- Paulsen, K. (2017). “2: Uncanny Confusion: Early Video and the Fantasy of Presence”. In *Here/There: Telepresence, Touch, and Art at the Interface*. The MIT Press.
- Pew Research Center. (2022). “AI and Human Enhancement: Americans’ Openness Is Tempered by a Range of Concerns”. March 2022.
- Pierson, E. (2017). “Gender differences in beliefs about algorithmic fairness”. *arXiv preprint arXiv:1712.09124*. 1-5.
- Polat, A. (2022). “Nitel araştırmalarda yarı-yapılandırılmış görüşme soruları: Soru form ve türleri, nitelikler ve sıralama”. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 22(Özel Sayı 2): 161-182.
- Pötzsch, H. (2019). “Critical Digital Literacy: Technology in Education Beyond Issues of User Competence and Labour-Market Qualifications”. *tripleC, Journal for a Global Sustainable Information Society*, 17(2): 221-240.
- Ricaurte, P. (2022). “Ethics for the majority world: AI and the question of violence at scale”. *Media, Culture & Society*, 44(4): 726–745.
- Roberts, R. E. (2020). “Qualitative interview questions: guidance for novice researchers”. *The Qualitative Report*, 25(9): 3185-3203.

- Roberts, H., Cows, J., Morley, J., Taddeo, M., Wang, V. ve Floridi, L. (2021). "The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation". *AI & Society*, 39: 59-77.
- Robertson, G. ve Watson, I. (2014). "A Review of Real-Time Strategy Game AI". *AI Magazine*, 35(4): 75-104.
- Rose, P., Beeby, J. ve Parker, D. (1995). "Academic rigour in the lived experience of researchers using phenomenological methods in nursing". *Journal of Advanced Nursing*, 21(6): 1123-1129.
- Rubin, H. J. ve Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data*. Sage, Los Angeles.
- Russell, S. J. ve Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. 3rd ed. Saddle River, Prentice Hall, NJ.
- Saenz, P. ve Spanger, A. Fundación Karisma. (2018). "Cámaras InDiscretas". Bogotá, Colombia.
- Samoili, S., López Cobo, M., Delipetrev, B. vd. AI Watch. (2021). "Defining Artificial Intelligence 2.0. Towards an operational definition and taxonomy for the AI landscape", EUR 30873 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-42648-6, JRC126426.
- Samuel, A. L. (1960). "Some Moral and Technical Consequences of Automation – A Refutation". *Science*, 132(3429): 741-742.
- Sancho-Gil, J. M. (2020). "Digital technology as a trigger for learning promises and realities" *Digital Education Review*, 37: 195-207.
- Sandbrink, J. B. (2023). "Artificial intelligence and biological misuse: Differentiating risks of language models and biological design tools". *arXiv*. 1-9.
- Sasse, M. A., Brostoff, S. ve Weirich, D. (2001). "Transforming the 'Weakest Link' – A Human/Computer Interaction Approach for Usable and Effective Security". *BT Technology Journal*, 19(3): 122-131.

- Sayan, P. (2023). "Gender Differences in Perceptions Towards the Use of Force". *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 25(3): 1328-1345.
- Sayler, K. M. (2020). "Artificial Intelligence and National Security". Congressional Research Service.
- Sayler, K. M. ve Harris, L. A. (2022). "Deep Fakes and National Security". *Congressional Research Service*.
- Schermelleh-Engel, K., Moosbrugger, H., ve Müller, H. (2003). "Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures". *Methods of Psychological Research*, 8(2): 23–74.
- Schmitt, M. N. ve Thurnher, J. S. (2013). "“Out of the loop”: Autonomous weapon systems and LOAC". *Harvard National Security Journal*, 4(2): 231-281.
- Schuett, J., Dreksler, N., Anderljung, M. vd. (2023). "Towards best practices in AGI safety and governance: A survey of expert opinion". Centre for the Governance of AI.
- Searle, J. (1980). "Minds, brains and programs". *Behavioral and Brain Sciences*, 3(3): 417-4274.
- Seçer, İ. (2015). *Psikolojik Test Geliştirme ve Uyarlama Süreci*. Ankara: Anı Yayıncılık.
- Sharma, S., Sato, K. ve Gautam, B. P. (2023). "A Methodological Literature Review of Acoustic Wildlife Monitoring Using Artificial Intelligence Tools and Techniques". *Sustainability*, 15(9): 7128.
- Shin, D. (2021). "The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI". *International Journal of Human-Computer Studies*, 146, 102551.
- Shokri, R., Strobel, M. ve Zick, Y. (2021). "On the Privacy Risks of Model Explanations". *arXiv*.

- Speer, A. (1970). *Inside the third Reich*. The Macmillan Company, New York.
- Stahl, B. C. (2022). “Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence”. *International Journal of Information Management*, 62: 102441. 1-13.
- Stahl, B. C. (2023). “Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems”. *Scientific Reports*, 13: 7586.
- Spitale, G., Biller-Andorno, N. ve Germani, F. (2023). “AI model GPT-3 (dis)informs us better than humans”. *Science Advances*, 9(26). eadh 1850.
- Stiegler, B. (1998). *Technics and Time, 1: The Fault of Epimetheus*. (Trans. R. Beardsworth & G. Collins), Stanford University Press, Stanford, California.
- Stiegler, B. (2016). *Automatic Society: Volume 1, the Future of Work*. Cambridge: Polity.
- Stiegler, B. (2018). “Artificial Stupidity and Artificial Intelligence in the Anthropocene”. (D. Ross trans.), Institute of Ereignis, Shanghai.
- Stiegler, B. (2019). *Antroposenden Çıkmak*. (Çev. N. Ökten), Cogito: Yerküre Krizi, Dönüşen İnsan, 93: 199-211.
- Stinson, C. (2022). “Algorithms are not neutral”. *AI Ethics*, 2: 763-770.
- Stolfi, D. H., Brust M. R., Danoy, G. ve Bouvry, P. (2021). “UAV-UGV-UMV Multi-Swarms for Cooperative Surveillance”. *Frontiers in Robotic AI*. 8: 616950.
- Suchman, L. (1998). “Human/Machine Reconsidered”. *Cognitive Studies*, 5(1): 5-13.
- Sümer, N. (2000). “Yapısal Eşitlik Modelleri”. *Türk Psikoloji Yazıları*, 3(6): 49-74.

- Şan, E. (2022). “Bernard Stiegler’in Teknoloji Felsefesi Problemleri: Algoritmik Yönetimsellik ve Bilişsel Proleterleşme”. *ViraVerita E-Journal: Interdisciplinary Encounters*, 15: 105-135.
- Taddeo, M. ve Floridi, L. (2018). “How AI can be a force for good”. *Science*, 361(6404): 751–752.
- Tal, E. (2023). “Target specification bias, counterfactual prediction, and algorithmic fairness in healthcare”. 1-19.
- T.C. Kişisel Verileri Koruma Kurumu. (2019). KVKK Kişisel Veri Saklama ve İmha Politikası. KVKK Veri Yönetimi Dairesi Başkanlığı.
- Tekindal, M. ve Uğuz Arsu, Ş. (2020). “Nitel Araştırma Yöntemi Olarak Fenomenolojik Yaklaşımın Kapsamı ve Sürecine Yönelik Bir Derleme”. *Ufkun Ötesi Bilim Dergisi*, 20(1): 153-182.
- Terzi, R. (2020). “An adaptation of Artificial Intelligence Anxiety Scale into Turkish: Reliability and validity study”. *International Online Journal of Education and Teaching*, 7(4): 1501-1515.
- Tiit, E. M. (2021). “Impact of voluntary sampling on estimates”, *Papers on Anthropology*, 30(2): 9-13.
- Toussaint, L. ve Webb, J. R. (2005). “Gender Differences in the Relationship Between Empathy and Forgiveness”. *Journal of Social Psychology*, 145(6): 673-685.
- Turing, A. M. (1950). “Computing machinery and intelligence”. *Mind*, 59: 433-460.
- Tversky, A. ve Kahneman, D. (1974). “Judgment under Uncertainty: Heuristics and Biases”. *Science*, 185(4157): 1124-1131.

- Umbrello, S. (2021). “AI Winter”. In Michael Klein & Philip Frana (eds.), *Encyclopedia of Artificial Intelligence: The Past, Present, and Future of AI*. Santa Barbara, USA: ABC-Clio. 7-8.
- UNESCO. (2021). “Recommendation on the ethics of artificial intelligence”. SHS/BIO/REC-AIETHICS/2021. UNESCO [6591].
- UNESCO. (2022). “Recommendation on the Ethics of Artificial Intelligence”. UNESCO, Paris, France.
- Urbina, F., Lentzos, F., Invernizzi, C. ve Ekins, S. (2022). “Dual use of artificial-intelligence powered-drug discovery”. *Nature Machine Intelligence*, 4: 189-191.
- Ursachi, G., Horodnic, A. ve Zait, A. (2015). “How reliable are measurement scales? External factors with indirect influence on reliability estimators”. *Procedia Economics and Finance*, 20: 679-686.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. ve Polosukhin, I. (2017). “Attention is All You Need”. *arXiv*, 1-15.
- Verma, S. ve Rubin, J. (2018). “Fairness Definitions Explained”. *FairWare’18: IEEE/ACM International Workshop on Software Fairness, Gothenburg, Sweden*. 1-7.
- Vinuesa, R., Azizpour, H., Leite, I. vd. (2020). “The role of artificial intelligence in achieving the Sustainable Development Goals”. *Nature Communications*, 11: 233.
- Vosoughi, S., Roy, D. ve Aral, S. (2018). “The spread of true and false news online”. *Science*, 359(6380): 1146–1151.
- Waldemarsson, C. (2020). “Disinformation, Deepfakes and Democracy: The European Response to Election Interference in the Digital Age”. *Alliance of Democracies Foundation, Copenhagen*.

- Wang, S., Lilienfeld, S. O. ve Rochat, P. (2015). "The Uncanny Valley: Existence and Explanations". *Review of General Psychology*, 19(4): 393: 407.
- Wardle, C. ve Derakhshan, H. (2017). "Information Disorder: Toward an interdisciplinary framework for research and policy making". Council of Europe report. DGI(2017)09. 1-109.
- Weinberger, M., Zhitomirsky-Geffet, M. ve Bouhnik, D. (2017). "Sex differences in attitudes towards online privacy and anonymity among Israeli students with different technical backgrounds". *Information Research*, 22(4): 1-23.
- Weiss, R. S. (1994). *Learning from strangers: the art and method of qualitative interview studies*. Free Press, New York.
- Weizenbaum, J. (1966). "ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine". *Communications of the ACM*. 9(1): 36-45.
- Westling, J. (2019). "Are Deep Fakes a Shallow Concern? A Critical Analysis of the Likely Societal Reaction to Deep Fakes". *TPRC47: The 47th Research Conference on Communication, Information, and Internet Policy 2019*.
- Whittemore, R., Chase, S. K. ve Mandle, C. L. (2001). "Validity in Qualitative Research". *Qual Health Res.*, 11(4): 522-537.
- Wiener, E. L. (1981). "Complacency: Is the term useful for air safety?". *In Proceedings of the 26th Corporate Aviation Safety Seminar*, 116–125. Denver, CO: Flight Safety Foundation.
- Winner, L. (1986). *The Whale and the Reactor*. Chicago University Press, Chicago.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. The Edinburg Press, Edinburg.

- Wong, P-H. (2009). "What Should We Share? Understanding the Aim of Intercultural Information Ethics". *ACM Sigcas Computers and Society*, 39(3): 50-58.
- Wong, P-H. (2016). "Responsible innovation for decent nonliberal peoples: a dilemma?". *Journal of Responsible Innovation*, 3(2): 154-168.
- Yam, K. C. (2018). "The Effects of Thought Suppression on Ethical Decision Making: Mental Rebound Versus Ego Depletion". *Journal of Business Ethics*, 147(1): 65-79.
- Yang, W., Wei, Y., Wei, H. vd. (2023). "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects". *Human-Centric Intelligent Systems*, 3: 161–188.
- Yeung, K. (2018). "Algorithmic regulation: A critical interrogation". *Regulation & Governance*, 12(4): 505-523.
- Yıldırım, A. ve Şimşek, H. (2011). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. Sekin Yayıncılık, Ankara.
- Yiu, E., Kosoy, E. ve Gopnik, A. (2023). "Imitation versus Innovation: What children can do that large language and language-and-vision models cannot (yet)?".
- Zahn-Waxler, C. ve Radke-Yarrow, M. (1990). "The origins of empathic concern". *Motivation and Emotion*, 14(2): 107-130.
- Zambak, A. ve Vergauwen, R. (2009). "Artificial Intelligence and Agentive Cognition: A Logico-Linguistic Approach". *Logique et Analyse*, 52(205): 57-96.
- Zerilli, J., Bhatt, U. ve Weller, A. (2022). "How transparency modulates trust in artificial intelligence". *Patterns*, 3(4): 1-10.
- Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Sellitto, M, Sakhaee, E., Shoham, Y., Clark, J. ve Perrault, R. (2022). "The

AI Index 2022 Annual Report”. *AI Index Steering Committee*, Stanford Institute for Human-Centered AI, Stanford University.

İnternet Kaynakları:

Adams, C. J. (2018). “The Feminist Ethics of Care”. <https://caroljadams.com/caregiving-intro>. (erişim tarihi: 17.07.2023).

AlSur. (2019). “Public consultation: Ethics and Data Protection in Artificial Intelligence: continuing the debate. A contribution from Latin America & the Caribbean” <https://www.alsur.lat/en/report/ethics-and-data-protection-artificial-intelligence-contribution-latin-america-caribbean>. (erişim tarihi: 09.06.2023).

Altman, S., Brockman, G. ve Sutskever, I. (2023). “Governance of superintelligence” <https://openai.com/blog/governance-of-superintelligence>. (erişim tarihi: 22.05.2023).

Anderljung, M. ve Scharre, P. (2023). “How to Prevent an AI Catastrophe: Society Must Get Ready for Very Powerful Artificial Intelligence”. *Foreign Affairs*. <https://www.foreignaffairs.com/world/how-prevent-ai-catastrophe-artificial-intelligence> (erişim tarihi: 14.08.2023).

Asociación por los Derechos Civiles. (2023). “ConMiCaraNo, Reconocimiento facial en la Ciudad de Buenos Aires”. <https://conmicarano.adc.org.ar/>. (erişim tarihi: 08.06.2023).

AWS. (2023). “Makine Öğrenimi ve Derin Öğrenme Arasındaki Fark Nedir?”. <https://aws.amazon.com/tr/compare/the-difference-between-machine-learning-and-deep-learning/>. (erişim tarihi: 05.06.2023).

Bengio, Y. (2023). “How Rogue AIs may Arise”. Yoshua Bengio. <https://yoshuabengio.org/2023/05/22/how-rogue-ais-may-arise/>. (erişim tarihi: 24.05.2023).

Bernstein, S. (2023). “Data Minimization: Centering Reasonable Consumer Expectation in the FTC’s Commercial Surveillance Rulemaking”. Electronic Privacy Information Center. <https://epic.org/data-minimization-centering-reasonable-consumer-expectation-in-the-ftcs-commercial-surveillance-rulemaking/> (erişim tarihi: 23.05.2023).

- Big Brother Watch. (2018). “Face Off: The lawless growth of facial recognition in UK policing”. <https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf>. (erişim tarihi: 05.06.2023).
- Borowski, A. (2022). “Philosopher of the apocalypse”. Aeon. <https://aeon.co/essays/gunther-anders-a-forgotten-prophet-for-the-21st-century>. (erişim tarihi: 28.05.2023).
- Bringsjord, S. ve Govindarajulu, N. S. (2022). “Artificial Intelligence”. The Stanford Encyclopedia of Philosophy (Fall 2022 ed.), Edward N. Zalta & Uri Nodelman (eds.) <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/>. (erişim tarihi: 03.06.2023).
- Britannica. (2023). “The Mechanical Turk: AI Marvel or Parlor Trick?”. <https://www.britannica.com/story/the-mechanical-turk-ai-marvel-or-parlor-trick>. (erişim tarihi: 04.08.2023).
- Burgess, M. (2023). “ChatGPT Has a Big Privacy Problem”. *Wired*. <https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/>. (erişim tarihi: 10.04.2023).
- Burt, A. (2019). “The AI Transparency Pradox”. *Harvard Business Review*. <https://hbr.org/2019/12/the-ai-transparency-paradox>. (erişim tarihi: 04.08.2023).
- Council of Europe Treaty Office. (2023). “Chart of signatures and ratifications of Treaty 223”. <https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treaty=223>. (erişim tarihi: 26.12.2023).
- Crabtree, M. (2023). “Machine Learning (ML) vs Deep Learning (DL): A Comparative Guide”. *Datacamp*. <https://www.datacamp.com/tutorial/machine-deep-learning>. (erişim tarihi: 29.09.2023).
- Datysoc. (2020). “Organizaciones de la sociedad civil y académicas expresan su preocupación por reconocimiento facial en el Proyecto de Ley de Presupuesto de Uruguay”.

<https://datysoc.org/2020/11/17/organizaciones-de-la-sociedad-civil-y-academicas-expresan-su-preocupacion-por-reconocimiento-facial-en-el-proyecto-de-ley-de-presupuesto-de-uruguay/> (erişim tarihi: 18.07.2023).

Davis, L. M. (2023). “The New Age of Hiring: AI Is Changing the Game for Job Seekers”. CNET. <https://www.cnet.com/tech/features/the-new-age-of-hiring-ai-is-changing-the-game-for-job-seekers/> (erişim tarihi: 11.06.2023).

De Vynck, G. (2023). “ChatGPT maker OpenAI faces a lawsuit over how it used people’s data”. *The Washington Post*. <https://www.washingtonpost.com/technology/2023/06/28/openai-chatgpt-lawsuit-class-action/>. (erişim tarihi: 28.06.2023).

Dittmer, J. (2023). “Applied Ethics”. *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/applied-ethics/#H1>. (erişim tarihi: 06.05.2023).

Elements of AI. (2023). “How should we define AI?”. University of Helsinki & MinnaLearn. <https://course.elementsofai.com/en-ie/1/1>. (erişim tarihi: 22.05.2023).

ESA. (2010). “ESA giving a space arm to Space Station”. *The European Space Agency*. https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/ESA_giving_a_spare_arm_to_Space_Station. (erişim tarihi: 12.06.2023).

Farid, H. (2023). “Deepfakes in the 2024 Presidential Election”. <https://farid.berkeley.edu/deepfakes2024election/>. (erişim tarihi: 09.08.2023).

Feingold, S. (2023). “The European Union’s Artificial Intelligence Act - explained”. *World Economic Forum*. <https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/>. (erişim tarihi: 03.07.2023).

Future of Life Institute. (2023). “Pause Giant AI Experiments”. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. (erişim tarihi: 29.05.2023).

- GPDP. (2023). “Provvedimento del 30 marzo 2023 [9870832]”. Garante Per La Protezione Dei Dati Personali. <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870832>. (erişim tarihi: 10.04.2023).
- Gulley, A. ve Hilliard, A. (2023). “Lost in Transl(A)t(I)on: Differing Definition of AI [Updated]”. Holistic AI. <https://www.holisticai.com/blog/comparing-definitions-of-ai>. (erişim tarihi: 15.11.2023).
- Hao, K. (2021). “How Facebook and Google fund global misinformation”. *MIT Technology Review*. <https://www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/>. (erişim tarihi: 27.07.2023).
- Hook, L. (2016). “The humans behind Mechanical Turk’s artificial intelligence” *Financial Times*. <https://www.ft.com/content/17518034-6f77-11e6-9ac1-1055824ca907> (erişim tarihi: 27.05.2023).
- Höne, K. (2022). “The Role of Trust in the German AI Strategy”. *Heinrich Böll Stiftung Tel Aviv*. <https://il.boell.org/en/2022/01/10/role-trust-german-ai-strategy>. (erişim tarihi: 03.06.2023).
- IBM. (2023a). “What is artificial intelligence (AI)”. <https://www.ibm.com/topics/artificial-intelligence>. (erişim tarihi: 05.08.2023).
- IBM. (2023b). “What is generative AI?” <https://research.ibm.com/blog/what-is-generative-AI/>. (erişim tarihi: 07.08.2023).
- Introna, L. (2017). “Phenomenological Approaches to Ethics and Information Technology”. *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.), Edward N. Zalta (ed.), <https://plato.stanford.edu/entries/ethics-it-phenomenology/>, (erişim tarihi: 03.04.2023).
- Jillson, E. (2021). “Aiming for truth, fairness, and equity in your company’s use of AI”. Federal Trade Commission. <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>. (erişim tarihi: 01.08.2023).

- Kaganer, E., Carmel, E., Hirschheim, R. ve Olsen, T. (2012). "Managing the Human Cloud". MIT Sloan Management Review. <https://sloanreview.mit.edu/article/managing-the-human-cloud/> (erişim tarihi: 03.06.2023).
- Keller, F., Schoch, D., Stier, S. ve Yang, J.-H. (2019). "It's not easy to spot disinformation on Twitter. Here's what we learned from 8 political 'astroturfing' campaigns". *The Washington Post*. <https://www.washingtonpost.com/politics/2019/10/28/its-not-easy-spot-disinformation-twitter-heres-what-we-learned-political-astroturfing-campaigns/>. (erişim tarihi: 18.05.2023).
- Knight, W. (2023). "Generative AI Is Making Companies Even More Thirsty for Your Data" Wired. https://www.wired.com/story/fast-forward-generative-ai-companies-thirsty-for-your-data/?bxdid=6432ce295b32040c3304e2a2&cndid=73501754&esrc=manage-page&mbid=mbid%3DCRMWIR012019%0A%0A&source=Email_0_EDT_WIR_NEWSLETTER_0_DAILY_ZZ&utm_brand=wired&utm_campaign=aud-dev&utm_content=WIR_Daily_081323&utm_mailing=WIR_Daily_081323&utm_medium=email&utm_source=nl&utm_term=P1 (erişim tarihi: 10.08.2023).
- Ladva, P. ve Grasso, A. (2023). "Benefits and use cases of AI insurance". Swiss Re. <https://www.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/opportunities-ai-insurance.html> (erişim tarihi: 27.06.2023).
- Lane, C. (2023). "Eye Scans Detect Parkinson's Years Before Symptoms Surface" Neuroscience News. <https://neurosciencenews.com/parkinsons-eye-scans-23796/>. (erişim tarihi: 26.08.2023).
- Mazzolin, R. (2020). "Artificial Intelligence and Keeping Humans "in the Loop"". The Centre for International Governance Innovation (CIGI). <https://www.cigionline.org/articles/artificial-intelligence-and-keeping-humans-loop/>. (erişim tarihi: 04.08.2023).
- Mori, M. (2012). "The Uncanny Valley: The Original Essay by Masahiro Mori". *IEEE Spectrum*. <https://spectrum.ieee.org/the-uncanny-valley>. (erişim tarihi: 17.06.2023).

- Murphy, H. ve Criddle, C. (2023). "Meta prepares chatbots with personas to try to retain users". *Financial Times*. <https://www.ft.com/content/fa76c8ce-cdfd-458c-baec-73dceb2d2ad5>. (erişim tarihi: 01.08.2023).
- Müller, V.C. (2023). "Ethics of Artificial Intelligence and Robotics", *The Stanford Encyclopedia of Philosophy* (Fall 2023 ed.), Edward N. Zalta & Uri Nodelman (eds.), <https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/>. (erişim tarihi: 20.06.2023).
- Nusca, A. (2013). "The 'largest AI program in history' led to Apple's Siri". *ZDNET*. <https://www.zdnet.com/article/the-largest-ai-program-in-history-led-to-apples-siri/>. (erişim tarihi: 20.05.2023).
- OECD. (2023). "OECD AI Principles overview". OECD.AI Policy Observatory. <https://oecd.ai/en/ai-principles>. (erişim tarihi: 09.11.2023).
- O'Neil, C. (2017). "Don't Grade Teachers with A Bad Algorithm". *Bloomberg*. <https://www.bloomberg.com/view/articles/2017-05-15/don-t-grade-teachers-with-a-bad-algorithm#xj4y7vzkg>. (erişim tarihi: 09.05.2023).
- OpenAI. (2023). "GPT-4". <https://openai.com/research/gpt-4>. (erişim tarihi: 05.06.2023).
- Pedersen, J. (2021). "The 5 Key Abilities of Mobile Manipulation Systems". *Robotic Business Review*. <https://www.roboticsbusinessreview.com/opinion/the-5-key-abilities-of-mobile-manipulation-systems/>. (erişim tarihi: 18.06.2023).
- Perrigo, B. (2023). "The Workers Behind AI Rarely See Its Rewards. This Indian Startup Wants to Fix That". *TIME*. <https://time.com/6297403/india-ai-karya-startup/>. (erişim tarihi: 28.07.2023).
- Poole, D. L. ve Mackworth, A. K. (2023). "Artificial Intelligence 3E: Foundations of Computational Agents". <https://artint.info/3e/html/ArtInt3e.C1.S2.html>. (erişim tarihi: 03.06.2023).

- PwC. (2017). "Sizing the price". <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. (erişim tarihi: 06.04.2023).
- Raicu, I. ve Kwan, J. (2023). "When It's About Power/Justice/Human Rights, It IS about Ethics". Markkula Center for Applied Ethics at Santa Clara University. <https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/when-its-about-powerjusticehuman-rights-it-is-about-ethics/>. (erişim tarihi: 29.09.2023).
- Ramani, A. ve Wang, Z. (2023). "Why transformative artificial intelligence is really, really hard to achieve" The Gradient. <https://thegradient.pub/why-transformative-artificial-intelligence-is-really-really-hard-to-achieve/> (erişim tarihi: 28.06.2023).
- Reese, H. (2022). "How to Survive the A.I. Revolution". Stanford Business Insights. <https://www.gsb.stanford.edu/insights/how-survive-artificial-intelligence-revolution>. (erişim tarihi: 04.04.2023).
- Reisner, A. (2023). "These 183,000 Books are Fueling the Biggest Fight in Publishing and Tech". The Atlantic. https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/?utm_source=copy-link&utm_medium=social&utm_campaign=share (erişim tarihi: 25.09.2023).
- Sayre-McCord, G. (2023). "Metaethics", The Stanford Encyclopedia of Philosophy (Spring 2023 ed.), Edward N. Zalta & Uri Nodelman (eds.). <https://plato.stanford.edu/archives/spr2023/entries/metaethics>. (erişim tarihi: 04.05.2023).
- Schwartz, O. (2019). "Untold History of AI: When Charles Babbage Played Chess With the Original Mechanical Turk". IEEE Spectrum. <https://spectrum.ieee.org/untold-history-of-ai-charles-babbage-and-the-turk>. (erişim tarihi: 03.05.2023).
- Seth, A. (2023). "Why Conscious AI Is a Bad Idea". *Nautilus*, <https://nautil.us/why-conscious-ai-is-a-bad-idea-302937/>. (erişim tarihi: 05.2023).

- Smith, G. (2018). “AI Lesson: 35 Years Ago, This Russian Saved the World”. *Hanson Robotics*.
<https://www.hansonrobotics.com/ai-lesson-stanislav-petrov/> (eriřim tarihi: 09.06.2023).
- Solon, O. (2023). “Trolls in Slovakian Election Tap AI Deepfakes to Spread Disinfo”. *Bloomberg*.
<https://www.bloomberg.com/news/articles/2023-09-29/trolls-in-slovakian-election-tap-ai-deepfakes-to-spread-disinfo?leadSource=uverify%20wall>. (eriřim tarihi: 03.10.2023).
- StatCounter. (2024). “Search Engine Market Share Worldwide”. *StatCounter, GlobalStats*.
<https://gs.statcounter.com/search-engine-market-share>. (eriřim tarihi: 22.01.2024).
- Stecklow, S. (2018). “Why facebook is losing the war on hate speech in Myanmar”. *Reuters*.
<https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>. (eriřim tarihi: 07.06.2023).
- Suchman, L. (2018). “Corporate accountability”. *Robot Futures. Lucy Suchman’s reflections on technocultures of humanlike machines*.
<https://robotfutures.wordpress.com/2018/06/10/corporate-accountability/>. (eriřim tarihi: 11. 05.2023).
- Suikkanen, J. (2023). “Normative Ethics”. *PhilPapers*.
<https://philpapers.org/browse/normative-ethics>. (eriřim tarihi: 08.06.2023).
- Tarnoff, B. (2023). “Weizenbaum’s nightmares: how the inventor of the first chatbot turned against AI”. *The Guardian*.
https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai?CMP=share_btn_tw. (eriřim tarihi: 02.08.2023).
- T.C. Dıřıřleri Bakanlıęı. (2022). “No: 350, 23 Kasım 2022, Trkiye’nin GPAI yelięi Hk.”
https://www.mfa.gov.tr/no_-350_-turkiye-nin-gpai-uyeligi-hk.tr.mfa. (eriřim tarihi: 24.06.2023).

Valentino-DeVries, J. (2020). “How the police use facial recognition, and where it falls short” *The New York Times*. <https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html>. (eriřim tarihi: 08.07.2023).

Ward, L. (2023). “How a Mechanical Chess-Playing Turk gave Birth to the AI Debate 250 Years Ago”. <https://www.messynessychic.com/2023/05/25/how-a-mechanical-chess-playing-turk-gave-birth-to-the-ai-debate-250-years-ago/>. (eriřim tarihi: 03.08.2023).

Winfield, A. (2018). “A round up of robotics and AI ethics: part 1 principles”. <https://robohub.org/a-round-up-of-robotics-and-ai-ethics-part-1-principles/>. (eriřim tarihi: 03.06.2023).



EKLER

EK 1 – Anket İzin Maili



RE: Permission to Use AT-EAI Questionnaire Tool
To: Neslihan Verda Özmen

28 March 2023 2:08 PM

Dear Neslihan Verda Özmen,

We grant you permission to use and modify the AT-EAI model for non-commercial research purposes. Please ensure you:

1. Reference our work in all publications using the model.
2. Do not use the model for commercial purposes.

Feel free to contact us with any questions.

Best regards,
Yeonju Jang
Creative Informatics & Computing Institute, Korea University



Neslihan Verda Özmen
Permission to Use AT-EAI Questionnaire Tool
To: [Redacted]

27 March 2023 6:54 PM

[Details](#)

Dear Yeonju Jang, Seongyune Choi and Hyeoncheol Kim,

I am a master's student from Akdeniz University completing a thesis in Management Information Systems.

I would like your permission to use the final version of AT-EAI (Attitudes toward the ethics of artificial intelligence) questionnaire in my research study to elicitate perceptions on ethics of AI. The research is being supervised by Dr. Tayfun Yörük who can be reached at tayfun@akdeniz.edu.tr. It is aimed to adapt the scale to Turkish and use it for research purposes in samples with a high level of education in Turkish culture. Considering that national and cultural factors play an important role in artificial intelligence ethics, it is also hoped that this study will contribute to increasing the generalizability of the results.

In addition to using the questionnaire only for my research study, I also ask your permission to reproduce it in my dissertation appendix.

If this is acceptable, please indicate so by replying to me through e-mail at ozmennv@gmail.com. If you have a specific statement of attribution that you would like for me to include, please provide it in your response as well.

Your kind consideration and positive response is deeply acknowledged in the fulfillment of this academic endeavor.

Sincerely,

Neslihan Verda Özmen

Note: The aforementioned questionnaire is presented in the attached article "Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of AI education" (2022), on "Appendix B - Table 14". *If there is an update other than "Author correction" that should be considered, we would be grateful if you let us know.*



s10639-022-11
086-5-1.pdf

EK 2 – Çevirileri Değerlendirme Formu

Madde no	Orijinal Metin	Değerlendirme	Türkçe çevirisi	Yorum
1	“When developing an AI that recognizes cats, it is preferable to use images of cats with typical appearances rather than images of cats with unusual appearances.*”	①②③④⑤	“Kedileri tanıyan bir yapay zekâ geliştirirken, alışılmadık görünümlere sahip kedilerin yerine tipik görünümlere sahip kedilerin görüntülerinin kullanılması tercih edilir.”	
2	“If the government develops AI robots to assist students in their studies and provides them for free, it should also provide them to elderly people.”	①②③④⑤	“Eğer hükümet, öğrencilere çalışmalarında yardımcı olmak için YZ destekli robotlar geliştiriyor ve bunları ücretsiz olarak sağlıyorsa, bunları yaşlı insanlara da sağlamalıdır.”	
3	“When developing an AI that recommends clothing that fits individuals, it should incorporate data from individuals with a variety of body types.”	①②③④⑤	“Bireylere uygun kıyafetler öneren bir YZ geliştirirken, çeşitli vücut tiplerine sahip bireylerden elde edilen veriler kullanılmalıdır.”	
4	“It is essential for AI to explain the reasons for its decisions.”	①②③④⑤	“YZ'nin kararlarının nedenlerini açıklaması çok önemlidir.”	
5	“Even if the AI does not explain why it made a particular decision, it is preferable to have higher accuracy.*”	①②③④⑤	“Yapay zekâ belirli bir kararı neden verdiğini açıklamasa bile, daha yüksek doğruluğa sahip olması tercih edilir.”	
6	“It is sufficient to provide an explanation to only experts as to why AI made its decision.*”	①②③④⑤	“YZ'nin kararını neden verdiğine dair yalnızca uzmanlara bir açıklama sunmak yeterlidir.”	
7	“AI that is unable to explain why it made a decision is likely to be untrustworthy.”	①②③④⑤	“Bir kararı neden verdiğini açıklayamayan YZ'nin güvenilirliği muhtemeldir.”	
8	“Individuals who utilize AI technology should make an effort to put it to good use.”	①②③④⑤	“YZ teknolojisini kullanan bireyler, bunu iyi bir şekilde kullanmak için çaba göstermelidir.”	
9	“It doesn't make much sense to be concerned about AI being misused right from the start.*”	①②③④⑤	“Yapay zekânın en başından itibaren kötüye kullanılmasından endişe duymak pek mantık değil.”	
10	“Before selling AI-based products, those who sell them should consider whether the product can be misused.”	①②③④⑤	“YZ tabanlı ürünleri satmadan önce, bunları satanlar ürünün kötüye kullanılıp kullanılmayacağını düşünmelidir.”	

11	"To avoid the potential misuse of AI, a lot of people should be educated on relevant ethical issues."	①②③④⑤	"YZ'nin potansiyel kötüye kullanımını önlemek için, birçok insan ilgili etik konular hakkında eğitilmelidir."	
12	"The government should regulate developers and companies that create AI to ensure that they do not freely use people's personal information."	①②③④⑤	"Hükümet, insanların kişisel bilgilerini özgürce kullanamadıklarından emin olmak için YZ yaratan geliştiricileri ve şirketleri düzenlemelidir."	
13	"When developing AI with people's personal information, causation must be exercised to avoid infringing on their privacy."	①②③④⑤	"İnsanların kişisel bilgileriyle YZ geliştirilirken, gizliliklerini ihlal etmekten kaçınmak için dikkatli olunmalıdır."	
14	"For the sake of our society's development, companies that create high-performance AI should be permitted to use personal information as they please.*"	①②③④⑤	"Toplumumuzun gelişimi için, yüksek performanslı yapay zekâ yaratan şirketlerin kişisel bilgileri istedikleri gibi kullanmalarına izin verilmelidir."	
15	"If any issues arise as a result of AI, the developer who created the AI must be held responsible.*"	①②③④⑤	"YZ'nin bir sonucu olarak herhangi bir sorun ortaya çıkarsa, YZ'yi oluşturan geliştirici sorumlu tutulmalıdır."	
16	"If an issue arises as a result of AI, the customer who purchased the product is not responsible.*"	①②③④⑤	"YZ'nin bir sonucu olarak bir sorun ortaya çıkarsa, ürünü alan müşteri sorumlu değildir."	
17	"In the event of any problems caused by AI, determining precisely who is to be held responsible is difficult; therefore, social consensus on who should compensate and how is required."	①②③④⑤	"YZ'nin neden olduğu herhangi bir sorun durumunda, tam olarak kimin sorumlu tutulacağını belirlemek zordur, bu nedenle kimin ve nasıl tazmin etmesi gerektiği konusunda toplumsal uzlaşma gereklidir."	

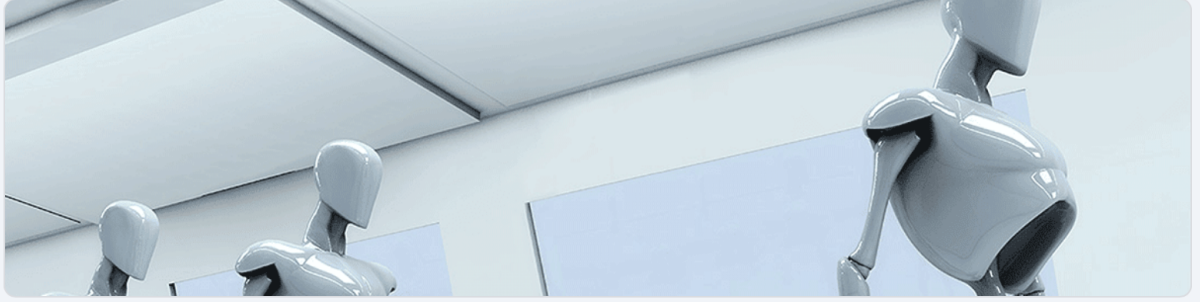
NOT: "Değerlendirme" kısmında yer alan seçenekler aşağıdaki gibidir:

- ① Hiç uygun değil (Maddenin uygun şekilde getirilmesi gerekir)
- ② Biraz uygun (Uygun ancak küçük değişiklik gerekir)
- ③ Uygun
- ④ Çok uygun
- ⑤ Tamamen uygun

İki metin birbiriyle tamamen örtüşüyor ise ⑤ kısmını işaretlemeniz (①②③④■), yapılan çeviride aksaklıklar tespit etmiş iseniz daha düşük puanlayarak “Yorum” kolonuna görülen aksaklıkları belirtmeniz ve ③ puanın altında puanlama yaptığınız maddeler için düzeltme önerisinde bulunmanız beklenmektedir.



EK 3 - Anket Formu (AT-EAI Ölçeği)



Bölüm 1/5

ÜNİVERSİTE ÖĞRENCİLERİNİN YAPAY ZEKA ETİĞİNE YÖNELİK TUTUMLARININ İNCELENMESİ KONULU ANKET



Bu anket formu, yüksek lisans tez araştırmasında kullanılmak üzere Doç. Dr. Tayfun YÖRÜK danışmanlığında Neslihan Verda ÖZMEN tarafından hazırlanmıştır. Söz konusu araştırma, üniversite öğrencilerinin yapay zeka etiğine yönelik tutumlarını incelemek amacıyla tasarlanmıştır. Anket gönüllülük esasına dayanmakla beraber, Demografik Bilgi Formu ve 17 maddeden oluşan Yapay Zeka Etiğine Yönelik Tutumlar Ölçeği'nin (AT-EAI) yanıtlanması yaklaşık 5 dakika sürmektedir. Araştırma ya da bulgular hakkında bilgi almak isterseniz aşağıda verilen adres üzerinden iletişime geçebilirsiniz. Değerli yardımlarınız ve katkılarınız için şimdiden teşekkürler.

E-mail: [REDACTED]

Neslihan Verda ÖZMEN, Yönetim Bilişim Sistemleri Tezli Yüksek Lisans Programı Öğrencisi

Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü

EK 4 – Yarı Yapılandırılmış Mülakat Formu

Aşağıdaki sorular; görüşmenin temel sorularını temsil etmektedir.

Genel olarak YZ nedir, nasıl tanımlarsınız?

Adalet

1. YZ uygulamalarının verdikleri kararlarda ve sergileyeceği davranışlarda “adaletli olması” ile sizce ne anlatılmak isteniyor?
 - a. Adaletli karar veren ve davranış sergileyen bir YZ uygulamasının özellikleri sizce nelerdir?
 - i. Bir önem sırasına koysanız, sizce bu özelliklerden hangisi/hangileri önceliklidir?
2. YZ uygulamalarının adaletli olması ya da olmamasının etkileri/doğuracağı sonuçlar sizce neler olabilir?
 - a. Buna örnek verebilir misiniz?

Şeffaflık

1. Kullanılan YZ sistemlerinin kararlarının arkasındaki etmenlerin sizin tarafınızdan bilinmesi veya bilinmemesi sizin açınızdan neden önemlidir?
2. Verilen kararların nedenlerinin açıklanmamasını gerektiren durumlar ve bağlamlar hakkındaki düşünceleriniz nelerdir?
3. Bu kararların arkasındaki nedenlerin size mantıksız gelmesi durumunda YZ sistemine karşı tutumunuzla ilgili neler söyleyebilirsiniz?

Kötü niyetli olmama

1. Kötü niyetli geliştirilmiş YZ sistemi denilince zihninizde canlananlarla ilgili bilgi verebilir misiniz?
 - a. Bu bir örnek olay, senaryo vb. olabilir.
2. Kötü niyetle bir YZ sistemi geliştirmenin arkasındaki temel amaç sizce ne/neler olabilir?
3. Kötü niyetli geliştirilmiş bir YZ sisteminin toplumsal sonuçları sizce neler olabilir?
 - a. Bireysel sonuçları neler olurdu?
 - b. Kişiler arası sonuçları neler olurdu?
 - c. Ülkeler bazında (siyaset, bürokrasi vb.) sonuçları neler olurdu?
 - d. Bunu bir süreç şeklinde açıklamanız gerekse, bu sürecin adımları nasıl olurdu?

Gizlilik

1. Genel olarak verilerinizin gizliliğine ilişkin yaklaşımınız nasıldır?
 - a. Size ait verilerinizi paylaşıırken, saklarken nelere dikkat ediyorsunuz?
 - b. Başkalarına ait verileri paylaşıırken, saklarken genel tutumunuz nasıldır?
2. Bir YZ sisteminin bilgi gizliliği ve güvenliği hakkında tutumu sizi, o sistemin kullanımı ile ilgili olarak nasıl etkiler?
 - a. Olumlu / Olumsuz tutumlarınız, kullanmaya devam etme, kullanma mecburiyetine karşı yaklaşımınız vb. açılarından değerlendirseniz, neler söylersiniz?
3. Bilgi gizliliği konusunda bir YZ sisteminin sahip olması gereken özellikler hakkındaki düşünceleriniz nelerdir?
 - a. Sizce bir YZ sistemi, hangi koşulda ya da davranışlarıyla gizliliği ihmal etmiş sayılmalıdır?
 - b. Hangi durumlarda gizlilik ihlali olağan karşılanabilir?

Sorumluluk

1. Bir YZ sistemi ile yaşanan bir sorunda sizce sorumluluk kimdedir? Geliştirici, kullanıcı, bizzat YZ sisteminin kendisi? Neden?
2. Bir YZ sisteminin uygulanması sonucunda istenmeyen bir durum söz konusu ise veri sağlayıcılar, geliştiriciler, tedarikçiler, tüzel ve bireysel kullanıcılar ve düzenleyiciler arasında bir sıralama yapmanız gerekse, sorumluluk öncelikle kimde aranmalıdır?
 - a. Farklı bağlamlar ve uygulamalar göz önünde bulundurulduğunda, bu konuda kimin sorumluluk alması gerektiğinin belirlenmesinin özellikle kritik olduğu bağlamlar aklınıza geliyor mu?

Kapanış

1. Tüm bu 5 boyut (Adalet, Şeffaflık, Kötü niyetli olmama, Gizlilik ve Sorumluluk) dikkate alındığında, tüm sorunlardan arınmış bir YZ sisteminin olanaklılığı hakkındaki görüşleriniz nelerdir?
 - a. Adalet, Şeffaflık, Kötü niyetli olmama, Gizlilik ve Sorumluluk boyutlarından birinin ihmal edilerek tasarlanacağı bir YZ sisteminde, hangi boyutun ihmal edilmesini tercih ederdiniz? Neden?

Bu konuşmaya eklemek istediğiniz bir şey var mı?

**Yarı yapılandırılmış mülakat sorularının, görüşme esnasında katılımcının yanıtlarına göre detaylandırılması öngörülmektedir.*

EK 5 – Gönüllü Katılım ve Bilgilendirilmiş Onam Formu

Sayın Katılımcı,

Bu araştırma, Akdeniz Üniversitesi Yönetim Bilişim Sistemleri Bölümü Yüksek Lisans öğrencisi Neslihan Verda Özmen tarafından Doç. Dr. Tayfun Yörük danışmanlığında, “Üniversite Öğrencilerinin Yapay Zekâ Etiğine Yönelik Tutumlarının İncelenmesi”ne yönelik tez çalışması kapsamında yürütülmektedir. Karma yöntemin benimsendiği çalışmanın nitel araştırma aşamasında yarı yapılandırılmış mülakat yöntemi ile görüşme yapılması planlanmıştır. Mülakatlar tahminen yaklaşık 30 dakika sürecektir ve katılımcıların izni dahilinde ses kaydı alınacaktır. Ayrıca araştırmacı, mülakat boyunca gözlemleri doğrultusunda gerekli gördüğü hususları not edebilir.

Çalışmaya katılım gönüllülük esasına dayanmaktadır. Çalışmanın amacına ulaşması için sizden beklenen, soruları eksiksiz, kimsenin baskısı veya telkini altında olmadan, içtenlikle cevaplamanızdır. Çalışmadan elde edilecek bilgiler tamamen araştırma amacı ile kullanılacak olup kişisel bilgileriniz gizli tutulacaktır; ancak verileriniz yayın amacı ile kullanılabilir. Mülakat esnasında istemediğiniz takdirde gerekçe belirtmeksizin konuşmama ve/veya çalışmayı bırakma hakkına sahipsiniz, bu durumda bilgileriniz kullanılmayacak ve imha edilecektir. Araştırmanın amacı ile ilgili verilen bu bilgiler dışında daha fazla bilgiye ihtiyaç duyarsanız şimdi sorabilir veya daha sonra araştırmacı ile iletişime geçebilirsiniz. Araştırma tamamlandığında genel/size özel sonuçların sizinle paylaşılmasını istiyorsanız lütfen araştırmacıya iletiniz.

“Yukarıda yer alan ve araştırmadan önce katılımcıya verilmesi gereken bilgileri okudum ve katılmam istenen çalışmanın kapsamını ve amacını, gönüllü olarak üzerime düşen sorumlulukları anladım. Çalışma hakkında yazılı ve sözlü açıklama aşağıda adı belirtilen araştırmacı/araştırmacılar tarafından yapıldı. Bu koşullarda söz konusu araştırmaya kendi isteğimle, hiçbir baskı ve telkin olmaksızın katılmayı kabul ediyorum.”

Katılımcının

Adı-Soyadı:

İmzası:

Araştırmacının

Adı-Soyadı:

İmzası:

Ö Z G E Ç M İ Ş

Adı ve SOYADI	Neslihan Verda ÖZMEN
EĞİTİM DURUMU	
Mezun Olduğu Lise	Metin Nuran Çakallıklı Anadolu Lisesi (Almanca), Antalya, 2009
Lisans Diploması	Yeditepe Üniversitesi / İktisadi ve İdari Bilimler Fakültesi / Siyaset Bilimi ve Uluslararası İlişkiler Bölümü (İngilizce), İstanbul, 2015
Yabancı Dil / Diller	İngilizce, Almanca
BİLİMSEL FAALİYETLER	
SCI, SSCI ve AHCI tarafından endekslenen yayınlanmış dergi makaleleri: Yörük, T., Akar, N. ve Özmen, N. V. (2023). “Research trends on guest experience with service robots in the hospitality industry: a bibliometric analysis”, <i>European Journal of Innovation Management</i>.	
Bildiri kitaplarındaki hakemli kongre/sempozyum yayınları: Özmen, N. V., Yörük, T. ve Tonguç, G. (2022). “Artırılmış Gerçeklik Teknolojisinin Eğitim Ortamlarında Kullanılabilirliğinin Değerlendirilmesi”. 5th International Open & Distance Learning Conference – IODL 2022 (pp. 1143-1154). Eskişehir, Türkiye.	
Kitaplar ve kitap bölümleri: Özmen, N. V. ve Tonguç, G. (2023). Otonom Araçlarda Etik, Güvenlik ve Hukuki Konular. Sosyal ve Beşerî Bilimlerde Uluslararası Teori, Araştırma ve Derlemeler (pp.201-224), Serüven Yayınevi, Ankara.	
İŞ DENEYİMİ	
Çalıştığı Kurumlar	- HEKTAŞ TİCARET T.A.Ş. / Marka ve İletişim Uzmanı, Gebze OBS, Kocaeli (2019-2020) - Caprinae Travel / Assistant Manager, Antalya (2017-2019)