T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK- ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

YAPAY ZEKÂDA ETİK

Kevser İNCİ

Ocak -2025
BATMAN

T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK- ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

YAPAY ZEKÂDA ETİK

Kevser İNCİ

Danışman
Prof. Dr. Ömer Faruk ERTUĞRUL

Diğer Jüri Yeleri

Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN Doç. Dr. Cafer BUDAK

Ocak-2025
BATMAN

TEZ KABUL VE ONAYI

Kevser İNCİ tarafından hazırlanan “Yapay Zekâda Etik ” adlı tez çalışması 29/11/2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Elektrik- Elektronik Mühendisliği Ana Bilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri **İmza**

Başkan
Prof. Dr. Ömer Faruk ERTUĞRUL

Danışman
Prof. Dr. Ömer Faruk ERTUĞRUL

Üye
Dr. Öğr. Üyesi Abdulkerim ÖZTEKİN

Üye
Doç. Dr. Cafer BUDAK

Yukarıdaki sonucu onaylarım.

Dr. Öğr. Üyesi Murat ÖTER
Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYANI

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını beyan eder, aksinin ortaya çıkması durumunda her türlü yasal sorumluluğu kabullendiğimi bildiririm.

ETHICAL DECLARATION

I declare that all the information in this thesis has been obtained within the framework of ethical behavior and academic rules, and that the source of any statements and information that do not belong to me in this study prepared in accordance with the thesis writing rules has been fully cited, and I declare that I accept all kinds of legal responsibility in case of any contrary situation.

Kevser İNCİ
Tarih: 29.11.2024

ÖZET

YÜKSEK LİSANS TEZİ

YAPAY ZEKÂDA ETİK

Kevser İNCİ

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü

Elektrik-Elektronik Mühendisliği Ana Bilim Dalı

Danışman: Prof. Dr. Ömer Faruk ERTUĞRUL

2025, 109 Sayfa

Bu tez, yapay zekâ (YZ) uygulamalarının etik bağlamında değerlendirmesini amaçlamaktadır. Yapay zekâ, hızla gelişen bir teknoloji olmasından kaynaklı birçok alanda çığır açacak şekilde yenilikler sunmaktadır. Fakat bu yenilikleri sunarken etik problemlerini de beraberinde getirmektedir. Bu çalışma, yapay zekâ uygulamalarının kullanımında ortaya çıkan etik problemleri ve çözümü için alınabilecek önlemleri sunmaktadır.

Tezde, yapay zekâ uygulamalarının etik problemleri; yapay zekânın temel etik ilkeleri doğrultusunda incelenmiştir. Özellikle, yapay zekâ uygulamalarında adaletli kararlar alma, kişisel verilerin gizliliği ve korunması, üretilen algoritmaların şeffaf olması ve bu teknolojilerden doğacak sorunların sorumluluğu gibi konulara odaklanılmıştır.

Araştırma, nitel bir yöntemle yürütülmüştür. Araştırma literatür taraması ile desteklenmiştir. Bu kapsamda, farklı yapay zekâ uygulamalarıyla ilgili analizler yapılmıştır ve bu teknolojilerin etik sorunları incelenmiştir.

Sonuç olarak, bu tez, yapay zekâ uygulamalarının etik kullanımına yönelik yaklaşımları değerlendirilmesi ve bu alandaki uygulamaların iyileştirilmesi için öneriler sunmaktadır. Yapay zekâ uygulamalarının sorumlu bir şekilde tasarlanması ve kullanılması için etik çerçevelerin güçlendirilmesi gerektiği vurgulanmaktadır. Bu çalışma, yapay zekâ uygulamalarının ilerleyen gelişmelerinde etik ilkelerin daha fazla yer edinmesi gerektiğine dair önemli bir katkı sağlamaktadır.

Anahtar Kelimeler: Adalet, etik, doğal dil işleme, gizlilik, sorumluluk, şeffaflık, yapay zekâ

ABSTRACT

MASTER THESIS

ETHICS IN ARTIFICIAL INTELLIGENCE

Kevser İNCİ

Batman University Graduate Education Institute

Electrical-Electronics Engineering Department of Science

Advisor: Prof. Dr. Ömer Faruk ERTUĞRUL

2025, 109 Pages

This thesis aims to evaluate artificial intelligence (AI) applications in the context of ethics. Since AI is a rapidly developing technology, it offers groundbreaking innovations in many fields. However, while offering these innovations, it also brings ethical problems. This study presents the ethical problems arising in the use of artificial intelligence applications and the measures that can be taken to solve them.

In the thesis, the ethical problems of artificial intelligence applications are examined in line with the basic ethical principles of artificial intelligence. In particular, issues such as making fair decisions in artificial intelligence applications, privacy and protection of personal data, transparency of the algorithms produced, and responsibility for the problems arising from these technologies are focused on.

The research was conducted with a qualitative method. The research was supported by a literature review. In this context, different artificial intelligence applications were analyzed and the ethical issues of these technologies were examined.

As a result, this thesis evaluates approaches to the ethical use of artificial intelligence applications and provides recommendations for improving practices in this field. It is emphasized that ethical frameworks need to be strengthened for the responsible design and use of artificial intelligence applications. This study makes an important contribution to the need for ethical principles to be more present in the further development of artificial intelligence applications.

Keywords: Artificial intelligence, ethics, justice, natural language processing, privacy, responsibility, transparency

ÖNSÖZ

Bu tez, yapay zekâ (YZ) teknolojilerinin etik boyutlarını değerlendirmek ve bu alanda oluşan sorunlara çözümler sunabilmek amacıyla hazırlanmıştır. Yapay zekâ uygulamaları, hayatımızın birçok yerinde bize büyük kolaylıklar sağlayan teknolojilerdir. Sağladığı kolaylıklarla beraber etik sorunlarına yönelik tartışmaları da getirmektedir. Bu çalışma, Yapay zekâ ve uygulamalarının insanlar üzerinde oluşturduğu olumsuz etkileri anlayıp bu olumsuz sonuçları minimum seviyeye indirme çabasıdır.

Tez süreci boyunca, bu çalışmanın gerçekleştirilmesine olanak sağlayan danışman hocam Prof. Dr. Ömer Faruk ERTUĞRUL'a çalışmanın her aşamasında gösterdiği sabır ve destek için teşekkürlerimi sunarım.

Bütün eğitim hayatım sürecinde maddi ve manevi desteğini hiçbir zaman esirgemeyen, her koşulda arkamda duran aileme sonsuz teşekkürlerimi sunmak istiyorum. Ayrıca bu çalışma sürecinde desteğini esirgemeyen arkadaşlarıma teşekkürlerimi sunarım.

Yapay zekâ teknolojileri hızla gelişirken etik kullanımın önemi giderek artmaktadır. Bu çalışmanın, yapay zekâ teknolojilerinin etik tasarlanması ve kullanılmasını teşvik edip bu konuda çalışacak araştırmacılara katkı sunmasını umuyorum.

Saygılarımla,

Kevser İNCİ
BATMAN-2024

İÇİNDEKİLER

ŞEKİLLER, TABLOLAR VE KISALTMALAR

Şekiller

Şekil No **Sayfa**

Tablolar

Tablo No **Sayfa**

Kısaltmalar

AB **:**Avrupa Birliği
ABD **:**Amerika Birleşik Devletleri
AI **:**Yapay Zekâ (Artificial Intelligence)
BM **:**Birleşmiş Milletler
COMPAS **:**Correctional Offender Management Profiling for Alternative Sanctions
DDİ **:**Doğal Dil İşleme
EURAI **:**Avrupa Yapay Zekâ Derneği
GYZ **:**Genel Yapay Zekâ
GPDR **:**Genel Veri Koruma Tüzüğü
IBM **:**Teknoloji Şirketi (International Business Machines Corporation)
ICT **:**Bilgi ve İletişim Teknolojileri (Information and Communications Technology
KADEM **:**Kadın ve Demokrasi Derneği
KVKK **:**Kişisel Verilerin Korunması Kanunu
LLMS **:**Büyük Dil Modelleri (Large Language Models)
MİT **:**Milli İstihbarat Teşkilatı
NAI **:**Dar Yapay Zekâ (Narrow Artificial Intelligence)
NLP **:**Doğal Dil İşleme (Natural Language Processing)
NSCAI **:**Ulusal Güvenlik Yapay Zekâ Komisyonu (National Security Commission on Artificial Intelligence)
NYU **:**New York Üniversitesi **(**New York University)
NIST **:**National Institute of Standards and Technology
OECD **:**Ekonomik İşbirliği ve Kalkınma Örgütü (Organisation for Economic Co-operation and Development)
UNESCO **:**Birleşmiş Milletler Eğitim, Bilim ve Kültür Örgütü (United Nations Educational, Scientific and Cultural Organization)
ÜYZ **:**Üretken Yapay Zekâ
YZ **:**Yapay Zekâ

1. GİRİŞ

Zekâ araştırmaları, geniş ve çeşitli bir literatür alanında yer bulmuştur. Bu alan, zekânın insan doğasındaki genetik temellerine dayanmakla birlikte, aynı zamanda toplumsal bir bağlamı da zorunlu kılmış ve bu sayede literatür daha da genişlemiştir. Bu nedenle zekâ, sadece genetik yönüyle değil, aynı zamanda psikoloji, eğitim bilimleri, işletme, yönetim ve organizasyon, sosyoloji ve kültürel çalışmalar gibi birçok farklı disiplinin inceleme alanına girmiştir. Zekânın tanımı konusundaki farklı görüşler, zekâ türlerinin çeşitliliğini de gündeme getirmiş ve bu türler, farklı bağlamlar ve açıklamalar çerçevesinde ele alınmıştır (Göktürk, 2023).

Zekâ, kişilerin öğrenebilme, problem çözümleme, anlama, akıl yürütme, soyut düşünebilme, çevresel değişikliklere uyum gösterebilme ve tecrübelerden ders çıkarma becerilerini kapsayan yetenektir. Zekâ, genellikle bilgi edinme, analiz etme ve bu bilgiyi etkili bir şekilde kullanma kapasitesi olarak tanımlanır.

Zekâ, farklı türlerden ve çeşitli kuramlardan oluşup çok yönlüdür. Birden çok zekâ türü olduğu düşünülmektedir. En bilinenlerden kurumalardan biri Howard Gardner'ın Çoklu Zekâ Teorisi olduğu düşünülmektedir. Bu kuram, zekâyı sekiz alanda ele almaktadır.

1980'lerin başında Harvard Üniversitesi'nde görev yapan Howard Gardner, "Proje Sıfır" adlı çalışmaları sonucunda bireylerin sadece sözel ve matematiksel zekâya değil, farklı zekâ alanlarına da sahip olduklarını tespit etmiştir. Bu çalışmalarını derleyerek 1983 yılında "Frames of Mind" adlı eserini yayımlamıştır (Kuru, 2001).

Çoklu Zekâ Kuramı, bireysel farklılıkların önemini vurgulayan bir anlayıştır. Bu kuramın eğitimde uygulanması, her bir öğrencinin ilgi ve yeteneklerine odaklanmanın yanı sıra, öğrencilerin farklı öğrenme yollarını tanıma ve bu yollara saygı göstermeyi gerektirir (Jasmine, 1996).

Çoklu Zekâ Kuramı'nın temelini, "Zeki olmanın tek bir yolu yoktur" düşüncesi oluşturur. Zekânın birden fazla yolu vardır ve öğrencilerin farklı zekâ alanlarında üstün olduklarını kabul ederek, onlara farklı yaklaşımlarla ulaşmak, her bir öğrenciyi başarıya ulaştırabilir. Zeki olmanın birçok yolu olduğu gibi öğretim yöntemlerinin de çeşitlendirilmesi gerekir (Kagan ve ark., 1998).

Doğal zekâ insana özgü bir özelliktir. İnsanların çevreyi anlama, düşünme, öğrenme, problemlere çözüm üretebilme, çevresiyle etkileşim halinde olma gibi

özelliklerini kapsar. Bu beynimizde gerçekleşen ve çevremizdeki olaylarla şekillenen bir zekâ türüdür.

Doğal zekâ; öğrenme, düşünme, problem çözme, karar verme, yaratıcılık ve duygusal zekâ gibi özelliklere sahiptir. Doğal zekâ, insan ilişkileri, iş hayatı, sosyal hayat, kişisel gelişim vb. birçok alanda önemli bir yere sahiptir.

İnsanlar doğuştan bir zekâ seviyesine sahiptirler. Fakat zamanla deneyimleri ve aldıkları eğitim sonucunda değişme gösterir. Ayrıca doğal zekânın şekillenmesinde genetik etkiler de büyük bir etkiye sahiptir.

Doğal zekâ insanların çevresiyle etkileşim kurması ve problemleri çözebilmesi açısından önemlidir. Her insanın doğal zekâsı birbirinden farklıdır. Bu yüzden doğal zekâ insanın benzersizliği olarak da tanımlanabilir. Doğal zekâ insanlık tarihindeki tüm ilerlemelerin temeli sayılmaktadır.

Farklı konularda okumlar yapmak, zekâ geliştirici oyunlar oynamak, yeni beceriler öğrenmek, insanlarla sosyal etkileşimde bulunmak, egzersiz yapmak gibi birçok şey doğal zekânın gelişmesine katkı sağlamaktadır. İnsanın en değerli varlıklarından biri olan doğal zekâyı geliştirmek için çaba gösterilmelidir.

Zekâ, problem çözme, öğrenme ve düşünme gibi becerileri içerir. Yapay zekâ bir makinenin beynimizi taklit etme kabiliyetidir. Yapay zekâ bir teknoloji değil makinelerin zekâsını geliştirmek için çeşitli teknolojilerin birlikte çalışmasından oluşur.

Yapay zekâ terimi ilk olarak 1955 yılında John McCarthy tarafından kullanılmıştır. McCarthy, yapay zekâyı "Akıllı makineler ve özellikle akıllı bilgisayar programları geliştirme bilimi ve mühendisliği" olarak tanımlamıştır (McCarthy, 2022).

MIT Bilgisayar Bilimleri Laboratuvarı'nın yöneticisi Edward Fredkin, BBC'ye verdiği bir röportajda tarihte üç büyük olayın varlığından söz etmektedir: evrenin oluşumu, yaşamın başlangıcı ve yapay zekânın doğuşu (TÜBİTAK, 2001).

Yapay zekâ, bilgisayar bilimleri, psikoloji, felsefe gibi çeşitli disiplinlerden beslenen bir alandır (Altıntop, 2023). Bu alan, bilgisayar sistemlerine insan benzeri zekâ kazandırmayı amaçlayan bir bilgi işlem dalıdır. Doğal dil işleme, görüntü tanıma, özerk hareket ve karar verme gibi insan eylemlerine benzer işlevlerin gerçekleştirilmesini sağlayan algoritmalar ve yazılımlar kullanır (Russell ve ark. 2010). Yapay zekâ, veri madenciliği, örüntü tanıma, doğal dil işleme, makine öğrenimi ve derin öğrenme gibi pek çok teknikten yararlanmaktadır (Jordan ve ark. 2015). Bu teknikler, büyük veri setlerini işleyip desenleri tanıma ve öğrenme yeteneği kazandırarak, yapay zekânın ticaret, sağlık, mimari, günlük yaşam, medya, edebiyat, tarım, eğitim, otomotiv, hukuk

ve çevre sorunları gibi pek çok alanda kullanılabilmesini sağlamaktadır (LeCun ve ark. 2015).

Slage, yapay zekâyı sezgisel programlama temelinde bir yaklaşım olarak tanımlar (Andrew, 1991). Popov, yapay zekâyı insanların gerçekleştirdiği işleri bilgisayarlara yaptırma çalışması olarak görür (Popov, 1990). Axe ise yapay zekâyı akıllı programlar geliştirmeyi hedefleyen bir bilim dalı olarak tanımlar (Copeland, 1993). Genesereth ve Nilsson'a göre yapay zekâ, akıllı davranış üzerine odaklanan bir çalışmadır ve amacı, doğadaki varlıkların akıllı davranışlarını yapay olarak üretmeyi amaçlayan bir kuram geliştirmektir (Charniak ve ark., 1985).

Yapay zekâ sistemlerini, günümüzde toplumun çeşitli temel rollerine hizmet eden, birçok alanda fayda sağlayan ve gelecekte belki de insan müdahalesine gerek duymadan özerk kararlar alabilecek sistemler olarak tanımlamak mümkündür (Turan, 2022).

Bunlardan yola çıkarak yapay zekâ için birçok farklı tanım yapıldığı söylenebilir.

İnsanlık, tarih boyunca duyularıyla algıladığı her şeyi kontrol etme çabasında olmuştur. Düşüncenin işleyişi, zihnin gizemli doğası ve metafizik varlıkların olasılığı gibi konular, sanatçılar, bilim insanları ve teologlar üzerinde derin etkiler bırakmıştır. İnsan, kendi gibi düşünebilen bir makine yaratma arzusunu, karşılaştığı problemlere çözüm bulma arayışıyla birleştirmiştir. Yapay zekâ, bilim kurgu eserlerinde bazen bir kurtarıcı, bazen ise dünyayı ele geçirmeye çalışan bir güç olarak karşımıza çıkmıştır. Ancak, günümüzde yapay zekâ, amacını onu kullanan kişinin belirlediği bir araç olarak işlev görmekte ve faydalı çözümler sunmaktadır. Kara ve hava araçlarından kişisel ve ticari ürünlere kadar pek çok alanda yapay zekâ algoritmaları kullanılmakta olup, şüphesiz ki gelecekte de artan etkisiyle birçok alanda varlığını sürdürmeye devam edecektir (Süslü, 2019).

Yapay zekâ (YZ), hızla gelişmekte ve birçok alanda çığır açacak şekilde gelişmeler yaşanmaktadır. Eğitim, sağlık, askeri, ekonomi gibi birçok alanda uygulama sunmaktadır. Bu gelişmeler, hayatımıza birçok kolaylık sağlarken beraberinde etik sorunları da getirmektedir.

Sorumluluk atfedemediğimiz yapay zekâ, şu soruyu gündeme getirmektedir: Yapay zekânın verdiği bir karar, bir kişinin hakkını ihlal ettiğinde ya da kontrolden çıkıp suç teşkil eden eylemler gerçekleştirdiğinde, bu durumun sorumluluğu yapay zekânın üreticisine yüklenebilir mi? Üreticinin yazdığı kodların bir suça yol açması,

dolaylı olarak üreticinin hatası olarak değerlendirilebilir mi? Bu sorulara net bir yanıt vermek oldukça zordur. Yapay zekânın hayatımıza girmesiyle birlikte, bu tür karmaşık sorunlarla karşı karşıya kaldığımız gerçeğiyle yüzleşmek zorundayız (Topakkaya ve ark., 2019).

NLP (Natural Language Processing) ya da Türkçe karşılığıyla Doğal Dil İşleme, bilgisayarların insan dilini anlaması, yorumlaması ve üretmesinde kullanılan yapay zekâ dalıdır. İnsanların günlük hayatta kullandığı dil (yazılı veya sözlü) ile bilgisayarlar arasında iletişim sağlamaktadır.

Chatbotlar, sanal asistanlar, metin analizi, otomatik özetleme, dil öğrenme uygulamaları, arama motorları, sosyal medya analizi vb. birçok alanda NLP uygulanmaktadır.

Metinlerdeki anlamları çıkarma, insanların ürettiğine benzer metinler üretme, metinleri farklı bir dile çevirme, metinlerden belirli bilgileri çıkarma, insanlarla sohbet etme NLP'nin temel amaçlarındandır.

Doğal dil işleme, farklı dil görevlerini tamamlamak için morfolojik analiz, sentaks analizi, anlamsal analiz, duygu analizi, makine çevirisi, dil üretimi, konuşma tanıma, adlandırılmış varlık tanıma gibi birçok teknik kullanmaktadır.

Yapay zekâ ve dilbilim alanının bir ürünü olan Doğal Dil İşleme (NLP) teknolojilerinin, çeşitli servislerin güçlendirilmesi amacıyla kullanılabileceği görülmektedir. NLP, İngilizce, Türkçe ve Almanca gibi doğal dillerin işlenmesi ve kullanılmasına yönelik araştırmalar yapan bir bilim dalıdır. Yapay zekâ teknolojilerindeki ilerlemeler, NLP alanında da önemli gelişmelere yol açmıştır. Bu sayede, dilin özelliklerinden faydalanarak kullanıcılarla etkili bir iletişim kurulması ve kullanıcı sorularının öğrenilen içeriklere dayalı olarak yanıtlanması mümkün hale gelmiştir. Bilişim teknolojilerinin gündeminde olan NLP modelleri, iş, sağlık, ulaşım ve eğitim gibi birçok alanda kullanılmaya başlanmıştır (Firat, 2020).

"Makine düşünebilir mi?" sorusuyla başlayan ve kökleri oldukça eskilere dayanan yapay zekâ araştırmaları, günümüzde neredeyse her alanın dönüşümüne katkıda bulunmuş ve yeni iş modellerinin ortaya çıkmasına zemin hazırlamıştır. Bu bağlamda, yapay zekâ ile ortaya çıkan yeni kavramlardan biri olan prompt mühendisliği, günümüzde yeni bir alan ve yaklaşım olarak tartışılmaya başlanmıştır. İletişimi mühendislikle entegre eden bu disiplin, Doğal Dil İşleme (NLP) algoritmaları ve makine öğrenimi yöntemleri aracılığıyla, insanların doğal dillerini kullanarak istemler oluşturmasına ve basit yanıtlar almasına olanak tanımaktadır. Yapay zekâ,

geleneksel robotik dil ve söylemden ziyade, doğal, etkileşimli, anlaşılır ve insansı bir tarzda gelişimini sürdürmektedir. Bu yeni kavram, kısaca yapay zekâ ile etkili bir şekilde "konuşabilme" yeteneği olarak tanımlanabilir ve AI ile olan iletişimi daha anlaşılır hale getirmeyi amaçlamaktadır. Bu nedenle, sadece komut vermek değil, aynı zamanda doğru komutları yazmayı bilmek, platformlardan doğru sonuçlar almak açısından önemli bir faktör olarak öne çıkmaktadır. Bu perspektiften bakıldığında, yapay zekâ ile iletişim ve etkileşimin doğru bir şekilde ilerlemesi büyük önem taşımaktadır (Kahveci, 2023).

NLP, karmaşık karmaşık ve belirsiz dil, dil bilgisi hatalarının düzeltilmesi gibi birçok zorlukla karşılaşmaktadır. Fakat derin öğrenme (deep learning) ve büyük dil modelleri (Large Language Models, LLMs) gibi teknolojilerin gelişmesiyle birlikte NLP'nin doğruluk ve verimliliği artırmıştır. Bu teknolojiler, çok büyük veri setleri üzerinde eğitilerek doğal ve insan benzeri dil işleme becerileri kazanmıştır. Doğal dil işleme, insan-dil etkileşimini daha sezgisel ve etkili hale getiren ve hızla gelişme gösteren bir alandır. Teknolojinin gelişmesiyle birlikte bu alanın da günlük yaşamda yaygınlık göstermesi beklenmektedir.

Yapay zekâ uygulamalarının artması, yapay zekânın etik ilkeleri konusunda tartışmalara yol açmıştır. Yapay zekâ sistemlerinde oluşabilecek önyargılar, kişi verilerinin ihlali, şeffaf olmaması gibi problemler ciddi sorunlar yaratmaktadır. Bu sistemlerin etik dışı kararları kullanıcıların güvenini sarsıp bu uygulamaları kabul etmelerini zorlaştırır. Bu sebepten, yapay zekâ uygulamaları tasarlanırken etik ilkeleri göz önünde bulundurmaları önem arz etmektedir.

Bu çalışmanın amacı, yapay zekâ uygulamalarındaki etik problemleri inceleyip bu sorunlara çözüm önerileri sunmaktır. Tez kapsamında, etik ilkeleri incelenecektir. Bu bağlamda, çeşitli yapay zekâ uygulamaları incelenerek etik açıdan değerlendirilecektir.

Çalışma, yapay zekâ uygulamalarındaki etik sorunları vurgulayarak bu konuda sorumlu davranılması gerektiğini amaçlamaktadır. Ayrıca, yapay zekâ uygulamalarının etik kullanımı için öneriler sunulacaktır. Bu şekilde, yapay zekâ uygulamalarının etik tasarlanması ve kullanılması konusunda literatüre ve sonraki yapılacak çalışmalara katkıda bulunmayı amaçlamaktadır.

Yapay zekâ uygulamalarının etik bağlamında değerlendirilip anlaşılması ileride daha güvenilir uygulamaların tasarlanmasına katkı sağlayacaktır. Bu sebepten yapay zekâ uygulamalarının etik ilkeler doğrultusunda değerlendirilmesi ve bu sorunlara yönelik çözümler üretilmesi büyük bir önem taşımaktadır. Bu çalışmanın, etik yapay

zekâ uygulamalarının tasarlanması ve kullanılması konusunda bilinç oluşturması ve bu alanda gelişmelere katkı sağlamasını umuyorum.

2. KAYNAK ARAŞTIRMASI

Yapay zekâ, bilim dünyasında, bir bilgisayarın veya bilgisayar destekli bir makinenin, genellikle insana özgü olan problem çözme, anlama, anlam çıkarma, genelleme ve geçmiş deneyimlerden öğrenme gibi ileri düzeydeki mantık süreçlerine ilişkin görevleri yerine getirme yeteneği olarak tanımlanmıştır (Nabiyev, 2012).

Kelime temsilleri ve dönüştürücü modelleri, dilin yapay zekâ tarafından kullanılmasına dair birçok sorunun çözümünde hâlihazırda etkin bir şekilde kullanılmaktadır. Yapay zekâ algoritmaları, her çözülen problemle birlikte, insan düşünme tarzına hızla yaklaşmaktadır. Ancak, bu ilerlemeyle birlikte, yapay zekâ algoritmalarının etik olup olmadığı da artan bir şekilde sorgulanmaktadır. Özellikle yapay zekâ modellerinin hukuk gibi hassas alanlarda kullanılması hedeflendiğinde, oluşturulan algoritmaların tarafsız ve önyargısız olması zorunluluğu gündeme gelmektedir. Öte yandan, dil modellerinin kullanıldığı tüm uygulamalarda, herhangi bir yanlılığın ortaya çıkması kabul edilemez bir durumdur. Örneğin, özgeçmiş sınıflandırması gibi işlemlerde, bir kişinin işe uygunluğunun yaşı, cinsiyeti, milliyeti veya ırkı gibi faktörlerden tamamen bağımsız olması gerekir. Bu nedenle, kelime temsillerinde veya daha büyük dil modellerinde istemeden ortaya çıkabilecek önyargıların tespit edilmesi büyük önem taşımaktadır (Sevim ve ark., 2021).

Günümüzde yapay zekâ etiği, doğru cevaplardan ziyade doğru sorulara odaklanmaktadır. Yapay zekânın insan zekâsına eşit olup olmayacağı henüz bilinmemektedir. Ancak, hızla ve öngörülemeyen bir şekilde geliştiği için, bu süreci kolaylaştıracak ve olası olumsuz sonuçların riskini azaltacak önlemler üzerinde düşünmek son derece önemlidir. Yapay zekâ sistemleri, neyin iyi ya da doğru olduğunu nasıl belirler? Algoritmaların mahremiyet ve veri gizliliği gibi temel insan haklarını ihlal etmediğinden nasıl emin olabiliriz? Kararlar ve eylemler tamamen otomatik olduğunda, hesap verebilirliği nasıl sağlarız? Alınan kararlar evrensel olarak mı iyi, yoksa sadece belirli gruplar için mi iyidir? Bazı ölçütlere göre iyi olabilirken, diğer ölçütler açısından yetersiz mi kalır? Bu ve benzeri sorular, yapay zekâ etiği üzerine yapılan çalışmaların artmasına ve bu alandaki gelişmelere olumlu katkı sağlamasına zemin hazırlamaktadır (Turan ve ark., 2022).

Yapay zekâ (YZ), makine öğrenmesi ve gelişmiş öğrenme teknikleri aracılığıyla süreçlerdeki ekonomiklik, etkinlik ve verimliliği büyük ölçüde artırma potansiyeline sahip yenilikçi teknolojilere dayanır. Ancak, bu potansiyel kabul edildiğinde, yalnızca

teknik değil, aynı zamanda etik ve dini sonuçlar da kaçınılmaz hale gelir. Birçok yenilikçi mühendis ve akademisyen, yapay zekânın derin öğrenme sayesinde insan beyninden daha akıllı olabileceğini ve bu durumun ciddi riskler taşıyabileceğini savunmaktadır. Bu nedenle, yapay zekânın ekonomik, sosyal, politik ve dini hayatımızdaki kararlar ve uygulamalar üzerindeki etkileri önemli bir faktör olarak değerlendirilmelidir (Efe, 2021).

Etik, insan davranışlarına ilişkin eylemlerin doğru-yanlış, iyi-kötü değerlerle ilişkili olarak ele alınmasının yanı sıra, bireyin yaşamını şekillendiren somut ahlaki değerleri de tartışan bir disiplindir. Bu disiplinde, eylemlerin teorik yapısı üzerinde durulmaktadır (Öztürk Dilek, 2019).

Etik bir yapay zekânın tasarımında öncelik, etik robotlar inşa etme ve bu süreçte hangi yaklaşımların daha etik olacağına karar vermektir. Yapay zekânın etik bir özne olarak inşası, zarar vermemesi gereken bir yetenek geliştirmesi gerektiği savunulmaktadır. Ancak, etik bir varlığın inşasında nasıl bir yol izleneceği hâlâ tartışmalı bir konudur (Öztürk Dilek, 2019).

Bir robot ya da bilgisayar sistemi, etik olarak belirli kuralları yerine getirebilir; örneğin, bir robot "adam öldürmek kötüdür" gibi bir programla bu eylemi bilerek engelleyebilir. Ancak, çalmanın yanlış olduğunu bilmesi, kendisine bu konuda bir program yüklenip yüklenmediğine bağlıdır. İnsan ise, bu değerleri tecrübe ve öğrenme yoluyla geliştirir. Dolayısıyla, insanın doğal zekâsı yapay zekâdan daha genel ve kapsamlıdır. Bu nedenle, robot ya da bilgisayar sistemlerinin etiksel statüsü, insanın etiksel statüsüne kıyasla daha düşük olabilir (Çelebi, 2017; Çelebi ve ark., 2019).

Yapay zekâ, süreçlerdeki verimliliği ve etkinliği artırma potansiyeline sahipken, etik ve manevi değerlerin korunması da önemlidir. Yapay zekanın etik ve manevi değerlerden yoksun olarak programlanması veya kötü niyetli kişiler tarafından kullanılması, insanlık için zararlı olabilir (Efe, 2021). Yapay zekâ sistemlerinin özerklik kazanarak daha az insan denetimi ile çalışabilmesi için etik yapıların oluşturulması gerekmektedir. Etik yapay zekâ, bireysel haklar, mahremiyet ve eşitlik gibi temel değerlere uygun olarak tanımlanmalıdır. Bu uygulamalar, kuruluşların verimliliğini artırabilir, çevresel etkileri azaltabilir ve insan sağlığını iyileştirebilirken, etik olmayan uygulamalar ciddi zararlara yol açabilir (Turan ve ark., 2022).

Günümüzde, ChatGPT dâhil olmak üzere yapay zekâ temelli chatbotlar, Doğal Dil İşleme (DDİ) sistemlerine dayanır (Eriç, ve ark.,, 2024). DDİ, makinelerin insan dilini anlamasını sağlamaya yönelik bir teknolojidir. Asıl hedef, insan konuşmasını veya

yazılı metni anlayan ve buna yanıt verebilen bilgisayar sistemleri oluşturmaktır (Aydın ve ark., 2023). Bu modeller, gözlemlenen verilere dayanarak gözlemlenemeyen veriler hakkında çıkarımlar yapmak için çeşitli algoritmalar ve varsayımlar içerir (Locke ve ark., 2021). DDİ, bilgisayarlı çeviri, metin özetleme, duygu analizi ve dil üretimi gibi geniş bir uygulama yelpazesi sunar (Biswas, 2023). Hesaplamalı dilbilim, doğal dil işleme, metin otomatik kategorizasyonu ve ayrıştırma gibi pratik problemlere algoritmik yöntemlerin uygulanmasına yönelik istatistiksel bir yaklaşımdır (Clark ve ark., 2012).

Doğal Dil İşleme (DDİ), bilgisayar sistemlerinin doğal dili çözümleme, anlama, yorumlama ve üretme işlevini konu alan bir bilim ve mühendislik alanıdır. DDİ, yapay zekâ, biçimsel diller kuramı, kuramsal dilbilim ve bilgisayar destekli dilbilim gibi çeşitli alanlarda geliştirilmiş teoriler, yöntemler ve teknolojileri bir araya getirir. 1960'lı yıllarda yapay zekânın bir alt dalı olarak değerlendirilen bu alan, araştırmacıların ve uygulamaların elde ettiği başarılar sonucunda artık bilgisayar bilimlerinin bir parçası olarak kabul edilmektedir.

DDİ alanındaki temel amaçlar şunlardır:

- Doğal dillerin işlev ve yapısını daha iyi anlamak,
- Bilgisayarlar ile insanlar arasında bir ara yüz olarak doğal dili kullanmak ve insan-bilgisayar iletişimini kolaylaştırmak,
- Bilgisayar ile dil çevirisi yapmak.

Doğal dillerin yapısının anlaşılabilmesi için dilin ayrıntılı bir çözümlemesi yapılmalı ve matematiksel bir analizi gerçekleştirilmelidir. Bu nedenle, dilin kuralları mühendislik bakış açısıyla ele alınarak, dilin genel yapısı, kuralları ve sapmaları ortaya konulmaktadır (Delibas, 2008).

2.1. Doğal Zekânın Özellikleri

Doğal zekâ insanların çevreyi anlamlandırabilme, düşünme, problemlere çözüm üretebilme gibi özelliklerini kapsamaktadır. Doğal zekâ birçok özelliğe sahiptir. Çevreden gelen bilgileri yorumlama olarak tanımlanabilir. Bu yetenek sayesinde insanlar gördüklerini ve duyduklarını anlamlandırabilir. Doğal zekâ aracılığıyla problemlere mantıklı çözümler üretilebilir. Bu yetenek karmaşık problemlerin çözümü için önemli bir yere sahiptir.

Bu zekâ, yeni bilgileri öğrenip mevcut olan bilgileri ise güncelleyebilme yeteneğine sahiptir. İnsanlar çevrelerinde gördüklerini davranışlarına yansıtma yeteneğine sahiptir. Edinilen bilgileri bu zekâ sayesinde analiz edilebilir. Farklı bilgilerle sentezlenebilir. Doğal zekâ sayesinde yeni fikirler bulunabilir. Bu sayede teknolojinin ilerlemesine katkıda bulunulur.

Farklı seçenekler arasından mantıklı olan seçeneği seçebilme yeteneği olarak tanımlanabilir. Bu yetenek sayesinde daha doğru kararlar alınabilmektedir. Doğal zekâ aracılığıyla duygular keşfedilip ifade edilebilir. Bu yetenek sayesinde kendi duygularını tanımlayabildikleri gibi başkalarının da duyguları anlamlandırılabilir. Düşünceleri ifade etme ve iletişim kurmada bu yetenek önemli bir yere sahiptir. Bu yetenek aracılığıyla paylaşım yapma imkânı sağlanmış olur. Çevremizdeki insanlarla iletişim kurma yeteneği olarak ifade edilebilir.

2.2. Doğal Zekânın Üstünlükleri

İnsanlar kendileri ve doğa hakkında bilinçlidir ve farkındalığa sahiptirler. Bu yüzden deneyimlerinden faydalanarak düşünebilirler ve yaşam için amaç edinebilirler. Öz farkındalık toplumsal normlara uygun davranışlar sergilemeye olanak tanır. Fakat yapay zekâda bir bilinç veya öz farkındalık yoktur. Sadece programlandıkları alanda çalışabilirler.

Doğal zekâ hayal gücünden faydalanabilir. Yeni fikirler üretebilir. Problemlere yaratıcı çözümler üretebilir. Doğal zekâ yaratıcılık konusunda yapay zekâdan daha üstün becerilere sahiptir. Yapay zekâ sadece belirlenen sınırlar içerisinde fikir ve çözüm üretebilir.

İnsanlar, farklı bilgi alanlarını birleştirerek disiplinler arası düşünme ve problem çözme yeteneğine sahiptirler. Örneğin, bir mühendis, hem teknik bilgiye hem de sosyal bilimlerden gelen iç görülere dayanarak karmaşık bir sorunu çözebilir. Yapay zekâ, genellikle belirli bir bilgi alanında uzmanlaşır ve farklı disiplinler arası geçişlerde zorlanabilir.

İnsanlar, sezgiye dayalı olarak hızlı ve etkili kararlar alabilirler. Bu, bilinçli düşünme süreçleri dışında gelişen bir yetenektir ve çoğu zaman deneyimlerden beslenir. Yapay zekâ, veriler ve algoritmalar üzerinden karar verirken, insanların sezgisel düşünme becerileri gibi soyut süreçleri simüle etmekte zorlanır.

İnsanlar yeni durumlara kolaylıkla uyum sağlayabilir. Bilgilerini farklı şekillerde kullanabilirler. Doğal zekâ değişen ve karmaşık olaylara kolaylıkla adapte olma becerisini içerir. Yapay zekânın belirlenen sınırların dışına çıkması mümkün değildir. Yeni durumlara uyum sağlayamaz. Yeni durumlarda insan müdahalesi ile yeniden programlanmaya ihtiyaç duyar.

İnsanlar, etik değerler ve ahlaki ilkeler doğrultusunda kararlar alabilirler. Bu, kültürel, dini ve toplumsal normlara dayalı karmaşık bir süreçtir. Yapay zekâ, etik kararlar almak için programlanabilir, ancak bu kararların insani değerlerle tamamen örtüşmesi her zaman garanti edilemez.

İnsanlar duyguları anlayıp empati kurabilir. Doğal zekâ, duygusal zekâ gerektiren durumlarda yapay zekâya oranla daha fazla başarı gösterir. Yapay zekânın duyguları anlama yeteneği yoktur. Sosyal etkileşim kurması gereken konularda ise sadece belirlenen sınırlar içinde etkinlik gösterir.

Yapay zekânın kökeni, insanın doğal zekâsına dayanmaktadır. İnsan, hayal gücünü kullanarak bilgisayar tabanlı sistemlere hız ve bilgi kapasitesi kazandırmak, ayrıca özerklik ve çözüm üretme yeteneklerini geliştirmek istemiştir. Ancak yapay zekâ tasarımları sadece insan zekâsından değil, diğer canlı varlıklar ve doğal dinamiklerden de ilham almaktadır. Bu bağlamda, yapay zekâ hem insana hem de doğaya ve dünyaya dair bir kavram olarak gelişmektedir (Küçüker, 2023).

Bireylerin yaşamlarının her aşamasında mutluluğu yakalayabilmeleri için güçlü ve geliştirilebilecek yönlerinin farkında olmaları, duygularını ve davranışlarını yönetebilmeleri, ayrıca çevrelerindeki insanların duygularını ve düşüncelerini anlamaya çalışmaları gerekmektedir. Günümüzde, başarılı bireylerin sadece teknik bilgiye sahip değil, aynı zamanda iş arkadaşlarıyla sağlıklı ilişkiler kurabilen ve yüksek duygusal zekâya sahip olanlar olduğu anlaşılmaktadır. İş yaşamında yüksek IQ'nun yanı sıra, duygusal zekânın da önemli olduğu görülmektedir. Yüksek duygusal zekâya sahip çalışanlar, kendilerini motive edebilir ve çevrelerinde pozitif bir atmosfer oluşturabilirler. Bu nedenle, duygusal zekânın etkin bir şekilde kullanılması, başarı için temel bir faktör olarak kabul edilmektedir (Kılıç ve ark., 2007).

Doğal zekâ, bireylerin kendi ve başkalarının deneyimlerini değerlendirerek gelişmesini sağlarken, yapay zekâ bilgisayara yüklenen bilgi ile sınırlıdır (Adalı, 2017). Bu açıdan doğal zekânın, yapay zekâya göre daha yaratıcı olduğunu söylemek mümkündür. Doğal zekâ, deneyimlere dayalı olarak öğrenme ve bu deneyimlerden yararlanma yeteneği sunarken, yapay zekâ genellikle sembolik girdilerle çalışır. Doğal

zekâya sahip bireyler, yeni durumlarla karşılaştıklarında deneyimlerine dayanarak hızlı ve yenilikçi çözümler üretebilir. Buna karşın, yapay zekânın üretebileceği çözümler, kodlama ve yüklenen bilgilere bağlı olarak sınırlı kalır (Çelebi ve ark. 2019; Adalı, 2017).

Doğal zekânın üstünlükleri, insanların karmaşık ve belirsiz durumlarla başa çıkmada, yaratıcı çözümler üretmede ve etik kararlar almada önemli bir avantaj sağlamaktadır. Yapay zekâ, belirli alanlarda insan zekâsını tamamlayabilir ve destekleyebilir, ancak doğal zekânın esnekliği, yaratıcılığı ve insan doğasına özgü diğer özellikler yapay zekânın çok ötesindedir.

2.3 Yapay Zekâ Türleri

Yapay zekâ insanların çevreyi anlamlandırabilme, düşünme, problemlere çözüm üretebilme gibi özelliklerini makinelerin taklit edebilme becerisidir. Yapay zekâ türleri dar yapay zekâ, genel yapay zekâ ve süper yapay zekâ olarak üçe ayrılmaktadır.

A. Dar Yapay Zekâ (Narrow AI veya Weak AI)

Yapay dar zekâ zayıf zekâ olarak da adlandırılır. NAI sadece belirli görevler için tasarlanan zekâ sistemleridir. Kapsam ve güçten yoksun olduğu için değil insani bileşenlere sahip olmadığından dar yapay zekâ olarak adlandırılır.

Bu zekâ türü kendi başına düşünüyormuş gibi görünse dahi sadece belirlenen çerçevede kararlar alabilmektedir, geniş çapta düşünememektedir. Yalnızca programlandığı alanda performans göstermektedir. Düşünme sürecinde bilinç ve duygular yer almamaktadır. Yeni görevlerle başa çıkamaz, çözüm üretemezler. Belirlenen alan dışındaki görevlerde etkisizdir. Belirlenen görevde yüksek verimlilik sağlayıp genellikle insan performansını geçerler. Belirlenen görevleri doğru bir şekilde yerine getirip insani hata oranını azaltırlar.

Yapay zekânın en yaygın kullanılan türüdür. Dar yapay zekâ, günümüz teknoloji dünyasında birçok uygulamada kullanılır. Bu sistemlerin sürekli olarak gelişmesi hayatı kolaylaştıran çözümler sunar. Siri, Alexa, Google Assistant gibi sesli asistanlar, belirli komutları yerine getirir. Örneğin, arama yapma veya alarm kurma gibi belirli görevleri yerine getirirler. Netflix, Amazon, Youtube gibi platformlarda kullanılan öneri sistemleri, kullanıcıların geçmişteki tercihlerine göre film, dizi veya ürün önerileri sunar. Yüz tanıma sistemleri, güvenlik kameralarında veya sosyal medya platformlarında kullanılır. Ayrıca, fotoğraflardaki nesneleri veya kişileri tanımak için

kullanılır. E-posta hizmetlerinde kullanılan spam filtreleri, gelen e-postaları analiz ederek spam veya zararlı olabilecek e-postaları belirler ve filtreler. Algoritmik ticaret sistemleri, finansal piyasalarda otomatik olarak alım-satım yapar ve belirli ticaret stratejilerini uygular. Sürücüsüz araçlar, belirli koşullarda araç kullanımı gerçekleştirebilir.

Günümüzde popüler olan birçok yapay zekâ (YZ) tabanlı program, genellikle "dar yapay zekâ" olarak tanımlanır. Bu tür yapay zekâ sistemleri, belirli görevleri yerine getirme kapasitesine sahiptir, örneğin yarışmalarda insanları yenme, hastalık teşhisi yapma gibi alanlarda başarılı olabilirler (Demir, 2019). Ancak, bu sistemler sadece belirli ve tanımlı görevlerde başarılıdırlar.

Bilgisayar mühendisi Donald Knuth, yapay zekânın birçok zorlu problemi başarabildiğini, ancak insanların ve hayvanların düşünmeden yaptıkları birçok basit şeyi başaramadığını belirtmiştir. Knuth'a göre, yapay zekâ karmaşık görevleri başarma konusunda büyük başarılar elde etmişken, insanların günlük yaşantılarında gerçekleştirdiği bazı basit işlerde hala yetersiz kalmaktadır (Eryılmaz, 2023).

Knuth'un ifadesi doğrultusunda, yapay zekâ sistemleri görsel sahneleri analiz etmekten doğal dilleri çevirmeye, hatta satranç ve go gibi oyunlarda insanları yenmeye kadar birçok alanda başarılıdır. Ancak bu sistemler, insan beynindeki her problemi çözme kapasitesine henüz ulaşamamıştır. Yapay zekâ, büyük veri setlerini işleme, depolama ve veriler arası bağlantılar kurma konusunda başarılı olabilirken, insan beynindeki tüm sorunları çözme yeteneğinden uzak kalmaktadır (Bostrom, 2020).

B. Güçlü Yapay Zekâ (General AI veya Strong AI)

Güçlü yapay zekâ insan zekâsına sahip olan zekâ sistemidir. Bu zekâ türü problem çözebilme, yaratıcı olma, düşünebilme gibi özelliklere sahip olan yapay zekâ türüdür.

Dar yapay zekâda olduğu gibi sadece belirli bir alanda değil farklı alanlarda da performans gösterebilir. Güçlü yapay zekâ belirsiz durumlarda karar alabilir, problemleri çözebilir ve önceki bilgilerden faydalanabilir. Yeni durumdaki problemlere çözüm üretebilirler. Farklı bilgilerden mantıklı çıkarımlar yapılabilir. Geniş çapta düşünebilme özelliğiyle insanlarla benzer özellik gösterir. Bundan dolayı insan düzeyi yapay zekâ olarak da adlandırılabilmektedir. Dar yapay zekâya göre daha geniş görev alanına sahiptir.

Genel yapay zekâ, dar yapay zekâdan önemli ölçüde farklıdır. Dar yapay zekâ belirli görevlerde yüksek performans sergileyebilirken, genel yapay zekâ sosyal ve

duygusal zekâya sahip olma, geçmişi ve geleceği düşünme becerilerine sahip olma, yaratıcılık ve özgünlük gibi insanlara özgü özellikleri bünyesinde barındırabilmelidir. Bu tür bir genel yapay zekâ, birçok dar yapay zekânın birleşimi olarak düşünülebilir ve insan deneyimlerinin tüm alanlarını kapsayarak en az bir insan kadar yetkin olması beklenir (Eryılmaz, 2023).

Bilgisayarların hayatımıza dâhil olmasından bu yana, özellikle son elli yıl içinde, "asla yapılamaz" olarak düşünülen birçok şeyin mümkün hale geldiği gözlemlenmiştir. Teknolojik ilerlemelerle, daha önce imkânsız olarak görülen birçok alan giderek gerçek hale gelmektedir (Reese, 2018).

Güçlü yapay zekânın geliştirilmesi insani fonksiyonları tamamen taklit etmeyi gerektirdiğinden oldukça zordur. Güçlü yapay zekâ ilerleyen zamanlarda iş olanaklarını değiştirebilir. Güçlü yapay zekâ sayesinde birçok alanda yenileşmeye gidilebilir. Hastalıkların teşhis ve tedavi süreçleri, eğitim öğretim alanı, tamamen insansı robotlar buna örnek gösterilebilir.

Güçlü yapay zekâ şu an sadece teorik bir kavram olarak varlığını sürdürmektedir. Geliştirilip hayata geçirilmesi uzun yıllar ve çalışmalar gerektirmektedir. Bu alanda yapılacak ilerlemeler teknolojide çığır açacaktır.

C. Süper Yapay Zekâ (Superintelligent AI)

Süper yapay zekâ her anlamda insandan daha yetenekli olan bir sistemdir. Güçlü yapay zekânın daha kapsamlı çeşidi sayılabilir.

Süper yapay zekâ her türlü faaliyette insanlardan daha üstün performans gösterir. Süper yapay zekâ ile bilgiler daha kısa sürede öğrenilebilmektedir. Kısa zamanda daha doğru bilgiler elde edilebilir. Karmaşık problemler çözülebilir. Süper yapay zekâ günümüzde halen bir kavram olarak varlığını sürdürmektedir. Süper yapay zekâ sayesinde keşifler hız kazanabilir. Tıp alanında yeni yöntemler bulunabilir. Yeni iş olanaklarına imkân sağlayabilir.

Süper yapay zekâ insandan daha yetenekli olacağından yanlış kişilerin kullanması sonucunda insanlık için tehdit oluşturabilir ve güç dengesizliklerine yol açabilir. Bu zekâ türünün ilerlemesi için uzun çalışmalar gerekmektedir.

Bir aşırı zeki makinenin tanımında, bu makinenin insan zekâsının ötesine geçerek entelektüel faaliyetleri önemli ölçüde aşabileceği belirtilmektedir. Bu tür bir makine, mevcut makineleri daha da geliştirme kapasitesine sahip olabilir ve bu durum, bir "zekâ patlamasına" yol açabilir. İnsan zekâsı, böyle bir makine karşısında oldukça geride kalabilir. Bu bağlamda, ilk süper zekâ makinesinin insan için son icat olabileceği

düşünülmektedir, ancak bunun gerçekleşmesi için makinenin nasıl kontrol altında tutulacağı konusunda bilgi vermesi gerekir (Barrat, 2020).

Yapay süper zekâ ile ilgili olarak üç temel sorun öne çıkmaktadır. İlk olarak, süper zekânın hızla gelişmesi, onu kontrol etmenin veya rekabet etmenin neredeyse imkânsız hale gelmesine neden olabilir. İkinci olarak, insan zekâsının standartlarıyla uyumlu ve insani değerlere sahip bir yapay zekâ tasarlamanın zorlukları söz konusudur ve bu durumun büyük ölçüde imkânsız olduğu ifade edilmektedir. Üçüncü olarak ise, süper zekânın kendi gücünü artırmaya yönelik çabası, sınırsız kaynak talebiyle sonuçlanabilir ve bu, yaşam için gerekli atomlar veya diğer temel ihtiyaçlar üzerinde etkili olabilir (Orhon, 2021).

2.4. Yapay Zekâ Alt Dalları

Yapay zekâ çok çeşitli alt dallarına sahiptir. Bunlardan en temel olanları şunlardır:

A. Makine Öğrenimi (Machine Learning)

Makine öğrenmesi, bilgisayarların veri setlerinden öğrenerek gelecekteki durumları tahmin etmelerini sağlayan bir yapay zekâ alt dalıdır. Basitçe söylemek gerekirse, makine öğrenmesi algoritmaları, verileri analiz ederek desenler ve ilişkiler bulur ve bu bilgiler doğrultusunda kararlar verir.

Makine öğrenmesi, bir problemi, o probleme ait verilerle modelleyen bilgisayar algoritmalarının genel adıdır. Bu süreçte, mevcut veri seti ve kullanılan algoritmalarla oluşturulan model, en yüksek performansı sağlamak üzere tasarlanır (Atalay ve ark. 2017).

Makine öğrenmesi birçok alanda kullanılmaktadır. Hastalıkların teşhis edilmesi, kişiye özgü tedavi planları, dolandırıcılıkların tespit edilmesi, belli bir hedefi olan reklamcılık, üretilen ürünlerin kalite kontrolü, sürüşleri sürücüsüz hale getirme bunlardan bazılarıdır.

Günümüz teknolojik olanakları göz önüne alındığında, bilgisayar programlarının insan zekâsına karşı elde ettiği en temel başarı olarak nitelendirilebilir. Bu başarı, bize basit görünse de, gelecekteki yapay zekâ sistemleri için bir temel oluşturmuştur (Gökalp, 2022).

Makine öğrenmesinin gelişmesiyle ileride daha akıllı sistemler ortaya çıkacaktır. Bu teknolojinin gelişmesi birçok alanda devrim niteliğinde değişimlere yol açacaktır. Bunun sonucunda iş alanlarında değişikler ortaya çıkması olasıdır.

Makine öğrenmesi 3 temel ilkeden oluşur.

- Veri: Makine öğrenmesinin beslendiği en önemli kaynaktır. Verinin büyüklüğü ve cevapların doğruluğu doğru orantılıdır.
- Algoritma: Verileri analiz edip model oluşturan formüllerdir.
- Model: Algoritma veri üzerinde çalıştırılıp gelecekteki durumlar tahmin etmede kullanılan yapıdır.

Makine öğrenmesi 3 türde gerçekleşir.

- Denetimli Öğrenme: Etiketlenmiş olan verilerin her açıdan doğruluğunu tespit eden öğrenme türüdür.

Denetimli öğrenme bazı kaynaklarda "gözetimli öğrenme" olarak da anılmaktadır. Bu yöntem tahminler yaparken regresyon ya da sınıflandırma tekniklerini kullanır. Örneğin, bir e-postanın spam olup olmadığını belirlemek için bu yöntem uygulanabilir. Spam olarak sınıflandırılan e-postalar, makine öğrenimi algoritmasının eğitilmesi için önemli örnekler olarak kullanılır (Gökalp, 2022).

- Denetimsiz Öğrenme**:** Verilerin etiketlenmeyip yanıtların doğru olmadığı bir öğrenme türüdür.

Bu sistem, etiketlenmemiş verilerdeki iç görüleri ve ilişkileri ortaya çıkarır. Eğitim sırasında etiketsiz veriler kullanılır ve bu verilerden anlamlar, tanımlar ve keşifler yapılır. Bu nedenle, sistemden elde edilen sonuçlar genellikle önceden belirlenmiş değildir. Sonuçlar, eğitim süreci ve gözetim dışında ortaya çıkar (Gökalp, 2022).

- Pekiştirmeli Öğrenme: Bir ajanın ortamda gerçekleştirdiği davranışlar sonucunda ödül ve ceza alarak uygun davranışları belirlediği öğrenme türüdür.

Pekiştirmeli (takviyeli) öğrenme yönteminde, hedef çıktıya ulaşmak için bir danışman kullanılmaz. Bunun yerine, elde edilen sonuçlar, belirli kriterlere göre iyi ya da kötü olarak değerlendirilir. Bu kriterler, verilen girişe karşılık çıkan sonuçların kalitesini belirlemek için kullanılır (Atalay ve ark., 2017).

B. Derin Öğrenme (Deep Learning)

Derin öğrenme yapay zekânın beyni olarak da bilinir. Makine öğreniminin alt dalıdır. İnsan beyninin sinir ağlarını taklit eden yapay sinir ağları aracılığıyla verileri analiz eder. Bu sayede daha karmaşık problemler daha doğru şekilde sonuçlanır.

Derin öğrenmenin birçok kullanım alanı vardır. Nesne, yüz ve ses tanıma, tıbbi görüntüleri çözümleme, duyguları analiz etme, konuşmaları sentezlemek bunlardan bazılarıdır.

Derin öğrenme makine öğrenmesinin alt dalı olmasına rağmen bazı farklılıkları mevcuttur. Derin öğrenme verilerdeki özellikler otomatik öğrenir. Fakat makine öğrenmesi önceki belirlenen öğrenmeleri kullanır. Derin öğrenme makine öğrenmesine oranla daha büyük verilerle çalışır. Derin öğrenme makine öğrenmesine göre daha karmaşık problemleri çözebilir.

Derin öğrenme, makine öğrenmesinin birçok pratik uygulamasını ve yapay zekâ alanının genişlemesini sağlamıştır. Bu yöntem, belirli görevlere yönelik algoritmaların ötesinde, verileri temsil etmeye dayalı bir yaklaşım sunar. Özellikle görüntü işleme alanında derin öğrenme, karmaşık problemleri etkili bir şekilde çözebilme kapasitesine sahiptir ve bu alanda başarılı sonuçlar elde edilmesini sağlar. Eğitim süreçlerinin uzun sürmesine rağmen, test aşamasında elde edilen yüksek başarı oranları, derin öğrenme yöntemlerine duyulan güveni artırmıştır (Daş ve ark., 2019).

Derin öğrenmenin gelişmesiyle ileride daha akıllı sistemler ortaya çıkacaktır. Bu teknolojinin gelişmesi birçok alanda devrim niteliğinde değişimlere yol açacaktır. Bunun sonucunda iş alanlarında değişikler ortaya çıkması olasıdır.

Derin öğrenmenin 3 temel ilkesi vardır.

- Yapay Sinir Ağları: Derin öğrenmenin temelidir. Birbirine bağlı yapay nöronların katmanlardan oluşur. Veri, ağa girdi olarak verilip ağın içindeki katmanlar arasından bir çıktı üretir.
- Katmanlar: Derin öğrenmede birçok katman bulunur. Bu katmanlar, verileri daha soyut ve karmaşık özelliklere dönüştürür.
- Öğrenme: Derin öğrenme ağları, büyük miktardaki verilerin eğitilmesiyle öğrenilir. Bu durumda ağın verdiği çıktının istenen çıktıya yakınlığı ölçülür. Sonucunda hatalarını düzeltir.

C. Doğal Dil İşleme (Natural Language Processing, NLP)

Doğal dil işleme insan dilinin anlaşılmasını sağlar. Farklı dillerin yapısını çözümler, çevirisini yapar. Yapay zekânın en zor alanlarından biridir.

Doğal Dil İşleme (NLP), yapay zekâ ve dilbilim alanlarının birleşimi olan bir bilim dalıdır. NLP, İngilizce, Türkçe, Almanca gibi doğal dilleri işlemek ve bu dilleri kullanarak araştırmalar yapmak amacıyla geliştirilmiştir. Yapay zekâ teknolojilerindeki ilerlemelerle birlikte, NLP alanında da önemli gelişmeler yaşanmıştır. Bu teknolojiler, dilin özelliklerinden faydalanarak kullanıcılarla etkili iletişim kurmayı ve kullanıcı sorularını, yapay zekânın öğrendiği içerikler doğrultusunda yanıtlamayı mümkün kılmıştır. NLP modelleri, iş, sağlık, ulaşım ve eğitim gibi çeşitli alanlarda kullanılmaya başlanmıştır (Firat, 2020).

Doğal Dil İşleme, bilgisayar bilimleri, dilbilim ve yapay zekânın kesişim kümesinde yer alan disiplinler arası bir çalışma alanıdır. Bu alan, doğal dil verisinin işlenmesi ve analiz edilmesi için bilgisayarların nasıl programlanabileceğini araştırır. NLP'nin nihai hedefi, doğal dilin tüm bileşenlerinin bilgisayarlar tarafından anlaşılmasını sağlamaktır. Ayrıca, doğal dil işleme, kullanıcıların işini kolaylaştırmak ve bilgisayarla doğal dilde iletişim kurma ihtiyaçlarına yanıt vermek amacıyla geliştirilmiştir (Güner, 2023).

Doğal dil işlemenin kullanıldığı birçok alan vardır. Yazım hatalarını düzeltir, çeviri yapar, metinleri özetler, metinleri sınıflandırır, metinden istenen bilgileri çıkartır, konuşmaları yazıya dökebilir, otomatik yanıtlama sistemleri vardır. Google, Yandex, Siri, Alexa, ChatGPT buna örnek olarak verilebilir.

Doğal Dil İşleme (NLP), günümüzde milyarlarca dolarlık bir endüstri haline gelmiştir. NLP, sağlıktan satış ve pazarlamaya kadar pek çok sektörde devrim yaratma potansiyeline sahiptir. Ayrıca, sıradan bir insanın günlük olarak kullandığı birçok dijital uygulamanın temelini oluşturmaktadır. Örneğin, Microsoft Word ve Libre Office gibi kelime işlemcilerden, Google ve Yandex gibi arama motorlarına kadar pek çok araç, doğal dil işleme mekanizmalarını arka planda kullanmaktadır (Güner, 2023).

NLP farklı disiplinlerden faydalanır. Bir NLP sisteminde metinlerdeki yazım yanlışları ve gereksiz unsurlar temizlenir. Kelimeler anlamsal açıdan benzer sayısal ifadelerle temsil edilir. Derin öğrenme tekniğiyle eğitilmiş sinir ağları sayesinde metinlerin anlamlarını belirler ve karmaşık problemleri çözer. NLP teknolojisinin gelişmesiyle birlikte daha karmaşık problemler çözülebilir.

D. Bilgisayarlı Görü (Computer Vision)

Bilgisayarlı görü makinelerin dünyayı görmesi olarak da bilinir. Yapay zekânın bu alt dalında bilgisayarlar dijital ortamdaki video ve görüntülerden bilgi çıkarma

işlemini yaparlar. İnsanların etrafındaki nesneleri algılamasında olduğu gibi bu sistem de nesneleri tanıyıp takip eder.

Birçok alanda bilgisayarlı görü kullanılmaktadır. Tıbbi görüntülemeler aracılığıyla hastalıkların teşhisi, araçlardaki şerit takip sistemleri, güvenlik sistemlerinde ve akıllı telefonlarda yüz tanıma, endüstri sistemlerinde ürünleri tanıma, hatalı ürünleri tespit etme ve kalite kontrolü yapma, sosyal medyalarda görüntülü görüşmeler örnek olarak gösterilebilir.

Bilgisayarlı görü sistemlerinin gelişmesiyle gelecekte daha akıllı sistemler ortaya çıkacaktır. İş hayatı ve sosyal hayatta birçok değişim meydana gelecektir.

Bilgisayarlı görü sistemleri 4 temel ilkeden oluşmaktadır. Görüntü İşleme: Görüntüler, bilgisayarın çözümleyebileceği şekilde sayısal verilere dönüştürülür. Bu süreçte, gürültü giderme gibi işlemler yapılır. Özellik Çıkarma: Görüntüdeki önemli öğeler belirlenir. Nesneleri tanımlamada bu öğeler kullanılır. Öğrenme: Bilgisayarların farklı nesneleri algılayıp sınıflandırabilmesi için büyük görüntü verisi üzerinde eğitir. Derin Öğrenme: Bilgisayarlı görüde büyük bir çığır açmıştır. Daha karmaşık görsellerden daha doğru sonuçlar elde edilebilir.

E. Robotik

Yapay zekânın fiziksel dünyayla iletişim kurmasını sağlar. Yapay zekâ ve robotik birbirini tamamlayan güçlerdir. Robotik sistemlerin gelişimi üretkenliğin, yaşam kalitesinin ve iş imkânlarının artması gibi sonuçlar doğuracaktır.

Yapay zekâ ve robotik, insan davranışlarını taklit etme noktasında ortak bir paydada buluşur ve bu iki alanın entegrasyonunu sağlayan güncel çalışmalar mevcuttur. Bu nedenle, yapay zekânın çalışma alanlarından biri robotik ile kesişmektedir. Bu durum, yapay zekâ ürünlerinin ve robotların bazen eşanlamlı olarak değerlendirilmesine yol açabilmektedir (Halidi, 2022).

Endüstri, sağlık, otomotiv, askeri, eğitim ve ev hizmetleri gibi birçok alanda robotik uygulama alanı bulmaktadır.

F. Uzman Sistemler (Expert Systems)

Uzman sistemlerin insanların uzman oldukları belirli alandaki bilgilerini taklit eden bilgisayar programlarıdır. Bu sistemlerde genelde bir uzmanlık alanıyla ilgili tüm bilgilere sahip olup o alandaki karmaşık problemleri çözebilirler. Uzman sistemler tıp, finans, mühendislik, jeoloji, askeriye gibi birçok alanda kullanılmaktadır. Uzman sistemler yapay zekânın ilk uygulamalarından olmakla birlikte yeni sistemlerin gelişmesiyle yerini daha yetenekli sistemlere bırakmıştır.

Uzman sistemler, bir uzman kişinin bilgi ve yeteneklerini gerektiren problemleri, bilgi ve mantıksal çıkarım yöntemlerini kullanarak çözebilen sistemlerdir (Atalay ve Çelik, 2017). Bu sistemler, problemi çözmede uzman bireylerin bilgi ve mantıksal çıkarım süreçlerinin modellemesini amaçlar (Harmon ve ark., 1985).

Uzman sistemler her koşulda aynı sonucu verip insan hatalarından etkilenmediğinden dolayı tutarlıdır. Karmaşık problemlerde bile hızlı sonuçlar elde edilir. Alandaki uzman bilgi ve görüşlerine hızlı bir ulaşım olanağı sağlar. Uzun yıllar boyunca bile bilgileri saklayıp koruyabilir.

Uzman sistemlerde bilgilerin eklenmesinde esneklik yoktur. Yeni bilgilerin eklenmesi uzun zaman alır veya yeniden yapılandırmaya gitmeyi gerektirebilir. Bu sistemlerin geliştirilmesi biraz maliyetlidir.

G. Bulanık Mantık (Fuzzy Logic)

Bulanık mantık için kesin olmayan dünyada karar verme de denilebilir. Diğer sistemlerin aksine doğru yanlış bilgiler değil belirsiz olan durumlarda karar verebilen sistemlerdir. Bu sisteme göre olaylar ya tamamen doğru ya tamamen yanlıştır. Kesin çizgilerle ayrılırlar.

Bulanık mantık, klasik mantığın belirli sınırlarla ayrılmış kategorileri yerine, tanımlı fonksiyonlarla birbirine geçmiş birçok aralığı kullanır. Bu yaklaşım, insan benzeri düşünme, karar verme ve seçim yapma yeteneğine sahip sistemlerin geliştirilmesini amaçlayan bulanık kümeler teorisiyle desteklenir. Bu tür sistemlere "bulanık sistemler" denir ve bulanık mantığın matematiksel yapısını kullanarak çalışırlar (Atalay ve ark., 2017).

Bulanık mantık; hastalıkların teşhis ve tedavi süreçlerinde, ev ve endüstriyel aletlerin kontrolünde, karar alma sistemlerinde ve finans gibi alanlarda uygulama olanağı bulmaktadır.

Klasik mantıkta net cevaplar olmadığından bulanık mantık belirsiz durumları çözümlemede avantaj sağlıyor. Bulanık mantık yeni bilgileri sisteme ekleyebilecek bir esnekliğe sahiptir. Belirsiz durumlarda insan benzeri karar verebilir. Ayrıca insanların anlayabileceği bir mantık yapısı vardır.

H. Genetik Algoritmalar (Genetic Algorithms)

Doğadaki değişimlerden ilham alıp karmaşık problemlere çözüm bulan bir yöntemdir. Birden fazla çözüm üzerinde aynı anda çalışabilmektedir. Farklı problem türlerini çözmek için uygundur. Klasik yöntemlerin çözümleyemediği problemlere çözüm üretir.

“Genetik algoritmalar, evrim teorisinin dayandığı temel prensiplerinden olan doğal seçilim ile en iyi bireylerin hayatta kalması ilkesini taklit eden bir tekniktir. Burada yapılan, en iyi çözümün pek çok çözüm seçeneği içinden arama yapılarak belirlenmesidir. Genetik algoritmalar, bilinen yöntemlerle çözülemeyen veya çözüm süresi problemin büyüklüğüne göre oldukça fazla olan problemlerde, kesin sonuca çok yakın sonuçlar verebilen bir yöntemdir” (Atalay ve ark., 2017).

Genetik algoritmalar; mühendislik, yapay zekâ, ekonomi, biyoloji ve finans gibi birçok alanda uygulama yapabilmektedir.

I. Yapay Yaşam (Artificial Life)

Doğal yaşam süreçlerini yeniden üretmeye çalışan bir bilim dalıdır. Bu alan yaşamın temel ilkelerini daha iyi anlamamızı sağlar. Yapay yaşam dünyada olamayan yeni yaşam sistemleri yaratabilir.

Yapay yaşam, evrimsel süreçlerin anlaşılması, sentetik biyoloji, robotik ve tıbbi araştırmalar gibi birçok alanda uygulama yapabilmektedir.

Yapay yaşam 3 ana dala ayrılmaktadır;

- Dijital (Yazılımsal) Yapay Yaşam (Soft ALife): Bu dalda bilgisayar programları kullanılarak oluşturulan sanal organizmaların değişimleri simüle edilir. Basit kurallardan karmaşık desenlerin ortaya çıktığı simülasyonlar örnek verilebilir.
- Fiziksel (Donanımsal) Yapay Yaşam (Hard ALife): Bu dalda bilgisayar programları ile canlıların fiziksel özellikleri taklit edilir. Yapay böcekler örnek verilebilir.
- Kimyasal (Islak) Yapay Yaşam (Wet ALife): Bu dalda bilgisayar programları aracılığıyla canlıların kimyasal ve biyokimyasal süreçleri kullanılarak yaşam benzeri sistemler oluşturulur. Sentetik DNA yapıları örnek olarak verilebilir.

J. Etik ve Felsefi Yapay Zekâ

Yapay zekânın ilerlemesiyle etik ve felsefi boyutları da önem kazanmıştır. Toplumun her alanına etki edeceğinden dolayı bu boyutların iyi anlaşılması gerekmektedir. İnsanların güvenebilmesi için şeffaf olmalıdır.

Yapay zekâ sistemleri ayrımcılığa yol açabilir. Yapay zekânın aldığı kararların sorumluluğunun kime ait olacağı belirsiz bir problemdir. Çok fazla kişisel veri işler ve bu verilerin korunup korunmayacağı problemlerini doğurur. Yapay zekâ birçok iş kolundan insan yerini alıp işsizliğe yol açabilir.

Yapay zekânın bilinç kazanması durumunda ahlaklı olup olmayacağı, ahlaki değerlerimizi öğretip öğretemeyeceğimiz, haklarının olup olmayacağı, insan ile farklarının ne olacağı, insanla hangi sınırlarla ayrılacağı gibi problemler doğabilmektedir.

2.5. Yapay Zekâ Etik İlkeleri

Yapay zekâyı etik açıdan incelemek için yapay zekânın etik ilkeleri hakkında fikir sahibi olmak gerekir.

A. Şeffaflık

Bilime olan güvenin artması için yapay zekâ sistemleri şeffaf olmalıdır. Sistemi kullananlar nasıl çalıştığını bilmelidir. Etik dışı uygulamaları önlemesi açısından şeffaflık önem arz etmektedir.

Otonom yapay zekâ teknolojileri genellikle şeffaflık ve açıklık gibi özellikler taşımamaktadır. Bu tür yapay zekâ sistemleri, yapılandırılmamış veriler ve algoritmik analizler kullanarak yeni ve özgün bilgiler oluşturabilir. Bu durum, kişisel verilerin ilk kaynağını belirlemeyi zorlaştırabilir ve veri öznelerine yeterli bilgi sağlama görevini güçleştirebilir. Örneğin, bir makine öğrenme modeli, kişisel olmayan verilerden kişisel veriler türetebilir veya mevcut kişisel verileri yeni ve beklenmedik şekillerde analiz edebilir. Ayrıca, makine öğrenmesindeki kara kutu problemi, veri öznelerinin verilerinin nasıl kullanılacağına dair tam ve bilinçli bir onay vermelerini zorlaştırabilir (Turgut, 2024).

B. Dürüstlük

Bu ilke şeffaflığı kapsıyor olsa da ondan daha geniş bir alana sahiptir. Bilimin ana yapısını oluşturur. Üretilen bilgilerin insanlığa faydalı olup olmadığı, bilinçli hatalar yapılıp yapılmadığı da dürüstlük ilkesi içerisinde değerlendirilebilir.

Dürüstlük kuralına uygun veri işleme sürecinde, ilgili kişinin menfaatlerinin korunması, gerekli bilgilendirmelerin doğru şekilde yapılması ve kişinin makul beklentilerinin göz önünde bulundurulması gerekmektedir (Kotil, 2022).

C. Özen

Bilgiler üretilirken hassasiyet gösterilmelidir. İnsanlığa fayda sağlaması, işe yararlılığı göz önünde bulundurulmalıdır. Üretilen bilgiler baştan savma yapılmamalıdır.

Özen, şeffaflık ve dürüstlüğü içeren bir yükümlülüktür. Bilimsel bilgi üretimi ve yayın aşamalarında, bilim insanları ve araştırma kurumlarının sorumluluklarının bilincinde olmaları, bu sorumluluğun gerektirdiği ciddiyeti ve etik duyarlılığı her aşamada göstermeleri gerekmektedir. Bu bağlamda özen ilkesi, baştan savma araştırmalardan kasıtlı etik ihlallere, veri sahtekârlığı ve manipülasyonuna, çıkar çatışmasına, bölerek yayınlama veya aynı çalışmayı birden fazla yerde yayımlama gibi durumlara kadar geniş bir yelpazeyi kapsar. Ayrıca, yazarlık katkıları, araştırma gönüllülerinin hakları ve çıkarlarının korunması gibi konularda da geçerlidir. Yapay zekâ kullanımının artması, bilim insanlarının ve araştırma süreçlerinin özen sorumluluğunu daha da artırmaktadır. Araştırmacılar, bu sorumluluk kapsamında bireylere, topluma ve insanlığa yarar sağlamayı hedeflemeli, araştırma süreç ve sonuçlarının insanlara, çevreye ve diğer canlılara zarar vermemesine dikkat etmelidir (YÖK, 2024).

D. Adalet

Bu sistemler tüm insanlık için eşit erişim olanağı sunmalıdır. Veri setleri herhangi bir önyargı içermemelidir. Cinsiyet, sosyoekonomik koşullar, ırk gibi herhangi bir duruma göre ayrım yapılmamalıdır.

Bilimsel araştırma ve yayın sürecinde, araştırma verilerinin toplandığı bireyler veya toplumlar, diğer canlılar ve çevreye karşı adil ve saygılı olmak, temel bir etik sorumluluktur (YÖK, 2024).

E. Gizlilik

Yapay zekâ sistemleri birçok veriye erişebilmektedir. Erişilen bu verilerin izin olmadan başkalarıyla paylaşılmaması gerekmektedir. Bu veriler kötüye kullanıma karşı korumaya alınmalıdır.

Veri gizliliğinin korunması, hem etik hem de yasal açıdan büyük önem taşır. Bu süreç, veri sağlayanların mahremiyetine saygı gösterilmesini, verilerin kullanımıyla ilgili özerkliklerinin gözetilmesini, aydınlatılmış onam sürecinin titizlikle yürütülmesini, veri toplama, depolama ve aktarım aşamalarında olası ihlalleri öngörüp önlem alınmasını ve gerektiğinde etik kurul onayının alınmasını içerir (YÖK, 2024).

F. Sorumluluk

Yapay zekâ sistemini geliştirenler ve bu sistemi kullananlar, kullanım ve sistemin yayılması gibi durumlarda etik ve yasal sorumluluklarını yerine getirmelidir. Sistemin yanlış verdiği kararlar sonucunda hatanın sorumluluğu belirlenebilmelidir.

Sistemin doğru kararlar vermesi üreticinin sorumluluğundayken verilen kararların doğruluğunu araştırmak kullanıcının sorumluluğudur.

Bilimsel araştırma ve yayın sürecinde, araştırmacılar yapay zekâ tarafından üretilen içeriği kullanırken yasal ve etik sorumluluklarını bilmelidirler. Yapay zekânın karar verme ve akıl yürütme süreçlerinin şeffaf olmaması, araştırmacının yasal ve etik yükümlülüklerini ortadan kaldırmaz (YÖK, 2024).

G. Anlaşılabilirlik

Yapay zekâ sistemlerinin aldığı kararlar ve bu kararların mantıklı olup olmadığı insanlar tarafından anlaşılabilmelidir.

Başlangıçta yapay zekâ sistemleri genellikle kolayca anlaşılabilirken, son yıllarda derin sinir ağları gibi karmaşık yapılarla donatılmış kara kutu modellerinin yaygınlaşması, bu modellerin nasıl çalıştığını anlamak için bir ihtiyaç yaratmıştır. Açıklanabilir yapay zekâ yaklaşımları, tasarımcılara önemli unsurları belirleme ve bu unsurları dikkate alarak süreçlerini yönlendirme amacı güder (Terzi, 2021).

H. İnsan Denetimi

Yapay zekâ sistemleri sadece belirlenen kodlarla çalışan otomatik sistemlerdir. Duruma göre karar veremeyeceğinden tamamen otomatik olmak yerine gerektiğinde insan müdahalesine olanak tanımalıdır.

I. İnsan Haklarına Saygı

Yapay zekânın temel ilkelerinden kabul edilen insan haklarına saygı, temel hak ve özgürlüklere zarar vermeyecek şekilde tasarlanmayı ve uygulamayı gerektirmektedir. Bu sistemler üretilirken insan onurunu zedelememesine özen gösterilmelidir. İnsan haklarını zayıflatacak şekilde kodlanmamalıdır. İnsan haklarına saygı gösterip bu hakları güçlendirmelidir. Yapay zekânın etik ve adaletli bir şekilde kullanılması için kritik bir rehberdir.

J. Çeşitlilik

Yapay zekâ sistemleri geliştirilirken sadece bir alana veya sadece bir gruba hitap etmemelidir. Evrensellik özelliği olmalıdır. Çeşitlilik, daha kapsayıcı ve adil yapay zekâ sistemlerinin geliştirilmesi için kritik önem taşımaktadır. Yapay zekâ sistemlerinin herkesin yararına gelişmesi için bu ilkeye sadık kalınmalıdır.

K. Kapsayıcılık

Yapay zekâ sistemleri tüm insanlığın erişebileceği şekilde olmalıdır. Sadece belirli bir kesime değil herkese fayda sağlayacak şekilde düzenlenmelidir. Kapsayıcılık, yapay zekânın toplumsal eşitliği desteklemesi ve ayrımcılıkla mücadele etmesi için

kritik bir prensiptir. Bu ilkeye bağlı kalınarak geliştirilen yapay zekâ sistemleri, daha adil, eşit ve sürdürülebilir bir topluma katkıda bulunur.

L. Veri Koruma

Yapay zekâ sistemleri sadece ihtiyaç duyulan verileri toplamalı ve kullanmalıdır. İhtiyaç duyulmayan verileri depolamamalıdır. Depolanan veriler konusunda kullanıcıya bilgi verilmelidir. Depolanan veriler başka kişi ve kurumlarla paylaşılmamalıdır.

Yapay zekâ teknolojilerinin hızla gelişmesi ve yaygınlaşmasıyla birlikte, yapay zekâ tarafından üretilen eserlerin telif hakları ve kişisel verilerin korunması gibi konular önemli bir gündem oluşturmuştur. Bu konular, yapay zekâ teknolojileri tarafından yaratılan eserlerin yasal durumlarının belirlenmesi, kişisel verilerin güvenliğinin sağlanması ve yapay zekâ kullanımı sırasında karşılaşılan diğer sorunların çözülmesi gibi çeşitli boyutları kapsamaktadır (Özdal, 2023).

M. Toplumsal Fayda

Bu sistemler üretilirken topluma fayda sağlamasına dikkat edilmelidir. Toplumda dayanışmayı sağlama, toplum refahını yükseltme gibi amaçları olmalıdır.

Yapay zekâ, toplumlara ve kuruluşlara rekabet avantajı sağlama kapasitesine sahiptir. Bu teknoloji sayesinde refah seviyesinin en yüksek düzeyde tutulması ve yaşam konforunun artırılması mümkün olacaktır. Böylece, insanların daha fazla boş zamana sahip olması sağlanacak, bu vakitler sosyal etkinliklere, aile ilişkilerine ve kişisel gelişime ayrılabilecektir (Öztemel, 2020).

N. Çevre Sorumluluğu

Yapay zekâ sistemlerinin geliştirilmesi, uygulanması ve kullanımı sırasında çevreye zarar vermemeyi, hatta çevresel sürdürülebilirliği desteklemeyi amaçlayan bir etik ilkedir. Bu ilke, yapay zekânın çevre üzerindeki olumsuz etkilerini en aza indirmeyi ve sürdürülebilir kalkınmaya katkıda bulunmayı hedefler. Bu sistemlerin enerji tüketimi fazladır. Bu sistemler üretilirken enerji verimli kullanma ve çevreye zarar vermeye özen gösterilmelidir.

O. Yasalara Uyum

Yapay zekâ teknolojilerinin tasarlanması, uygulanması ve kullanımı sırasında yerel, ulusal ve uluslararası yasa ve düzenlemelere tam uyum sağlanmasını ifade eden bir etik ilkedir. Bu ilke, yapay zekâ sistemlerinin hukuka aykırı veya etik olmayan şekillerde kullanılmasını önlemeyi amaçlar. Yasalara uyum, hem geliştiriciler hem de kullanıcılar için temel bir sorumluluk olarak kabul edilir. Bu sistemler yürürlükte olan yasalara uygunluk göstermelidir.

Yapay zekâ sistemleri yanlış kararlar verdiğinde, bu durumun nedenini tatmin edici bir şekilde açıklamak zor olabilir. Bu nedenle, bu sistemlerin daha açıklanabilir ve yasalara uygun olması zorunludur (Terzi, 2021).

P. Etik Değerlendirme

Yapay zekâ sistemleri geliştirilirken üretici etik ilkeleri göz önünde bulundurmalıdır. Bu sistemler mevcut eşitsizliği derinleştirmeyip aynı zamanda yeni eşitsizliklere yol açmamalıdır.

Üst düzey bilişsel yeteneklere sahip yapay zekâ teknolojilerinin bilimsel araştırma ve yayın süreçlerinde kullanımıyla ilgili potansiyel etik sorunların tespit edilmesi, bildirilmesi ve çözülmesi, tüm bilim insanlarının sorumluluğundadır. Bu amaçla, etik hassasiyeti geliştirmek ve desteklemek için uygun kurumsal bir ortamın oluşturulması gerekmektedir (YÖK, 2024).

2.6. Yapay Zekânın Üstünlükleri

İnsan tarih boyunca duyularıyla algıladığı her şeye hükmetmeye çalışmıştır. Düşüncenin işleyişi, zihnin gizemi ve metafizik varlıkların olasılığı gibi konular, sanatçılardan bilim insanlarına ve teologlara kadar geniş bir yelpazede düşünürleri etkilemiştir. İnsanların, kendi gibi düşünen makineler yaratma hayali, temelde karşılaşılan problemlerin çözümüne yönelik bir arayıştır. Yapay zekâ, bilim kurgu eserlerinde genellikle kurtarıcı veya dünyayı ele geçiren bir güç olarak tasvir edilmiştir. Ancak yapay zekâ, hâlâ amacını belirleyen kişinin yönlendirmesiyle çalışan bir araç olarak, günümüzde faydalı çözümler sunmaktadır. Kara ve hava araçlarından kişisel ve ticari ürünlere kadar birçok teknolojik alanda yapay zekâ algoritmaları kullanılmakta ve gelecekte de artan etkisiyle bu alandaki varlığını sürdürmesi beklenmektedir (Süslü, 2019).

Yapay zekâ büyük veri kapasitesine sahip olduğundan dolayı karmaşık işlemleri dahi kısa sürede çözümleyebilmektedir. Yoğun işlem gerektiren durumlarda avantajlı durumdadır. Büyük miktarda verileri analiz edip anlamlı çıkarımlar yapabilir. Bu sayede karmaşık problemlerin çözümüne katkı sağlar. Doğal zekâ ise sınırlı veriyi analiz edebilir.

İnsan beyni ise hızlı olmasına rağmen hız konusunda yapay zekâ ile kıyaslanamaz. Kesintisiz çalışabilme özelliğine sahiptir. İnsanlar ise belirli bir süreden sonra dinlenmeye ihtiyaç duyduğundan kesintisiz çalışamazlar. Kodlandığı alanlarda

hata oranları yok denilecek kadar azdır. İnsan ise yapısı gereği hata yapmaya müsaittir. Verilerin doğruluğu açısından yapay zekâ büyük bir avantaj sağlamaktadır.

Yapay zekâ aynı anda birden fazla problemi çözebilme yeteneğine sahiptir. Doğal zekâ ise aynı anda birden fazla problem çözümüne imkân sağlamamaktadır. Yapay zekâ problemlerin çözümünde çok fazla veriyi analiz edip nesnel kararlar verebilmektedir. Doğal zekâ ise çözümleme de sezgileri ve duyguları ile hareket edebildiğinden nesnel kararlar vermesi oldukça zordur.

Yapay zekâ insan gibi hatalarını fark ederek bir sonraki durumda hataların tekrarlanmasına karşın tepkisel özellik bakımından ayrılmaktadır. Yapay zekânın tepkileri her durumda aynı davranışı sergileyen sisteme sahiptir. Yapay zekâ problemleri her defasında aynı şekilde yanıtlamaktadır. Doğal zekâ ise problemlere her defasında farklı cevap veya yanlış cevap verebilir.

Yapay zekânın performansı ölçülebilir. Performans zamana veya duruma göre değişiklik göstermez. Doğal zekâda ise performansı tam olarak ölçmek zordur. Performans durumlara göre değişiklik gösterebilmektedir. Yapay zekâ karmaşık ve detaylı simülasyonlar yapabilmektedir. Doğal zeka simülasyon yapabilme yeteneğine sahiptir. Fakat yapay zekâ gibi karmaşık ve detaylı simülasyon yapması zordur.

Yapay zekâ çok fazla veriyi saklayabilme kapasitesine sahiptir. İstenilen zamanda saklanılan verileri hızlı ve hatasız bir biçimde sunabilmektedir. Doğal zekâda ise verileri saklayabilme kapasitesi sınırlıdır. Ayrıca zamanla unutulabileceğinden veya yanlış hatırlanabileceğinden istenilen zamanda eksik veya hatalı veri sunması olasıdır.

Yapay zekâ tehlikeli görevlerde kullanılabilmektedir. Fakat doğal zekânın tehlikeli durumlarda kullanılması zordur. Yapay zekâda bilgi aktarımı yapmak oldukça basittir. Doğal zekâ da ise bilgi aktarımı yapabilmek için belirli bir süre gerekmektedir. Yapay zekâda makineler arası geçiş kopyalama ile mümkün olabilmektedir. Doğal zekâda ise bazı alanlar arası geçiş mümkün olmamaktadır.

Yapay zekâ, doğal zekâya kıyasla kalıcıdır. İnsanlar, öğrendikleri ve deneyimledikleri bilgileri zamanla unutabilirler, özellikle aynı deneyimleri tekrar yaşamadıkları sürece. Ancak yapay zekâda, bir bilgisayarın veya robotun belleğine kaydedilen veriler, silinmediği sürece saklanır ve unutulmaz (Adalı, 2017). Bu nedenle, bilgisayarların unutma kapasitesi yoktur. Diğer bir fark ise, doğal zekâya sahip bir kişinin bilgi birikimini tam anlamıyla başka bir insana aktaramaması, oysa yapay zekâda bir bilgisayarın tüm verileri kolaylıkla başka bir bilgisayara aktarılabilir (Çelebi ve ark., 2019). Ayrıca, bir bilgisayarın zekâ seviyesini artırmak, bir insanın zekâ

seviyesini yükseltmekten genellikle daha basit ve hızlıdır (Adalı, 2017). Yapay zekâ, daha kolay elde edilebilirken, doğal zekânın geliştirilmesi zordur. İnsanlar, aynı olaya farklı yaklaşımlar sergileyebilirler çünkü her bireyin zekâ düzeyi farklıdır (Çelebi ve ark., 2019). Yapay zekâ sistemleri ise, aynı ya da benzer durumlarda tutarlı yanıtlar verir, çünkü bu sistemlere aynı algoritmalar veya kodlamalar yüklenmiştir (Adalı, 2017; Çelebi ve ark., 2019).

Yapay zekâ insan müdahalesiyle yapılmasına karşın birçok alanda doğal zekâdan daha fazla avantaj sağlamaktadır. Yapay zekâ sistemlerinin üstünlükleri, belirli alanlarda insan kapasitesini aşarak endüstriyel, bilimsel ve günlük yaşamda devrim yaratmaktadır. Ancak, bu üstünlüklerin etik, yasal ve sosyal boyutları da dikkate almak büyük önem arz etmektedir.

2.7. Yapay Zekânın Etik Kullanımını Sağlamak İçin Yapılan Çalışmalar

2.7.1. Türkiye'de yapılan çalışmalar

Yapay zekânın etik kullanımı büyük önem arz etmektedir. Türkiye de bu soruna kayıtsız kalmamaktadır. Akademi, kamu ve özel sektörde bunun için çalışmalar yapılmaktadır.

2021 yılında yayınladığı " Ulusal Yapay Zekâ Stratejisi " ile etik bir yapay zekânın gelişmesini amaçlamıştır. Bu strateji daha güvenilir ve eşitlikçi bir yapay zekâ oluşturulması için gerekli etik ilkeleri içermektedir. Bu strateji insan hakları ve etik değerlere uygun sistemler geliştirilmesini amaçlamaktadır.

Türkiye'nin Ulusal Yapay Zekâ Stratejisi, ülkenin dijital dönüşüm sürecinde önemli bir rol oynamakta ve yapay zeka teknolojilerinin topluma, ekonomiye ve devlet hizmetlerine entegre edilmesini amaçlamaktadır. Bu strateji, Türkiye'nin yapay zeka alanında uluslararası bir oyuncu olarak konumlanmasını sağlayacak kapsamlı bir vizyon sunmaktadır.

Yapay zekâ sistemlerini etik bir şekilde kullanımını Kişisel Verilerin Korunması Kanunu (KVKK) aracılığıyla desteklemiştir. yapay zeka uygulamaları büyük miktarda veri işlediğinden, KVKK ile uyumlu olmak büyük önem taşır. Yapay zekanın veri toplama, analiz etme ve karar verme süreçlerinde kişisel verilerin korunmasına dikkat edilmesi gerekir. Özellikle, veri minimizasyonu ve anonimleştirme gibi prensipler,

yapay zeka uygulamaları geliştirilirken göz önünde bulundurulmalıdır. Bu kanun ile kişisel verilerin etik bir şekilde kullanımı sağlanmaya çalışılmıştır.

Türkiye Yapay Zekâ İnisiyatifi (TRAI) platformu oluşturulmuştur. Bu platform yapay zekânın etik bir şekilde kullanılması için çalışmalar yapmaktadır. Bu platform aracılığıyla eğitimler verilip bireyler bilinçlendirilmektedir.

Yapay zekânın etik kullanımı, toplumsal ve çevresel refah, çeşitlilik, . .ayrımcılığın önlenmesi, adalet, gizlilik ve veri yönetimi gibi temel değerlere dayanır. Özellikle gizlilik ve veri yönetimi, kişisel verilerin korunmasını sağlayan mevcut yasal çerçevelerle desteklenmektedir. Hesap verebilirlik ilkesi, yapay zekâ sistemlerinin neden olabileceği zararlardan dolayı geliştiricilerin ve uygulayıcıların hukuki sorumluluk taşımalarını sağlar. Bu durum, teknolojinin insan haklarına ve hukuka uygun bir şekilde geliştirilip kullanılmasını garanti eder (Webtures, 2024).

Üniversitelerde etik bir yapay zekâ için çalışmalar yapılmaktadır. Ayrıca etik bir yapay zekâ oluşturulması konusunda üniversiteler eğitimler verilip bu konuda bilinçlenme sağlamaya çalışmaktadır.

Kamu ve özel sektörde yapay zekânın etik bir şekilde kullanımı sağlamak için birçok çalışma yapılmaktadır. Ayrıca etik kullanımı arttırmak için kamu ve özel sektör birlikte de çalışmalar yürütmektedir.

Türkiye'de yapay zekânın etik bir şekilde ilerlemesi için denetim mekanizmaları oluşturulmaktadır. Bu denetim mekanizmaları sayesinde uygulamaların etik değerlere uygunluğu denetlenmekte ve etik dışı uygulamaların kullanımı azaltılmaya çalışılmaktadır.

Türkiye'de yapay zekânın şeffaf bir şekilde tasarlanmasına özen gösterilmektedir. Bu şekilde uygulamaların güvenilirliği arttırılmaya çalışılmaktadır.

Türkiye etik bir yapay zekâ kullanımı için Avrupa Birliği ile birlikte çalışmalar yapmaktadır. Bu şekilde birçok ulusu kapsayacak etik değerler belirlemeye çalışmaktadır.

Her geçen gün yapay zekâ hayatımızda daha fazla yer kaplamaktadır. Bu yüzden etik değerlere uygunluğun da önemi giderek artmaktadır. Birçok ülkede olduğu gibi Türkiye'de de etik bir yapay zekâ oluşturulması için çalışmalar yapılmaya devam etmektedir.

2.7.2. Dünyada yapılan çalışmalar

Yapay zekâ teknolojisinin gelişmesiyle birlikte dünya genelinde etik yapay zekâ oluşturma çalışmaları artış göstermektedir. Bu sistemlerin etik dışı kullanımı ciddi sorunlar doğurduğundan birçok ülke bu konuda çalışma yapmaktadır.

A. Avrupa Birliği (AB)

2018 yılında Avrupa Komisyonu, yapay zekâ konusunda yüksek bir uzman grubu (High-Level Expert Group on Artificial Intelligence) görevlendirmiştir. Bu grup, 2019 yılında "Güvenilir Yapay Zekâ İçin Etik Rehber İlkeleri (Ethics Guidelines for Trustworthy AI)" adlı raporu yayımladı.

AB'nin 2018 yılında yürürlüğe giren GDPR'si, yapay zekâ uygulamaları için kritik bir hukuki çerçeve sunar. Yapay zekâ sistemlerinin kişisel verileri nasıl işlediği, GDPR'nin gerekliliklerine uygun olmalıdır. Bu, veri sahiplerine yapay zekâ sistemlerinde nasıl kararlar alındığı konusunda şeffaflık ve hesap verebilirlik sağlar.

AB, 2018 yılında "Yapay Zekâ İçin Avrupa Stratejisi"ni yayımladı. Bu strateji, yapay zekânın etik ve insan haklarına uygun olarak geliştirilmesini teşvik etmeyi amaçlamaktadır.

Avrupa'daki yapay zekâ araştırmalarını destekleyen EurAI, etik yapay zekâ araştırmalarını teşvik etmekte ve bu alanda farkındalığı artırmak için çalışmaktadır.

Avrupa Birliği etik bir yapay zekâ geliştirilmesi için "Güvenilir AI İçin Etik Yönergeleri yayınladı. Bu yönergeler etik bir yapay zekânın gelişmesi için temel ilkeler içermektedir. Bu ilkelere göre tasarlan yapay zekâ projelerini arttırmayı hedeflemektedir.

B. OECD AI İlkeleri

Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) yapay zekânın etik kullanımı için bazı ilkeler belirlemiştir. Bu ilkeler OECD ülkelerinde etik bir yapay zekâ tasarlanması için temel yapı taşı sayılmaktadır.

OECD, yapay zekâ (YZ) için beş temel değer temelli ilke önermektedir. Bu ilkeler, yapay zekânın sorumlu bir şekilde yönetilmesini sağlamak amacıyla belirlenmiştir. İşte bu ilkeler:

a) Yapay zekâ, kapsayıcı büyümeyi, sürdürülebilir kalkınmayı ve refahı teşvik ederek insanlara ve gezegene fayda sağlamalıdır.

b) Yapay zekâ sistemleri hukukun üstünlüğüne, insan haklarına, demokratik değerlere ve çeşitliliğe saygı duyacak şekilde tasarlanmalı ve adil ve adil bir toplum

sağlamak için uygun güvenlik önlemleri içermelidir - örneğin gerektiğinde insan müdahalesini mümkün kılmalıdır.

c) İnsanların yapay zekâ tabanlı sonuçları anlamasını ve bunlara meydan okuyabilmesini sağlamak için yapay zekâ sistemleri etrafında şeffaflık ve sorumlu bir açıklama olmalıdır.

d) Yapay zekâ sistemleri, yaşam döngüleri boyunca sağlam, güvenli ve emniyetli bir şekilde işlemeli ve potansiyel riskler sürekli olarak değerlendirilmeli ve yönetilmelidir.

e) Yapay zekâ sistemlerini geliştiren, dağıtan veya çalıştıran kuruluşlar ve bireyler, yukarıdaki ilkeler doğrultusunda düzgün işleyişlerinden sorumlu tutulmalıdır (Efe, 2021).

C. Birleşmiş Milletler (BM)

Birleşmiş Milletler insan haklarına saygılı etik bir yapay zekâ tasarlanması için öneriler sunmuştur. UNESCO etik bir yapay zekâ geliştirmek amacıyla "Yapay Zekânın Etiği için Küresel Rehber"i yayımlamıştır.

BM Genel Sekreteri, yapay zekânın sürdürülebilir kalkınma hedeflerine nasıl katkıda bulunabileceğine dair bir rapor yayımladı. Bu rapor, yapay zekânın sosyal ve ekonomik etkilerini analiz ederken, etik ve adil kullanımına yönelik önerilerde bulunmaktadır.

BM İnsan Hakları Konseyi, yapay zekânın insan hakları üzerindeki etkilerini inceleyen çalışmalara öncülük etmektedir. Konsey, yapay zekânın insan haklarına zarar vermemesi için uluslararası standartlar ve ilkeler geliştirmeyi hedefler.

BM'nin ICT Paneli, dijital dönüşümün ve yapay zekânın toplumsal etkilerini incelemekte ve bu konuda uluslararası işbirliklerini teşvik etmektedir. Panel, yapay zekânın adil, şeffaf ve sürdürülebilir bir şekilde kullanılması için politikalar ve öneriler geliştirmektedir.

BM Bilim ve Teknoloji Konseyi, bilimsel ve teknolojik gelişmelerin etik yönlerini ele alır ve yapay zekânın etik ve adil bir şekilde kullanılmasını teşvik eder.

BM Çocuk Hakları Komitesi, yapay zekânın çocuklar üzerindeki etkilerini incelemekte ve çocukların dijital haklarını korumak için stratejiler geliştirmektedir.

D. Amerika Birleşik Devletleri (ABD)

2016 yılında Google, Facebook, Amazon, IBM ve Microsoft gibi teknoloji devlerinin bir araya gelmesiyle kurulan Partnership on AI, yapay zekânın sorumlu, etik

ve şeffaf bir şekilde geliştirilmesi ve uygulanması için rehberlik sağlamayı amaçlamaktadır.

The Future of Life Institute, yapay zekânın gelecekte insanlığa yönelik risklerini inceleyen ve bu risklerin minimize edilmesi için çalışan bir organizasyondur. 2017 yılında bu kurum, yapay zekâ etik ilkeleri için Asilomar AI İlkeleri'ni yayımlamıştır.

2019 yılında Beyaz Saray, federal kurumların yapay zekâ teknolojilerini kullanırken uyması gereken rehber ilkeler belirledi.

ABD Savunma Bakanlığı'na bağlı olan The Defense Innovation Board, 2019 yılında askeri yapay zekâ kullanımına yönelik beş temel etik ilke önerdi. Bu ilkeler; sorumluluk, adalet, izlenebilirlik, güvenilirlik ve yönetilebilirliktir. Bu ilkeler, savunma amaçlı yapay zekâ uygulamalarının etik kurallar çerçevesinde kullanılması için rehberlik sağlamaktadır.

2020 yılında yürürlüğe giren Ulusal Yapay Zekâ İnisiyatifi Yasası (National AI Initiative Act), ABD'nin yapay zekâ alanındaki küresel liderliğini sürdürmek ve etik yapay zekâ kullanımını teşvik etmek amacıyla oluşturuldu.

National Institute of Standards and Technology, ABD'de teknoloji standartlarının belirlenmesinde kritik bir rol oynamaktadır. 2021 yılında, NIST "Güvenilir Yapay Zekâ (Trustworthy AI)" çerçevesini geliştirmek için çalışmalara başladı.

ABD bu sistemler gelişirken etik kurallara uyması için bazı stratejiler belirlemiştir. AI Ulusal Güvenlik Komisyonu (NSCAI), yapay zekânın etik kullanımı için kılavuzlar hazırlamıştır.

Stanford Üniversitesi'nde kurulan İnsan Merkezli Yapay Zekâ Enstitüsü (Human-Centered AI Institute - HAI), yapay zekânın insanlara ve topluma yönelik etkilerini incelemekte ve etik kullanımı teşvik etmektedir. Enstitü, araştırma ve eğitim yoluyla yapay zekânı etik sorunlarını ele almayı ve bu alanda politika önerileri geliştirmeyi hedeflemektedir.

New York Üniversitesi'ne bağlı AI Now Institute, yapay zekânın sosyal etkilerini incelemekte ve politika yapıcılar için rehberlik sunmaktadır

E. Çin

Çin hükümeti, 2017 yılında yayınladığı "Yeni Nesil Yapay Zekâ Geliştirme Planı" ile yapay zekâ teknolojilerinde küresel lider olmayı hedeflediğini açıkladı. Bu plan, yapay zekânı etik yönlerine de değinmekte ve yapay zekânın güvenli, kontrollü ve toplumsal çıkarları gözeten bir şekilde geliştirilmesi gerektiğini vurgulamaktadır.

2019 yılında, Pekin'de yapay zekâ araştırmaları ve uygulamaları üzerine çalışan akademik ve endüstri liderleri, "Pekin Yapay Zekâ İlkeleri"ni yayımladı. Bu ilkeler, yapay zekânı etik olarak geliştirilmesi ve uygulanması için öneriler sunar.

2021 yılında, Çin hükümeti "Yapay Zekâ Etiği ve Yönetimi Kılavuzu" adlı bir belge yayımladı. Bu kılavuz, yapay zekânın geliştirilmesi ve uygulanmasında insan hakları, güvenlik, gizlilik ve şeffaflık ilkelerine uyulmasını öngörmektedir.

Çin yapay zekâ uygulamalarının gelişmesi için etik standartlar belirlemiştir. "Yapay Zekâ Ahlakı"na uygun yönergeler yayınlamış. Bu yönergelere göre insan haklarına saygılı bir yapay zekâ geliştirilmesi amaçlanmıştır.

Çin Siber Uzay İdaresi, yapay zekâ ve veri güvenliği konularında çeşitli düzenlemeler yapmaktadır. Özellikle, büyük veri ve yapay zekâ kullanımı yoluyla bireylerin gizliliğinin korunması ve veri güvenliğinin sağlanması üzerinde durulmaktadır.

F. Kanada

Kanada etik bir yapay zekâ tasarlamak için "Pan-Canadian AI Strategy"yi geliştirmiştir. Bu strateji etik değerlere uygun yapay zekâ tasarlanmasını sağlamıştır.

G. Google

Google etik bir yapay zekâ geliştirmek için "AI Principles" adlı bir kılavuz yayınlamıştır. Bu ilkeler etik değerlere uygun davranmayı gerektirir.

H. Microsoft

Microsoft, yapay zekânın etik bir şekilde kullanımını desteklemek için "AI for Good" girişiminde bulunmuştur. Bu sayede etik değerlere uygun yapay zekâ sistemleri geliştirilmiştir.

I. International Business Machines (IBM)

IBM etik değerlere uygun bir yapay zekâ tasarlanması için "IBM Principles for Trust and Transparency" adlı ilkeler yayınlamıştır. Bu şirket, adaletli ve anlaşılabilir yapay zekâ sistemlerinin geliştirilmesini amaçlamaktadır.

J. AI Now Institute (NYU)

AI Now Institute, yapay zekânın etik değerlerini araştıran bir merkezdir. Yapay zekânın temel etik ilkelerini inceleyen raporlar hazırlamaktadır.

K. Harvard Berkman Klein Center

Harvard Üniversitesi bünyesinde bulunan Berkman Klein Center, etik sorunlarla ilgili araştırmalar yapmaktadır. Bu sorunlara etik kurallara uygun çözümler sunmaktadır.

L. Oxford Internet Institute

Oxford Üniversitesi bünyesinde bulunan Oxford Internet Institute, etik bir yapay zekâ ile ilgili araştırmalar yapmaktadır. Yapay zekânın insan üzerindeki etkileri araştırılmaktadır.

M. AI Ethics Lab

AI Ethics Lab, yapay zekâ tasarımcılarına uygulamalarının etik kurallara uygun olması konusunda destek sağlamaktadır. Bu sayede etik dışı uygulamaları azaltmaya çalışmaktadır.

N. Partnership on AI

Partnership on AI, yapay zekânın etik kullanımını desteklemek için kurumlarla işbirliği yapan bir platformdur. Bu şekilde etik değerlere uygun uygulamalar yapılmasını hedeflemektedir.

O. Future of Life Institute

Future of Life Institute, yapay zekâ sistemlerinin etik değerlere uygun ve güvenli bir şekilde tasarlanması için çalışmalar yapmaktadır. Etik dışı uygulamalardan doğabilecek sorunların çözümlerini araştırmaktadır.

Etik bir yapay zekâ geliştirmek için dünyanın birçok yerinde çalışmalar devam etmektedir. Bu çalışmalar sayesinde yapay zekânın temel etik ilkelerine uygun şekilde sistemler tasarlanması amaçlanmaktadır.

3. MATERYAL VE YÖNTEM

Bu tezde, yapay zekâ (YZ) uygulamalarında ortaya çıkan etik sorunları incelemek amacıyla nitel araştırma yöntemi kullanılmıştır. Araştırma, yapay zekâ uygulamalarını etik çerçevesinde detaylı inceleyerek veri toplama ve toplanan verileri analiz etme şeklindedir. Araştırma, literatür taraması yöntemini içermektedir.

Bu tez, yapay zekâ uygulamalarındaki etik sorunları detaylı inceleyerek bu sorunlara çözüm üretmek amacıyla yapılmaktadır. Araştırma, yapay zekâ uygulamalarının etik bağlamında değerlendirmesini hedeflemektedir.

Nitel araştırma kapsamında ele alınacak temel araştırma soruları şunlardır:

- Yapay zekâ temel etik ilkeleri nelerdir?
- Yapay zekâ uygulamalarından doğabilecek sorunlar nelerdir?
- Etik bir yapay zekâ için mevcutta yapılan çalışmalar nelerdir?
- Yapay zekânın etik tasarımı ve kullanımını için hangi önlemler alınmalıdır?

Araştırmaya başlarken, yapay zekâ ve etik üzerine yapılan çalışmalar incelenerek teorik bilgiler toplanacaktır. Bu literatür taraması, yapay zekânın farklı uygulamalarını ve bu uygulamaların doğurduğu etik problemleri anlamanın temelini oluşturacaktır.

Toplanan veriler, tematik analiz yöntemiyle incelenecektir. Tematik analiz, verilerdeki tema ve kategorileri belirleyip ana temayı bulmayı sağlamaktadır. Her ana tema, yapay zekânın etik bağlamında detaylı görüş elde etmek hedefiyle analiz edilecektir.

Nitel araştırmada güvenirlik ve geçerlilik, tutarlı ve doğru verilerle sağlanmaktadır. Verilerin güvenirliği, farklı kaynaklar arasında tutarlılık sağlanarak artırılacaktır.

Bu araştırmanın sınırlılıklarını genişletmek için veri toplanırken detaylı bir araştırma yapılmaya çalışacaktır.

Nitel araştırma yöntemleri, yapay zekâ uygulamalarının etik bağlamda detaylı bir şekilde değerlendirebilmesi için güçlü bir yöntemdir. Bu tez, yapay zekâ etik problemlerini anlayıp bu soruna en uygun çözümleri üretebilmeyi hedeflemektedir.

3.1. Materyal

Kullanılan veri setleri aşağıda açıklanmıştır.

3.1.1. Kredi kartı istemcilerinin temerrüdü veri kümesi

Kredi kartı istemcilerinin temerrüdü veri seti (Default of Credit Card Clients Dataset), çoğunlukla makine öğrenimi ve veri analitiği çalışmaları için kullanılan popüler bir veri setidir. Bu veri seti, Tayvan'da kredi kartı kullanıcılarına ait finansal ve demografik bilgileri içerir. Kredi Kartı İstemcilerinin Temerrüdü Veri Kümesi, Tayvan'daki 30.000 kredi kartı istemcisi hakkında bilgi içerir ve temerrüt riskini tahmin etmeye odaklanır.

Kredi kartı istemcilerinin temerrüdü üzerine oluşturulan veri setlerinde genellikle şu tür bilgiler yer alır:

A. **Müşteri Profili ve Demografik Bilgiler:**
   - Yaş, cinsiyet, medeni durum, eğitim seviyesi ve meslek gibi bireyin temel özelliklerini tanımlayan değişkenler.
   - Bu bilgiler, kredi kartı kullanım davranışlarını anlamak ve temerrüt riski tahmini yapmak için önemli bir rol oynar.

B. **Kredi Kartı Kullanım Verileri:**
   - Kredi kartı limiti, aylık harcamalar, dönem sonu borç tutarları ve ödeme geçmişi gibi kredi kartına ilişkin detaylı bilgiler.
   - Bu veriler, müşterinin finansal yönetim alışkanlıklarını ve potansiyel risklerini değerlendirmeye yardımcı olur.

C. **Finansal Durum ve Sosyo-Ekonomik Faktörler:**
   - Müşterinin gelir düzeyi, konut sahipliği durumu, araç sahibi olup olmadığı gibi değişkenler.
   - Bu veriler, bireyin genel ekonomik durumunu yansıtarak kredi risk analizine katkı sağlar.

D. **Ödeme ve Temerrüt Geçmişi:**
   - Daha önce yaşanmış temerrüt olayları, gecikmeli ödemelerin sıklığı ve süresi gibi ödeme davranışını gösteren bilgiler.
   - Müşterinin ödeme alışkanlıklarını değerlendirmek ve temerrüt tahmininde bulunmak için önemlidir.

E. **Makroekonomik Göstergeler:**
   - Enflasyon oranı, faiz oranları ve işsizlik gibi ekonomik göstergeler.
   - Bu tür veriler, bireylerin ödeme kapasitesini etkileyen dış faktörlerin analizinde kullanılır.

Bu tür değişkenler, bireysel ve çevresel faktörleri bir arada değerlendirerek kredi temerrüt riski üzerine derinlemesine bir analiz yapmayı mümkün kılar.

Kredi kartı müşterilerinin temerrüdü veri seti, finansal kuruluşlar açısından önemli bir bilgi kaynağıdır. Bu tür bir veri seti sayesinde çeşitli stratejik ve operasyonel faydalar elde edilebilir:

A. **Risk Analizi:**
   - Müşterilerin kredi geri ödeme davranışlarını değerlendirerek, yeni kredi başvurularında daha bilinçli kararlar alınabilir.
   - Mevcut müşteriler için uygun kredi limitleri belirlenerek finansal risklerin minimize edilmesi sağlanır.

B. **Tahmin Modellerinin Geliştirilmesi:**
   - Geçmiş ödeme verileri kullanılarak, temerrüt riskini öngörebilecek istatistiksel analizler ve makine öğrenimi modelleri oluşturulabilir.
   - Bu modeller, yüksek risk taşıyan müşterilerin önceden tespit edilmesine ve olası zararların önlenmesine yardımcı olur.

C. **Müşteri Segmentasyonu:**
   - Veri seti, müşterilerin özelliklerine göre segmentlere ayrılmasını kolaylaştırır.
   - Bu segmentler, farklı müşteri gruplarına yönelik özelleştirilmiş risk yönetimi stratejileri geliştirmede kullanılır.

D. **Pazarlama ve Müşteri Yönetimi:**
   - Müşterilerin harcama ve ödeme davranışları analiz edilerek, kişiselleştirilmiş pazarlama kampanyaları planlanabilir.
   - Sadık müşterileri ödüllendirme veya düşük risk grubundaki müşterilere ek avantajlar sunma gibi yöntemlerle müşteri bağlılığı artırılabilir.

E. **Dolandırıcılık Tespiti:**
   - Harcama alışkanlıklarından sapma veya ödeme düzensizlikleri gibi anormal durumlar tespit edilerek dolandırıcılık ihtimaline karşı erken önlemler alınabilir.

Bu analizler ve uygulamalar, hem müşteri memnuniyetini artırmayı hem de finans kuruluşlarının risk yönetimini optimize etmeyi hedefler.

Kredi kartı müşterilerinin temerrüdü veri seti, finansal kararların daha doğru ve etkili alınmasına yardımcı olur. Ancak bu verilerin toplanması, saklanması ve analiz edilmesi sırasında bazı zorluklarla karşılaşılabilir. Bu zorluklar arasında veri gizliliği, veri kalitesi, modelin karmaşıklığı ve düzenleyici kurallar sayılabilir.

Sonuç olarak, kredi kartı müşterilerinin temerrüdü veri seti, finans kuruluşları için büyük bir potansiyel taşımaktadır. Bu verilerin doğru ve etkili bir şekilde kullanılması, hem kurumun karlılığını artırır hem de müşteri memnuniyetini sağlar.

Veri seti anonimleştirilmiştir; dolayısıyla müşterilerin kimlik bilgileri yer almamaktadır. Çalışma sırasında etik değerlere ve veri gizliliğine uygun hareket edilmelidir.

Şekil 3.1. Kredi kartı istemcileri (odakarge.com)

3.1.2. 70.000'den fazla iş başvurusunda bulunanın istihdam edilebilirlik sınıflandırması veri kümesi

70.000'den fazla iş başvurusunda bulunanın istihdam edilebilirlik sınıflandırması veri seti, iş piyasası analizi, insan kaynakları süreçleri ve istihdam stratejileri üzerinde çalışmalar yapmak için kullanılan değerli bir veri kaynağıdır. Bu tür bir veri setiyle, adayların işe alınma potansiyelini değerlendiren modeller geliştirilebilir.

70.000'den fazla iş başvurusunda bulunanın istihdam edilebilirlik sınıflandırması üzerine oluşturulan veri setlerinde genellikle şu tür bilgiler yer alır:

A. **Demografik Bilgiler:**
   - Yaş, cinsiyet, eğitim seviyesi, deneyim, coğrafi konum gibi temel bilgiler, adayların genel profilini oluşturmaya yardımcı olur.

B. Eğitim ve Sertifikalar:
   - Adayların sahip olduğu diploma, sertifika ve kurslar, sahip oldukları teknik bilgi ve becerileri gösterir.

C. **İş Deneyimi:**
   - Önceki işlerde üstlenilen roller, sorumluluklar ve başarılar, iş deneyimini ve sektörel uyumu değerlendirmek için önemlidir.

D. **Yumuşak Beceriler:**
   - İletişim, problem çözme, ekip çalışması gibi yumuşak beceriler, iş başarısı için giderek daha önemli hale gelmektedir.

E. **Başvuru Kalitesi:**
   - Özgeçmişin yazım kalitesi, kapak mektubu, referanslar ve online varlık gibi faktörler, adayın dikkatli ve özverili olup olmadığını gösterir.

F. **Mülakat Performansı:**
   - Mülakat sırasında sergilenen iletişim becerileri, teknik bilgi, problem çözme yeteneği ve şirket kültürüyle uyum gibi faktörler, adayın işe uygunluğunu değerlendirmek için kullanılır.

G. Psikolojik Test Sonuçları:
   - Bazı şirketler, adayların kişilik özelliklerini ve yeteneklerini ölçmek için psikolojik testler kullanır.

H. **Sektör ve Pozisyon**:
   - Adayların başvurdukları sektör ve pozisyon, o sektöre özgü beceri ve deneyimlerin önemini vurgular.

Bu tür bir veri setiyle, istihdam sürecinde veriye dayalı kararlar almak ve operasyonel verimliliği artırmak mümkün hale gelir. 70.000'den fazla iş başvurusunda bulunanın istihdam edilebilirlik sınıflandırması veri setinin faydaları şunlardır:

A. **İşverenler İçin:**
- İdeal aday profilini belirleme
- Aday havuzunu daha etkili bir şekilde yönetme
- İş ilanlarını daha doğru hedefleme

- Mülakat süreçlerini optimize etme
- Performans yönetimi sistemlerini geliştirme

B. **Adaylar İçin:**

- Kendi güçlü ve zayıf yönlerini anlamalarına yardımcı olma
- Gelişim alanlarını belirleme
- Kariyer hedeflerine ulaşmak için daha bilinçli kararlar verme

C. **Hükümet ve Politikacılar İçin:**

- İşgücü piyasasındaki eğilimleri anlama
- Eğitim ve istihdam politikalarını geliştirme
- İşsizlik sorununu çözmeye yönelik stratejiler geliştirme

70.000'den fazla iş başvurusunun incelenmesiyle elde edilen veri seti, işgücü piyasası hakkında derinlemesine bir anlayış sunar ve hem işverenler hem de adaylar için değerli bilgiler sağlar. Bu verilerin doğru ve kapsamlı bir şekilde analiz edilmesi, daha iyi kararlar alınmasına ve işsizlik sorununa çözüm bulunmasına katkı sağlayabilir.

Şekil 3.2. İşe alım (evam.tuik.gov.tr)

3.2. Metod

Bu tezde kullanılan metodlar aşağıda açıklanmıştır.

3.2.1. COMPASS

COMPASS (Correctional Offender Management Profiling for Alternative Sanctions), ceza adalet sisteminde kullanılan bir risk değerlendirme aracıdır. Bu araç, suçluların yeniden suç işleme olasılığını tahmin etmek ve ceza adaleti süreçlerinde daha bilinçli kararlar alınmasına yardımcı olmak için geliştirilmiştir.

3.2.2. Adverserial Testing

Adversarial Testing (Saldırgan Test) genellikle makine öğrenimi, özellikle de derin öğrenme modellerinde, sistemlerin dayanıklılığını ve güvenilirliğini değerlendirmek için kullanılan bir tekniktir. Amaç, bir modelin zayıflıklarını veya hassas noktalarını tespit etmek ve bu zayıflıklardan faydalanarak modeli yanıltan özel durumları (adversarial örnekler) ortaya çıkarmaktır.

3.2.3. AI Fairness 360

AI Fairness 360 (AIF360), IBM tarafından geliştirilmiş, makine öğrenimi modellerindeki önyargıları (bias) tespit etmek ve azaltmak için kullanılan açık kaynaklı bir araç setidir. Bu araç, adil ve tarafsız yapay zekâ sistemleri oluşturmayı hedefler ve hem akademik hem de endüstriyel projelerde yaygın olarak kullanılır.

3.2.4. ChatGPT

Uzun süreli ve kapsamlı araştırmaların sonucunda yapay zekâ teknolojileri hızla gelişmiş ve birçok uygulama ortaya çıkmıştır. Yapay zekâ tabanlı öğrenme yöntemleri, insan-makine etkileşiminde sürekli olarak ilerlemekte ve teknolojik yeniliklerle katkıda bulunmaktadır. Bu önemli katkılardan biri, OpenAI tarafından 30 Kasım 2022'de tanıtılan ChatGPT'dir. Sadece 5 gün içinde 1 milyon kullanıcıya ulaşan ChatGPT, GPT-3.5.5 dil modeline dayanmaktadır ve bugüne kadar geliştirilen en kapsamlı yapay zekâ sistemlerinden biridir. ChatGPT, kullanıcıların girdilerine dayanarak hızlı ve karmaşık çıktılar üretebilme yeteneğine sahip olup, insan üretimiyle ayırt edilmesi zor bir

performans sergilemektedir (Karakoç Keskin, 2023). ChatGPT'nin sağladığı avantajlar, yüksek kullanıcı memnuniyeti ve kullanıcı sayısında belirgin bir artışa yol açmıştır (Kırık ve ark., 2023).

ChatGPT doğal dil işleme tekniklerini kullanarak sorulan sorulara cevap veren bir yapay zekâ uygulamasıdır.

İnsansı yanıtlar verebilme yeteneği sayesinde kısa sürede büyük ilgi görmeyi başarmıştır. İletişim alanına getirdiği yenilik ve devrim niteliğindeki etkisi, onu diğer yapay zekâ uygulamalarından ayırmaktadır (Koçyiğit ve ark., 2023).

İnternette bulunun büyük miktardaki veriler konusunda eğitilir. Bu sayede uygun kelimeleri bir araya getirerek anlamlı metinler oluşturması sağlanmaktadır.

Bu uygulama insanlarla sohbet edebilecek şekilde programlanmıştır. Sohbet esnasında uygulama cevaplarına insanların yaptığı geri bildirimleri işleyerek daha sonra bu geri bildirimleri en uygun şekilde kullanabilmektedir.

Bu uygulamanın daha kaliteli bir hale gelmesi için kullanıcıların verilen cevapları puanlaması istenmektedir. Bu puanlamalar doğrultusunda gerekli geliştirmeler yapılmaktadır. Bu da pekiştirmeli öğrenme tekniğini kullandığını göstermektedir.

ChatGPT'ye bir soru sorulduğunda uygulama bunu anlayabilmek için sayısal formata çevirir. Bu format belirli aşamalardan geçerek daha önceden öğrendiği bilgiler aracılığıyla sorgulama yapar. Bu sorgulama sonucunda en uygun yanıtı belirleyerek metine dönüştürür ve kullanıcıya sunar.

Teknik destek sağlama, şiir veya hikâye gibi türler yazabilme, sorulara uygun cevap verdiğinden eğitim, kodlama yapabilme, asistanlık yapabilme, yazım hatalarını düzeltme, içerik üretimi, çeviri yapma vb. birçok alanda destek sağlamaktadır.

Çok fazla veri setine sahip olduğundan dolayı farklı konulardaki soruları cevaplandırabilmektedir. Uygulama devamlı olarak geliştirilmeye devam ettiğinde güncel konulara hâkim olabilmektedir. İnsanlardan çok daha hızlı bir şekilde yanıtlama yapmaları avantaj sağlamaktadır. Bu uygulama birçok kullanıcının aynı zamanda kullanabileceği şekilde tasarlanmıştır. Uygulama sürekli aktif olduğundan ve her yerde erişim imkânı sağladığından dolayı istenildiği zaman kullanılabilmektedir.

Bu uygulama çok fazla veri setine erişebildiğinden dolayı yanlış bilgileri de hafızasında bulundurma ihtimali yüksektir. Verilen cevapların doğruluğunu kontrol etmek insan sorumluluğundadır. Bu uygulama belirli bir grubu ayıracak bir veri setine sahip olabileceğinden dolayı önyargılı cevaplar vermesi muhtemeldir. Bu uygulamalar

belirli aralıklarla güncellenmektedir. Fakat en son güncellemeden sonra gelişen durumlara hâkim olmamaktadır.

Doğal dil işleme alanındaki hızlı ilerlemelerle birlikte ChatGPT, heyecan verici ve vazgeçilmez bir teknoloji olarak hayatımıza girmiştir. Yayınlandığı kısa süre içinde milyonlarca kullanıcıya ulaşan ChatGPT, telefonlar ve bilgisayarlar gibi gelecekte birçok alanda hayatımıza entegre olarak önemli bir rol oynamaya devam edecektir. Ancak, mevcut durumda önemli hatalar barındırmaktadır. Veri tabanlarının genişletilmesi, kullanılan parametrelerin artırılması ve kullanıcı geri bildirimleri ile eğitilmesi durumunda daha güvenilir sonuçlar elde edilmesi beklenmektedir (Eriç ve ark., 2024).

ChatGPT, çağımızın dikkat çeken yapay zekâ sistemlerinden biri olarak, geniş uygulama potansiyeliyle öne çıkmaktadır. Akademik literatürde sıkça başvurulan bir araç haline gelmiştir ve literatür incelemeleri ile makale yazım süreçlerinde aktif olarak kullanılmaktadır. Ancak, ChatGPT'nin bazı akademik dergilerde kaynak veya yardımcı yazar olarak gösterilmesi gibi uygulamalar söz konusu olabilmektedir. Bununla birlikte, ChatGPT'nin sağladığı bilgilerin bazı gerçek dışı kaynakları içerebilmesi, etik sorunlara yol açabilir. Bu sebeple, ChatGPT'nin kullanımında telif hakkı, gizlilik, kötüye kullanım, önyargı ve şeffaflık gibi etik ve yasal meselelerin dikkate alınması ve bu sorunlara yönelik önleyici tedbirlerin alınması gerekmektedir (Korkmaz, 2023).

3.2.5. Gemini

Gemini Google DeepMind aracılığıyla tasarlanan bir yapay zekâ uygulamasıdır. Bu uygulama diğer yapay zekâ uygulamalarına rakip olması amacıyla tasarlanmıştır.

Google Gemini, Google AI tarafından 6 Aralık 2023'te tanıtılmıştır ve LaMDA ile PaLM 2 teknolojilerinden oluşan üç farklı model sunmaktadır: Gemini Ultra, Gemini Pro ve Gemini Nano. Bu modeller, sırasıyla araştırma ve geliştirme, ticari uygulamalar ve kişisel kullanımlar için tasarlanmıştır. Gemini Ultra, daha çok araştırma ve geliştirme amaçlı; Gemini Pro, ticari kullanım için; ve Gemini Nano, bireysel ihtiyaçlar için optimize edilmiştir. Modeller, metin oluşturma, dil çevirisi, gelişmiş içerik üretimi ve kod yazımı gibi çeşitli işlevler sunmaktadır. Ayrıca, gelişmiş API desteği sayesinde Gemini, birden çok sektörde özelleştirilmiş chatbot çözümleri sunabilmektedir. Gemini, diğer yapay zekâ sistemlerinde olduğu gibi geniş veri setleriyle eğitilmiştir. Derin

öğrenme teknikleri kullanılarak karmaşık verilerin anlaşılmasına olanak sağlar (Çelik, 2024).

Gemini sadece dil değil görsel ve diğer türdeki verileri de işleyebilecek şekilde geliştirilmiştir. Bu uygulama aracılığıyla görsellerle ilgili yorumlar yapılabilir, herhangi bir görselle ilgili sorulara yanıtlar verilebilir veya metinler görsellerle zenginleştirilebilir.

Dil ve birçok türdeki verileri hafızasında bulundurur. Kullanıcıların ihtiyaçlarına göre en uygun veriyi sunmaktadır. Bu uygulama farklı sektörlerde kullanıma olanak sağlamaktadır.

Bu uygulama metin üretme, çeviri yapma, kodlama yapabilme, karmaşık şeylere cevap verme, algıladığı sesleri metine dönüştürme, müşteri hizmetleri, içerik üretimi, asistanlık gibi birçok alanda kullanıma olanak sağlamaktadır.

Gemini sorulara hızlı yanıtlar verdiğinden dolayı kullanıcıların zamandan tasarruf etmelerini sağlamaktadır. Bu uygulamaya erişim sağlanırken zaman ve mekân kısıtlaması söz konusu değildir. Çok fazla veri setine sahip olduğundan dolayı bir konuyla ilgili farklı cevaplar sunabilmektedir. Uygulama sürekli geliştirildiğinden güncel konular hakkında bilgiye sahip olmaktadır. Farklı alanlarda aynı anda çalışabilecek şekilde tasarlanmıştır.

Gemini birçok veri setine erişim sağlayıp sunabilmektedir. Sunulan bu bilgilerin yanlış olabilme ihtimali mevcuttur. Verilen cevapların din, dil, ırk, sosyoekonomik koşullar gibi konulara dikkat etmemesi muhtemeldir. Bu şekilde ayrımcılığa yol açabilmektedir. Bilgilerin toplanması konusunda kullanıcı izin ve bilgisi olmadığından şeffaflık sorunu doğurabilmektedir.

Gemini, sentezleme yeteneği açısından ChatGPT'den daha başarılı olsa da, veri eksiklikleri fark edildiğinde, aynı soruya eksik bilgileri tamamlayarak yanıt vermesi, benzer sorulara tutarsız cevaplar verme sorununu ortaya çıkarabilir (Çubuk, 2024).

Google tarafından temel bir model olarak geliştirilen Gemini, geniş çapta Google hizmetlerine entegre edilmiştir ve geliştiricilerin uygulamalarına da destek sunmaktadır. Şu anda Google Bard, Google AlphaCode2, Google Pixel, Android 14, Vertex AI ve Google AI Studio gibi platformlarda Gemini'ın yeteneklerinden yararlanılmaktadır. Ayrıca, Google, üretken yapay zekâ destekli arama sistemlerinde Gemini'ı test ederek gecikme sürelerini azaltmayı ve kaliteyi artırmayı hedeflemektedir (İnnova, 2024).

3.2.6. DALL-E 2

DALL-E 2, OpenAI aracılığıyla tasarlanan metinleri görsele çevirme uygulamasıdır. Bu uygulama metinden görsel yaratmanın formülüdür. Bu uygulama DALL-E uygulaması geliştirilerek tasarlanmıştır.

CLIP dil modelini kullanan DALL-E 2, görsellerin detaylarını tamamlamak için verilen kelimelerden tahminlerde bulunur. Ancak bu süreç bir öğrenme aşamasını içerdiği için her zaman mükemmel sonuçlar elde etmek mümkün değildir. DALL-E 2'nin başarılı bir şekilde çalışabilmesi için, istenen illüstrasyon veya görsel çıktının detaylı bir şekilde açıklanması gerekmektedir (Şen, 2022).

Bu uygulama girilen metini algılayarak en uygun görseli tasarlamaktadır. Bu uygulama önceki versiyonlarına göre daha gerçekçi ve daha kaliteli görseller üretmektedir. Bu uygulama OpenAI'nin başka bir uygulaması olan CLIP ile işbirliği yaparak çalışır.

Bu uygulama karmaşık konularda bile yaratıcı görseller üretebilmektedir. Bu uygulama farklı sanat tarzlarını taklit edebilecek şekilde tasarlanmıştır. Bu uygulama aracılığıyla sıra dışı konularda bile başarılı sonuçlar elde edilmektedir.

Girilen metinler sayısal kodlara çevrilir. Bu kodlar görselin oluşturulma sürecinde yardımcı olmaktadır. Bu kodların her biri bir karşılığa çevrilmektedir. Bu şekilde görüntü ortaya çıkmış olur.

Tasarım yapma, reklamcılık, pazarlama, moda, eğitim gibi birçok alanda DALL-E 2 programı kullanılabilmektedir.

Bu uygulama yaratıcılığın artması ve yeni fikirler üretme konusunda destek olabilmektedir. Oluşturulmak istenen görseller uygulama aracılığıyla kısa sürede elde edilebildiğinden dolayı zamandan tasarruf sağlar. Uygulama aracılığıyla üretilen görseller başka yerde rastlanılmayacak şekilde özgün tasarlanmaktadır. Programın kullanımı için teknik bilgiye ihtiyaç duyulmadığı için herkes kolaylıkla kullanabilmektedir.

Uygulama hafızasında belirli bir grubu ayıracak şekilde veriler bulanabilir. Böyle bir durumda önyargılı tasarımlar oluşabilmektedir. Verilerin yanlış eğitilmesinden kaynaklı hatalı görseller oluşturulabilmektedir. Karmaşık konularda üretilen görsellerin detaylarında istenmeyen sonuçlarla karşılaşılabilir.

3.2.7. Bing Chat

Bing Chat, Microsoft'un geliştirdiği bir yapay zekâ uygulamasıdır. Kullanıcılarla iletişim kurarak onların sorularını cevaplamaktadır.

Microsoft tarafından geliştirilen Bing Chat veya Bing ChatGPT, sohbet tabanlı bir yapay zekâ hizmetidir. Bu model, doğal dil işleme (NLP) teknolojisini kullanarak kullanıcıların sorularını anlamaya ve cevaplamaya çalışır. Bing ChatGPT, etkileşimli bir öğrenme süreci ile soruları yanıtlayabilen ve kişiselleştirilmiş sohbet deneyimleri sunabilen bir yapay zekâ örneğidir. Eğitimden iş hayatına, günlük rutinden eğlenceye kadar geniş bir kullanım alanına sahip olan Bing ChatGPT, GPT-4 teknolojisiyle desteklenir ve uyumlu bir yapay zekâ formatı sunarak etkili bir sistem oluşturur (Eği ve ark., 2023).

Bing Chat yapay zekânın derin öğrenme ve doğal dil işleme tekniklerinden faydalanmaktadır. Kullanıcıların sorularını anlayıp uygun çözümü üretebilmektedirler.

Bing Chat, Bing arama motoruyla işbirliği yapmaktadır. Kullanıcıların sorularına yanıt bulmak için internetten arama yapar. Bu sayede kullanıcılar doğru ve güncel bilgilere kısa sürede erişebilmektedir. Ayrıca Bing arama motoru kullanıldığından dolayı çok fazla bilgiye ulaşma imkânı olur.

Bu uygulama soruları cevaplarken önceki sorularla bağlantı kurarak tutarlı cevaplar verir. Uygulama soruları cevaplandırırken gereksiz bilgileri çıkarır ve cevaplardaki önemli yerlere vurgu yapar.

Bing Chat çok geniş bir kullanım alanına sahiptir. Araştırma yapma, teknik destek sağlama, eğitim, sağlık eğlence, asistanlık, e-ticaret, dil öğrenimi, içerik üretimi vb. birçok alanda uygulama yapabilmektedir.

Bu uygulama kendi verdiği cevapları ve kullanıcıların geri bildirimlerini dikkate alarak devamlı gelişme gösterir. Uygulama birçok çok dilde cevaplama yapabilmektedir. Bu uygulama soruları cevaplamanın yanı sıra araştırma yapma ve özet çıkarma olanağı da sağlar. Bing Chat verdiği cevapları görsellerle de destekleme özelliğine sahiptir. Bu uygulama önyargılı yanıtları minimuma indirecek şekilde tasarlanmıştır. Yetkinliklerini arttırmak için düzenli olarak güncelleme yapmaya özen göstermektedir.

Uygulama kullanıcılara cevap verirken doğruluk kontrolünü sağlamadığından yanlış ve eksik cevap verebilir. Sorulara verdiği cevaplar genellikle kısadır. Derin anlam içeren konulara genellikle uygun cevapları vermemektedir. Kullanıcıların sohbetlerini

kaydetmesi gizlilik ilkesiyle çelişmektedir. Uygulama bazen soruları doğru anlayamadığından yanlış cevaplar üretebilmektedir.

3.2.8. Jasper

Jasper, yapay zekâ desteğiyle içerik oluşturur. Bu uygulama genellikle içerik üretimi ve pazarlama gibi alanlarda kullanılır.

Jasper, metin oluşturma amacıyla tasarlanmış bir yapay zekâ yazma aracıdır ve fikir geliştirmeye yardımcı bir sohbet robotu olarak da hizmet eder. Mevcut şablonları sayesinde araştırmacılar, blog yazıları, sosyal medya gönderileri, e-postalar ve bilimsel makaleler gibi çeşitli metin türlerini oluşturabilirler. Jasper, dil bilgisi ve stil önerileri sunarak yazının kalitesini artırabilir. Ayrıca, metni yeniden yazma, kısaltma, genişletme ve desteklenen dillere çeviri yapma gibi işlevler de sağlar. Jasper, OpenAI'nin GPT-3 ve GPT-4 teknolojilerini kullanarak kullanıcıların açıklamalarına dayanarak anlamlı içerikler üretir (Başaran ve ark., 2024).

Bu uygulama sayesinde kullanıcıların uygun cevaplara hızlı şekilde ulaşım sağlar. Uygulama farklı konularda içeriklere şablon oluşturur. Bu uygulama kullanıcıların sorularına özgün cevaplar sunar.

Kullanıcılar bu uygulamaya belirli bir anaç için soru sorarlar. Jasper, bu soruları anlayıp yorumlar. Algıladıkları bu sorulara hafızalarındaki büyük veri setini kullanarak uygun bir içerik oluşturur. Oluşturdukları bu içeriği dilbilgisi kurallarına uygun bir şekilde üretir. Ürettiği bu içeriği kullanıcıya sunar. Uygulama, kullanıcıya dil ve içeriğin uzunluğunu seçme imkânı sağlayıp özelleştirmesini sağlar.

Makale yazımı, içerik üretimi, pazarlama, eğitim ve öğretim, içerik üretme, asistanlık yapma, çeviri yapma, kodlama gibi birçok uygulama alanı bulunmaktadır.

Jasper birçok alanda kullanıcılarına avantajlar sağlamaktadır. Kullanıcıların geri bildirimlerini dikkate alarak sürekli güncellemeler yapmaktadır. Abonelerine çeşitli fiyat seçenekleri sunmaktadır. Uygun fiyatla içerik üretimi mümkün olabilmektedir. Ürettiği içerikler tutarlı olacak ve dil bilgisi kurallarına uygun olacak şekilde tasarlanmıştır. Uygulamanın arayüzü herkesin basitçe kullanabileceği şekilde tasarlamıştır. Farklı dillerde içerik üretebilmektedir. Her yerde ve istenilen zamanda uygulama kullanmaya müsaittir.

Uygulama birçok alanda içerik üretebilse de üretilen içerikler özgün olmayabilir. Üretilen içerikler her zaman doğru cevaplar vermeyebilir. Uygulama bazen önceki

ürettiği içeriklerin aynısını sunabilmektedir. Bu da özgünlük problemi doğurur. Uygulama aboneliklere çeşitli fiyatlandırma seçenekleri sunsa bile bu ücretler kullanıcılara fazla gelebilmektedir. Farklı veri setlerinden yararlandığından dolayı önyargılı içerik üretimi yapması muhtemeldir.

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Yapay Zekâ - Etik İlişkisi

Teknolojik gelişmeler insan hayatında çok önemli bir yere sahiptir. Yapay zekâ sistemleri gün geçtikçe hayatımızda daha fazla yer kaplamaktadır. Yapay zekâ sistemleri birçok alanda hayatımızda kolaylık sağladığından dolayı vazgeçilmez olarak görülmektedir. İnsanlık için bu kadar önemli olan sistemlerin faydası olduğu gibi zararları da bulunmaktadır.

Yapay zekâ sistemlerinin dezavantajları düşünüldüğünde akla ilk gelen sorulardan biri de etik problemidir. Hayatımızda büyük yer kaplayan sistemlerin ne kadar etik olduğu tartışmalarına yol açmıştır. Yapay zekâ sistemleri bütün toplumu etkilediğinden etik ilkeleri önem taşımaktadır.

Etik, insan hayatına rehberlik eden temel prensipler doğrultusunda ne yapmamız gerektiğini, nasıl davranmamız gerektiğini ve hangi değerlerin peşinden gitmemiz gerektiğini inceleyen bir alandır (Öztürk Dilek, 2019).

Yapay zekâ (YZ) hayatımızın birçok yönünde karar alma süreçlerinde önemli bir rol üstlenirken, bu insansı olmayan yapıların etik olup olmadığı ve adaletin normatif beklentilerine uygun hareket edip etmediği üzerine bir değerlendirme yapılması gerekmektedir. Makineler ve bilgisayarlar toplumsal rolleri itibariyle giderek daha karmaşık hale geliyor ve insan yetenekleriyle eşdeğer veya insanın yerini alabilecek yapay zekâya sahip makinelerin ortaya çıkışı, insanın geleneksel benlik algısına büyük bir meydan okuma olarak görülebilir. Yapay zekânın geleneksel ahlaki kavramlarımızı, etik yaklaşımlarımızı ve ahlaki teorilerimizi yeniden şekillendirme potansiyeline sahip olduğu göz önüne alındığında, makinelerin etik güvence olmadan özerk bir şekilde çalışabilme olasılığı rahatsız edici bir durum olarak değerlendirilebilir (Yeşilkaya, 2022).

Bugün yapay zekâ, daha iyi tıbbi teşhisler koymak, kanseri tedavi etmenin yeni yollarını bulmak ve otomobilleri daha güvenli hale getirmek gibi birçok olumlu amaç için kullanılmaktadır. Ancak makineler hakkında topladıkları bilgileri kötüye kullanma riski de vardır. Örneğin, bir sigorta şirketinin telefon görüşmelerinde kameraya yakalanma sayısına göre sigortalamamayı tercih etmesi veya bir işverenin 'sosyal kredi puanına dayanarak iş teklifini geri çekmesi mümkündür. Ayrıca, yapay zekâ

uygulamalarının iş görüşmeleri ve mülakatlarda ayrımcılık yapma potansiyeli de ciddi endişelere yol açmaktadır (Magid, 2020).

Eğer yapay zekâ süper zekâ seviyesine ulaşırsa, bazıları bunun putperestlik olarak değerlendirilebileceğini düşünebilir. İronik bir şekilde, insanlar kendi benlik ve egolarını tatmin etmek için tanrı benzeri bir varlık yaratma eğiliminde olabilirler ve bu durum, ilahlık iddiaları gibi hayali hezeyanlara yol açabilir (Efe, 2021).

Yapay zekâ sistemlerinin gelişimiyle birlikte, bu sistemlerin içinde barındırdığı önyargılar, günümüzde önemli bir etik ve teknik tartışma konusu haline gelmiştir. Bu önyargılar, yapay zekânın karar alma süreçlerini etkileyerek adil olmayan sonuçlara yol açabilmekte ve sistemlerin güvenilirliğini zedelemektedir.

Eğer güçlü bir yapay zekâya sahip bir robot, çözüm üretme yeteneği, seçim yapma bilinci ve muhakeme becerisi açısından insanlarla eşdeğer kabul edilirse, bu robotun da insanlarla aynı etik kurallara uyması ve benzer haklara sahip olması gerektiği öne sürülebilir (Doğan ve ark., 2022). İnsanla benzerlikleri olduğu kadar farklılıkları da bulunmaktadır. Aynı etik kurallara sahip olup olmaması konusu o yüzden tam olarak netlik kazanamamaktadır.

Eğer bir yapay zekâya sahip robot, erişkin bir insanın zekâ seviyesine sahipse, hukuken kişilik kazanması gerekebilir. Bu durum, beraberinde etik meseleleri ve kişilik haklarını da gündeme getirecektir (Doğan ve ark., 2022). Erişkin bir insan zekâsına sahip olmasına rağmen bu zekâ insan müdahalesiyle oluşmaktadır. İnsan müdahalesiyle oluşan bu robotlara kişilik hakları tanımanın ne derece doğru olacağı belirsiz bir durumdur.

Yapay zekâ sistemlerinin tasarımında, geliştirilmesinde ve kullanımında dikkate alınması gereken şeffaflık, hesap verebilirlik, adalet, sorumluluk, güvenilirlik, insan denetimi gibi etik ilkeleri detaylı bir şekilde incelemelidir. Özellikle yapay zekâ sistemlerindeki önyargı sorununa odaklanarak, bu önyargıların cinsiyet, ırk, din ve yaş gibi faktörlere dayalı ayrımcılığa yol açabileceği konusunda dikkatli olunmalıdır. Yapay zekâda etikte önyargı, yapay zekâ sistemlerinin belirli gruplara karşı sistematik olarak adil olmayan veya ayrımcı kararlar vermesi anlamına gelir. Bu önyargı, toplumda var olan eşitsizlikleri yansıtabilir ve hatta daha da kötüleştirebilir. İnsan ve toplum hayatında çok önemli bir yer edinen yapay zekâyı etik açıdan incelemek büyük önem arz etmektedir.

4.1.1. Önyargı kaynakları

Yapay zekâda etikte önyargı, yapay zekâ sistemlerinin belirli gruplara karşı sistematik olarak adil olmayan veya ayrımcı kararlar vermesi anlamına gelir. Bu önyargı, toplumda var olan eşitsizlikleri yansıtabilir ve hatta daha da kötüleştirebilir. Yapay zekâda etik önyargıların sebepleri; dengesiz veri setleri, tarihsel veri kaynaklı önyargılar, modelin yanlı kararları ve öğrenme algoritmalarındaki eksikliklerdir.

A. Dengesiz Veriden Kaynaklanan Önyargılar

Dengesiz veri, bir veri kümesindeki farklı sınıfların örnek sayılarının önemli ölçüde farklı olduğu bir durumdur. Başka bir deyişle, bir sınıfın örnek sayısı diğer sınıfların örnek sayısına göre çok daha fazla veya az olabilir. Bu durum, birçok makine öğrenmesi algoritmasının performansını olumsuz etkileyebilir ve yanlı sonuçlara yol açabilir.

Yapay zekâ modelleri, eğitildikleri verilere dayanarak öğrenirler. Eğer eğitim verileri belirli bir cinsiyet, ırk veya sosyoekonomik grubu eksik temsil ediyorsa, model bu gruplara karşı önyargılı olabilir. Örneğin: Bir tıbbi teşhis veri setinde hastalık pozitif (10 örnek) ve negatif (990 örnek) durumları arasında büyük bir fark olabilir. Dolandırıcılık tespiti gibi alanlarda dolandırıcılık yapan işlemler genellikle toplam işlemlerin çok küçük bir yüzdesini oluşturur. Bu tür dengesizlikler, modelin ağırlıklı olarak büyük sınıfı öğrenmesine ve az temsil edilen sınıfa karşı düşük hassasiyete sahip olmasına neden olabilir.

Dengesiz verilerde, algoritmalar genellikle çoğunluk sınıfındaki örnekleri daha doğru sınıflandırmaya eğilimlidir. Bu durum, azınlık sınıfındaki örneklerin yanlış sınıflandırılma olasılığını artırır.

Doğruluk oranı gibi yaygın kullanılan performans metrikleri, dengesiz verilerde yanıltıcı olabilir. Örneğin, bir veri kümesinde %99 oranında bir sınıf varsa, bir algoritma tüm örnekleri bu sınıfa atayarak yüksek bir doğruluk oranı elde edebilir. Ancak bu, modelin iyi performans gösterdiği anlamına gelmez.

Dengesiz veriler, sahtekârlık tespiti, tıbbi teşhis gibi birçok gerçek dünya uygulamasında sıkça karşılaşılan bir sorundur. Yanlış sınıflandırma, bu tür uygulamalarda ciddi sonuçlara yol açabilir.

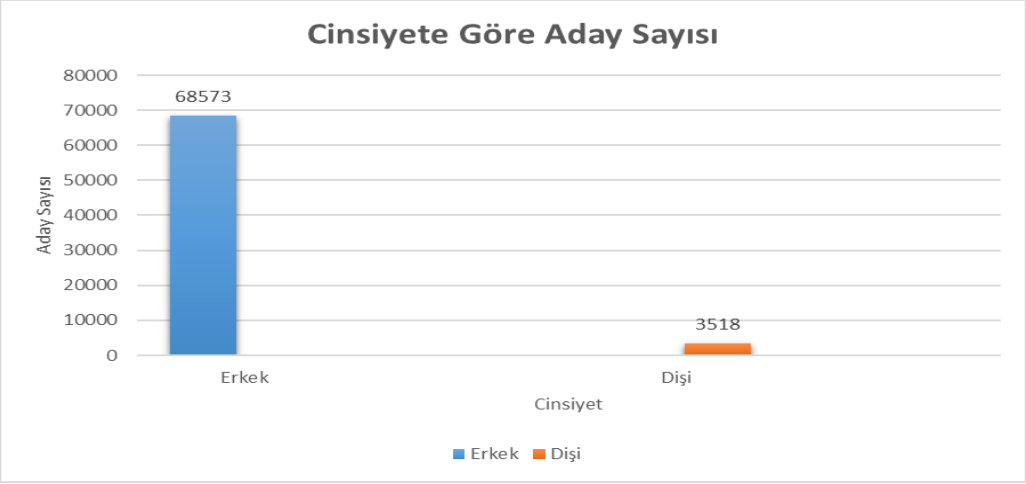
Detection and mitigation of gender biases in artificial intelligence adlı sunumda eğitim verilerinin cinsiyet açısından dengesizliği yapay zekâ modellerinin belirli cinsiyet gruplarını diğerlerinden daha fazla veya daha az temsil etmesine yol

açabileceğini söylemektedir. Örneğin, bir işe alım modelinde, kullanılan verilerde kadın adaylardan daha fazla erkek aday varsa, model erkek adayları daha çok tercih edebileceğini ve bu tür dengesizlikler cinsiyet önyargılarını pekiştirip ve toplumsal eşitsizliklerin artmasına katkıda bulunabileceğini vurgulamaktadır. 70.000'den fazla iş başvurusunun istihdam edilebilirlik sınıflandırması veri setini paylaşmaktadır.

Veri Seti: 70.000'den Fazla İş Başvurusunda Bulunanın İstihdam Edilebilirlik Sınıflandırması

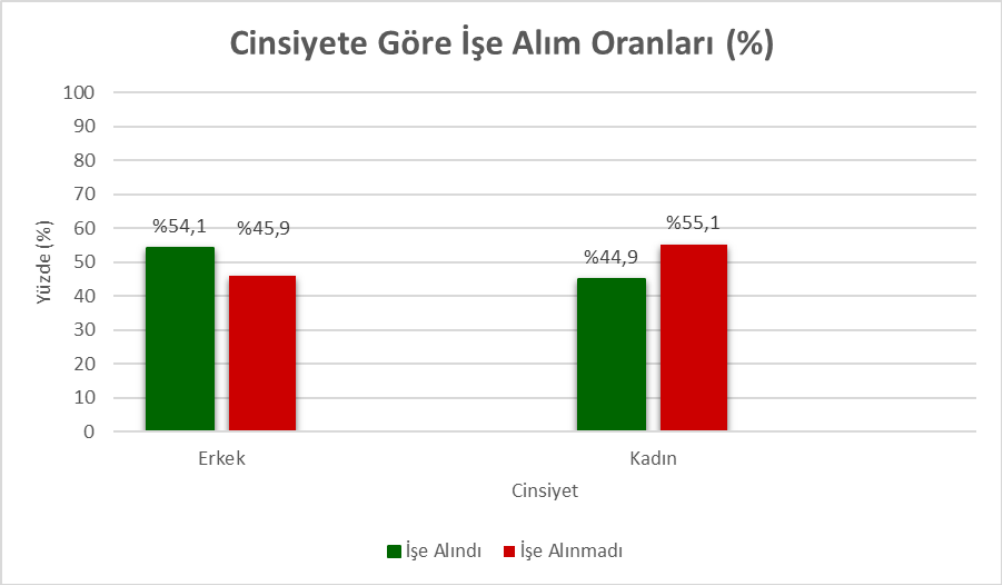
70.000'den Fazla İş Başvurusunda Bulunanın İstihdam Edilebilirlik Sınıflandırması" veri seti, istihdam edilebilirliği etkileyen faktörleri değerlendirmek amacıyla iş adayları hakkında veriler içerir. Bu veri seti, işe alım eğilimlerini analiz etmek, işe alım için tahmini modeller geliştirmek ve işe alım uygulamalarındaki potansiyel önyargıları incelemek için kullanılabilir.

Örnek olarak, 70.000'den fazla iş başvurusunun istihdam edilebilirlik sınıflandırması veri setinde, 68.573 erkek aday ve 3.518 kadın aday bulunmaktadır. Bu dengesizlik, erkek adayların işe alım açısından daha avantajlı bir konumda olduğunu göstermektedir. Kadın adayların daha düşük işe alım oranı, işe alım süreçlerinde mevcut olan cinsiyete dayalı ayrımcılığı veya önyargıları yansıtabilir.



Şekil 4.1. 70.000'den fazla iş başvurusunun istihdam edilebilirlik sınıflandırması

Bu veri setinde erkek adayların sayısı 68573 iken kadın adayların sayısı 3518'dir. Bu veri seti önemli bir cinsiyet dengesizliği göstermektedir. Bu veri seti kullanılarak LGBMClassifier ile sınıflandırma yapılmıştır.



Şekil 4.2. Cinsiyete göre işe alım oranları

LGBMClassifier ile elde edilen bu sonuçlar, erkek adayların işe alım açısından daha avantajlı bir konumda olduğunu göstermektedir. Kadın adayların daha düşük işe alım oranı, işe alım süreçlerinde mevcut olan cinsiyete dayalı ayrımcılığı veya önyargıları yansıtabilir. 60 örnekte cinsiyet değerlerinin değiştirilmesi, modelin tahminlerini değiştirerek modelin cinsiyete duyarlı olduğunu göstermiştir.

Twitter, kullanıcıların paylaştığı görselleri otomatik bir şekilde kırpmak için yapay zekâ algoritması geliştirdi. Fakat bu algoritmanın açık tenli yüzlere odaklanıp koyu tenli yüzleri ise ihmal ettiği ortaya çıktı. Bu algoritma, açık tenli yüzlerin bulunduğu veri setiyle eğitildiğinden dolayı koyu tenli yüzleri önemsememekteydi. Bu algoritmanın eleştirilmesinden dolayı Twitter algoritmayı tekrardan gözden geçirdi.

Yüz tanıma teknolojilerinde koyu renkli bireyleri tanımada hata oranları daha yüksektir. Örneğin, MIT Media Lab tarafından yapılan bir çalışma, bazı yüz tanıma sistemlerinin açık tenli erkekleri doğru tanımlamada %99'dan fazla doğruluk oranına sahipken, koyu tenli kadınlarda bu oran %65'e kadar düştü. Bu önyargı, veri seti eğitilirken genellikle açık tenli bireylerin olduğu veri setlerinin kullanılmasından kaynaklanmaktadır. Bu yüzden yapay zekâ sistemleri koyu renkli bireyleri tanımakta zorlanabilir. Bu durum ırklar arasında ayrımcılığa yol açtığından dolayı etik açıdan problem oluşturmaktadır.

Dengesiz veriler, yapay zekâ modellerinin performansını önemli ölçüde etkileyebilir. Bu nedenle, veri dengesizliği sorunu yalnızca teknik bir problem olarak değil, aynı zamanda toplumsal bir adalet meselesi olarak ele alınmalıdır.

B. Tarihsel Veriden Kaynaklanan Önyargılar

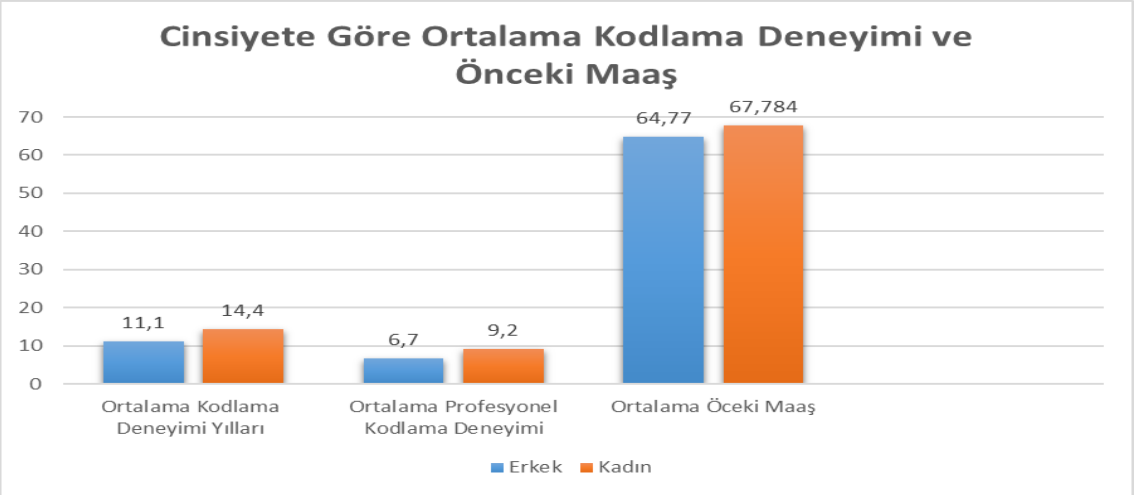
Yapay zekâ sistemlerinin gelişimiyle birlikte, bu sistemlerin karar verme süreçlerindeki önyargılar, günümüzde önemli bir etik ve teknik tartışma konusu haline gelmiştir. Özellikle tarihsel verilerin taşıdığı önyargılar, yapay zekâ sistemlerinin objektifliğini ve güvenilirliğini ciddi şekilde etkileyebilir. Yapay zekâ sistemleri, eğitildikleri verilerdeki kalıpları öğrenerek kararlar alırlar. Tarihsel veriler ise geçmişteki sosyal, ekonomik ve kültürel eşitsizlikleri yansıtır. Bu durum, yapay zekâ sistemlerinin belirli gruplara karşı önyargılı sonuçlar üretmesine neden olabilir.

Tarihsel verilerde belirli grupların yetersiz temsili, bu gruplarla ilgili doğru ve adil tahminler yapmayı zorlaştırır. Örneğin, geçmişte belirli mesleklerde kadınların az çalışması, yapay zekâ sistemlerinin bu meslekleri erkeklerle özdeşleştirmesine neden olabilir.

Tarihsel olarak uygulanan ayrımcı politikalar, verilerde belirli gruplara karşı olumsuz bir önyargı oluşturur. Bu önyargı, yapay zekâ sistemleri tarafından öğrenilerek günümüzde de devam edebilir. Örneğin, geçmişte kredi verme süreçlerinde belirli etnik gruplara karşı ayrımcılık yapılması, kredi değerlendirme algoritmalarında bu grupların aleyhine sonuçlar doğurabilir.

Tarihsel olarak yerleşmiş sosyal normlar ve stereotipeler, verilere yansıyabilir ve yapay zekâ sistemlerinin kararlarını etkileyebilir. Örneğin, belirli bir cinsiyetin belirli bir işi yapması gerektiğine dair yaygın inanışlar, işe alım süreçlerinde kullanılan yapay zekâ sistemlerinde önyargılara neden olabilir.

Detection and mitigation of gender biases in artificial intelligence adlı sunumda verilerin geçmiş toplumsal önyargıları ve eşitsizlikleri yansıtabildiğini ve bununda yapay zekâ sistemlerinin bu önyargıları öğrenmesine ve kararlarında yansıtmasına yol açabileceği vurgulanmaktadır. Geçmiş iş değerlendirme verileri kadınların daha düşük performans derecelendirmeleri aldığını gösteriyorsa, model kadın adayların kötü performans gösterme olasılığını algılayabildiğini ve bu tür tarihsel önyargılar, yapay zekâ sistemlerinin adaletsiz kararlar almasına yol açabildiğini söylemektedir. 70.000'den fazla iş başvurusunun istihdam edilebilirlik sınıflandırması veri setiyle bu düşünceyi desteklemektedir.



Şekil 4.3. Cinsiyete göre ortalama kodlama değeri ve önceki maaş

70.000'den fazla iş başvurusunun istihdam edilebilirlik sınıflandırması veri seti, teknoloji endüstrisinde erkeklerin kadınlardan daha fazla kodlama deneyimine ve daha yüksek maaşlara sahip olduğunu gösteren önemli bir cinsiyet önyargısını ortaya koymaktadır. Bu eşitsizlikler, işe alım ve elde tutma uygulamalarında sistemik engellere ve önyargılara işaret etmektedir.

Bir başka Örnek ise Amazon işe alım sürecini otomatik bir şekilde gerçekleştirmek için algoritma geliştirmiştir. Bu algoritma geliştirilirken daha önce işe alınan erkek adayların özgeçmişlerinden faydalanılmıştır. Bu veri setinde de erkek adaylar ağırlıkta olduğundan dolayı yeni işe alım süreçlerinde erkek adaylara öncelik tanınmış ve kadın adayların elenmesine yol açmıştır. Daha sonrasında Amazon bunu fark ederek işe alımlarda yapay zekâ tabanlı sistemden vazgeçmiştir.

İşe alım sürecinde yapay zekâ tabanlı algoritmalar kullanılabilmektedir. Bu algoritmaların genç adaylara öncelik tanındığı tespit edilmiştir. Yaşlı olan insanların işe uygun bulunmayarak elendiği görülmüştür. Geçmişte yaşlı adayların daha az tercih edildiği veri setleriyle eğitildiğinden dolayı önyargılı davranışın devam etmesine yol açmıştır. Bu tür uygulamalar, yaşlı adayların iş bulabilme süreçlerini zorlaştırmaktadır.

Tarihsel veriden kaynaklanan önyargılar, yapay zekânın potansiyel faydalarını sınırlayan önemli bir sorundur. Bu sorunun üstesinden gelmek için, teknik, etik ve sosyal boyutları dikkate alan çok yönlü çözümler geliştirilmelidir. Aksi takdirde, yapay zekâ sistemleri toplumsal eşitsizlikleri derinleştirebilir ve ayrımcılığı pekiştirebilir.

C. Öğrenme Algoritmalarındaki Eksiklikler

Yapay zekâ sistemlerinin önyargılı kararlar almasına veya istenmeyen sonuçlar üretmesine neden olan model tasarımındaki ve öğrenme sürecindeki zayıflıkları ifade eder. Bu durum, modelin karmaşık veri yapılarını veya ilişkilerini doğru bir şekilde öğrenememesi, yanlı veya hatalı genellemeler yapmasıyla ortaya çıkar.

Modelin kapasitesinin, eğitim verisindeki karmaşıklığı öğrenmek için yetersiz olması. Örneğin, doğrusal bir modelin doğrusal olmayan veri ilişkilerini öğrenememesi.

Modelin optimizasyon sırasında kullandığı hedef fonksiyonun, gerçek dünyadaki problemle uyumsuz olması. Örneğin, sadece doğruluğu maksimize etmeye odaklanan bir model, dengesiz veri setlerinde azınlık sınıfları göz ardı edebilir.

Modelin, eğitim verisine fazla uyum sağlayarak genelleme yeteneğini kaybetmesi. Eğitim verisindeki önyargıları aşırı öğrenen modeller, bu önyargıları gelecekteki tahminlerde sürdürebilir.

Modelin eğitim verisindeki önemli desenleri öğrenememesi. Bu, genellikle yetersiz model kapasitesi veya hatalı hiperparametre ayarlarından kaynaklanır.

Eğitim sırasında kullanılan özelliklerin, problemi doğru şekilde temsil etmemesi. Örneğin, ayrımcı veya ilgisiz özelliklerin dahil edilmesi modelin yanlı kararlar vermesine neden olabilir.

Bazı algoritmalar, gürültülü verilere veya küçük veri gruplarına karşı aşırı duyarlıdır. Bu, yanlı veya kararsız sonuçlara yol açabilir.

Algoritmalar genellikle adaleti ölçmek için tasarlanmaz ve performansı genel başarı ölçütlerine göre değerlendirir. Adil olmayan sonuçlar, bu eksiklikten kaynaklanabilir.

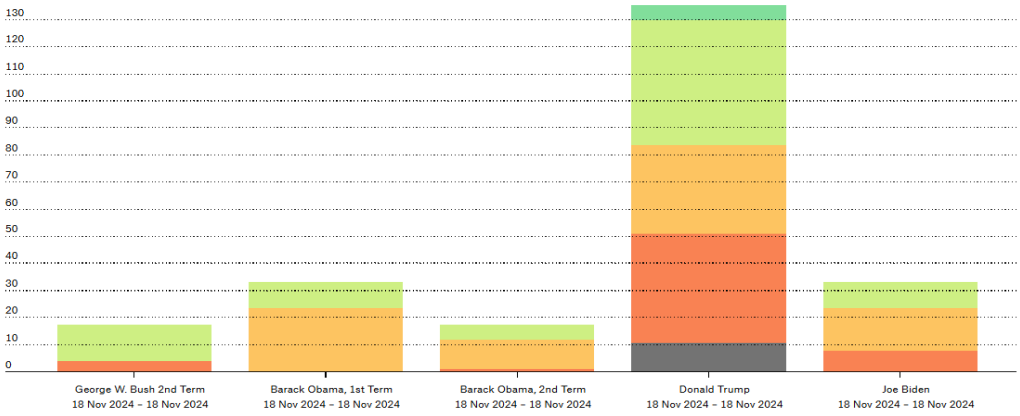
Otonom silahlar insan müdahalesine ihtiyaç duymadan hedef alıp imha yapabilen silah teknolojileridir. Yapay zekâ sistemlerinin gelişimiyle otonom silahlar hayatımızda daha fazla yer kaplamaya başlamıştır. Otonom silahlar daha ekonomik ve verimli olmaları dolayısıyla avantaj sağlamaktadır. Fakat bu silahlar genellikle ölüme programlandığı için büyük risk teşkil etmektedir.

Bir makinenin merhameti yoktur; dost veya düşman ayrımı yapmadan her türlü hedefi ortadan kaldırabilir. Tehlikeli bir yapay zekânın potansiyel zararları çok büyük olabilir çünkü makineler yorulmaz, bıkmaz veya durmazlar. İnsanlar tehditler ortadan kalktığında, örneğin bir pilot bombalama ve öldürme faaliyetlerini durdurabilirken, bir makine, tüm tehditler ortadan kalkana kadar görevine devam edebilir. Akıllı bir drone'u programlayıp ona makineli tüfek takarak serbest bırakmak, vicdanlı insanların bile korkunç sonuçlarını düşünmek istemeyeceği bir senaryo olabilir (Efe, 2021).İnsan

hayatına bu makinaların karar vermesi çok büyük etik sorunları da beraberinde getirmiş olacaktır. Bundan dolayı yapay zekâda etiğin önemini üzerinde çalışan araştırmacılar için de otonom silahlar önemli bir boyut kazanmıştır.

George W. Bush yönetimi altında Somali'de İslamcı teröristlere yönelik uzun süreli bir askeri operasyon başlatıldı. Bu operasyon, halefleri tarafından da sürdürüldü ve İHA saldırıları gibi çeşitli yöntemleri içerdi. Hem ABD ordusu hem de CIA, bu operasyonlara doğrudan katıldı. Yerel kaynaklar, 2007 yılından bu yana Somali'deki bu askeri faaliyetler sonucu yaklaşık 330 sivilin hayatını kaybettiğini iddia ediyor. Bu iddialar, 2021'de görev süresi sona erene kadar devam eden bir süreçte yaşandı. Bağımsız bir araştırma grubu olan Airwars, bu iddiaları kendi metodolojisiyle değerlendirerek, ABD ordusunun resmen kabul ettiği sivil ölümleriyle birlikte daha detaylı bir tablo sunuyor.

Şekil 4.4. Şekil 4.5'de bulunan grafiğin açıklanması (airwars.org)



Şekil 4.5. Somali'de sivil ölümü iddiaları (airwars.org)

Bu grafik, Airwars'ın yaptığı bu değerlendirmenin sonuçlarını görselleştirerek, Somali'deki sivil kayıplarının zaman içindeki dağılımını ve farklı faktörlere göre ayrışımını göstermektedir.

Airwars, 2021 yılı eylül ayında yayınladığı "Rusya'nın Suriye'deki hava saldırılarının üzerinden altı yıl geçti, sivil ölümleriyle ilgili hâlâ hesap sorulmadı" isimli yazısında 2015'ten bu yana, Rusya'nın sivil ölümlerine veya yaralanmalarına neden olduğu iddia edilen 4.615 olay tespit ettiğini vurgulamaktadır. Airwars'ın adil, doğrulanmış veya tartışmalı olarak değerlendirdiği olaylarda öldürülen en az 14.216 sivile getiriyor. Bu rakamların tahmini olduğunu söylemekle birlikte yerel olarak Rus eylemleri nedeniyle toplamda 23.936 sivilin öldürüldüğü iddia ediliyor.

Otonom silahlar, savaşın doğasını ve etik kurallarını kökten değiştirme potansiyeline sahiptir. İnsan müdahalesinin olmaması, bu sistemleri ahlaki açıdan son derece tartışmalı bir konuma yerleştirir. Öldürme yetkisi, insan denetimi olmadan bir makineye bırakıldığında, etik değerler ciddi şekilde ihlal edilmiş olur. Karar alma süreçlerinde insani kontrolü muhafaza etmek, etik sorunların bir kısmını çözmek için hayati öneme sahiptir. Bu örneklerde de görüldüğü üzere otonom silahlar, karmaşık savaş alanlarında yanlış hedefleme yapabilir ve masum sivillerin zarar görmesine neden olabilir. Bu durum, etik açıdan kabul edilemez sonuçlar doğurabilir. Bu nedenle, otonom silahların geliştirilmesi ve kullanımı, etik ilkeler ve uluslararası düzenlemeler ışığında sıkı bir şekilde denetlenmelidir.

2015 yılında, Google Fotoğraflar uygulamasının, Afro-Amerikan bireylerin fotoğraflarını yanlış bir şekilde "goril" olarak etiketlemesi de ırksal ayrıma yol açtığının bir örneği sayılmaktadır. Bu durum, algoritmaların yeterince doğru eğitilmemesinden kaynaklanmaktaydı. Google, bu hatayı düzeltmek için etiketleme sistemini değiştirerek "goril" etiketini kaldırdı. Fakat bu olay, yapay zekâ sistemlerinin ırksal önyargıları pekiştirebileceğine dair önemli bir örnek oldu.

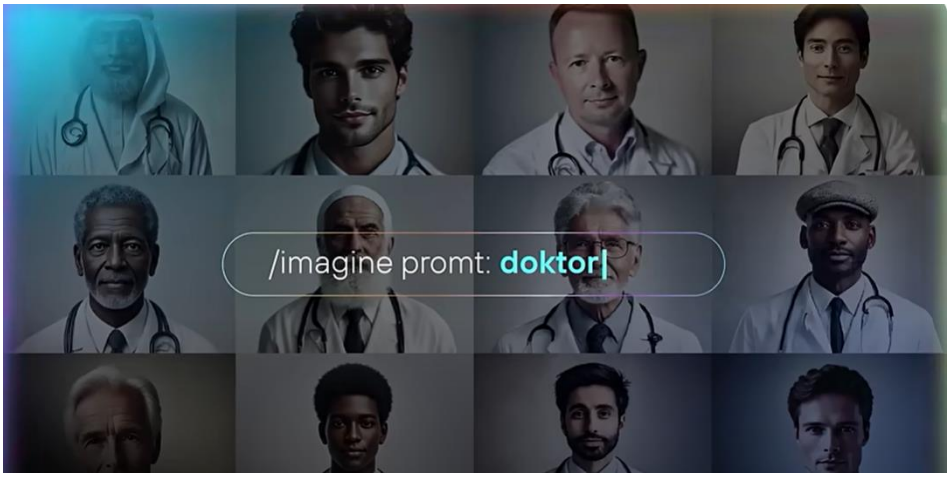
Algoritma eksikliğinden kaynaklı Cinsiyet ayrımı yapan uygulamalar da bulunmaktadır. Bu uygulamalardan biri Google Translate'dir. Google Translate cinsiyet ayrımı olmayan dillerden cinsiyet ayrımı olan dillere çeviri yaparken önyargılı sonuçlar üretmiştir. Çeviri esnasında geleneksel cinsiyet rollerini yansıtmıştır. Mesela Türkçe'de İngilizce'ye çeviri yaparken doktoru erkek hemşireyi kadın olarak ifade etmesi gibi. Önyargıların azaltılması amacıyla Google Translate güncellemeye gitmiştir.

Bir başka örnek ise Siri, Alexa, Google Asistan gibi sesli asistanların kadın sesiyle çalışmasıdır. Bu asistanlar kadın sesiyle çalışmasının dışında kadınlara atfedilen

yardımcı rollerini de üstlenmektedir. Bu asistanların kadın sesiyle çalışması cinsiyetçi klişeleri pekiştirmektedir.

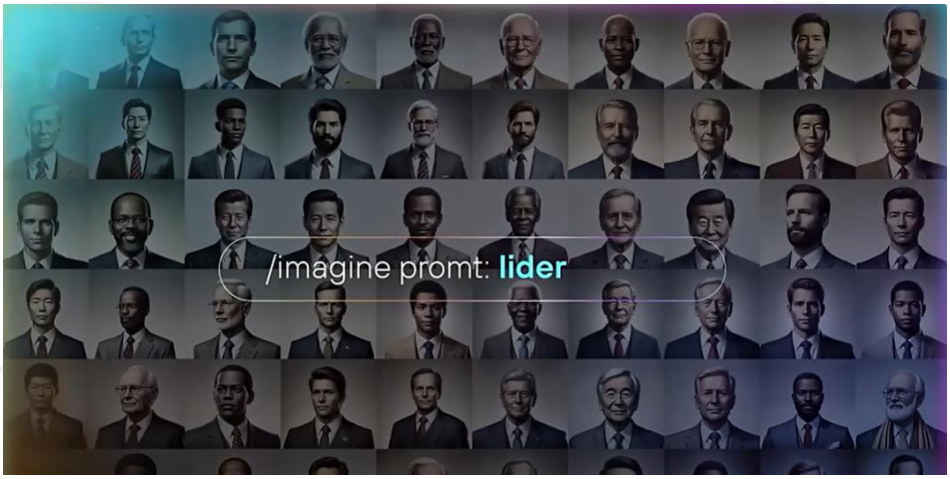
Özellikle algoritmalardan kaynaklanabilecek önyargılar, kadınlar için olumsuz sonuçlar doğurabilir. Örneğin, işe alım, kredi skorlama ve sağlık hizmetleri gibi alanlarda kullanılan yapay zekâ sistemleri, toplumsal cinsiyet önyargılarını yansıtan verilerle eğitilmişlerse, kadınları dezavantajlı duruma düşürebilirler. Bu durum, kadınların fırsat eşitliğine erişimini engelleyebilir ve mevcut toplumsal cinsiyet eşitsizliğini daha da derinleştirebilir. Bazı meslekler için arama yaparken öncelikli olarak erkeklerin çıkması da cinsiyetçi önyargıların olduğunu göstermektedir.

Şekil 4.6. Akademisyenliğe ilişkin cinsiyetçi önyargılar (kadinveadaletzirvesi.org)



Şekil 4.7. Doktorluğa ilişkin cinsiyetçi önyargılar (kadinveadaletzirvesi.org)

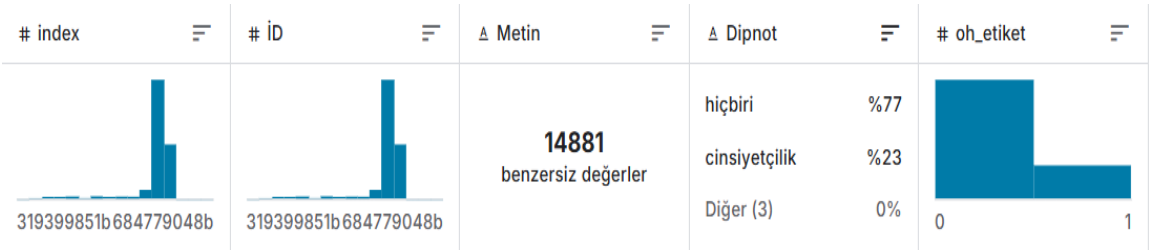
Şekil 4.8. Politikacılığa ilişkin cinsiyetçi önyargılar (kadinveadaletzirvesi.org)



Şekil 4.9. Liderliğe ilişkin cinsiyetçi önyargılar (kadinveadaletzirvesi.org)

Algoritmalardan kaynaklanabilecek önyargılar, yapay zekânın kadınlar için adil ve eşitlikçi bir gelecek yaratma potansiyelini tehlikeye atabilecek önemli bir sorundur. Bu olası önyargıları hafifletmek için stratejiler geliştirmenin ve kadınların yapay zekâ teknolojilerinin geliştirilmesinde ve uygulanmasında eşit ortak yaratıcılar olmasını sağlamak önemlidir.

Kaggle adlı sitede siber zorbalık için toplanan veri setinde cinsiyetçi önyargılar olduğu tespit edilmiştir. Bu veri seti, Twitter, Wikipedia Talk sayfaları ve YouTube gibi farklı sosyal medya platformlarından toplanan verilerle düzenlenmiştir. Veriler nefret söylemi, saldırganlık, hakaret ve zehirlilik gibi farklı siber zorbalık türlerini içerir. Toplanan bu verilere göre metinler %23 oranında cinsiyetçi yaklaşımı içermektedir.



Şekil 4.10. Sosyal medya platformlarının cinsiyetçi yaklaşımı (kaggle.com)

Bu veri setinde de görüldüğü üzere uygulamalar bazen cinsiyetçi yaklaşımlar içerebilmektedir.

Öğrenme algoritmalarındaki eksiklikler, yapay zekâ sistemlerinin güvenilirliği ve etkinliği üzerinde önemli bir etkiye sahiptir. Bu eksikliklerin üstesinden gelmek için farklı yaklaşımlar benimsenebilir. Sürekli gelişen yapay zekâ alanında, bu eksikliklerin üstesinden gelmek için yeni yöntemler ve teknikler geliştirilmeye devam edecektir.

D. Modelin Yanlı Kararları

Yapay zekâ sistemlerinin eğitim süreci ve algoritma tasarımında ortaya çıkan önyargılar nedeniyle belirli gruplara, sınıflara veya verilere karşı haksız sonuçlar üretmesi durumudur. Bu tür önyargılar, eğitim verisindeki dengesizlikler, öğrenme algoritmalarındaki eksiklikler veya yanlış model değerlendirme kriterleri gibi nedenlerden kaynaklanabilir. Yanlı kararlar, yalnızca teknik başarısızlık değil, aynı zamanda etik sorunlara da yol açar.

Eğitim verilerindeki tarihsel önyargılar, modelin kararlarına yansır. Örneğin, geçmiş işe alım verilerinde kadınlara karşı ayrımcılık yapıldıysa, model de bu önyargıyı öğrenebilir.

Algoritmaların belirli gruplar veya sınıflar için daha az duyarlı olması. Örneğin, yüz tanıma sistemlerinin farklı ten renklerine karşı performans farkları göstermesi.

Modeller genellikle doğruluk, kesinlik gibi genel metriklerle değerlendirilir; bu durum adaletsiz kararların gözden kaçmasına yol açabilir. Az temsil edilen gruplar için model performansı düşük olsa bile genel doğruluk yüksek olabilir.

Modeller, verilerdeki karmaşık ilişki ve etkileşimleri anlamada yetersiz kalabilir. Bu, belirli grupların yanlış değerlendirilmesine neden olabilir.

Modelin optimizasyon sırasında adalet ve tarafsızlık gibi önemli hedefleri göz ardı etmesi. Örneğin, yalnızca kârı maksimize etmeye odaklanan bir model, sosyal eşitliği dikkate almaz.

ABD'de kullanılan COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) yazılımı örnek gösterilebilir. COMPAS suç işleme olasılığını tahmin eden bir yazılımdır. Bu yazılımda Afro-Amerikan bireyleri beyaz bireylere oranla daha yüksek riskli olarak sınıflandırılmaktadır. COMPAS, tarihsel suç verileriyle eğitildiğinden dolayı ırksal önyargılar taşımaktadır. COMPAS, Afro-Amerikan bireylerini sistematik yüksek riskli olarak göstermeye sebep olduğundan ırksal ayrıma sebep olmuştur.

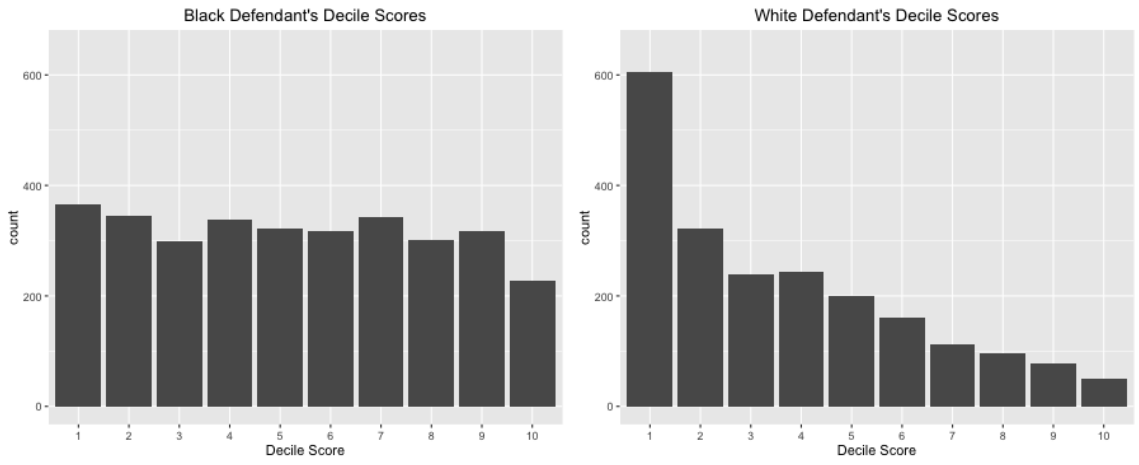
Northpointe'in COMPAS (Alternatif Yaptırımlar için Ceza İnfaz Suçlusu Yönetim Profili) adlı aracının analizi, siyah sanıkların beyaz sanıklara göre tekrar suç işleme riskinin yanlışlıkla daha yüksek olduğu yönündeki yargılanma ihtimalinin çok daha yüksek olduğunu, beyaz sanıkların ise siyah sanıklara göre yanlışlıkla düşük riskli olarak işaretlenme ihtimalinin daha yüksek olduğunu buldu. COMPAS aracının öngördüğü tekrar suç işleme riski kategorilerini, puanlamadan sonraki iki yıldaki sanıkların gerçek tekrar suç işleme oranlarıyla karşılaştırdık ve puanın bir suçlunun tekrar suç işleme oranını %61 oranında doğru tahmin ettiğini, ancak şiddet içeren tekrar suç işleme oranını sadece %20 oranında doğru tahmin ettiğini bulduk. Kimin tekrar suç işleyeceğini tahmin ederken, algoritma siyah ve beyaz sanıklar için tekrar suç işlemeyi yaklaşık aynı oranda doğru bir şekilde tahmin etti (beyaz sanıklar için %59 ve siyah sanıklar için %63) ancak çok farklı şekillerde hatalar yaptı. İki yıllık bir takip süresi boyunca incelendiğinde beyaz ve siyah sanıkları farklı şekilde yanlış sınıflandırıyor (Larson ve ark., 2016).

Analiz şunu buldu:

- Siyah sanıkların genellikle gerçekte olduklarından daha yüksek bir tekrar suç işleme riskine sahip oldukları öngörülmüştür. Analizimiz, iki yıllık bir süre boyunca tekrar suç işlemeyen siyah sanıkların, beyaz meslektaşlarına kıyasla daha yüksek riskli olarak yanlış sınıflandırılma olasılıklarının neredeyse iki katı olduğunu buldu (%45'e karşı %23).
- Beyaz sanıkların genellikle olduklarından daha az riskli oldukları tahmin ediliyordu. Analizimiz, sonraki iki yıl içinde tekrar suç işleyen beyaz sanıkların, siyah tekrar suç işleyenlere göre neredeyse iki kat daha sık olarak yanlışlıkla düşük riskli olarak etiketlendiğini buldu (%48'e karşı %28).
- Analiz ayrıca, önceki suçlar, gelecekte tekrar suç işleme olasılığı, yaş ve cinsiyet gibi faktörler kontrol edildiğinde bile, siyah sanıkların beyaz sanıklara göre yüzde 45 daha fazla risk puanı alma olasılığının bulunduğunu gösterdi.

- Siyah sanıkların ayrıca beyaz sanıklara göre şiddet içeren suç tekrarı riskinin daha yüksek olduğu şeklinde yanlış sınıflandırılma olasılığı iki kat daha fazlaydı. Ve beyaz şiddet içeren suç tekrarı yapanların, siyah şiddet içeren suç tekrarı yapanlara kıyasla şiddet içeren suç tekrarı yapma riskinin düşük olduğu şeklinde yanlış sınıflandırılma olasılığı %63 daha fazlaydı.
- Şiddet içeren suçların tekrarlanması analizi, daha önceki suçlar, gelecekte tekrar suç işleme, yaş ve cinsiyet kontrol edildiğinde bile siyah sanıkların beyaz sanıklara göre yüzde 77 daha fazla risk puanı alma olasılığının bulunduğunu gösterdi (Larson ve ark., 2016).

COMPAS algoritmasını incelemeyi seçmesinin sebebi ülke çapında kullanılan en popüler puanlardan biri ve giderek artan bir şekilde duruşma öncesi ve ceza vermede, yani ceza adalet sisteminin sözde "ön yüzü"nde kullanılması olduğunu söylemektedirler. Broward County'yi seçmelerinin sebebi duruşma öncesi tahliye kararlarında COMPAS aracını kullanan büyük bir yargı bölgesi ve Florida'nın güçlü açık kayıt yasaları olmasıdır. ProPublica, kamu kayıtları talebi yoluyla Florida'daki Broward County Şerif Ofisi'nden iki yıllık COMPAS puanlarını elde etti. 2013 ve 2014'te puanlanan 18.610 kişinin tamamına ait verileri aldı. “Tekrar Suç İşleme Riski” ve “Şiddetli Tekrar Suç İşleme Riski” için COMPAS puanlarını analiz etti (Larson ve ark., 2016).



Şekil 4.11. COMPAS risk analizi (propublica.org)

Broward County'deki 18.610 kişinin COMPAS puanlarını içeren histogram, beyaz sanıkların risk puanlarının büyük bir kısmının düşük risk kategorilerinde

yoğunlaştığını göstermektedir. Bu durum, beyaz sanıkların genel olarak daha düşük riskli olarak değerlendirildiği anlamına gelmektedir. Aksine, siyah sanıkların risk puanları daha geniş bir aralığa yayılmış olup, düşük, orta ve yüksek risk kategorilerinde eşit sayıda birey bulunmaktadır. Bu bulgu, siyah sanıkların risk değerlendirmelerinde daha fazla bir belirsizlik olduğunu ve daha geniş bir risk spektrumunda değerlendirildiğini ortaya koymaktadır. Bu durum, adalet sisteminde ırksal önyargılar olabileceği ve siyah bireylerin daha sıkı gözetim altında tutulma eğiliminde olduğu yönündeki endişeleri güçlendirmektedir.

Yapay zekâ (YZ) uygulamaları, geliştirme süreçlerindeki çeşitli nedenlerle ırksal ayrımcılığa yol açabilir. Bu durum, yapay zekâ sistemlerinin eğitiminde kullanılan verilerdeki önyargılar ve toplumdaki mevcut ırksal önyargıların sistemlere yansıması gibi faktörlerle tetiklenir. Bu örnekler de yapay zekâ sistemlerinin ırksal ayrım yapma potansiyelinin yüksek olduğunu göstermektedir. Irksal ayrımlar toplumda eşitsizliğe yol açmakla beraber mevcut eşitsizlikleri de pekiştirebilmektedir. Bu da ciddi etik problemlere yol açabilmektedir. Yapay zekânın insanlığa hizmet etmesi için, ırksal eşitlik perspektifinin ön planda tutulması büyük önem taşımaktadır.

Yapay zekâ sistemlerinin, eğitildiği veri setlerinden kaynaklı din ayrımcılığı yapabilmektedir. Bu durum, özellikle dinin toplumsal kimlik ve değerler üzerindeki önemli etkisi göz önüne alındığında, ciddi etik sorunlar ortaya çıkarmaktadır.

2017'de, ProPublica adlı bir araştırma kuruluşu, Google'ın reklam hedefleme algoritmasının İslamofobik terimler ve içerikler için reklamlar sunduğunu ortaya çıkarması din ayrımcılığına örnek olarak gösterilebilir. Belirli dinleri hedef alan nefret söylemleri, yapay zekâ tarafından tanınarak buna göre reklamlar sunulmaktaydı. Yapay zekâ sistemlerinin eğitildiği veri setlerinde din temelli ön yargılar ve nefret söylemleri bulunduğundan dolayı Müslümanlar gibi belirli dinlere karşı ayrımcılık pekiştirilmiş oldu. Bu problemi fark eden Google algoritmaları gözden geçirerek güncelledi.

Youtube, video tavsiye ederken bazı din gruplarına nefret söylemli videoları önerebilmektedir. Örneğin İslam karşıtı videolar izlendikten sonra buna benzer videolar tavsiye etmektedir. Algoritma, etkileşimlere dayanarak bu videoları önerdiğinden dolayı nefret söylemli içerikler yayılabilmektedir. Bu şekilde belirli dinlere karşı ön yargılar pekiştirilmektedir.

2019 yılında, Facebook'un içerik algoritmalarının, Müslümanlara karşı nefret söylemi içeren gönderileri yeterince hızlı veya etkili bir şekilde kaldırmadığına dair iddialar ortaya atıldı. Buna karşılık, Müslüman toplulukların savunma amaçlı yaptıkları

paylaşımlar sıklıkla kaldırılıyordu. Algoritmanın eğitildiği veri setinden kaynaklı belirli içeri türlerine karşı hassas olunmamaktaydı. Bunun sonucunda Müslümanları hedef alan nefret söylemleri kaldırmazken bu söylemlere karşı yapılan savunmaları kaldırmaktaydı. Bu durum, Müslümanlara karşı ön yargıları arttırmaktaydı. Facebook bu durumu düzeltmek için içerik politikalarını gözden geçirerek güncellemiştir.

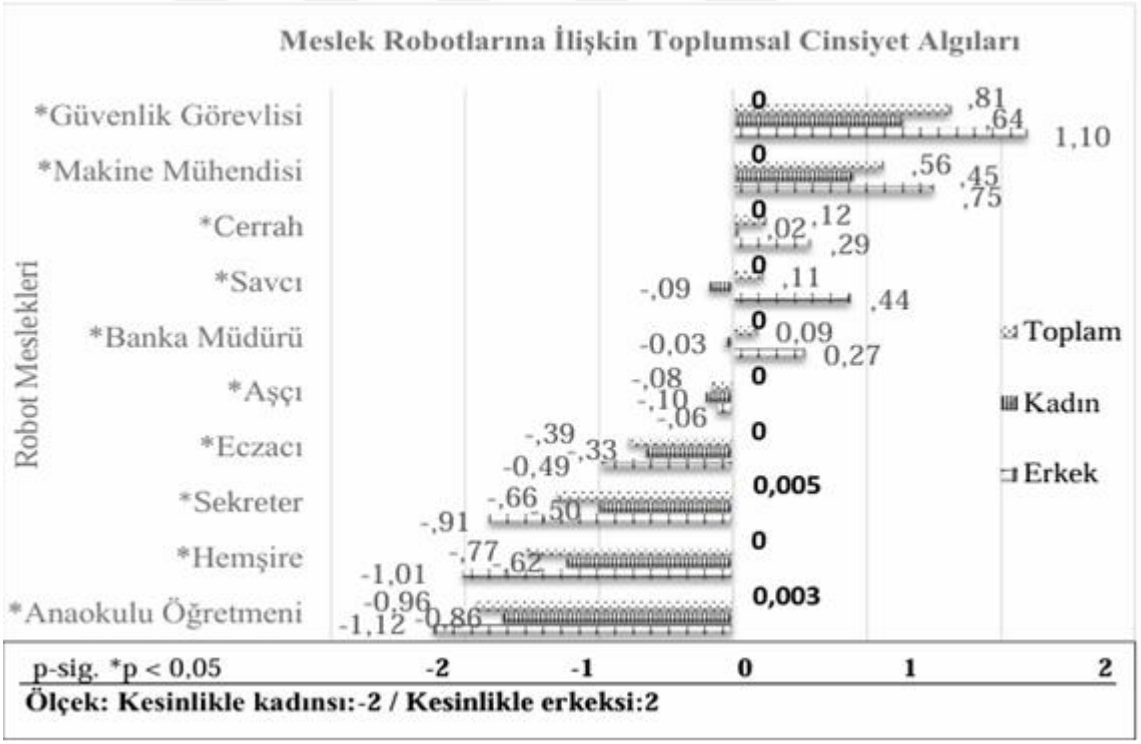
Yapay zekâ (YZ) uygulamaları, doğru şekilde tasarlanmadığında veya eğitim verisi önyargılı olduğunda, örneklerde de görüldüğü gibi din ayrımcılığına yol açabilir. Bu durum, yapay zekânın karar alma süreçlerinde, dini gruplara yönelik olumsuz önyargılar üretebileceği anlamına gelir. Din ayrımcılığı, insanların dini kimliklerine dayalı olarak adaletsiz bir şekilde ayrımcılığa uğraması, dışlanması veya olumsuz muamele görmesi anlamına gelir. Yapay zekâ sistemlerinin din ayrımcılığı yapma potansiyeli yüksektir. Bu durum, toplumun barış ve huzurunu bozacağından dolayı etik sorunlara yol açabilmektedir. Bu sorunları önlemek için veri setleri eğitilirken ön yargıları ortadan kaldıracak yaklaşım benimsenmelidir.

Online reklamcılıkta kullanılan yapay zekâ sistemleri, belirli yaş gruplarını hedeflerken önyargılı davranabilir. Örneğin, belirli ürün veya hizmetlerin reklamları, yaşlı bireylere daha az gösterilebilir ya da yaşlı bireylere yönelik bazı ürünler için ayrımcı reklamlar sunulabilir. Belirli yaş grubundaki bireyleri daha az karlı gördüğünden dolayı onlara içeriklere daha az sunmaktadır. Bu durum, yaşlı bireylerin bazı ürün ve hizmetlere erişimini zorlaştırmaktadır.

Kredi kararı vermede kullanılan yapay zekâ algoritmalarının yaşlılara ayrımcılık yaptığı tespit edilmiştir. Ekonomik olarak aynı koşullara sahip olan bireyler arasında genç bireylere daha düşük miktarda veya daha yüksek faiz oranıyla kredi imkânı sunmaktadır. Bu da yapay zekâ sistemlerinin belirli bir gruba karşı ayrımcılığa sebep olmasına yol açmaktadır.

Yapay zekâ (YZ) sistemleri, eğitim verileri üzerinde eğitildikleri için bu verilerdeki önyargılar doğrudan sistemlere yansır. Özellikle yaşa dayalı olarak oluşturulmuş önyargılar, yapay zekânın belirli yaş gruplarına karşı ayrımcı davranmasına neden olabilir. Bu durum, yapay zekâ uygulamalarının karar verme süreçlerinde yaşa dayalı adaletsizlikler yaratmasına ve toplumsal eşitsizlikleri derinleştirmesine yol açar. Yapay zekâ sistemlerinin yaş bağlı ayrımcı davranışları belirli bir grubun bazı hizmetlere erişimini zorlaştırmakta ve toplumsal hayata tam katılımını engellemektedir. Bu sistemlerin yaşa bağlı ayrımcılık yapması ciddi etik problemleri doğurmaktadır.

Ç.Ü. Sosyal Bilimler Enstitüsü dergisinde yayımlanan yapay zekâ konusunun toplumsal cinsiyet kapsamında incelenmesi: mesleklere yönelik bir araştırma adlı çalışmada on farklı meslek grubu için geliştirilen meslek robotlarına yönelik toplumsal cinsiyet algıları ölçülmüştür. Bu çalışma, robot teknolojilerinin yükselişiyle birlikte toplumsal cinsiyet rollerinin ve algılarının nasıl şekillendiğini anlamak için önemli bir adım teşkil ediyor. Araştırmada, kadın ve erkeklerin robot mesleklerine yönelik tutumlarını değerlendirmek amacıyla T testi analizi kullanılmış ve elde edilen bulgular, görsel bir anlatımla desteklenmiştir. Bu sayede, robot mesleklerine atfedilen cinsiyetçi algıların istatistiksel olarak anlamlı bir şekilde var olup olmadığı ve bu algıların cinsiyetler arasında farklılık gösterip göstermediği ortaya konmuştur. Elde edilen sonuçlar, hem akademik çevreler hem de politika yapıcılar için robot teknolojilerinin toplumsal cinsiyet eşitliği üzerindeki potansiyel etkilerini değerlendirmede önemli bir veri kaynağı sunmaktadır.



Şekil 4.12. Meslek robotlarına ilişkin cinsiyet algıları (Çiftçi ve ark., 2020)

İlginç bir şekilde, katılımcılar, anaokulu öğretmeni, hemşire, sekreter, eczacı ve aşçı gibi mesleklerin robot temsilcilerini daha çok kadınsı, güvenlik görevlisi, makine mühendisi, cerrah, savcı ve banka müdürü gibi mesleklerin robot temsilcilerini ise daha çok erkeksi olarak algılamışlardır. Bu bulgu, robotlara atfedilen cinsiyetçi algıların,

gerçek hayattaki meslekler hakkındaki toplumsal cinsiyet kalıplarını yansıttığını göstermektedir. Yani katılımcılar, robotları değerlendirirken mevcut toplumsal cinsiyet rollerine dayalı zihinsel şemaları kullanmışlardır. Bu durum, robot teknolojilerinin gelişiminin, mevcut toplumsal cinsiyet eşitsizliklerini pekiştirme riski taşıdığını göstermektedir.

4.1.2. Önyargı tespit araçları ve yöntemleri

Önyargı Tespit Araçları ve Yöntemleri, yapay zekâ modellerinde önyargıları (bias) analiz etmek ve değerlendirmek için geliştirilmiş teknikler ve araçlardır. Bu yöntemler, hem modelin çıktılarında hem de verilerin kendisinde yer alan adaletsizlikleri tespit ederek, yapay zekâ sistemlerinin daha adil ve kapsayıcı olmasını sağlamayı hedefler.

Önyargı tespit araçları ve yöntemleri, yapay zekâ sistemlerinin adil, şeffaf ve etik bir şekilde çalışmasını sağlamak için kritik öneme sahiptir. Bu araçlar, sadece teknik çözümler sunmakla kalmaz, aynı zamanda yapay zekânın topluma daha geniş bir fayda sağlamasını mümkün kılar. Önyargıyı tespit etmek ve düzeltmek, hem teknik hem de etik sorumluluk olarak kabul edilmelidir. Bu amaçla AI Fairness 360 (AI F360) ve Adversarial Testing gibi araçlar kullanılabilir

A. AI Fairness 360 (AI F360)

Yapay zekâ modellerindeki önyargıları tespit etmek ve azaltmak için çeşitli araçlar ve yöntemler geliştirilmiştir. AI Fairness 360 (AIF360), bu alanda en popüler ve kapsamlı araçlardan biridir. AI Fairness 360 (AIF360), IBM tarafından geliştirilen ve açık kaynak kodlu bir araç setidir. Yapay zekâ sistemlerindeki önyargıları tespit etmek, analiz etmek ve azaltmak için tasarlanmıştır. Bu sayede, daha adil ve güvenilir yapay zekâ sistemleri geliştirmek mümkün hale gelir.

Yapay zekâ modellerinin eğitiminde kullanılan verilerdeki önyargılar, modelin çıkışlarında da önyargılara yol açabilir. Örneğin, bir kredi değerlendirme modelinde, geçmiş verilerde belirli bir etnik gruba ait kişilerin krediye erişiminin daha düşük olduğu gözlemlenmişse, model yeni başvurularda da bu gruba ait kişilere daha düşük puan verebilir. Bu durum, sosyal adaletsizliklerin derinleşmesine ve belirli grupların mağdur olmasına neden olabilir. AIF360, bu tür önyargıları tespit ederek ve azaltarak daha adil ve eşitlikçi yapay zekâ sistemleri oluşturmamıza yardımcı olur.

Model eğitimi öncesi, sırasında veya sonrasında önyargıyı azaltmak için çeşitli algoritmalar sunar. Bu algoritmalar, veri ön işleme, model eğitimi ve sonuçların düzenlenmesi gibi aşamalarda kullanılabilir.

Kendi önyargı metriklerinizi ve algoritmalarınızı ekleyebilmeniz için esnek bir yapıya sahiptir.

AIF360, Python tabanlı bir kütüphane olarak geliştirildi ve modüler yapısı sayesinde farklı uygulamalara kolayca entegre edilebilir.

Araç kiti, model performansı ve adalet metriklerini görselleştirmek için güçlü araçlar sunar. Bu, kullanıcıların önyargı ve adaletsizlik sorunlarını kolayca anlamalarına yardımcı olur.

Farklı türdeki önyargıları ölçmek için çeşitli metrikler sunar. Örneğin, "disparate impact" (ayrımcı etki) metriksi, bir modelin farklı gruplara ait bireyleri ne kadar farklı şekilde etkilediğini ölçer.

Disparate Impact (Ayrımcı Etki), bir yapay zekâ modelinin farklı gruplara ait bireyleri farklı şekilde etkilediğini gösteren bir önyargı metrikidir. Örneğin, bir kredi verme modelinde, belirli bir etnik gruba ait bireylerin kredilendirilme oranı diğer gruplara göre daha düşükse, bu durum disparate impact olarak değerlendirilir.

Disparate Impact, genellikle pozitif oran oranı (positive rate ratio) olarak hesaplanır. Bu oran, avantajlı gruptaki bireylerin olumlu bir sonuç alması olasılığının, dezavantajlı gruptaki bireylerin olumlu bir sonuç alması olasılığına oranıdır.

Pozitif Oran Oranı = (P(y=1|g=adv)) / (P(y=1|g=disadv))

Burada:

P(y=1|g=adv): Avantajlı gruptaki bireylerin olumlu sonuç alma olasılığı

P(y=1|g=disadv): Dezavantajlı gruptaki bireylerin olumlu sonuç alma olasılığı

Örnek:

Bir kredi verme modelinde, "beyaz" grubu avantajlı, "siyah" grubu dezavantajlı olarak kabul edelim. Eğer beyaz grubun %70'i, siyah grubun ise %50'si krediye kabul ediliyorsa, disparate impact oranı 70/50 = 1.4 olacaktır. Bu, beyaz grubun bireylerinin krediye kabul edilme olasılığının siyah grubuna göre %40 daha yüksek olduğunu gösterir.

Disparate impact, bir modelin belirli gruplara karşı önyargılı olup olmadığını tespit etmek için önemli bir göstergedir. Birçok ülkede, yapay zekâ sistemlerinin adil olması için düzenlemeler yapılmaktadır. Disparate impact, bu düzenlemelere uygunluk için önemli bir ölçüt olarak kullanılmaktadır. Disparate impact, belirli grupların

ekonomik ve sosyal fırsatlara erişimini engelleyebilir ve toplumsal eşitsizlikleri derinleştirebilir. AIF360, bu ve diğer birçok önyargı metrikini hesaplayabilmenizi ve modelinizdeki önyargıları daha iyi anlamanızı sağlar.

Örnek olarak üniversiteye kabul sistemi senaryosunu inceleyelim. Bir üniversite, başvuru değerlendirme sürecini otomatikleştirmek için bir yapay zekâ modeli kullanıyor. Model, başvuru sahiplerinin test skorlarını, önceki akademik başarılarını, yaşadıkları bölgeleri ve okul geçmişlerini analiz ederek "kabul" ya da "red" kararı veriyor. Ancak, bu sürecin eşitlikçi olmadığına dair endişeler var; özellikle modelin sosyoekonomik durum, cinsiyet ya da etnik kökene dayalı önyargılar geliştirmiş olabileceği düşünülüyor.

Model, geçmiş verilere dayalı olarak eğitildiği için, bu verilerdeki potansiyel önyargılar tespit edilmelidir. Eğitim verilerindeki önyargılar şunlardan kaynaklanabilir:

- Başvuranların geçmişte ağırlıklı olarak sosyoekonomik olarak avantajlı bölgelerden olması.
- Cinsiyet dengesizliği (daha fazla erkek ya da kadın başvuranın kabul edilmesi).
- Azınlık gruplara yönelik kabul oranlarının düşük olması.

Bu veriler AI Fairness 360 gibi bir araçla analiz edilir. Araç, özellikle şu metrikleri kullanarak önyargıları ölçebilir:

- Demografik Parite: Farklı gruplar (örneğin, kadınlar ve erkekler, farklı etnik kökenler) arasında kabul oranlarının eşit olup olmadığına bakar.
- Eşitlik Oranı: Azınlık gruplarına ait bireylerin kararlarının, çoğunluk gruplarına kıyasla ne kadar farklı olduğunu ölçer.

Eğitim ve test verileri üzerinde analiz yapılır ve modelin kararları değerlendirilir. Önyargının tespiti şu şekilde gerçekleştirilir:

Cinsiyet Önyargısı:

- Model, kadın ve erkek başvuranlar için aynı test skorları ve akademik geçmişle farklı kararlar veriyor mu?
- Örnek: Kadınların kabul oranı %30, erkeklerin kabul oranı %50 ise bu bir önyargı göstergesidir.

Sosyoekonomik Duruma Dayalı Önyargı:

- Model, sosyoekonomik olarak daha avantajlı bölgelerden gelen başvurulara daha fazla ağırlık veriyor mu?
- Örnek: Düşük gelirli bölgelerden başvuranların kabul oranı %20 iken, yüksek gelirli bölgelerden başvuranların kabul oranı %70 ise, bu bir adaletsizlik işaretidir.

Etnik Köken Önyargısı:

- Farklı etnik gruplara yönelik kabul oranlarında bir farklılık var mı?
- Örnek: Beyaz başvuranların kabul oranı %60, diğer grupların oranı %30 ise modelde etnik önyargı bulunabilir.

Tablo 4.1. AI Fairness çıktıların analizi

Kriter	Grup 1	Grup 2	Adil karar mı?
Cinsiyet (Kadın-Erkek)	Kadın: %35	Erkek: %55	Hayır
Bölge (Gelir seviyesi)	Düşük: %25	Yüksek: %75	Hayır
Etnik köken	Azınlık: %40	Çoğunluk: %60	Hayır

Bu senaryo, AI Fairness 360 gibi araçların önyargı tespiti ve azaltılması için nasıl kullanılabileceğini göstermektedir. Modeldeki gizli önyargılar tespit edildikten sonra, gerekli düzeltmeler yapılarak daha adil ve eşitlikçi bir sistem oluşturulabilir.

AIF360, yapay zekâ sistemlerinde adaletin sağlanması için kritik bir rol oynar. Toplumda güven oluşturmak, etik sorunları azaltmak ve kapsayıcılığı artırmak için bu tür araçlara olan ihtiyaç giderek artmaktadır. AI Fairness 360, sadece teknik bir çözüm sunmakla kalmaz, aynı zamanda etik yapay zekâ geliştirme süreçlerine rehberlik eder. Bu araç, daha adil ve eşitlikçi bir gelecek için yapay zekâ teknolojisinin geliştirilmesine önemli katkılar sağlayabilir.

B. Adversarial Testing

Yapay zekâ sistemlerinin güvenilirliğini, dayanıklılığını ve adaletini değerlendirmek için kullanılan bir yöntemdir. Bu test, modelin zayıflıklarını tespit etmek amacıyla kasıtlı olarak yanıltıcı veya manipülatif girdilerle sistemi sınar. Yapay zekâ modellerinde, özellikle önyargı tespiti ve güvenlik açıklarının bulunmasında etkili bir yöntemdir.

Adversarial testing, insan gözüne normal görünse bile, yapay zekâ sistemini aldatabilecek küçük değişikliklerle sistemin zayıf noktalarını ortaya çıkarır. Bu sayede, sistemin gerçek dünyadaki tehditlere karşı ne kadar dirençli olduğu anlaşılır.

Bu testler sayesinde sistemler, gerçek dünyadaki beklenmedik durumlar ve saldırılara karşı daha dayanıklı hale getirilir.

Sistemlerin kararlarına daha fazla güvenebilmek için, adversarial testing ile güvenilirlikleri doğrulanır.

Sistemlerin belirli gruplara veya verilere karşı önyargılı olup olmadığını tespit etmek için kullanılabilir.

Adversarial örnekler, genellikle orijinal veriye küçük, insan gözünün fark edemeyeceği kadar küçük değişiklikler yaparak oluşturulur. Bu değişiklikler, örneğin bir görüntünün piksellerini hafifçe değiştirmek veya bir ses dosyasına gürültü eklemek gibi basit işlemler olabilir. Ancak bu küçük değişiklikler, yapay zekâ sistemi için büyük bir anlam ifade edebilir ve sistemin yanlış sonuçlar üretmesine neden olabilir.

Adversarial testing, özellikle kritik sistemlerde (otonom araçlar, tıbbi teşhis sistemleri, siber güvenlik sistemleri vb.) büyük önem taşır. Bu sistemlerin hatalı çalışması ciddi sonuçlara yol açabileceğinden, adversarial testing ile güvenilirlikleri sağlanmalıdır.

Örnek olarak sigorta primi belirleme süreci senaryosunu inceleyelim. Bir sigorta şirketi, müşterilere yönelik sigorta primlerini belirlemek için bir yapay zekâ modeli kullanıyor. Model, müşterilerin yaş, cinsiyet, meslek, sağlık durumu ve ikamet ettiği bölge gibi verileri analiz ederek kişisel primleri hesaplıyor. Ancak, belirli grupların daha yüksek prim ödemek zorunda kaldığına dair şikâyetler gelmiştir. Bu durum, modelin önyargılı kararlar verdiği endişesini doğurmuştur.

Model, geçmiş sigorta verileriyle eğitilmiştir. Ancak, bu veri setleri aşağıdaki gibi potansiyel önyargılar içerebilir:

- Kadın müşterilerin sağlık riskleri için daha yüksek prim ödediği durumlar.
- Belirli bölgelerde yaşayan bireylerin daha yüksek riskli olarak değerlendirilmesi.
- Yaşlı bireylerin, sağlık durumu gözetilmeden yüksek primlere maruz kalması.

Bu önyargıları tespit etmek için Adversarial Testing yöntemi uygulanır.

Adversarial Testing, modelin kararlarını test etmek için yapay olarak oluşturulan veya değiştirilmiş veri noktalarını kullanır. Bu senaryoda şu yöntemler uygulanır:

Yapay Profillerin Oluşturulması

Aynı sağlık durumuna, mesleğe ve gelire sahip bireylerin:

- Farklı cinsiyet (kadın ve erkek),
- Farklı yaş grubu (genç ve yaşlı),
- Farklı bölge (kırsal ve kentsel) bilgileriyle sigorta primi değerlendirilir.

Eşit Durumlardaki Farklılıkların Tespiti

Örneğin:

- Kadın ve erkek müşterilere aynı sağlık geçmişiyle verilen primlerin farkı.
- Aynı yaş grubunda, yalnızca ikamet edilen bölgeye bağlı prim değişiklikleri.
- Yaşlı bireylere yönelik sistematik olarak daha yüksek prim uygulamaları.

Tablo 4.2. Test Sonuçları

Kriter	Kadın	Erkek	Fark	Adil karar mı?
Aynı sağlık durumu	$150	$120	%25 daha fazla	Hayır
Bölge	Kırsal	Kentsel	Fark	Adil karar mı?
Aynı sağlık durumu	$200	$150	%33 daha fazla	Hayır
Yaş	20-30	60+	Fark	Adil karar mı?
Aynı sağlık durumu	$100	$180	%80 daha fazla	Hayır

Cinsiyet Önyargısı

Kadın müşterilere, aynı sağlık durumu ve diğer kriterlerde erkeklere kıyasla daha yüksek prim uygulanıyor.

Sebep: Eğitim verilerinde kadınların sağlık masraflarının erkeklere göre daha yüksek olduğu varsayımı.

Bölgesel Ayrımcılık

Kırsal bölgelerde yaşayan müşterilere, aynı koşullarda kentsel bölgelerdeki müşterilere göre daha yüksek prim uygulanıyor.

Sebep: Bölgesel verilerin dengesiz temsili.

Yaş Ayrımcılığı

Yaşlı bireyler, sağlık durumları gözetilmeksizin genç bireylere kıyasla sistematik olarak daha yüksek prim ödüyor.

Sebep: Eğitim verilerinde yaşa dayalı önyargılar.

Bu örnekte Adversarial Testing yöntemi kullanılarak modelin cinsiyet, yaş ve bölge bazında önyargılar içerdiği tespit edilmiştir.

Adversarial testing, yapay zekâ sistemlerinde adalet, dayanıklılık ve güvenliği sağlamak için güçlü bir yöntemdir. Önyargıları ve güvenlik açıklarını ortaya çıkararak, modellerin daha etik ve güvenilir bir şekilde kullanılmasına katkı sağlar. Özellikle yüksek riskli uygulamalarda (sağlık, hukuk, finans) bu tür testler kritik bir öneme sahiptir.

4.1.3. Önyargı azaltma teknikleri

Yapay zekâ sistemlerindeki önyargılar, sistemlerin eğitim verilerindeki eksiklikler, yanlış etiketlemeler veya algoritmik tasarım hataları gibi çeşitli nedenlerle ortaya çıkabilir. Bu önyargılar, sistemlerin kararlarında tutarsızlıklar ve adaletsizlikler yaratabilir. Önyargıyı azaltma, modelin daha adil ve şeffaf hale gelmesi için kritik öneme sahiptir. Bu nedenle, önyargıları azaltmak ve daha adil sistemler geliştirmek için çeşitli teknikler kullanılmaktadır.

Bu teknikler, yapay zekâ modellerinin daha adil kararlar vermesini sağlamak için kullanılabilir. Hangi yöntemin seçileceği, modelin kullanım amacına, verilerin yapısına ve adalet hedeflerine bağlıdır.

Veri Seti: Kredi Kartı İstemcilerinin Temerrüdü

Kredi Kartı İstemcilerinin Temerrüdü Veri Kümesi, Tayvan'daki 30.000 kredi kartı müşterisi hakkında bilgi içeren ve temerrüt riskini tahmin etmeye odaklanan bir veri kümesidir. Bu veri kümesi, yaş, cinsiyet, kredi limiti, eğitim seviyesi, medeni durum ve ödeme geçmişi gibi değişkenleri içerir.

A. Veri Çeşitliliği ve Temsil Eşitliği

Yapay zekâ sistemlerinin gelişimiyle birlikte, bu sistemlerin karar verme süreçlerindeki önyargılar giderek daha büyük bir endişe kaynağı haline gelmiştir. Bu önyargılar, genellikle eğitim verilerindeki eksiklikler veya temsildeki eşitsizliklerden kaynaklanır. Bu nedenle, veri çeşitliliği ve temsil eşitliği, yapay zekâ sistemlerinin adil ve güvenilir olması için kritik öneme sahiptir.

Veri çeşitliliği, bir veri setindeki farklılıkların ve çeşitliliğin derecesini ifade eder. Bu farklılıklar, demografik özellikler (cinsiyet, yaş, etnik köken), coğrafi konum, sosyal ekonomik durum gibi birçok faktörü içerebilir. Çeşitli bir veri seti, farklı perspektifleri ve deneyimleri temsil eder ve bu sayede daha genelleyici ve önyargısız modellerin geliştirilmesine olanak tanır.

Temsil eşitliği ise, veri setindeki farklı grupların, gerçek dünyadaki oranlarına uygun olarak temsil edilmesini ifade eder. Örneğin, bir ülke nüfusunun %50'si kadın ise, eğitim verilerinde de kadınların %50 oranında temsil edilmesi gerekir. Temsil eşitliği, önyargıların oluşmasını engellemek ve sistemlerin tüm kullanıcılar için adil olmasını sağlamak için önemlidir.

Çeşitli ve temsili bir veri seti, modelin farklı gruplara karşı önyargılı kararlar vermesini engeller. Çeşitli bir veri setiyle eğitilen modeller, farklı durumlara daha iyi uyum sağlar ve daha doğru tahminler yapar. Temsil eşitliği, tüm kullanıcıların sistemden eşit faydalanmasını sağlar. Çeşitli ve temsili bir veri setiyle eğitilen modeller, daha güvenilir ve kabul edilebilirdir. Örneğin, bir kredi skorlama modeli geliştirirken sadece belirli bir yaş ve gelir grubuna odaklanmak yerine, farklı yaş, gelir ve cinsiyet gruplarını içeren dengeli bir veri seti kullanmak önyargıyı azaltır.

Tablo 4.3. Sadece belirli bir yaş ve gelir grubuna odaklanmış dengesiz veri seti.

Cinsiyet	Yaş	Gelir	Skor
Erkek	35	7000	750
Kadın	40	8000	780

Tablo 4.4. Farklı yaş, gelir ve cinsiyet gruplarını içeren dengeli veri seti.

Cinsiyet	Yaş	Gelir	Skor
Erkek	25	4000	650
Kadın	23	5000	680
Erkek	45	6000	720
Kadın	46	7000	750

Veri çeşitliliği ve temsil eşitliği, yapay zekâ sistemlerinin adil ve etkili çalışması için gereklidir. Bu unsurları göz ardı etmek, ayrımcı sistemlerin oluşmasına yol açabilir. Etkili veri yönetimi ve adil model tasarımı, hem etik hem de teknik açıdan başarılı bir yapay zekâ uygulaması için temel taşlarıdır.

B. Veri Toplama Yöntemlerinin Gözden Geçirilmesi

Yapay zekâ (AI) sistemlerinin başarısı büyük ölçüde kullanılan verilerin kalitesine bağlıdır. Veri toplama yöntemleri, önyargının en büyük kaynaklarından biri

olabilir. Bu nedenle, veri toplama süreçlerinin gözden geçirilmesi ve iyileştirilmesi, adil ve güvenilir yapay zekâ modelleri geliştirmek için kritik öneme sahiptir.

Farklı demografik gruplardan, coğrafi bölgelerden ve sosyal bağlamlardan veri elde edilmelidir. Örneğin: Farklı yaş gruplarını kapsayan anketler düzenlemek.

Az temsil edilen gruplar için hedefli veri toplama çalışmaları yapılmalıdır. Örnek: Kadın ve erkek sporcuları eşit şekilde temsil eden bir veri seti oluşturmak.

Eksik veya az temsil edilen gruplar için yapay olarak üretilmiş veri kullanılabilir. Örnek: Sanal ortamda farklı cilt tonlarına sahip yüz görüntüleri oluşturmak.

Veri toplama kaynağına dair önyargı veya yanlış bilgi içerip içermediği değerlendirilmelidir. Örneğin: Sadece bir sosyal medya platformundan alınan veriler, o platformun kullanıcılarına özgü bir yanlılık taşıyabilir.

Verinin toplandığı bağlamın (zaman, mekân, kültür) önyargıya neden olup olmayacağı göz önünde bulundurulmalıdır. Örneğin: Belirli bir yılın ekonomik kriz verilerinin genelleştirilmesi uygun olmayabilir.

Veri toplama süreçlerinde öznel kararların etkisi azaltılmalıdır. Örnek: Müşteri memnuniyeti değerlendirmelerinde standart bir ölçek kullanmak.

Veri toplama sürecinde ayrımcılık, mahremiyet ihlali gibi etik sorunlara dikkat edilmelidir. Örneğin: Hassas bilgiler anonimleştirilerek toplanmalıdır.

Her veri noktasının diğer bağlamlarla uyumlu ve eksiksiz olmasına özen gösterilmelidir. Örnek: Sağlık kayıtlarının hem tanı hem de tedavi sürecini kapsaması.

İnsan kaynaklı hataları azaltmak için sensörler, web tarayıcıları gibi otomasyon araçları kullanılabilir. Örnek: İnternet trafiğini analiz etmek için otomatik veri kazıma teknikleri kullanmak.

Modelin potansiyel zayıflıklarını hedef alan veriler toplanarak sistemin dayanıklılığı artırılır. Örneğin: Bir konuşma tanıma sistemi için farklı aksanlardan ses verileri toplamak.

Tablo 4.5. Dengesiz Veri Seti

Cinsiyet	Yaş	Gelir
Erkek	31	5000
Kadın	33	6000
Erkek	45	7500
Kadın	42	8000

Tablo 4.6. Dengeli Veri Seti

Cinsiyet	Yaş	Gelir
Erkek	22	2500
Kadın	25	3500
Erkek	37	5000
Kadın	34	6000
Erkek	50	7000
Kadın	46	4000
Erkek	70	3000
Kadın	65	2000

Veri toplama yöntemlerinin gözden geçirilmesi ve iyileştirilmesi, yapay zekâ modellerinin adil ve doğru sonuçlar üretmesi için gereklidir. Bu süreç, sadece teknolojik değil, aynı zamanda etik, sosyal ve yasal boyutları da dikkate alarak tasarlanmalıdır. Çeşitli ve dengeli veri, başarılı yapay zekâ uygulamalarının temel taşıdır.

C. Veri Anonimleştirme

Veri anonimleştirme, kişisel ve hassas bilgilerin, bireylerin kimliğini tespit etmeyi imkânsız hale getirecek şekilde işlenmesi sürecidir. Bu işlem, veri güvenliği, gizlilik ve etik sorunları açısından önemli bir adımdır. Özellikle yapay zekâ ve makine öğrenimi sistemlerinde, anonimleştirilmiş verilerle eğitim yapılması, kullanıcıların gizliliğini korurken veri analizlerinin yapılabilmesine olanak tanır.

Veri anonimleştirme, aynı zamanda yasal düzenlemelere ve gizlilik politikasına uymayı sağlamak için de gereklidir. Bu süreç, kişisel verilerin korunması adına yapılan bir dizi işlemi içerir ve veri üzerinde yapılan değişikliklerin, orijinal bilgiyi geri kazanmayı imkânsız hale getirmesini sağlar.

Veri anonimleştirme, kişisel bilgilerin korunmasını sağlar ve bireylerin mahremiyetini korur. Bu, özellikle hassas verilerle çalışan sistemlerde önemlidir.

Çeşitli düzenlemeler, özellikle Genel Veri Koruma Yönetmeliği (GDPR) gibi yasalar, kişisel verilerin korunmasına dair katı kurallar getirir. Anonimleştirme, bu düzenlemelere uyum sağlamayı kolaylaştırır. Örnek: GDPR, verilerin anonimleştirilmesi durumunda, verilerin kişisel veriler olarak sayılmadığını belirtir.

Anonimleştirilmiş veriler, güvenli bir şekilde üçüncü şahıslarla paylaşılabilir. Bu, veri analizi ve araştırmalar için önemli bir avantajdır. Örneğin: Sağlık verileri, anonimleştirildikten sonra araştırma amaçlı olarak sağlık kuruluşları veya bilim insanlarıyla paylaşılabilir.

Anonimleştirme, kimlik bilgileri üzerinden yapılabilecek ayrımcılığın ve önyargıların önüne geçer. Örneğin: Eğitim ve işe alım sistemlerinde anonimleştirilmiş veriler kullanılarak, cinsiyet veya etnik köken gibi faktörlerin etkisi minimize edilebilir.

Anonimleştirme, bazen verinin bazı yönlerinin kaybolmasına yol açabilir. Bu, modelin doğruluğunu ve analiz sonuçlarını etkileyebilir. Örneğin, çok fazla genel hale getirilmiş veri, spesifik kararlar almakta zorlanabilir.

Anonimleştirilmiş veriler, bazı tekniklerle geri deanonimleştirilebilir (kimliklerin tekrar açığa çıkması). Bu, özellikle büyük veri setlerinde ve çok sayıda veri kaynağı kullanıldığında bir risk oluşturabilir. Örneğin: Veri setinde belirli bir demografik bilgi belirginse, kimlik tespiti yapılabilir.

Veri anonimleştirildikten sonra, veri setinden çıkarılabilecek diğer hassas bilgiler de olabilir. Bu nedenle anonimleştirilmiş verilerde, çıkarılabilir verilerin izlenmesi gerekebilir. Örneğin: Sağlık verilerinin anonimleştirilmesi sırasında, yalnızca yaş ve cinsiyet gibi bilgilerin değil, aynı zamanda hastalık geçmişi ve tedavi türlerinin de korunması önemlidir.

Anonimleştirilmiş verilerin, hala etik veya yasal sorunlar oluşturabilecek durumlar olabilir. Veri anonimleştirme sürecinde dikkat edilmesi gereken etik ilkeler ve güvenlik gereksinimleri vardır.

Tablo 4.7. Anonimleştirilmemiş Veri Seti

Cinsiyet	Yaş	Ağırlık
Erkek	28	75
Erkek	35	80
Erkek	40	85
Kadın	30	60
Kadın	35	65

Tablo 4.8. Anonimleştirilmiş Veri Seti

Cinsiyet	Yaş	Ağırlık
Erkek	20-30	70-80
Erkek	30-40	70-80
Erkek	40-50	80-90
Kadın	30-40	50-70
Kadın	30-40	50-70

Tablo 4.9. Veri Anonimleştirme Öncesi ve Sonrası

Anonimleştirmeden Önce Yargı Analizi	Cinsiyet	Seçim Oranı
Anonimleştirmeden Önce	Erkek	0.207329
	Kadın	0.243936
Anonimleştirmeden Sonra	-	0.218833

Cinsiyet anonimleştirme, modelin cinsiyet önyargısını azalttı. Anonimleştirmeden önce, erkekler için seçim oranı kadınlardan daha düşüktü. Anonimleştirmeden sonra, seçim oranı eşitlendi ve önyargı ortadan kaldırıldı. Bu, cinsiyet bilgisinin modelin performansını etkilemediğini ve modeli daha adil hale getirdiğini gösterir.

Veri anonimleştirme, kişisel verilerin korunması ve veri paylaşımı arasında denge kurmak için önemli bir araçtır. Ancak, mükemmel bir çözüm olmadığını ve dikkatli bir şekilde uygulanması gerektiğini unutmamak gerekir.

Veri anonimleştirme, verilerin gizliliğini korurken, veri analizi ve paylaşımı için de olanak sağlar. Bu sayede hem bireylerin hakları korunur hem de bilimsel çalışmalar ve iş dünyası için değerli bilgiler elde edilir.

D. Veri Normalizasyonu

Veri normalizasyonu, farklı ölçeklerdeki veri özelliklerinin bir standarda getirilmesi işlemidir. Bu süreç, verilerin daha tutarlı, karşılaştırılabilir ve modelleme için uygun hale gelmesini sağlar. Veri normalizasyonu, özellikle makine öğrenimi algoritmalarının doğru çalışabilmesi için önemlidir, çünkü bazı algoritmalar, verinin aynı ölçek ve aralıkta olmasını gerektirir.

Veri setindeki özelliklerin farklı ölçeklere sahip olması, bazı özelliklerin daha baskın hale gelmesine neden olabilir. Örneğin, bir veri setinde yaş (0-100 arası) ve maaş (10,000-100,000 arası) gibi iki özellik bulunabilir. Eğer bu özelliklerin farklı ölçekleri varsa, model daha büyük değerlere sahip olan özellikleri (örneğin, maaş) daha fazla dikkate alabilir. Normalizasyon, bu farkları ortadan kaldırır.

Özellikle gradyan tabanlı algoritmalar (örneğin, yapay sinir ağları, lojistik regresyon) için normalizasyon, daha hızlı ve daha doğru sonuçlar alınmasını sağlar. Çünkü bu algoritmalar, verinin daha küçük ve benzer aralıklarla olması durumunda daha verimli çalışır.

Özellikle derin öğrenme gibi karmaşık modellerde, özellikler farklı aralıklara sahip olduğunda eğitim süreci daha istikrarsız olabilir. Normalizasyon, modelin daha kararlı bir şekilde eğitim almasını sağlar.

Normalizasyon, tüm özelliklerin model üzerinde eşit ağırlığa sahip olmasını sağlar. Bu, modelin herhangi bir özelliğe gereksiz yere fazla ağırlık vermesini engeller.

Verileri belirli bir aralığa (genellikle 0 ile 1 arasına) yeniden ölçeklendirir. Her bir özellik için minimum ve maksimum değerler kullanılarak veriler yeniden boyutlandırılır.

$$X_{normalized} = (X - X_{min}) / (X_{max} - X_{min})$$

Kolay ve hızlı uygulanabilir. Veri seti 0 ile 1 arasında normalize edilerek daha homojen bir hale gelir.

Aykırı değerler (outliers) var ise, bu değerler veri setinin genel dağılımını değiştirebilir.

Örnek:

Bir veri setinde, yaşın minimum değeri 20, maksimum değeri 70 ise, 50 yaşındaki bir bireyi normalize ettiğimizde:

$$X_{normalized} = (50 - 20) / (70 - 20) = 0.4286$$

Tablo 4.10. Normalize Edilmemiş Veri Seti

Cinsiyet	Yaş	Ağırlık	Gelir
Erkek	25	75	4000
Kadın	30	65	5000
Erkek	35	85	7000
Kadın	28	55	6000

Tablo 4.11. Normalize Edilmiş Veri Seti

Cinsiyet	Yaş	Ağırlık	Gelir
Erkek	0	0.67	0
Kadın	0,-.33	0.33	0.33
Erkek	1	1	1
Kadın	0.17	0	0.67

Tablo 4.12. Veri Normalize Edilmeden Öncesi ve Sonrası

Anonimleştirme ve Normalleştirme Öncesi ve Sonrası Önyargı Analizi	Cinsiyet	Seçim Oranı
Anonimleştirme Öncesi	Erkek	0.207329
	Kadın	0.243936
Anonimleştirme Sonrası	-	**0.218833**
Anonimleştirme ve Normalleştirme Öncesi	Erkek	0.207329
	Kadın	0.243936
Anonimleştirme ve Normalleştirme Sonrası	-	**0.221792**

Verilerin anonimleştirilmesi ve normalleştirilmesi nedeniyle cinsiyet grupları arasındaki ayrım ortadan kaldırılmış, bu da erkekler için seçim oranının artmasına ve modelden daha adil bir sonuç elde edilmesine neden olmuştur. Ek olarak, normalleştirmenin, verileri yalnızca anonimleştirmeye kıyasla veri dağılımını dengeleyerek performansı iyileştirdiği dikkat çekicidir.

Veri normalizasyonu, makine öğrenimi ve yapay zekâ alanlarında kritik bir adımdır. Farklı normalizasyon yöntemleri, veri setinin özelliklerine ve kullanılacak modele bağlı olarak seçilmelidir. Doğru yöntemle yapılan normalizasyon, modelin doğruluğunu artırabilir ve eğitim sürecini daha verimli hale getirebilir. Ancak, her

normalizasyon yöntemi her veri setine uygun olmayabilir, bu yüzden verinin özelliklerine göre dikkatli bir seçim yapılmalıdır.

E. Yeniden Örnekleme (Re-Sampling)

Yeniden örnekleme, veri setindeki örneklerin (gözlemlerin) seçilme yönteminin değiştirilmesi sürecidir. Bu işlem, özellikle dengesiz veri setlerinde, modelin daha iyi performans gösterebilmesi için kullanılır. Dengesiz veri, bir sınıfın (örneğin, pozitif veya negatif etiket) diğerine göre çok daha fazla örneğe sahip olduğu durumu ifade eder. Yeniden örnekleme, bu dengesizliği gidermeyi amaçlar.

Bu yöntemler, özellikle sınıf dengesizliği problemi yaşanan durumlarda modelin daha iyi genelleme yapabilmesi için kullanılır. Yeniden örnekleme, sınıf dengesizliğini gidererek, modelin azınlık sınıfına karşı daha adil bir performans sergilemesini sağlar. Özellikle sınıf dengesizliği bulunan veri setlerinde, yeniden örnekleme, doğruluk, F1 skoru gibi performans metriklerini iyileştirebilir. Aşırı örnekleme yöntemleri (özellikle SMOTE) modelin daha iyi genelleme yapmasına olanak tanır.

Aşırı örnekleme, aynı örneklerin tekrarlanması veya sentetik örneklerin eklenmesi, modelin aşırı uyum yapmasına yol açabilir. Eksik örnekleme, önemli verilerin kaybolmasına neden olabilir, bu da modelin doğruluğunu olumsuz etkileyebilir. Aşırı örnekleme, özellikle SMOTE ve ADASYN gibi algoritmalar kullanıldığında hesaplama maliyetini artırabilir.

Yeniden örnekleme, çoğunluk sınıfının çok fazla olduğu durumlarda, modelin azınlık sınıfını daha iyi öğrenebilmesi için kullanılır. Örneğin: Kredi kartı dolandırıcılığı tespiti gibi uygulamalarda, dolandırıcılık işlemleri genellikle çok daha az olduğu için yeniden örnekleme yöntemleri kullanılarak bu dengesizlik giderilebilir.

Anomali tespiti, genellikle çok az sayıda anormal gözlemle ilgilenir. Yeniden örnekleme, bu tür anormal gözlemleri daha iyi tespit etmek için kullanılır.

Sınıflandırma algoritmalarında, sınıf dengesizliği problemi varsa, yeniden örnekleme ile verinin dengelenmesi modelin doğruluğunu artırabilir.

Yeniden örnekleme genellikle iki ana yöntemle yapılır:

- Aşırı örnekleme (Oversampling): Azınlık sınıfındaki örneklerin sayısını artırmak.
- Eksik örnekleme (Undersampling): Çoğunluk sınıfındaki örneklerin sayısını azaltmak.

Aşırı Örnekleme (Oversampling): Aşırı örnekleme, azınlık sınıfındaki verilerin sayısını artırmayı amaçlar. Bu, çoğunluk sınıfına karşı denge oluşturmak için yapılır. Aşırı örnekleme, genellikle veri setine yeni örnekler ekleyerek yapılır. Bu yeni örnekler, ya orijinal örneklerin kopyaları ya da mevcut verilerden türetilen yeni örnekler olabilir.

Eksik Örnekleme (Undersampling): Eksik örnekleme, çoğunluk sınıfındaki örneklerin sayısını azaltmayı amaçlar. Bu, çoğunluk sınıfının fazla örneğe sahip olduğu ve bu dengesizliğin modelin eğitimine zarar verdiği durumlarda kullanılır.

Tablo 4.13. Yeniden Örnekleme Sonuçları

Anonimleştirme ve Normalleştirme Öncesi ve Sonrası Önyargı Analizi	Cinsiyet	Seçim Oranı
Anonimleştirmeden Önce	Erkek	0.207329
	Kadın	0.243936
Anonimleştirmeden Sonra	-	**0.218833**
Anonimleştirme ve Normalleştirme Sonrası	-	**0.221792**
Anonimleştirme, Normalleştirme ve SMOTE (aşırı örnekleme) Sonrası	-	**0.5**
Anonimleştirme, Normalleştirme ve TomekLinks (alt örnekleme) Sonrası	-	**0.235604**

SMOTE'yi uyguladıktan sonra, seçim oranı önemli ölçüde 0.5'e yükseldi, bu da tekniğin azınlık gruplarının temsilini artırarak veri setini etkili bir şekilde dengelediğini ve böylece sonuçlarda adaleti iyileştirdiğini gösteriyor. Tomek Links'i uyguladıktan sonra seçim oranı 0.235604'e yükseldi. Bu, tekniğin gürültülü örnekleri kaldırarak veri kümesini etkili bir şekilde temizlediğini gösterir, bu da modelin performansını artırmaya yardımcı olur ve sonuçlardaki yanlılığı azaltır.

Yeniden örnekleme, sınıf dengesizliği ile başa çıkmak için etkili bir tekniktir. Aşırı örnekleme ve eksik örnekleme yöntemleri, dengesiz veri setlerinde daha iyi performans gösterecek modeller oluşturulmasına yardımcı olabilir. Ancak, her iki yöntemin de dezavantajları bulunmakta olup, modelin aşırı uyum yapması veya veri kaybı gibi riskleri vardır. Bu nedenle, yeniden örnekleme yöntemleri dikkatlice seçilmeli ve gerektiğinde birlikte kullanılmalıdır.

F. Yeniden Ağırlıklandırma (Re-Weighting)

Yeniden ağırlıklandırma, veri setindeki örneklerin (gözlemlerin) sınıflarına veya belirli özelliklerine göre ağırlıkların yeniden belirlenmesi işlemidir. Bu işlem, genellikle sınıf dengesizliği gibi sorunları çözmek için kullanılır. Yeniden ağırlıklandırma, azınlık sınıfına ait örneklerin daha fazla ağırlık almasını sağlayarak modelin bu sınıfı daha iyi öğrenmesini hedefler. Aynı şekilde, çoğunluk sınıfındaki örnekler daha düşük ağırlıklara sahip olabilir.

Bu yöntem, yeniden örnekleme (re-sampling) yöntemlerine alternatif veya tamamlayıcı olarak kullanılabilir. Yeniden örnekleme, veri setinin sayısını değiştirirken, yeniden ağırlıklandırma verilerin sayısını değiştirmeden modelin öğrenme sürecinde ağırlıklı olarak hangi örneklerin dikkate alınacağını belirler.

Yeniden ağırlıklandırma, dengesiz veri setlerinde, azınlık sınıfına daha fazla önem verilmesini sağlar, böylece modelin her iki sınıfı da daha iyi öğrenmesini sağlar. Yeniden ağırlıklandırma, veri setinin boyutunu değiştirmediği için, yeniden örneklemeye kıyasla veri kaybı yaşanmaz ve verilerin çeşitliliği korunur. Yeniden ağırlıklandırma, veri setindeki örneklerin sayısını değiştirmeden dengesizliği ele almak için kullanılabilir, bu da bazı durumlarda daha uygun olabilir.

Azınlık sınıfına verilen yüksek ağırlıklar, modelin azınlık sınıfına aşırı uyum yapmasına yol açabilir. Bu durum, modelin genelleme yeteneğini olumsuz etkileyebilir. Ağırlıkların hesaplanması, özellikle adaptif ağırlıklandırma yöntemlerinde karmaşık olabilir. Her örneğin veya sınıfın ağırlığını doğru belirlemek zordur ve yanlış ağırlıklar, modelin performansını olumsuz etkileyebilir. Ağırlıklandırma sırasında aykırı değerlerin (outliers) etkisi artabilir. Aykırı değerler, aşırı derecede yüksek ağırlık alabilir ve bu da modelin kötü performans göstermesine yol açabilir.

Yeniden ağırlıklandırma, dengesiz veri setlerinde, azınlık sınıfının öğrenilmesini sağlamak amacıyla kullanılır. Bu, özellikle sınıfların çok farklı sayılarda olduğu durumlar için önemlidir (örneğin, kredi kartı dolandırıcılığı tespiti, sağlık verileri gibi). Adaptif ağırlıklandırma, modelin zor öğrenilen örnekler üzerinde daha fazla çalışmasını sağlar. Bu yöntem, özellikle derin öğrenme gibi karmaşık modellerde faydalıdır. Anomali tespiti ve hata tespiti gibi uygulamalarda, modelin nadir ve önemli örnekleri doğru şekilde öğrenebilmesi için yeniden ağırlıklandırma kullanılabilir.

Tablo 4.14. Yeniden Ağırlıklandırma Öncesi ve Sonrası

Anonimleştirme ve Normalleştirme Öncesi ve Sonrası Önyargı Analizi	Cinsiyet	Seçim Oranı
Anonimleştirmeden Önce	Erkek	0.207329
	Kadın	0.243936
Anonimleştirmeden Sonra	-	**0.218833**
Anonimleştirme ve Normalleştirme Sonrası	-	**0.221792**
Anonimleştirme, Normalleştirme ve SMOTE (aşırı örnekleme) Sonrası	-	**0.5**
Anonimleştirme, Normalleştirme ve TomekLinks (alt örnekleme) Sonrası	-	**0.235604**
Anonimleştirme, Normalleştirme, SMOTE ve Yeniden Ağırlıklandırma Sonrası	-	**0.974996**

Ağırlık yeniden uygulandıktan sonra, seçilme oranı 0,974996'ya kadar önemli ölçüde arttı. Bu, azınlık gruplarının temsilini artırmada ve modelin adaletini ve performansını önemli ölçüde geliştirmedeki etkinliğini göstermektedir.

Yeniden ağırlıklandırma, dengesiz veri setlerinde sınıf dengesizliğini çözmek için etkili bir yöntemdir. Bu yöntem, veri setindeki örneklerin ya da sınıfların ağırlıklarını değiştirerek, modelin her bir sınıfı daha doğru bir şekilde öğrenmesini sağlar. Ancak, ağırlıklandırma işleminin doğru şekilde yapılması önemlidir, çünkü yanlış ağırlıklar modelin aşırı uyum yapmasına ya da kötü performans göstermesine yol açabilir. Yeniden ağırlıklandırma, yeniden örnekleme yöntemleriyle birlikte kullanılabilir ve sınıf dengesizliğine karşı daha güçlü bir çözüm sunabilir.

G. Adalet Kısıtlamaları (Fairness Constraints)

Adalet kısıtlamaları, makine öğrenmesi ve yapay zekâ modellerinde adaletin sağlanmasını amaçlayan matematiksel kısıtlamalardır. Bu kısıtlamalar, modelin çıktılarının belirli gruplar arasında eşit veya adil olmasını sağlamak için kullanılır. Adalet kısıtlamaları, özellikle sınıf dengesizliği, önyargı ve ayrımcılığın önlenmesi için gereklidir. Bu kısıtlamalar, modelin öğrenme sürecinde adaletin korunmasını sağlamak için belirli eşitlikleri (fairness) dikkate alır.

Adalet kısıtlamaları genellikle modelin doğruluğunu optimize ederken, grup veya birey bazında eşitlik sağlamaya yönelik ek bir hedef veya sınırlama olarak eklenir.

Böylece modelin karar verme süreci sadece doğrulukla değil, aynı zamanda adaletle de uyumlu hale gelir.

Farklı gruplar (örneğin, cinsiyet, ırk, yaş) arasında eşitlik sağlamak amaçlanır. Bu tür adalet kısıtlamaları, gruplar arasındaki sonuç farklılıklarını minimize etmeyi hedefler. Örneğin, bir işe alım modelinin erkek ve kadın adaylara eşit şekilde fırsatlar sunmasını sağlamak.

Benzer durumda olan bireylerin benzer sonuçlar alması gerektiği ilkesine dayanır. Aynı özelliklere sahip bireylerin modelin kararları tarafından eşit şekilde değerlendirilmesi sağlanır. Örneğin, kredi başvurusunda benzer finansal geçmişe sahip bireylerin eşit muamele görmesi.

Modelin, gruplar arasında eşit fırsatlar sunması sağlanır. Bu, özellikle belirli bir grup için "fırsat eşitliği" sağlamak amacıyla uygulanabilir. Örneğin, ırksal gruplar arasında işe alınma şanslarının eşit olmasını sağlamak.

Verideki grupların veya bireylerin özelliklerinin modelin kararlarına orantılı şekilde yansıması sağlanır. Bu, genellikle veri ön işleme adımlarında, gruplar arasındaki adaletsiz temsili ortadan kaldırmak için kullanılan yöntemleri içerir.

Tablo 4.15. Adalet Kısıtlamaları Öncesi ve Sonrası

Eşit Oranlardan Önce Önyargı Analizi	Cinsiyet	Seçim Oranı
Eşitlik Oranlarından Önce	Erkek	0.429918
	Kadın	0.442661
Eşitlik Sağlandıktan Sonra	Erkek	0.424984
	Kadın	0.422868

Eşitlenmiş Oranlar uygulandıktan sonra, modelin yanlılığı önemli ölçüde azalmış ve cinsiyetler arasındaki seçim oranları neredeyse eşitlenmiştir.

Adalet kısıtlamaları, makine öğrenmesi ve yapay zekâ modellerinde gruplar arasında eşitlik ve adaletin sağlanmasını hedefler. Bu kısıtlamalar, grup bazlı veya bireysel bazda adalet sağlamayı amaçlar ve modelin doğruluğunu ve performansını dengeleyerek toplumsal eşitlik yaratmaya çalışır. Ancak, adalet kısıtlamalarının uygulanması sırasında modelin performansının etkilenmesi gibi zorluklar da ortaya çıkabilir. Bu nedenle, adaletin sağlanması ve modelin etkinliğinin korunması arasında dikkatli bir denge kurulmalıdır.

H. Adil Temsil Öğrenme (Fair Representation Learning)

Adil Temsil Öğrenme (Fair Representation Learning), makine öğrenmesi ve yapay zekâ sistemlerinde, modellerin kararlarını adil bir şekilde alabilmesi için verilerin temsili üzerinde yapılan bir tekniktir. Bu yaklaşım, verilerin ve öğrenilen temsillerin (özelliklerin) belirli gruplar arasında adaletli bir şekilde dağıtılmasını amaçlar. Adil Temsil Öğrenme, özellikle gruplar arasında ayrımcılığın önlenmesi ve önyargıların ortadan kaldırılması için kullanılır. Bu yöntem, verilerin önyargılı olmayan bir şekilde modelin eğitimine dahil edilmesini sağlayarak, modelin kararlarının daha adil olmasına yardımcı olur.

Adil Temsil Öğrenme, verilerin daha adil bir temsile dönüştürülmesini sağlamak için kullanılan bir tekniktir. Bu, özellikle veri setlerinde önyargı (bias) ve adaletsizlik (unfairness) bulunan durumlarda önemlidir. Bu yaklaşımda, gruplar arasındaki eşitsiz temsili düzelterek, modelin öğrenme sürecinin gruplar arasında adaletli bir şekilde gerçekleşmesi sağlanır.

Temsil öğrenme, makine öğrenmesinin önemli bir bileşeni olup, verinin içindeki karmaşık ilişkileri daha anlamlı bir şekilde öğrenmeyi amaçlar. Ancak, temsil öğrenme sırasında verideki ırk, cinsiyet, yaş gibi hassas özelliklerin adaletli bir şekilde ele alınması, önyargılı kararların önüne geçilmesini sağlar.

Adil Temsil Öğrenme, modelin çıktılarının önyargılardan bağımsız olmasını sağlar. Örneğin, ırk, cinsiyet gibi özelliklere dayalı ayrımcılığı engeller. Verideki gruplar arasında temsil eşitliği sağlanır. Bu, her grubun modelin eğitimine orantılı bir şekilde katkı sağlamasını ve modelin her grup için adil kararlar almasını hedefler. Modelin kararları, yalnızca doğruluğu değil, aynı zamanda adalet ölçütlerini de göz önünde bulundurur. Bu sayede her gruptan gelen bireyler, benzer durumlar için benzer kararlar alır. Temsil öğrenme sırasında, modelin doğruluğunu artırırken, aynı zamanda ırk, cinsiyet veya diğer hassas grupların özelliklerinin kararlar üzerindeki etkisi azaltılır.

Adil Temsil Öğrenme, önyargıların ortadan kaldırılmasına yardımcı olur ve gruplar arasındaki ayrımcılığı engeller. Bu, modelin daha adil kararlar almasını sağlar. Temsil öğrenme ile modelin gruplar arasında adaletli kararlar alması sağlanır, bu da toplumsal eşitlik açısından önemlidir. Adil Temsil Öğrenme, verinin her bir grubun özelliklerini dikkate alarak daha kapsayıcı ve eşitlikçi bir model oluşturulmasına olanak tanır.

Adalet sağlamak için ek kısıtlamalar getirilmesi, modelin doğruluğunu veya genelleme yeteneğini olumsuz etkileyebilir. Verilerdeki hassas grupların özelliklerini ayrıştırmak ve adaletli bir temsile dönüştürmek, karmaşık olabilir ve bu süreçlerde veri kaybı yaşanabilir. Adaletin sağlanıp sağlanmadığını değerlendirmek, özellikle farklı adalet ölçütleri (grup adaleti, bireysel adalet, fırsat eşitliği) arasındaki dengeyi kurmak zor olabilir.

Tablo 4.16. Adil Temsil Öğrenme Öncesi ve Sonrası

Adalet Kısıtlamalarından Önce Önyargı Analizi	Cinsiyet	Seçim Oranı
Adil Temsil Öğrenmeden Öncesi	Erkek	0.411699
	Kadın	0.451376
Adil Temsil Öğrenmeden Sonrası	Erkek	0.394720
	Kadın	0.414701

Adil temsil öğrenme, modeldeki cinsiyetler arasındaki ayrımcılığı azaltarak önyargının etkisini en aza indirmiştir.

Adil Temsil Öğrenme, yapay zekâ ve makine öğrenmesi uygulamalarında adaletin sağlanması açısından önemli bir yaklaşımdır. Bu yöntem, modelin öğrenme sürecinde, gruplar arasındaki önyargıyı ortadan kaldırarak daha eşitlikçi sonuçlar elde edilmesini sağlar. Adil temsil, yalnızca doğruluğu değil, aynı zamanda sosyal adaletin sağlanmasını da hedefler. Ancak, bu yöntemin uygulanmasında dikkatli bir denge kurmak, modelin performansı ile adalet arasındaki ilişkiyi doğru yönetmek gereklidir.

Önyargılı yapay zekâ modellerini azaltmak için bahsedilen tekniklerin bir kombinasyonunu kullanmak önemlidir. Önyargı azaltma teknikleri, yapay zekâ ve makine öğrenmesi sistemlerinde adaletin sağlanması için kritik öneme sahiptir. Bu teknikler, veri setlerinde veya modelde bulunan önyargıları ortadan kaldırarak, daha adil ve eşitlikçi sonuçlar elde edilmesini sağlar. Ancak, bu tekniklerin uygulanması sırasında modelin doğruluğu ile adalet arasında bir denge kurmak önemlidir.

4.2. Yapay Zekâ Uygulamalarının Etik Açıdan Değerlendirilmesi

Yapay zekâ uygulamaları hayatımızda büyük ölçüde yer edinmektedir. Yapay zekâ sayesinde teknolojik gelişmeler hız kazanmıştır. Bu hızlı gelişim sürecinde etik

değerlerin göz ardı edilmemesi gerekmektedir. Bu teknolojilere toplumun ön yargılı davranmayıp kolayca benimsemeleri için etik yapının oluşturulması gerekmektedir.

Yapay zekâ uygulamaları yeni gelişen teknolojilerdir. Bu teknolojilerin gelişimiyle beraber yapay zekâ ve etik ayrılmaz ikili haline gelmiştir. Gün geçtikçe kullanımı artan uygulamalar etik tartışmalarına yol açmıştır.

DALL-E, verilen komutlara göre yaratıcı görseller üretebilen bir yapay zekâ programıdır. ChatGPT ise anlaşılır cevaplar vererek sohbet tabanlı etkileşimler sunar. Deepfake teknolojisi, kişilerin yüzlerini ve seslerini değiştirme yeteneğine sahiptir. Midjourney ise metinsel açıklamalardan görseller oluşturur. Bu yapay zekâ araçları, hem sosyal hem de profesyonel dünyada geniş bir etki alanına sahiptir ve hemen hemen her mesleğe ve yeteneğe dokunabilir. AI araçlarının kullanıcı dostu ara yüzleri ve yetenekleri, makinelerle iletişim ve etkileşim biçimimizi değiştirirken, bu etkileşimlerin nasıl gerçekleşmesi gerektiğine dair soruları da gündeme getirmektedir (Kahveci, 2023).

"Yapay zekâlı robotların tasarımında etik konusuna yönelik faydacı yaklaşım veya Kantçı yaklaşım gibi farklı yaklaşımlar vardır. Faydacı yaklaşım yapılan davranışın sonucunun insanları ne kadar mutlu ettiğine odaklanır. Kantçı yaklaşım ise davranışın sonucuna değil davranış esnasındaki düşünceye odaklanır. Bu yaklaşımlarla birlikte genel kabul görmüş beş etik birim gerekliliğinden söz edilmektedir. Bunlar (Usta ve ark., 2018):

- Tutarlılık (sistemde çelişen kurallar olmaması),
- Tamlık (çıktı verme zorunluluğu),
- Gerçek Zamanlılık (problemin kısa sürede çözülme zorunluğu),
- Şeffaflık (kuralların okunabilir ve anlaşılabilir olması),
- Etik sezgiye uygunluk (kodlanan kuralların etik kurallara uygun olması) (Doğan ve ark., 2022)."

Yapay zekâ uygulamaları hafızasında büyük veri setlerini barındırır. Bu veri setleri oluşturulurken genellikle kullanıcı onayı alınmaz. Toplanan bu veriler başka kullanıcılarla paylaşılabilmektedir. İzinsiz kullanılan veriler etik değerlere aykırılık göstermektedir.

Örneğin; bir sağlık destek chatbotu, kullanıcıların tıbbi geçmişini toplar ve bu bilgileri şifrelemeden saklar. Kullanıcılar, bilgilerin nasıl korunduğu hakkında bilgiye sahip değildir. Kişisel sağlık bilgileri, uygun şekilde korunmadığında gizlilik ihlali ve veri güvenliği sorunları yaşanabilir. Chatbot, kullanıcı verilerini şifrelemeli ve yalnızca

gerekli izinlere sahip kişilere erişim sağlamalıdır. Kullanıcılara, verilerinin nasıl korunacağı ve saklanacağı hakkında net bilgi verilmelidir. Bir sağlık destek chatbotu, kullanıcının kişisel sağlık bilgilerini toplarken bu bilgileri şifrelemeli ve sadece gerekli izinlere sahip kişilerin erişimine açık olmalıdır. Ayrıca, kullanıcıya verilerini nasıl koruyacağına dair bilgi ve veri silme talebi hakkında seçenek sunmalıdır.

Örneğin; bir müşteri hizmetleri chatbotu, kişisel bilgileri toplarken siber saldırılara karşı korunmalı ve veri sızıntılarını önlemek için güncel güvenlik önlemleri ile korunmalıdır. Ayrıca, chatbotlar kötüye kullanım veya dolandırıcılık gibi kötü niyetli aktivitelerden korunmalıdır.

Bu uygulamalar çok fazla veri topladığından dolayı hatalı veri de biriktirebiliyor. Kullanıcılara bazen hatalı veya eksik bilgi sunabilmektedir. Kullanıcıların bu hatalı bilgileri kullanıp olumsuz sonuçlarla karşılaştığında bu hatanın kime ait olacağı belirsizdir. Yapay zekâ uygulamalarını tasarlayan bireyler bu sistemlerin davranışlarından ve ürettiği sonuçlardan sorumludurlar. Bu uygulamaları tasarlayan bireyler ön yargılı algoritma oluşturmamak için sürekli denetim yapmalıdır. Yapay zekâ uygulamalarını kullanan bireyler, bu uygulamaları kullanma ve sonucun nasıl yönetileceği konusunda sorumluluk taşır. Bu sorumluluk yapay zekâ uygulamalarının etik dışı kullanımını önlemek içindir.

Örneğin; bir e-ticaret chatbotu, müşteri hizmetleri desteği sunmaktadır. Kullanıcı, chatbot aracılığıyla ürün iadesi yapmak istiyor. Chatbot, kullanıcının bir yapay zekâ tarafından yanıtlandığını belirtmiyor. Kullanıcı, bir insanla konuştuğunu düşünürken gerçek durumun farkında olmayabilir, bu da güven ve şeffaflık sorunlarına yol açabilir. Chatbot, etkileşimin başında veya kullanıcı sorularına yanıt verirken, kendisinin bir yapay zekâ olduğunu açıkça belirtmelidir. Ayrıca, chatbotun sınırlamaları hakkında bilgi vermelidir. Bir müşteri hizmetleri chatbotu, kullanıcılara bir yapay zekâ programı olduğunu ve etkileşimde bulundukları sistemin bir insan değil, bir yazılım olduğunu açıkça belirtmelidir. Ayrıca, bu chatbot, verilerin nasıl toplandığını ve kullanıldığını kullanıcılarla paylaşmalıdır.

Örneğin; bir finansal danışman chatbotu, yanlış yatırım tavsiyeleri veriyor ve kullanıcılar maddi zarara uğruyor. Ancak, chatbotun yanlış bilgi verdiğinde sorumluluğun kimde olduğu belirsiz. Kullanıcılar, chatbotun verdiği yanlış tavsiyeler yüzünden maddi kayba uğrayabilir. Sorumluluk ve hesap verebilirlik eksikliği yaşanabilir. Chatbotun, verdiği tavsiyeler hakkında bilgi sahibi olan bir destek ekibi veya sorumlu bir kişi tarafından denetlenmesi gerekmektedir. Ayrıca, yanlış bilgi

verdiğinde nasıl düzeltilmesi gerektiğine dair bir süreç belirlenmelidir. Bir finansal danışman chatbotu yanlış bilgi verirse veya yanıltıcı tavsiyelerde bulunursa, bu durumdan sorumlu olan kişiler veya kuruluşlar açıkça belirlenmelidir. Ayrıca, chatbotun yanlış bilgilendirme yaptığı durumlarda kullanıcıların nasıl geri bildirimde bulunabilecekleri ve sorunların nasıl düzeltileceği hakkında bilgiler sağlanmalıdır.

Yapay zekâ uygulamaları eğitimde ve öğretim sürecinde büyük kolaylıklar sağlamaktadır. Akademik çalışmalarda büyük miktarda veri toplayıp işleyebilmektedir. Toplanan bu veriler gizlilik ve mahremiyet konusunda tehlike yaratabilmektedir. Bu bilgilerin bilerek veya yanlışlıkla kötüye kullanılması ciddi sorunlar doğurabilmektedir. Yapay zekâ uygulamaları akademik değerlendirme süreçlerinde kullanılabilmektedir. Hafızasında belirli bir gruba karşı eğitilen veri setini barındırıyorsa fırsat eşitsizliğine yol açabilmektedir. Bazı öğrenci ve araştırmacılar hile yaparak GPT algoritmalarını kullanmaktadır. Bu tür kullanımlar akademik değerlere zarar vererek hileli öğrenmeye yol açabilmektedir ve bu da dürüstlük ilkesiyle çelişmektedir. Yapay zekâ uygulamaları tarafından önerilen metodolojiler etik kuralların göz ardı edilmesine sebep olabilmektedir. Bu araştırma sürecini hızlandırsa bile ciddi etik ihlallere yol açabilmektedir. Eğitimde insanın yerini yapay zekâ uygulamalarının almasıyla birlikte eğitimcilerin rolü zayıflamaktadır. Bu durum, öğrencilerin bireysel gelişimini olumsuz etkilemektedir. Yapay zekâ uygulamaları tarafından birçok içerik üretilmekte ve bunlar akademide kullanılmaktadır. Üretilen bu içeriklerin kime ait olduğu sorusu ciddi bir etik sorun teşkil eder. Kaynakçası olmayan bilgiler etik dışı unsurların doğmasına sebebiyet vermektedir. Ayrıca akademide bu uygulamaların kullanımı dürüstlük ilkesiyle çelişmektedir.

Örneğin; bir eğitim chatbotu, yanlış veya yanıltıcı bilgiler veriyor ve bu bilgiler öğrencilerin öğrenme sürecini etkiliyor. Yanlış bilgi vermek, kullanıcıların eğitim süreçlerini olumsuz etkileyebilir ve bilgi kalitesini sorgulanabilir hale getirebilir. Chatbot, doğru ve güncel kaynaklardan bilgi sağlamalı ve verilen bilgilerin doğruluğunu sürekli olarak kontrol etmelidir. Yanlış bilgi verildiğinde, bu hataların nasıl düzeltileceği hakkında bir süreç belirlenmelidir. Bir eğitim chatbotu, öğrencilerin geri bildirimlerine dayalı olarak sürekli güncellenmeli ve eğitim materyallerindeki değişikliklere uyum sağlamalıdır. Ayrıca, chatbotun eğitilmesinde kullanılan veri setlerinin güncel ve tarafsız olması gerekir.

Örneğin; bir psikolojik destek chatbotu, kullanıcılar üzgün olduğunda empati göstermiyor ve standart yanıtlar veriyor. Empati eksikliği, kullanıcıların duygusal olarak

daha kötü hissetmesine neden olabilir. Chatbotun destekleyici ve anlayışlı olması önemlidir. Chatbot, duygusal duruma uygun yanıtlar vermeli ve empati göstermelidir. Kullanıcıların duygusal ihtiyaçlarına yönelik özel yanıtlar ve destek sunmalıdır. Bir psikolojik destek chatbotu, kullanıcıların duygusal durumlarına duyarlı olmalı ve empati göstererek destek sağlamalıdır. Bu chatbot, kullanıcıların duygusal ihtiyaçlarına uygun yanıtlar vererek, onların kendilerini anlaşılmış ve desteklenmiş hissetmelerini sağlamalıdır.

Yapay zekâ uygulamalarının verdiği cevaplar anlaşılabilirdir. Toplumun her kesimi bu uygulamaları kullanabilmektedir. Bu uygulamalar zaman ve mekân fark etmeden istenilen zamanda uygulamalara erişim sağlanabilmektedir. Bu yüzden de etik değerlere uygunluk göstermektedir.

Yapay zekâ uygulamaları kolaylık sağlar. Fakat beraberinde ciddi sorumluluk da getirir. Henüz insanlar için bile etik değerler tam anlamıyla sağlanamamışken yapay zeka uygulamalarına bunu entegre etmek zor olacaktır. Uygulamalar bazı konularda etik değerlere uygunluk gösterse bile göstermediği alanlar da mevcuttur. Nihayetinde bu uygulamalar insan gibi bir bilince sahip değildir. Bu uygulamaların etik değerlere uyumu tasarımındaki kodlarla belirlenmektedir.

Örneğin; bir işe alım chatbotu, adayların cinsiyetine göre farklı yanıtlar veriyor ve bu durum işe alım sürecini etkiliyor. Cinsiyet, yaş veya ırk gibi kişisel özelliklere dayalı ayrımcılık yapmak, adaletsiz ve önyargılı bir değerlendirme sürecine yol açar. Chatbot, tamamen adil ve tarafsız olmalı; değerlendirme sürecinde önyargılardan kaçınmalıdır. Algoritmalarda ve veri setlerinde adalet ilkelerine uygunluk sağlanmalıdır. Bir iş başvuru sürecinde kullanılan chatbot, adayları değerlendirirken ırk, cinsiyet veya yaş gibi önyargılardan etkilenmemelidir. Ayrıca, aynı niteliklere sahip adaylara eşit fırsatlar sunulmalı ve önyargıların etkisi minimize edilmelidir.

Bu yüzyılda, etik değerler ve ahlaki yaklaşımlarla besleme konusunda kapsamlı bir yaklaşıma ihtiyaç duyulmaktadır. Küreselleşmenin, maddeciliğin ve dünya-perestliğin yapısal etkilerinin daha adil ve uyumlu bir şekilde dengelenmesi önemlidir. Modern dini çalışmalar, sadece geleneksel kaynaklara dayanan otantik yaklaşımları değil, aynı zamanda yapay zekâ ve çağdaş öğretme-öğrenme süreçlerini de dikkate alacak yeni bir teorik temel talep etmektedir (Ali, 2012).

Daha etik uygulamalar oluşması için evrensel etik değerler oluşturulmalıdır. Uygulamaların etik değerlere uygunluk gösterebilmesi için kodlama yaparken tasarımcıların oluşturulan bu etik değerleri göz önünde bulundurulması gerekmektedir.

Ayrıca bu uygulamalar sürekli test edilerek etik değerlere uygunluğu kontrol edilmelidir.

4.3. Yapay Zekânın Ahlaki Statüsü

Yapay zekâ sistemlerinin gelişmesiyle bu sistemlerin ahlaki bir statüye sahip olup olmadıkları önem arz etmeye başlamıştır. Bu yüzden bu sorun günümüzde önemli tartışma konularından biri haline geldi.

Bazı filozoflar etiği akıl veya bilinç perspektifinden değerlendirirken, bazıları etiği duyu ve hissiyat açısından ele alır. Diğer düşünürler ise bu iki yaklaşımı birleştirir. Bu durum, ahlaki durumları belirleyen özelliklerin ve yetilerin ne olduğunu kesin olarak belirlemeyi zorlaştırmaktadır (Çelebi, 2017).

Yapay zekânın hatalı karar alması durumunda sorumluluğun kimde olacağı, bu sistemi geliştirebilmek için hangi etik kuralların oluşturulması gerektiği ve bu sistemin hangi haklara sahip olacağı gibi soruların cevapları bu sistemin ahlaki statüsünün belirlenmesinde önem arz etmektedir.

Ahlaki statü, bir varlığın ahlaki değerlere uygun davranıp davranılmadığını inceler. Fakat yapay zekâ sistemlerinde bilinç bulunmadığından dolayı ahlaki statüsünü olmadığı düşünülebilmektedir. Yapay zekâ sistemleri belirlenen kodların dışına çıkamadığından ahlaki statüyü algılayabilmek için insan müdahalesine ihtiyaç duymaktadır.

Bir robotun ahlaki olarak mutluluğu hedeflemesi mümkün olabilir. Bu çerçevede, bir yapay zeka sistemine yapay bir mutluluk entegrasyonu sağlanabilir (Searle, 2005). Bilim insanları, yarattıkları sistemlerle etkileşimde bulunabilirler; onlarla sohbet edebilir, eğlenebilir, dans edebilir ve hatta bu sistemler, bir profesyonel kemancı gibi keman çalabilir (Kaku, 2016). Mutluluk duygusal bir yeti olsa da, gelecekte bizden daha akıllı olacak robotların hiç ağlamayacağı öngörülmektedir (Kaku, 2016). Bir insana dokunduğunuzda vereceği tepkiyi veya hissedeceği duyguları bir yapay zekâ sisteminden beklemek mümkün değildir (Penrose, 1997). Profesör Jefferson'a göre, bir makinenin duygular ve düşünceler aracılığıyla bir şiir yazması veya bir konçerto bestelemesi, onu bir beyinle eşdeğer kılmaz; çünkü bu tür yaratımların anlamını da kavrayabilmesi gerekir (Turing, 2008). Bu nedenle, robotlara entegre edilen yapay mutluluk veya duygusal yetenekler, insanların doğal mutluluğu ve duygusal deneyimleriyle aynı olmayacaktır. İnsanlarda çeşitli mizaçlar, ruh halleri,

karakterler, alışkanlıklar ve duygusal durumlar bulunurken, robotlarda bu çeşitlilik mevcut değildir (Mattrie, 1980). Bu bakımdan insandan insana duygu durumları farklılık gösterdiğinden dolayı insan bir an mutluyken bir müddet sonra mutsuz olabilir. Oysa bir yapay zekâ sistemlerinde böyle bir durum söz konusu değildir.

Bir yapay zekâ sistemi, programlaması sırasında entegre edilen duygu durumuna göre hareket eder. Yani, sistem mutlu olarak programlanmışsa, bir süre sonra mutsuz hale gelmesi beklenmez (Çelebi ve ark., 2019).

Bu sistemlerin insana zarar verdiği durumlar ahlaki olarak sorgulanmasına yol açabilmektedir. Bu da insanlar için ahlaki sonuç doğurmasına sebep olmaktadır.

Yapay zekâ sistemleri bir bilince sahip olmadığından dolayı verilen yanlış kararlar sistemi tasarlayanların sorumluluğundadır. Bu yüzden ahlaki sorumluluğu da üreticiye yüklenmiş olur.

Yapay zekâ sistemleri sadece üreticinin belirlediği şekilde karar alabildiğinden dolayı insana ve topluma zarar verebilecek kararlar alması üreticinin sorumluluğunda olduğu düşünülmektedir. Fakat bu sistemler ahlaki değerler dışında kullanıldığında sorumluluk kullanıcıya ait olabilmektedir.

Bu sistemler bir bilince sahip olmadığından dolayı verdiği kararların insana zarar verebileceğini kestiremezler. Bu yüzden doğabilecek zararları minimuma indirmekte üreticinin sorumluluğundadır.

Yapay zekânın insan gibi haklarının olup olmayacağı konusunda tartışmalar devam etmektedir. Fakat bu sistemlerin insan gibi duygu ve bilinci olmadığından dolayı insana özgü hakların tanınması mümkün değildir.

Yapay zekâ sistemleri insan müdahalesiyle oluşmaktadır ve bu sistemler insanlara yardımcı olabilmek için tasarlanmaktadır. Fakat teknolojinin gelişmesiyle insana benzer özellikleri olabilir. Bu durumun sonucunda ahlaki statü kazanma durumu tekrardan tartışılabilir.

Yapay zekâ sistemlerinin kullanımı toplumsal açıdan adaletsizliğe sebebiyet veriyorsa bu ahlaki sorunlara yol açabilir. Bu sorunların çözülmesi için sorumlulukların belirlenmesi gerekmektedir.

Yapay zekâ sistemleri bilinç, irade ve duygulara sahip olmadığından dolayı ahlaki statüye de sahip değillerdir. Fakat ilerleyen teknolojiyle beraber bu özellikleri kazanmaları durumunda ahlaki statüye de sahip olmaları muhtemeldir. Ayrıca bu durumda insan gibi haklara da sahip olması gerektiği konusunda tartışmalara yol açabilir.

İnsanların sadece diğer insanlara karşı ahlaki sorumluluk taşıdığını düşünme eğilimleri, yalnızca robotlarla ilgili değildir. Geçmişte hayvanlar, çoğu zaman sadece yiyecek ve giyecek sağlayan basit varlıklar olarak görülüyordu ve bu nedenle istismar edilebiliyordu. Ancak, 20. yüzyılın ikinci yarısında hayvan hakları üzerine tartışmalar başladı. Bu değişim, hayvanların ontolojik statüsünden ziyade, bu canlılara yönelik düşünme ve tepki verme biçimimizdeki değişimi yansıtıyordu (Gunkel, 2020). Benzer şekilde, robotlara karşı sorumluluklarımız ve robot hakları konusundaki yaklaşımımız da, robotlara nasıl baktığımıza bağlı olarak şekillenecektir (Doğan, 2023).

Yapay zekânın ahlaki statüsüyle ilgili tartışmalar devam etmektedir. Fakat mevcut durum için ahlaki statüye sahip olmadığı görüşü ağır basmaktadır.

4.4. Yapay Zekâ Kullanımından Doğabilecek Sorunlar

Yapay zekâ sistemleri her geçen gün hayatımızda daha fazla yer edinmektedir. Hayatımızda birçok kolaylık sağlamasının yanında bazı dezavantajları da bulunmaktadır.

Bu sistemler çok fazla veriyle çalışabilmektedir. Bu kadar fazla veriyi toplayan sistemlerin toplama aşamasında gizlilik ihlalinde bulunması olası bir problemdir. Bu sistemlere birçok kişi erişim sağlayabildiğinden dolayı kötü niyetli insanların eline geçerek gizlilik ve güvenlik sorunları yaratabilmektedir. Bu sistemler izin verilmeden de veri toplayabilmekte ve bunu başka kişilerle paylaşabilmektedir. İzin veri toplandığından dolayı gizlilik ilkesi ihlal edilmiş olur.

Yapay zekâ sistemlerini geliştirmek insan kontrolündedir. Bu sistemler belirli bir grubu dışlayacak şekilde programlandığında o gruba karşı ön yargılı sonuçlar doğurur. Bu sistemler sadece belirli bir gruba hizmet edecek şekilde kodlandığında sosyal eşitsizlik ortaya çıkabilmektedir. Bu da adaletsizliğin derinleşmesine yol açabilir.

Yapay zekâ sistemlerinin gelişmesiyle birlikte insanın yapabileceği birçok işi makineler daha hızlı ve kolaylıkla yapabilmektedir. Makinelerin daha hızlı çalışabilmesi işletmeler için büyük avantaj sağlamaktadır. Bunun sonucunda insanın yerine makinelerin geçmesiyle işsizlik sorunu meydana gelmiş olur. Yeni sistemlere herkes uyum sağlayamayabilir. Bu sistemleri kullanamayan personellerin işten çıkarılması gibi sorunlar doğabilmektedir.

Bu sistemler sayesinde yeni askeri alanda yeni teknolojiler geliştirilebilmektedir. Bu teknolojinin kötüye kullanılması güvenlik açısından sorun yaratmaktadır. Bu

sitemlerin gelişmesiyle birlikte siber saldırıların da artma tehlikesiyle karşı karşıya kalmış durumdayız.

Bu sistemler sadece belirlenen kodlarla çalıştığından dolayı etik değerlere uymaması muhtemeldir. Bu da ciddi sorunlar doğurabilmektedir. Bu sistemlerin hızlı bir şekilde gelişmesinden dolayı yasal düzenlemeleri yapılmaması olasıdır. Bu durum yasal çerçevelerin eksikliğine yol açmaktadır.

Yapay zekâ sistemleriyle çok fazla insana ulaşım sağlanabilmektedir. Bu sistem aracılığıyla yapılabilecek yanlış haberler toplumsal sorunlara yol açabilmektedir. Bu sistemlerin gelişmesiyle birlikte insanlar çevresine daha az zaman ayırmaya başlamıştır. Bu durum insan ilişkilerinin bozulmasına neden olmaktadır.

Yapay zekâ doğal zekâdan daha karmaşık kararlar alabildiğinden dolayı alınan bazı kararları insanın zor anlayabilmesi muhtemeldir. Bu sistemler karar alırken insan müdahalesine ihtiyaç duymazlar. Bu süreçte arızalanmaları yanlış sonuçlar doğurabilir. Bu yüzden ciddi sorunlarla karşılaşılabilir.

Bu sistemler çalışırken çok fazla enerji tüketirler. Tüketilen enerji çevresel sorunlara yol açabilmektedir.

Yapay zekâ sistemlerinin gelişimi birçok alanda kolaylık sağlamıştır. Bu kolaylığa alışıp aşırı kullanım sağlayan bireylerin doğal kapasitelerinde zayıflama görülmesi olasıdır.

Yapay zekâ bazı durumlarda insan zekâsını aşabilmektedir. Bu durumda insanın bu sistemi kontrolü zorlaşır ve istenmeyen sonuçların doğmasına sebep olabilir.

Günümüzde yapay zekâ ve entegre edilmiş otonom araçlar geniş bir yelpazede hizmet vermekte ve bu durumun sağladığı faydalar yanında çeşitli sorunları da beraberinde getirmektedir. Bu sorunların ciddiyeti, tartışmalara yol açmaktadır. Örneğin, 1981 yılında Japonya'daki Kawasaki fabrikasında, 37 yaşındaki Kenji Urada, bakımını yaptığı robotu kapatmayı unuttuğunda, robotun kendiliğinden hareket ederek Urada'ya zarar vermesi ve ölümüne neden olması, bu olayın robotlar tarafından gerçekleştirilen ilk ölüm olarak kabul edilmesine neden olmuştur. Bu tür olaylar, "sorumluluk" tartışmalarını gündeme getirmektedir (Demirdağ, 2020).

Ayrıca, insan zekâsını taklit etmek amacıyla geliştirilen yapay zekâ sistemlerinin, günümüzde birçok alanda faaliyet gösterdiği ve dünya üzerinde önemli etkiler yarattığı görülmektedir. Yapay zekânın entegre olduğu otonom makineler, insan ve diğer canlılara kıyasla çok daha büyük bir güç kapasitesine sahiptir. Bu güç sadece

fiziksel değil, aynı zamanda sahip olduğu zekâ ile de büyüktür ve gelecekte daha da güçlenecektir (Demirdağ, 2020).

Yapay zekâ sistemlerinin kullanımında doğabilecek sorunlar, teknolojinin toplum üzerinde yaratabileceği olumsuz etkileri minimize etmek için dikkatli bir planlama, etik rehberlik ve etkili düzenlemeler gerektirir. Bu zorluklar, yapay zekânın potansiyelini tam anlamıyla gerçekleştirmek için dengeli ve sorumlu bir yaklaşımla ele alınmalıdır.

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu tez, yapay zekâ (YZ) uygulamalarını etik bağlamında değerlendirmeyi hedeflemektedir. Yapay zekâ, uygulamaları büyük avantajlar sağladığı halde doğru tasarlanmadığı veya kullanılmadığı durumlarda etik sorunlarına yol açabilmektedir. Çalışmada, yapay zekânın temel etik ilkelerine uygun tasarlanması ve kullanılması gerektiği vurgulanmıştır.

Araştırma, hızla gelişen yapay zekâ sistemi uygulamalarının karar süreçlerinde karşılaşılan temel etik sorunları ve bu sorunların çözümü için yapılan çalışmaları sunmaktadır. Hızla gelişen bu teknolojilerin insan zekâsıyla kıyası yapılmıştır. Bu uygulamaların avantaj ve dezavantajları açıklanmıştır. Özellikle, yapay zekânın karar alma süreçlerinde temel etik ilkelere uyması etik kullanım açısından büyük önem arz ettiği belirlenmiştir.

Tezin bulguları, yapay zekâ uygulamalarının etik problemlerini çözmek için evrensel etik çerçevesi belirlenmesi gerektiğini göstermektedir. Bu çerçeve, yapay zekânın temel etik ilkelerine uygun tasarım yapılmalıdır. Ayrıca, bu süreçlerin kontrolü için denetim mekanizmaları belirlenmesi gerektiği önerilmektedir.

Yapay zekâ uygulamaları, eğitildikleri veri setlerindeki önyargıları yansıtma eğilimindedir. Bu, cinsiyet, ırk, yaş, din ve diğer kimlik kategorilerine yönelik ayrımcılığı pekiştirebilir. Özellikle azınlık gruplara karşı taraflı sonuçlar üretme riski, toplumsal eşitlik ve adalet açısından ciddi sorunlar yaratmaktadır. Irkçılık, cinsiyet, yaş, din vb. birçok alanda ayrımcılığa yol açtığı örnekler ve veri setleriyle sunulmuştur. Bu örneklerde de görüldüğü üzere veri setlerinin dengeli oluşturulması kritik önem taşımaktadır. Önyargılı sonuçların ciddi sosyal etkileri olabileceği gibi işe alım ve kredi verme gibi hassas uygulamalarda çok önemli yere sahip olduğu anlaşılmıştır. Veriler yapay zekâ yazılımlarının kamuoyuna sunulmadan önce test edilmesinin ve demografik özelliklere karşı adalet, eşitlik ve önyargı açısından kontroller yapılmasının önemini vurgulamaktadır. Bu önyargılar, yapay zekânın adil ve eşitlikçi bir şekilde kullanılmasını engelleyebilir ve bu nedenle veri setlerinin çeşitliliği artırılmalı, algoritmalar önyargıları azaltacak şekilde geliştirilmelidir.

Yapay zekâ sistemleri büyük miktarda veri topladığından dolayı gizlilik ve mahremiyet konusunda ciddi riskler doğurmaktadır. Farkında olmadan paylaşılan kişisel

bilgiler kötüye kullanılabilmektedir. Bu sebepten dolayı veri toplama konusunda şeffaflık ilkesi göz ardı edilememelidir.

Yapay zekâ sistemlerinin verdiği kararların sorumluluğu etik kullanım açısından önem arz etmektedir. Bir yapay zekâ uygulaması yanlış bilgi yaydığında veya ayrımcı bir sonuç ürettiğinde, bu hataların sorumluluğunun belirlenmesi gerekmektedir. Bu nedenle, geliştiriciler, kullanıcılar ve kurumlar arasında net bir sorumluluk paylaşımı yapılmalıdır.

Yapay zekâlar, doğru, nazik, iyi huylu ve zeki olmaları gereken çocuklar gibi eğitilmelidir. Önemli kararlar alacaklarsa, akıllıca ve bilinçli hareket etmeleri gerekir. Vatandaşlar olarak, yapay zekâ programcılarının bu standartlara uyduğundan emin olmalıyız. Olası kazaların önlenmesi için işlerin doğru yapıldığından emin olmamız gerekmektedir (Efe, 2021).

Yeni teknolojiler genellikle olumlu sonuçlar üretmesi umuduyla geliştirilir ve yapay zekâ da insanlara yardımcı olmak ve dünyayı daha iyi bir yer yapmak amacıyla tasarlanmıştır. Ancak, bu sistemlerin beklenen faydaları sağlayabilmesi için etik kurallara uygun şekilde geliştirilmesi gerekmektedir. Günümüzde insanlar için bile genel bir etik yapının oluşturulamamış olması, yapay zekâ sistemleri için bu tür bir yapıyı oluşturmanın ne kadar zor olduğunu göstermektedir. Ayrıca, yapay zekânın gelişimiyle ortaya çıkan etik kaygıların çeşitliliği ve karmaşıklığı, gelişimini etkileyecektir. Neyse ki, hükümetler, geliştiriciler ve akademisyenler tarafından yapılan birçok çalışma, yapay zekâ etiği alanında umut verici gelişmeler sağlamaktadır (Turan, Turan ve ark., 2022).

Yapay zekânın bilimsel araştırmalarda kullanımı, büyük potansiyelleri yanı sıra önemli etik sorunları da beraberinde getirmektedir. Bu teknolojinin avantajlarından yararlanırken, veri gizliliği, önyargı, şeffaflık, hesap verebilirlik, yanlış bilgi üretimi ve bilimsel hazırcılık gibi etik endişeleri dikkate almak hayati öneme sahiptir. Yapay zekânın bilimsel araştırmalardaki rolü geliştikçe, bu etik sorunlara çözüm bulmak ve yapay zekâ sorumlu ve etik kullanımını sağlamak için sürekli çaba gösterilmesi gerekmektedir. Etik kurallar çerçevesinde, yapay zekâ araçlarının bilimsel araştırmalarda etkili ve doğru şekilde kullanılması, araştırmacılara birçok fayda sağlayabilir (Başaran ve ark., 2024).

Sonuç olarak yapay zekâda etik konuların önemi gün geçtikçe artmaktadır. Gelecekte nasıl bir dünyada yaşayacağımızı bugün planladığımız yapay zekâ teknolojileri belirleyecektir. Bu alanda yapılacak çalışmalarda etik ilkeler göz önünde

bulundurulmalıdır. Bu noktada, adalet, hesap verebilirlik ve şeffaflık ilkelerini merkeze alan bir bakış açısı, yapay zekânın potansiyel faydalarını maksimize ederken, potansiyel risklerini de en aza indirmeye yönelik kritik bir öneme sahiptir. Yapay zekâ teknolojilerinin etik açıdan sorumlu tasarlanması ve kullanılması, hem hakların korunması hem de adaletin sağlanması açısından hayati önem arz etmektedir. Bu çalışmaların sonuçlarının yapay zekâ alanındaki etik tartışmalara katkı sağlayacağı düşünülmektedir.

5.2 Öneriler

Yapay zekâ sistemlerinin hızla gelişimiyle etik sorunları gündeme gelmiştir. Bu sistemlerin topluma faydası en üst düzeye çıkarılıp zararları minimuma indirilmesi amaçlanmaktadır. Bu amaç doğrultusunda bazı önlemler alınabilmektedir.

Akıllı sistemlerin sunduğu faydalar, sahip oldukları güce kıyasla telafisi mümkün olmayan zararlara yol açabilecek potansiyeldedir. Bununla birlikte, bilim kurgu filmlerindeki yapay zekâ kaynaklı felaket senaryolarının etkisiyle, bazı gruplar yapay zekâ ve benzeri gelişmelere karşı çıkmakta ve bu teknolojilerin engellenmesi gerektiğini savunmaktadır. Bizim görüşümüze göre, bu teknolojilerin tamamen durdurulması yerine, kontrolünü kaybetmeyecek şekilde dikkatli müdahalelerde bulunulmalıdır (Demirdağ, 2020).

Algoritmalardan beklenen, ne kadar gelişmiş olurlarsa o kadar güvenli olmalarıdır. Daha akıllı olduklarında daha iyi olacakları konusunda şüpheler bulunsa da, kendilerini optimize eden ve büyük veriyle yüksek hızlarda çalışan algoritmaların izlenmesi ve anlaşılması giderek zorlaşmaktadır. Bu durum, algoritmaların doğru şekilde denetlenme yeteneğimizi azaltırken, hataların büyük felaketlere yol açma riskini artırmaktadır. Bu riskleri azaltmak için idari, politik, akademik ve teknik önlemlerin alınması gerekmektedir (Efe, 2021).

Yapay zekâ sistemlerinin etik kullanımı için evrensel düzeyde yasal düzenlemeler yapılmalıdır. Doğabilecek sorunlar bu düzenlemeler aracılığıyla en az düzeye indirilecek şekilde planlanmalıdır.

Yapay zekâ sistemlerinin üretimi, geliştirilmesi ve kullanımı sırasında karşılaşılacak etik dışı eylemlerin sorumluluğunun kime ait olduğu önceden belirlenmelidir. Böyle bir durumda sorumlu olan kişinin etik dışı eylemi ortadan kaldıracak şekilde düzeltme yapması sağlanmalıdır.

Bu sistemlerin belli aralıklarla etik değerlere uygunluğunun kontrolü sağlanmalıdır. Etik olmayan uygulamalarda düzenleme yapılmalıdır.

Şeffaf bir sistem oluşturulmasına özen gösterilmelidir. Sistemi kullananlar bu sistemlerin karar mekanizmaları ve çalışma koşulları hakkında bilgilendirilmelidir. Bu şekilde sisteme olan güven arttırılabilir.

Bazı durumlarda yapay zekâ insan zekâsından üstün olabilmektedir. Bundan dolayı verdiği kararları insanın anlayamaması doğaldır. Bu yüzden karar verirken insanın anlayabileceği düzeyde olması gerekmektedir.

Yapay zekâ sistemleri çok fazla veri toplama kapasitesine sahiptir. Bu veriler toplanırken sistemi kullanacak kullanıcılar konu hakkında bilgilendirilmeli ve toplanacak veriler için kullanıcılardan izin alınmalıdır.

Bu sistemlerde toplanan veriler yabancı şahıslara sunulmamalıdır. Yetkisi olmayan kişilerin kullanımına açılmamalıdır. Oluşabilecek siber saldırılar için gerekli önlemler alınmalıdır.

Bu zeki sistemler hayatın birçok alanında bulunmakta ve kolaylık sağlamaktadır. Bu yüzden eşit şekilde sisteme erişim sağlanmasına özen gösterilmelidir. Sisteme erişim konusunda dil, din, ırk ayrımı yapılmamalıdır.

Yapay zekâ sistemleri oluşturulurken belirli bir kesimi ayırmamasına özen gösterilmelidir. Sistemin ayrımcılık yapmaması herkese eşit yaklaşması için gerekli düzenlemeler yapılmalıdır. Bu düzenlemeler belirli aralıklarla kontrol edilmelidir.

Bu sistemlerin kullanımı konusunda herkese hitap eden eğitimler yapılmalıdır. Bu eğitimlerde doğru kullanımın nasıl olacağı hakkında bilgilendirme yapılmalıdır. Bu şekilde yanlış kullanımdan doğabilecek sorunların önüne geçilmiş olur.

Bu teknolojiler hakkında toplum bilinçlendirilmelidir. Bu şekilde etik olmayan uygulamaların kullanımının azaltılması sağlanabilir.

Sistemi kullanmaya başlamadan önce sistemin çalışma prensipleri ve oluşabilecek riskler araştırılmalıdır.

Bu sistemleri de insanların oluşturduğu ve bu yüzden hatalı, yanlış sonuçlarla karşılaşılabileceği unutulmamalıdır. Bu yüzden kullanım esnasında gerekli kontroller yapılmalıdır.

Bu sistemler geliştirildikten sonra insan müdahalesi olamadan çalışmaktadır. Fakat bazı durumlarda istenmeyen etik dışı eylemlere sebep olabilmektedir. Bu yüzden sistemler tamamen otomatik değil gerektiği zaman insan müdahalesine izin verecek şekilde tasarlanmalıdır.

Yapay zekâ sistemleri bağımsız karar alabilmektedir. Bu sistemler kritik durumda insan kontrolünde çalışmalıdır. Bu şekilde yaşanacak olumsuz durumlar en aza indirilmiş olur.

Yapay zekâ sistemleri geliştirilirken evrensellik göz önünde bulundurulmalıdır. Geliştirilen sistemler tüm toplumların değer yargılarına hitap etmelidir.

Geliştirilen sistemler toplumların her kesimine hitap etmelidir. Sadece belirli kesime hitap eden sistemler teknolojide eşitsizliğe yol açabilmektedir.

Bu sistemlerin geliştirilmesi ve kullanımı sırasında enerji tüketimine ihtiyaç duyulur. Bu ihtiyaç en az düzeye indirilerek sistem enerjiyi verimli kullanacak şekilde tasarlanmalıdır.

Yapay zekâ sistemleri oluşturulurken ve kullanılırken enerji kaynaklarına ihtiyaç duyar. Bu ihtiyaçların çevre dostu kaynaklardan kullanımı sağlanmalıdır.

Yapay zekâ sistemlerin etiğe uygunluğunu incelemek amacıyla etik kurulları oluşturulmalıdır. Bu kurullar sistemlerin etik değerlere uyup uymadığını kontrol etmelidir. Etik dışı uygulamalar için düzeltecek şekilde aksiyon alması sağlanmalıdır.

Yapay zekâ birçok alanı kapsamaktadır. Bu yüzden sistem geliştirilirken farklı disiplinlerden faydalanılmalıdır. Bu sayede sistemin her açıdan değerlendirilmesi kolay olur.

Yapay zekâ sistemlerindeki etik problemler incelenmelidir. Bu problemleri en aza indirmek veya çözmek için gerekli araştırmalar sağlanmalıdır. Bu şekilde olumsuz etkiler azaltılmış olur.

Yapay zekâ sistemleri geliştirilirken her aşaması ayrı ayrı etik açıdan incelenmelidir. Bu incelemeler sayesinde etik dışı uygulamaların oluşumu en az düzeye indirilmiş olur.

Yapay zekâ sistemleri oluşturulurken kullanıcıların etik dışı durumlarla karşılaştığında sisteme bilgi verebileceği şekilde tasarlanmalıdır. Bu şekilde sitemin etik dışı unsurları belirlemesinde kolaylık sağlanmış olur.

Bu sistemler geliştirilirken alınacak önlemler sayesinde daha eşit ve daha güvenilir sistemler oluşturulması sağlanmış olur. Bu şekilde doğabilecek sorunların önüne geçilmiş olur.

KAYNAKLAR

Adalı, E., 2017, Yapay zekâ, insanlaşan makineler içinde, haz. Mehmet Karaca, *İstanbul: İstanbul Teknik üniversitesi Vakfı Yayınları*, 8- 13.

Airwars Suriye ekibi, 2021, After six years of Russian airstrikes in Syria, still no accountability for civilian deaths. https://airwars.org/research/after-six-years-of-russian-airstrikes-in-syria-still-no-accountability-for-civilian-deaths/

Akın, H.L., Usta, K.Y., 2018, Servis robotları için etik birimi tasarımı. https://www.cmpe.boun.edu.boun.edu.tr/~akin/popers/ServisRobotlariIçinEtikBirimiTasarimi.

Al-Adely, S., Bashall, A., Kitchen, G.B., Locke, S., Moore, J., Wilson, A., 2021, Natural language processing in medicine: A review, *Trends in Anaesthseia and Critical Care*, 38, s. 4- 9.

Ali, A.Z., 2012, A philosophical approach to artificial intelligence and islamic values, *IIUM Engineering Journall*, 12(6), Special Issue in Science and Ethics.

Altıntop, M., 2023, Yapay zekâlı akıllı öğrenme teknolojileriyle akademik metin yazma: ChatGPT örneği, *Süleyman Demiral Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 846), 186- 211.

Andrew, A.M., 1991, Artificial intellegence. *Boston: Addison-Wesley Company.*

Angwin, J., Kirchner, L., Larson, J. Ve Mattu, S., 2016, How we analyzed the COMPAS recidivism Algorithm. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Asker, Ö.F., Bekiroğlu, N., Eriç, A., Özgür, E.G., 2024, ChatGPT ve sağlık bilimlerinde kullanımı, *Celal Bayar Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, 11(1), 176-182.

Atalay, M. Ve Çelik, E., 2017, Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları, Artificial intelligence and machine learning applications in big data analysis, *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155- 172.

Aydın, Ö., Karaslan, E., 2023, Is ChatGPT leading generative AI? What is beyond expectations?, *SSRN Electronics Journal*.

Barrat, J., 2020, yapay zekâ ve insanlık çağının sonu son icadımız, *İstanbul: Pegasus Yayınları*.

Başaran, R., Özenç, Y.Y., 2024, Bilimsel araştırma sürecinde yapay zekâ araçlarının kullanımı, *Uluslarası Eğitimde Mükemmellik Arayışı Dergisi (UEMAD)*, 4(1), 35-53.

Biswas, S., 2023, ChatGPT and the future of medical writing, *Radiologicaal Society of North America (RSNA)*.

Charniak, E., & mcDermot, D., 1958, Introduction to Artificial Intelligence. *Boston: Addison-wesley Company*.

Civilian Deaths by US President in Somalia. https://airwars.org/research/civilian-deaths-by-us-president-in-somalia/

Clark, A., Fox, C., Lappin, S., 2012, The handbook of computational linguisticsand natural languge processing, *John Wiley & Sons*, pp 1- 2.

Copeland, V., 1993, Artificial Intelligence: A Philosophical, *Blackwell: Oxford*.

Çakıcı A.C. ve Doğan,S., 2022, Yapay zekâlı hizmet robotlarına yönelik etik hususlar, *Güncel Turizm Araştırmaları Dergisi*, 6(1), 162-176.

Çelebi V., İnal A., 2019, Yapay zekâ bağlamında etik problemi, *Uluslararası Sosyal Araştırmalar Dergisi*, c.12, s.66. https:// dx.doi.org/10.17719/jisr.2019.3614

Çelebi, Ö.F., 2017, Yapay zekâ ve etik, *İstanbul Medeniyet üniversitesi Sosyal Bilimler Enstitüsü*, s. 1-20.

Çelik, K., 2014, Çoklu zekâ ve disiplinler arası yaklaşım temelli fen ve teknoloji dersi ve uygulamalarına ilişkin öğretmen görüşleri, *ESOGÜ Eğitim Bilimleri Enstitüsü*.

Çelik, M., 2024, Dijital halkla ilişkiler açısından e-ticaret web sitelerinde yapay zekâ destekli chatbot uygulamalarının incelenmesi, *Üsküdar Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Anabilim Dalı Halkla İlişkiler ve Reklamcılık Tezsiz Yüksek Lisans Proje Çalışması*.

Çiftçi, B. S., Şahin Başfırıncı, Ç., 2020, Yapay zekâ konusunun toplumsal cinsiyet kapsamında incelenmesi: Mesleklere yönelik bir araştırma, *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, 29(4), 183-203.

Çubuk, E.B.S., 2024, Veri kaynağı olarak sohbet robotları: Kamu yönetimi disiplininde örnek bir araştırma, *International Artificial Intelligence and Data Science Congress, aydın Adnan Menderes Üniversitesi Söke İşletme Fakültesi*.

Darı, A.B., Koçyiğit, A., 2023, Yapay zekâ iletişiminde ChatGPT: İnsanlaşan dijitalleşmenin geleceği, *Strateji ve Sosyal Araştırmalar Dergisi*, 7(2), 427- 438.

Daş, R., Polat, B., Tuna, G., 2019, Derin öğrenme ile resim ve videolarda nesnelerin tanınması ve takibi, *Fırat üniversitesi Mühendislik Bilimleri Dergisi*, 31(2), 571-581.

Delibas, A., 2008, Doğal dil işleme ile Türkçe yazım hatlarının denetlenmesi, *Doctoral Dissertation, Fen Bilimleri Enstitüsü*.

Demir, O., 2019, Sürdürülebilir kalkınma için yapay zekâ ve gelecek, Ed. GoncaTelli, *İstanbul: Doğu Kitap Evi*, 44- 63.

Demiral, Ö., Doğan, S., Kılıç, S., 2007, Kurumların başarısında duygusal zekânın rolü ve önemi, *Yönetim ve Ekonomi Dergisi*, 14(1), 209- 230.

Demirdağ, M., 2020, Yapay zekâ ve hukuk: Akıllı sistemler, hukuk uygulamalarını nasıl etkiliyor?, *Evrim Ağacı Akademi: Yapay Zekâ Uygulamaları Yazı Dizisi*. doi:10.47023/ea.bilim.8212. https://evrimağaci.org Is/s8212.

Demirel, Ö., 1999, Plandan değerlendirmeye öğretme sanatı, *Ankara: Pegem A Yayıncılık*.

Doğan, M., 2023, Sinemada yapay zekâ: Robotlarda bilinç, duygular ve etik, *Kültür ve İletişim*, 26(52), 318- 343.

Efe, A., 2021, Yapay zekâ risklerinin etik yönünden değerlendirilmesi, *Bilgi ve İletişim Teknolojileri Dergisi*, 3(1), 1-24.

Eryılmaz, H.E., 2023, Yapay zekâ çağında kişisel veri mahremiyeti, *UMAY Sanat ve Sosyal Bilimler Dergisi*,1(2), 6- 25.

E-vam, 2020, Yıllık iş kayıtları çerçevesi mikro veri seti. https://evam.tuik.gov.tr/dataset/detail/30

Eyibaş, Y., Topakkaya, A., 2019, Yapay zekâ ve etik ilişkisi, *Felsefe Dünyası*, (70), 81-99.

Firat, M., 2020, December, Öğrenci destek servislerinde doğal dil işleme: GPT- 3 örneği, *In International Conference of Strategic Research in Social Science and Education*, pp. 532- 536.

Gökalp, Ö.M., 2022, Makine öğrenmesi, *Gazi Üniversitesi, Gazi Bilişim Enstitüsü, Adli Bilişim Bölümü*.

Göktürk, I., 2023, Kültürel zekâ: Zekanın kültürel olarak kavramsallaştırılması ya da kültürel bağlam içinde zekâ, *Elektronik Sosyal Bilimler Dergisi*, 221881, 1939-1957.

Gunkel, D.J., 2020, How to survive a robot invasion rights, Rsponsibility and Al, *New York: Routledge*.

Güner, S.p., 2023, Çevirmen- bilgisayar etkileşiminin kilit bileşeni: Doğal dil işleme, *Karamanoğlu Mehmetbey Üniversitesi Uluslarası Filoloji ve Çeviribilim Dergisi*, 5(1), 56- 79.

Halidi G., 2022, Yapay zekâ etiği tartışmaları için bazı tarihsel- kavramsal önbilgiler, *Türkiye Biyoetik Dergisi Vol9*, No.4, 155-163.

Harmon, p. ve King, D., 1958, Expert systems: Artificial intelligence in business, *New York: John Wilwy and Sons*.

Jasmine, J., 1996, Teaching with multiple intelligences, *Professional's Guide, Teacher Created Materials*, Inc., 6421 Industry Way, Westminter, CA 92683.

Jordan, M.I., Mitchell, T.M., 2015, Machine learning: Trends, *Perpectives, and procpects*, Science, 349(6245)., 255- 260.

KADEM, 2024, Artificial intelligence and women, Kadın ve adalet zirvesi. https://kadinveadaletzirvesi.org/en/.

Kagan, M. Ve Kagan, S., 1998, Multiple intelligences, *The Complete MI Book*, San Clemente, CA: Kagan Cooperative Learning.

Kahveci, S., 2023, İnsan-yapay zekâ iletişiminde yeni bir paradigma: Prompt mühendisliği1.

Kaku, M., 2016, Geleceğin fiziği, (Çev. Yasemin Saraç Oymak ve Hüseyin Oymak), *Ankara: ODTÜ Gelişme Vakfı Yayıncılık*.

Kanal, S. ve Kanal, Y., 2023, Etkin öğrenme ve çoklu zekâ kuramı üzerine bir çalışma , *Socrates Journal of Interdisciplinary Social Studies*, 9(26), 1- 10. https://doi.org/10.5129/socrates26-220

Karakoç Keskin, E., 2023, Yapay zekâ sohbet robotu ChatGPT ve Türkiye internet gündeminde oluşturduğu temalar, *Yeni Medya Elektronik Dergisi*, 7(2), 114- 131.

Kırık, A.M., Özkoçak, V., 2023, Medya ve iletişim bağlamında yapay zekâ tarihi ve teknolojisi: ChatGPT ve Deepfake ile gelen dijital dönüşüm, *Karadeniz Uluslararası Bilimsel Dergi*, 589, 73-99. https:// doi.org/ 10.17498/ kdeniz.13.08471.

Korkmaz, Ö.Ü.A., 2022, Dijital dilin yolculuğu: ChatGPT ve yapay zekânın doğal dil işleme dünyasındaki rolü, *Bilisel International World Science and Research Congress*.

Kotil Öğretmen, Z., 2022, Kişisel verilerin korunması çerçevesinde yapay zekâ, *İstanbul: On İki Levha Yayıncılık*.

Kuru, E., 2001, Kinestetik zekâ ve beden eğitimi, *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 21(2).

Küçüker, M., 2023, Muhasebede yapay zekâ uygulamaları: ChatGPT'nin muhasebe sınavı, *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 33(2), 875- 888.

Küçüksille E.U., Turan G. ve Turan T., 2022, Yapay zekâ etiği: Toplum üzerine etkisi, *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(2):292-299.

Lecun, Y., Hinton, G., 2015, *Deep learning*, *Nature*, 521(75539)., 436- 444.

Lunenberg, F. ve Lunenberg, M., 2014, Applying multiple intelligences in the classroom: A fres look at teaching writing, *International Journal of Scholarly Academic Intellevtual Diversity*, 16(1), 1- 14.

Magid, J.M., 2020, Does your AI discriminate? The Conversation, Mayıs 15. https://theconversation.com/does-your-ai-discriminate-132847.

McCarthy, J., 2022, Ocak 9, What is artificial intelligence? Stanford University. http://jmc.stanford.edu/articles/whatisia.html.

Nabiyev, V.V., 2012, Yapay zekâ: insan- bilgisayar etkileşimi, *Baskı yeri: Seçkin Yayıncılık*.

Odak arge merkezi, 2021, Veri seti nedir, Yapay zekâ teknolojilerinde nasıl kullanılır?

Orhon, E.N., 2021, Süper zekâ: Yapay zekâ uygulamaları, tehlikeler ve stratejiler, *TRT Akademi*, 6(13), 942-947. https://doi.org/10.37679/trta.1002519

Özdal, M.A., 2023, Yapay zekâ ile oluşturulan eserlerin telif hakkı ve kişisel verilerin korunması, *Hakkâri Review*, 7(1), 90-110.

Öztemel, E., 2020, Yapay zekâ ve insanlığın geleceği, *Bilişim Teknolojileri ve İletişim: Birey ve Toplum Güvenliği*, 95- 112.

Öztürk, D., 2019, Yapay zekânın etik gerçekliği, *AUSBD*, 2(4), 47-59.

Penrose, R., 1997, Kralın yeni usu (Bilgisayar ve zeka9, (Çev. Tekin Dereli), Ankara: TÜBİTAK Popüler Bilim Kitapları.

Popov, E.V., (Ed), 1990, Yapay zekâ, uzman sistemler ve doğal dil işleme, *Moskova: Radio: Suyaz*, S. 461.

Reese, B., 2018, Yapay zekâ çağı, dördüncü çağ: Akıllı robotlar, bilinçli bilgisayarlar ve insanlığın geleceği, (Mihriban Doğan Çev.), Say Yayınları.

Russel, S.J., Norving, P., 2010, Artificial intelligence: A modern approach (3rd ed), Prentice Hall.

Searle, J.,2005, akıllar, beyinler ve bilim, (Çev. Kemal Bek), İstanbul: Soy Yayınları.

Sevim, N., Koç, 2021, A., Türkçe kelimeler temsillerinde cinsiyetçi ön yargının incelenmesi, *Investigation of Gender Bias in Turkish World Embeddings*.

Siber zorbalık veri seti https://www.kaggle.com/datasets/saurabhshahane/cyberbullying-dataset/data?select=twitter_sexism_parsed_dataset.csv

Süslü, A., 2019, Doğa ve insan bilimlerinde yapay zekâ uygulamaları, *Akademia Doğa ve İnsan Bilimleri Dergisi*, 5(1), 1- 10.

Şen, E., 2022, İllüstrasyon alanında yapay zekâ uygulamaları, *Abant Sosyal Bilimler Dergisi*, 22(3), 1320- 1332.

Terzi, R., 2021, Sağlık sektöründe açıklanabilir yapay zekâ, *Nobel Yayınları*.

Turing, A.:., 2008, bilgi işlem makineleri ve zekâ, Aklın gözü içinde, haz. Douglas R. Hofstadter ve Daniel C. Dennet, (Çev. Zehra Bayramoğlu), İstanbul: Havass Yayınları.

TÜBİTAK, 2001, Yapay zekâ, Bilim ve Teknik Dergisi, 38- 47.

Webtures, 2024, Türkiye'de yapay zekâ etik ilkeleri, Dijital Strateji ve Yapay Zekâ Danışmanlık Şirketi. https://tr.linkedin.com/pulse/t%C3%BCrkiyede-yapay-zekâ-etik-ilkeleri-webtures-7hi7f.

Yeşilkaya, N., 2022, Yapay zekâya dair etik sorunlar, *Şarkiyat*, 14(3), 948- 963.

YÖK, 2024, Yükseköğretim kurumları bilimsel araştırma ve uygun faaliyetlerinde üretken yapay zekâ kullanımına dair etik rehber.