



**ÜRETKEN YAPAY ZEKA TEKNOLOJİSİNDE
ÜRETİLEN SENTETİK SESLERİN
GERÇEK SESLER İLE
KARŞILAŞTIRILMASI**

Erdem ÖZKAN
Yüksek Lisans Tezi
Danışman: Doç. Dr. Seyhan CANYAKAN
Haziran, 2024
Afyonkarahisar

T.C.
AFYON KOCATEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
MÜZİK ANASANAT DALI
YÜKSEK LİSANS TEZİ

ÜRETKEN YAPAY ZEKA TEKNOLOJİSİNDE ÜRETİLEN
SENTETİK SESLERİN GERÇEK SESLER İLE
KARŞILAŞTIRILMASI

Hazırlayan
Erdem ÖZKAN

Danışman
Doç. Dr. Seyhan CANYAKAN

AFYONKARAHİSAR 2024

ETİK VE BİLİMSEL İLKELER SORUMLULUK BEYANI

Yüksek Lisans Tezi olarak sunduğum “**Üretken Yapay Zeka Teknolojisinde Üretilen Sentetik Seslerin Gerçek Sesler İle Karşılaştırılması**” adlı çalışmanın, tarafımdan bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin Kaynakça’da gösterilen eserlerden oluştuğunu, bunlara atıf yapılarak yararlanmış olduğumu belirtir ve bunu onurumla doğrularım.

23/06/2024

İmza

Erdem ÖZKAN



T.C.
AFYON KOCATEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

ENSTİTÜ ONAYI

Öğrencinin	Adı- Soyadı	Erdem ÖZKAN
	Numarası	2200651512
	Anabilim Dalı	Müzik
	Programı	Müzik
	Program Düzeyi	☒ Yüksek Lisans ☐ Doktora ☐ Sanatta Yeterlik
Tezin Başlığı		Üretken Yapay Zeka Teknolojisinde Üretilen Sentetik Seslerin Gerçek Sesler İle Karşılaştırılması
Tez Savunma Sınav Tarihi		23.06.2024
Tez Savunma Sınav Saati		14.00

Yukarıda bilgileri verilen öğrenciye ait tez, Afyon Kocatepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca jüri üyeleri tarafından değerlendirilerek ☒oy birliği – ☐oy çokluğu ile kabul edilmiştir.

Prof. Dr. Hacı İbrahim DELİCE
MÜDÜR

Bu tez, Enstitü Müdürlüğünce kontrol edilerek, elektronik imza kullanılarak onaylanmıştır.

ÖZET

ÜRETKEN YAPAY ZEKA TEKNOLOJİSİNDE ÜRETİLEN SENTETİK SESLERİN GERÇEK SESLER İLE KARŞILAŞTIRILMASI

Erdem ÖZKAN

AFYON KOCATEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
MÜZİK ANASANAT DALI

Haziran, 2024

Danışman: Doç. Dr. Seyhan CANYAKAN

Bu araştırma, yapay zeka teknolojisi kullanılarak üretilen sentetik seslerin, gerçek insan sesleriyle olan benzerlikleri ve farklılıklarını belirlemeyi amaçlamaktadır. Araştırmada, betimsel ve karşılaştırmalı bir metodoloji kullanarak, sentetik ve gerçek seslerin algılanış biçimleri, doğallıkları, ifade yetenekleri ve güvenilirlik algıları incelenmektedir. Katılımcılar, yapay zeka teknolojisi ile üretilmiş sentetik sesler ve gerçek insan sesleri arasındaki farkları değerlendirmek için dinleme testlerine ve yarı yapılandırılmış görüşmelere katılmışlardır. Araştırma, sentetik seslerin kullanıcı deneyimi üzerindeki etkilerini ve sosyal kabul düzeylerini de ele almaktadır. Bulgular, teknolojinin ses üretiminde sağladığı avantajlar ile duygusal ifade ve doğal ses kalitesi gibi alanlarda insan sesinin hala daha üstün olduğunu göstermektedir. Araştırma, sentetik ses teknolojilerinin gelecekteki gelişimine ışık tutmayı hedeflemekte ve bu seslerin daha geniş bir kabul ve uygulama alanı bulmasında yardımcı olabilecek bilimsel veriler sunmaktadır.

Anahtar Kelimeler: Yapay zeka, sentetik sesler, gerçek sesler, ses teknolojileri, kullanıcı deneyimi, sosyal kabul.

ABSTRACT

COMPARISON OF SYNTHETIC VOICES PRODUCED IN GENERATIVE ARTIFICIAL INTELLIGENCE TECHNOLOGY WITH REAL VOICES

Erdem ÖZKAN

**AFYON KOCATEPE UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF MUSIC**

June, 2024

Advisor: Assoc. Doç. Dr. Seyhan CANYAKAN

This research aims to identify the similarities and differences between synthetic voices produced using artificial intelligence technology and real human voices. Employing descriptive and comparative methodologies, the study examines the perceptions, naturalness, expressiveness, and reliability perceptions of synthetic and real voices. Participants have engaged in listening tests and semi-structured interviews to assess the differences between AI-generated synthetic voices and real human voices. The research also addresses the impact of synthetic voices on user experience and their social acceptance levels. The findings indicate that while technology provides advantages in sound production, human voices still surpass in areas such as emotional expression and natural sound quality. This study aims to illuminate the future development of synthetic voice technologies and provide scientific data that may help these voices gain broader acceptance and application.

Keywords: Artificial intelligence, synthetic voices, real voices, sound technologies, user experience, social acceptance.

ÖN SÖZ

Yüksek lisans eğitimim ve tezim sürecinde ilgi ve desteği ile yanımda olan, engin bilgi ve tecrübesinden yararlanma olanağı sağlayan, bu tezin hazırlanıp takdim edilmesinde tecrübe ve katkılarını esirgemeyen tez danışmanım Doç.Dr.Seyhan CANYAKAN 'a, tüm akademik hayatım boyunca maddi manevi her şekilde bana destek olan rahmetli babam İsmail ÖZKAN, annem Reyhan ÖZKAN ve abim Doç. Dr. Erkan ÖZKAN'a sonsuz sevgi, saygı ve teşekkürlerimi sunarım.Yüksek lisans eğitimi gibi zorlu bir süreçte beni her şekilde anlayışla karşılayan, desteğini hep yanımda hissettiğim sevgili eşim Dr. Merve ÖZKAN 'a sonsuz teşekkürlerimi sunarım..

Erdem ÖZKAN
2024, Afyonkarahisar



İÇİNDEKİLER

	<u>Sayfa</u>
ETİK VE BİLİMSEL İLKELER SORUMLULUK BEYANI	ii
ENSTİTÜ ONAYI	iii
ÖZET	iv
ABSTRACT	v
ÖN SÖZ	vi
İÇİNDEKİLER.....	vii
TABLOLAR LİSTESİ	ix
ŞEKİLLER LİSTESİ	x
GİRİŞ.....	1

BİRİNCİ BÖLÜM

YAPAY ZEKA VE SENTETİK SES ÜRETİMİ

1. YAPAY ZEKA.....	5
1.1. YAPAY ZEKA (AI) KAVRAMI.....	5
1.2. YAPAY ZEKA UYGULAMA ALANLARI.....	8
1.2.1. Makine Öğrenimi.....	8
1.2.2. Derin Öğrenme	11
1.3. YAPAY ZEKA ÜZERİNE ETİK TARTIŞMALAR	13
1.3.1. Bağımsızlık ve Sorumluluk	14
1.3.2. Adalet, Önyargı ve Yorumlanabilirlik	16
1.3.3. Tehlikeler ve Riskler	19
2. SENTETİK SES ÜRETİMİ.....	20
2.1. KONUŞMA SENTEZİ (TEXT-TO-SPEECH-TTS)	21
2.2. MÜZİK VE SES ÜRETİMİ	23
2.2.1. Veri Analizi, Ritim ve Melodi Oluşturma	24
2.2.2. Armoni, Enstrümantasyon ve Duygusal İçerik	25
2.3. SES EFEKTLERİ VE TASARIMI	26
2.3.1. Doğal Ses Taklidi ve Hayali Ses Efektleri	27
2.3.2. Ses Ortamı Oluşturma ve Ses İyileştirme	28
3. GELİŞMİŞ SES ÜRETİM MODELLERİ.....	29
3.1. WAVENET	30
3.2. MELNET	31
4. SES KLONLAMA TEKNOLOJİLERİ	32
4.1. VERİ TOPLAMA VE YÜKSEK KALİTE ÜRETİMİ.....	33
4.2. HOPARLÖR KODLAMASI VE UYARLAMASI.....	34
5. MAKİNE ÖĞRENİMİ VE SES ÜRETİMİ	35
5.1. REKÜRRENSİNİR AĞLARI (RNN) VE UZUN-KISA VADELİ BELLEK (LSTM)	36
5.2. EĞİTİM SÜREÇLERİ VE OPTİMİZASYON	37
5.3. VERİ SETİ HAZIRLIĞI VE ETİKETLEME	38
5.4. OBJEKTİF METRİKLER VE SES KLONLAMANIN UYGULAMA ALANLARI	39
6. YAPAY ZEKA VE SES İŞLEME.....	40
6.1. SES DALGALARININ TEMEL PRENSİPLERİ VE İŞLEME TERİMLERİ	40
6.2. GENLİK, FREKANS VE FAZ ANALİZLERİ	41

6.3. NYQUIST TEOREMİ VE UYGULAMALARI.....	42
---	----

İKİNCİ BÖLÜM

İLGİLİ ARAŞTIRMALAR

1. MAKALELER.....	47
2. BİLDİRİLER.....	51
3. KİTAPLAR.....	54
4. KİTAP İÇİ BÖLÜMLER.....	55

ÜÇÜNCÜ BÖLÜM

ÜRETKEN YAPAY ZEKA TEKNOLOJİSİNDE ÜRETİLEN SENTETİK SESLERİN GERÇEK SESLER İLE KARŞILAŞTIRILMASI

1. ARAŞTIRMANIN PROBLEM CÜMLESİ	57
2. ARAŞTIRMANIN AMACI	57
3. ARAŞTIRMANIN ÖNEMİ	58
4. ARAŞTIRMANIN SAYILTILARI.....	58
5. ARAŞTIRMANIN SINIRLILIKLARI	59
6. YÖNTEM	59
6.1. ARAŞTIRMA YÖNTEMİ	59
6.2. ÇALIŞMA GRUBU	60
6.3. VERİLERİN TOPLAMA ARAÇLARI	61
6.4. VERİLERİN TOPLANMASI VE ÇÖZÜMLENMESİ	61
7. ARAŞTIRMANIN BULGULARI VE YORUMLAR	65
7.1. BİRİNCİ ALT PROBLEME YÖNELİK BULGULAR.....	65
7.2. İKİNCİ ALT PROBLEME YÖNELİK BULGULAR	72
7.3. ÜÇÜNCÜ ALT PROBLEME YÖNELİK BULGULAR.....	77
7.4. DÖRDÜNCÜ ALT PROBLEME YÖNELİK BULGULAR.....	84
7.5. BEŞİNCİ ALT PROBLEME YÖNELİK BULGULAR	89
TARTIŞMA, SONUÇ VE ÖNERİLER.....	95
KAYNAKÇA.....	104
EKLER	110

TABLÖLAR LİSTESİ

	<u>Sayfa</u>
Tablo 1. Katılımcıların Demografik Yapısı	60
Tablo 2. Katılımcıların "Algı" Temasına İlişkin Görüşleri	66
Tablo 3. Katılımcıların "Doğallık" Temasına İlişkin Görüşleri.....	73
Tablo 4. Katılımcıların "Doğallık" Temasına İlişkin Görüşleri.....	77
Tablo 5. Katılımcıların "Güvenilirlik" Temasına İlişkin Görüşleri	84
Tablo 6. Katılımcıların "Sosyal Kabul ve Etkileşim" Temasına İlişkin Görüşleri	89



ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 1. Hazırlık Aşaması Alt Basamakları.....	63
Şekil 2. Veri Toplama İşlemi Alt Basamakları	63
Şekil 3. Veri Analizi İşlemi Alt Basamakları.....	64
Şekil 4. Bulguların Raporlanması İşlemi Alt Basamakları	64
Şekil 5. Tartışma, Sonuç ve Öneriler İşlemi Alt Basamakları	65



GİRİŞ

Yapay zeka (AI), insan düşünce süreçlerini taklit eden bir teknoloji olup üç temel kategoriye ayrılır: dar, genel ve üstün yapay zeka. Dar yapay zeka yalnızca belirli alanlarda etkilidir, genel yapay zeka ise insanların yürüttüğü karmaşık işleri yapabilir ve üstün yapay zeka insan zekasını geçen performanslar sergileyebilir. Teknolojik gelişmeler, hızlı veri işleme yeteneği sağlamasına rağmen, bağımsız düşünme ve kritik karar alma yetileri konusunda geliştirilmeye ihtiyaç duymaktadır. Yapay zeka insan duygularını anlama ve tepkileri taklit etme yeteneğine sahip olsa da, bu alanda tam anlamıyla insan benzeri bir farkındalık elde etmek şu an için mümkün değildir.

Yapay zekanın uygulama alanları oldukça geniştir ve özellikle makine öğrenimi, büyük veri setleri üzerinden kalıpları tanıyarak ve yeni senaryolara uyum sağlayarak önemli bir rol oynar. Makine öğrenimi, verileri analiz edip sonuçlar üretebilirken, derin öğrenme daha karmaşık veri setlerini işleyerek daha detaylı tahminlerde bulunabilir. Yapay zeka aynı zamanda etik tartışmaları da beraberinde getirmekte, özerklik, adalet, önyargı ve şeffaflık gibi konularda çeşitli soruları gündeme taşımaktadır. Algoritmaların karar verme süreçleri, gizlilik ihlalleri ve çevresel etkiler gibi potansiyel riskleri de içermektedir, bu nedenle yapay zekanın etik ve güvenli kullanımı önem taşımaktadır.

Literatürde sentetik ses üretimi, özellikle medya endüstrisindeki çalışmalar dikkat çekmektedir. Bu alandaki ilerlemeler, ses klonlama teknolojileri ve derin öğrenme tekniklerinin kullanılmasıyla, gerçek bireylerin ses özelliklerine benzer sentetik seslerin üretilmesine olanak tanımaktadır. Bu sesler, metin okuma, sesli kitap anlatımı ve dublaj gibi çeşitli alanlarda kullanılmaktadır. NVIDIA gibi firmalar, kullanıcıların metinleri belirli bir ses tonunda ve duygusal ifadeyle sentezleyebilmeleri için RAD-TTS gibi modeller geliştirmiştir. Konuşma sentezi (Text-to-Speech, TTS) teknolojisi, metin tabanlı bilgileri otomatik olarak konuşmaya dönüştürmektedir ve bilgisayar destekli dil öğrenimi gibi alanlarda yaygın olarak kullanılmaktadır. Yapay zeka destekli müzik üretimi, müzik yapımcılarının yeni müzikal fikirler üretmesine olanak tanırken, aynı zamanda müzik üretim sürecinde zaman ve kaynak tasarrufu sağlamaktadır. Ayrıca, yapay zeka, ses efektleri ve tasarımı alanında da kullanılmakta olup, ses ortamlarını yaratma ve ses kalitesini artırma işlemlerini kolaylaştırmakta ve medya prodüksiyon süreçlerini dönüştürmektedir.

Gelişmiş ses üretim modelleri, yapay zeka teknolojileri aracılığıyla ses üretiminde sağlanan devrimsel yenilikleri temsil etmektedir. Bu modeller, sentetik seslerin gerçek seslerle karşılaştırılmasını ve analiz edilmesini sağlayarak önemli araçlar haline gelmiştir. Özellikle, WaveNet ve MelNet gibi modeller, gerçek zamanlı konuşma sentezi ve müzikal kompozisyonlar üretebilme yetenekleri ile dikkat çekerler. Bu teknolojiler, eğlence ve medya endüstrilerinde kullanıcı deneyimini gerçekçilik ve etkileşim açısından iyileştirmekte, yapay zekanın sanatsal ifade ve yaratıcı süreçler üzerindeki etkisini artırmaktadır. Derin öğrenme ve büyük veri setlerinden elde edilen bilgilerle beslenen bu modeller, robotik ve interaktif sistemlerde de kullanılarak, makinelerin çevreleriyle daha doğal ve anlamlı bir şekilde etkileşimde bulunmalarını sağlamaktadır.

Ses klonlama teknolojileri, yapay zeka ve makine öğrenimi araçlarını kullanarak, kişilerin özgün seslerini sentetik olarak yeniden oluşturmayı amaçlar. Derin öğrenme teknikleriyle, hedeflenen sesin karakteristiklerini modelleyerek doğrulukla ses üretimi yapılmasını sağlar. Bu süreç, geniş ses veri setlerinin toplanması ve analiziyle başlar, ki burada sesin ton, pitch ve timbre gibi özellikleri incelenir. Sonrasında, derin sinir ağları ve GAN'lar gibi teknolojilerle ses sentezi gerçekleştirilir. Ses klonlama, konuşma yetisini kaybetmiş bireyler için iletişim araçları geliştirmeden, medya ve eğlence sektörlerinde seslendirme yapmaya kadar çeşitli alanlarda kullanılır. Ancak, bu teknolojinin etik ve yasal sorunları, özellikle izinsiz kullanım ve kimlik hırsızlığı riskleri içerir. Bu nedenle, teknolojinin kullanımında şeffaflık ve bireysel rıza büyük önem taşır. Özetle, ses klonlama teknolojileri, doğru ve etik kullanıldığında geniş bir yelpazede faydalı uygulamalar sunabilir, ancak gelişimi teknolojik yeniliklerle etik standartların dengeli ilerlemesine bağlıdır.

Makine öğrenimi ve yapay zeka, ses üretimi ve klonlama alanlarında öncü teknolojilerdir. Bu alanda, özellikle rekürrensiz sinir ağları (RNN) ve uzun kısa vadeli bellek (LSTM) ağları gibi yapay sinir ağları kullanılır. Bu ağlar, sesin zaman içerisindeki değişimlerini modelleyebilir ve ses özelliklerini gerçekçi bir şekilde yeniden üretebilir. Eğitim sürecinde, bu modeller çeşitli ses örnekleri üzerinde çalışarak ses üretimini öğrenirler ve eğitimden sonra belirli metinleri seslendirmek veya belirli bir tonlama ile konuşmak gibi görevlerde kullanılabilirler. Modelin sürekli olarak iyileştirilmesi ve optimize edilmesi, ses klonlama teknolojisinde önemli bir yere sahiptir. Bu sürekli gelişim, ses klonlama ve üretimi teknolojilerinin daha etkili ve doğal hale gelmesini sağlar, geniş bir kullanım alanına hitap eder.

Yapay zeka (AI), insan benzeri yetenekler sergileyen ve ses işleme alanında etkili uygulamaları bulunan teknolojileri kapsar. Ses işleme, ses ve müzik üretimi, konuşma tanıma ve interaktif dijital asistanlar gibi çeşitli alanlarda yapay zekanın kullanımını içerir. Derin öğrenme ve yapay sinir ağları gibi teknikler, ses dalgalarının karmaşık analizlerini başarılı bir şekilde gerçekleştirir ve bilgisayar sistemlerinin öğrenmesini sağlar. Ancak, yapay zeka kullanımı, veri ihlalleri gibi güvenlik endişelerini de beraberinde getirebilir. Yapay zeka, müzik alanında yenilikçi çalışmalara imkan tanırken, kişisel verilerin kötüye kullanılması potansiyeline dikkat çekmektedir. Temel ses işleme kavramları olan frekans, genlik ve faz gibi özellikler, sesin üretimini ve algılanmasını anlamada kritik role sahiptir. FFT ve spektrogramlar gibi tekniklerle ses sinyalleri detaylı bir şekilde işlenir ve yapay zeka teknolojileri bu analizlerde önemli bir rol oynar. Özetle, yapay zeka ve ses işleme teknolojileri, akademik araştırmalar ve günlük kullanım için büyük potansiyeller sunarken, etik kullanım ve güvenlik konularında ilerleme sağlanması beklenmektedir.

Yapay zeka teknolojilerinin ilerlemesi ile birlikte, sentetik seslerin üretimi ve kullanımı, bilgi iletişiminden eğitime, sanayiden günlük hayata kadar geniş bir alanda etkilerini göstermeye başlamıştır. Gerçek insan sesleri ile üretken yapay zeka teknolojisi kullanılarak üretilen sentetik sesler arasındaki benzerlikler ve farklar, bu seslerin algılanış biçimi, doğallık ve ifade yeteneği, güvenilirlik algıları ve sosyal etkileşime etkileri, teknolojik ve sosyal dinamikler açısından önemli bir araştırma alanını oluşturmaktadır. Bu çalışma, sentetik seslerin gerçek seslerle karşılaştırılmasını ve değerlendirilmesini merkezine alarak, ses teknolojilerinin gelecekteki gelişimine ışık tutmayı amaçlamaktadır.

Bu araştırmanın temel problemi, yapay zeka ile üretilen sentetik seslerin, gerçek seslerle olan benzerlikleri ve farklılıklarının neler olduğunu detaylı bir şekilde ortaya koymaktır. Bu bağlamda, araştırmada; katılımcıların sentetik ve gerçek sesleri nasıl algıladıkları, bu sesler arasındaki tercihlerinin nedenleri ve bu seslerin kullanıcı deneyimi üzerindeki etkileri analiz edilmektedir. Ayrıca, sentetik seslerin güvenilirlik algıları ve sosyal kabul düzeyleri gibi önemli boyutları da incelenmektedir.

Bu tez çalışması, sentetik ses teknolojilerinin etkilerini anlamak ve bu teknolojilerin insan hayatındaki yerini belirlemek için kritik bir adım oluşturmaktadır. Ses teknolojilerindeki gelişmelerin yanı sıra sosyal ve kültürel dinamikleri de dikkate

arak, teknolojik inovasyonların toplumsal kabulüne ve etkileşimlerine odaklanarak, ses teknolojilerinin gelecekteki yönünü şekillendirmede önemli bir rol oynamaktadır.



BİRİNCİ BÖLÜM

YAPAY ZEKA VE SENTETİK SES ÜRETİMİ

1. YAPAY ZEKA

1.1. YAPAY ZEKA (AI) KAVRAMI

Yapay zeka, bilişsel işlevleri taklit etmek için tasarlanmış bir teknoloji teorisidir. Bu bilim dalı, insanların gerçekleştirdiği işlerde kullanılmak üzere, büyük miktarda veri ve deneyimlerden öğrenme yeteneğine sahip akıllı bilgisayar sistemlerinin geliştirilmesine odaklanır. Yapay zekanın öncüsü John McCarthy, sinir ağlarının temellerini atmıştır. Birçok kişi, doğal insan zekasını taklit eden robotların geliştirilmesini insanlık tarihinin en büyük ilerlemelerinden biri olarak görür. Yapay zeka, yetenek seviyelerine göre üstün, genel ve dar olarak üç kategoriye ayrılır. Dar yapay zeka, aynı zamanda zayıf yapay zeka olarak da bilinir ve yalnızca belirli işlerde kullanılabilir, çünkü sınırlamalara sahiptir (Borano, 2016: 67).

Yapay genel zeka, insanların yerine getirebildiği karmaşık görevleri anlayabilme veya öğrenebilme kapasitesine sahip hipotetik bir makine zekasıdır. Bu yapay zeka türünün, belirsiz durumlar altında sorun çözme, planlama yapma, bilgiye dayalı kararlar alma ve bilişsel muhakeme yeteneği gibi potansiyel becerileri bulunmaktadır. Ancak, mevcut teknoloji gelişmeleri esas olarak veri işlemeyi hızlandırmaktadır ve bağımsız ve eleştirel düşünme yeteneğine ulaşabilmek için daha fazla zamana ihtiyaç olduğunu belirtmek gerekir (Semel, 2022: 267).

Yapay zeka, teorik olarak insanlardan daha üstün kabul edilen bir uzman sistemdir. Brundage (2016) tarafından, "neredeyse tüm alanlarda insan bilişinden çok daha ileri düzeyde olan zekâ" olarak tanımlanmaktadır. Bu tür sistemlerin insan anlayışını aşacak şekilde uygarlığa etki etmesi, pek çok kişide endişe uyandırmaktadır.

Yapay zeka türlerinin sınıflandırılması yapılırken, bir uzman sistemin insan yeteneklerini taklit etme kapasitesi bir ölçüt olarak ele alınır. Etkileşimli makineler, mevcut verilere dayanarak çalışan ve yeni bilgi birikimi oluşturmayan daha basit yapay zeka türleridir. Bu tür makineler, yalnızca son durumları değerlendirir ve bu nedenle bu yaklaşıma "tepki" denir. Etkileşimli makineler çevresel bilgiye sahip değildir ve her zaman programlanan talimatları izler. IBM'in 1980'lerin ortasında geliştirdiği ve satranç oynayan süper bilgisayar Deep Blue, etkileşimli makinelerin tanınmış bir örneğidir (Korinek & Stiglitz, 2018: 360).

Bu sistem, bir insan rakibe karşı yarışacak şekilde tasarlanmıştır ve gereken tahminler ile hamleleri yapabilme kapasitesine sahiptir. Etkileşimli makineler belirli komutları yerine getirir, fakat daha karmaşık işlevler için programlanmamışlardır. Sınırlı hafızalı zeka sistemleri, topladıkları verileri eylemlere dönüştürmeyi öğrenmiştir ve bu nedenle yalnızca tekrarlanan davranışlar üzerine programlanmışlardır. Örneğin, otonom araçlar bu sistemin tipik bir örneğidir; bu araçlar hareket halindeyken önceden belirlenmiş gözlemler ve bilgileri kullanır.

Bu sistemler aynı zamanda çevreyi gözlemlemek, potansiyel desenleri ve tepkileri tespit etmek için programlanır. Bu, yapay zekanın toplanan bilgileri nasıl en iyi şekilde kullanacağına karar vermesine yardımcı olur (Reynoso, 2019: 16).

Zihin teorisi, makinelerin insan davranışlarını anlama yeteneğidir. Makinelerde öz-farkındalık henüz varsayımsal bir durumdur. Zihin teorisine en yakın örnekler, insan benzeri kararlar alabilen makinelerdir. Siri ve Alexa gibi sesli asistanlar, insan yeteneklerini taklit ederken, tam ve doğal insan konuşmasına yakın performans sergileyen yaygın araçlardır. Robot Sophia, insan yüz ifadelerini okuma yeteneği ile insanlara verbal olmayan yardım sağlayabilir ve bu, bilimin zihin teorisi konusundaki ilerlemesinin belirgin bir örneğidir (Lane & Martin, 2021). Ancak, makinelerin insan duygularını tam olarak simüle edebilmeleri için daha fazla gelişmeye ihtiyaç vardır. Şu anda tam anlamıyla öz-farkındalığa sahip bir yapay zeka bulunmamaktadır, ancak sürekli araştırmalarla en gelişmiş yapay zeka türlerine ulaşılması mümkündür (Lane & Martin, 2021: 13).

Yapay zeka (AI), genel olarak insan zekasının gerektirdiği fonksiyonları yerine getirebilen bilgisayar sistemlerini tanımlar. Bu fonksiyonlar arasında konuşma ve görüntü tanıma, büyük veri işleme ve öneri sistemleri bulunur. Ancak yapay zeka teknolojileri üzerine genel bir fikir birliği olmaması, insan zekasının ne gerektirdiğinin belirsizliği ve zamanla değişebilir olması nedeniyle, yapay zekanın kapsamlı bir tanımı henüz netleşmemiştir (Petrooulos, 2018).

Avrupa Komisyonu'nun ilk yapay zeka uzman grubu tarafından önerilen tanıma göre, yapay zeka çevresel faktörleri analiz edip, akıllıca davranışlar sergileyebilen ve önceden belirlenen görevleri bağımsız olarak yerine getirebilen sistemleri ifade eder. Bu tanım, akıllı davranışın ne anlama geldiğine dair daha fazla açıklık getirirken, yapay zeka

sistemlerinin sadece dijital ortamlarda değil, robotlar, dronlar veya otonom araçlar gibi fiziksel makinelerle de nasıl entegre edilebileceğini açıklar (Petropoulos, 2018: 122).

Yapay zeka araştırmaları, genellikle sembolik ve sembolik olmayan yapay zeka olarak iki ana kategoriye ayrılır. Sembolik yapay zeka, insanların oluşturduğu modeller, mantıksal kurallar ve hesaplamalara dayanır ve bu nedenle kural tabanlı yapay zeka olarak da bilinir. Bu tür sistemler, belirlenen gerçekleri ve kuralları dinamik olarak değiştirmekte zorlanır ve yeni verilere uyum sağlamaları zaman alabilir.

Yapay zeka (AI), ham verileri kullanarak kalıpları tanımlayabilen ve kendi modellerini oluşturabilen sistemler için makine öğrenimi adı da verilir. Bu sistemler, öğrenme yeteneği sayesinde değişen koşullara ve çelişkili verilere adapte olabilirler ve oluşturdukları modeller ve temsiller genellikle insanların anlaması zor, karmaşık veya soyut olabilir. Bu sistemler, verimli çalışabilmek için büyük miktarda veri gerektirir. Sembolik ve sembolik olmayan yapay zekaları bütünleştirerek, insanların anlayabileceği temsiller makine öğrenme algoritmalarına dahil edilebilir (Joshi, 2019: 58).

Yapay zekanın kategorilendirilmesi literatürde üç ana şekilde yapılır: analitik, insanlardan ilham alan ve geçmiş deneyimlerden öğrenerek gelecekteki seçimlere rehberlik eden insanlaştırılmış yapay zeka. İnsanlardan ilham alan yapay zeka, duygusal ve bilişsel öğelerle donatılmıştır ve bu öğeleri karar verme süreçlerinde kullanır. İnsanlaştırılmış yapay zeka, etkileşimde bulunduğu diğerleriyle bilinçli ve kendinin farkında bir şekilde iletişim kurabilme yeteneğini içerir, bu da bilişsel, duygusal ve sosyal zeka özelliklerini birleştirir. Bazı araştırmacılar, dört tür yapay zeka olduğunu ileri sürmektedirler: En basit tür olan reaktif makineler, geçmiş deneyimleri veya anıları kullanmadan mevcut duruma göre kararlar alır ve gerçek dünyanın bilgisayar simülasyonlarını doğru bir şekilde oluşturmak zorlayıcıdır (Joshi, 2019: 58).

Yapay zekanın "dünyayı temsil etme" yeteneği, kısıtlı ve misyonlarına göre uyarlanmış dünya devletlerini içermez, bu yüzden mevcut ayarlar AI'nın aynı oyunları daha etkin oynamasına olanak tanır ancak bunlar kolayca modifiye edilemez veya farklı bağlamlarda kullanılamaz. Yapay zekanın dış dünya hakkında sınırlı bir anlayışı vardır ve bu, onun kendisine tahsis edilen görevlerin dışındaki işleri yapmasını imkansız kılar (Paschen, Pitt & Kitzmann, 2020: 147).

Örneğin, sınırlı teori kategorisindeki sürücüsüz otomobiller, geçen araçların yönünü ve hızını takip eder. Bu, sadece bir kez tanıma yerine sürekli izleme gerektirir.

Bu veriler, trafik işaretleri, şerit işaretleri ve yoldaki virajlar gibi öğeleri içeren otonom araçların önceden programlanmış modellerinin oluşturulmasına yardımcı olur. Bir araç şerit değiştirmeye karar verdiğinde, çevresel bilgiler dikkate alınarak diğer sürücülerin yola çıkışını veya yakındaki araçlarla çarpışmayı önlemeye çalışır. Ancak bu bilgiler genellikle geçicidir ve bir sürücünün yıllar süren deneyimi, bir aracın öğrenebileceği şekilde tam olarak kaydedilemez (Paschen, Pitt & Kitzmann, 2020: 152).

Üçüncü kategori olarak bilinen 'zihin teorisi' kavramı, bilgisayarların ne tür temsiller üretmesi gerektiğine dair net bir yaklaşım gerektirir. Bu daha ileri düzeydeki bilgisayarlar, sadece dünyanın kendisini değil, aynı zamanda araçlar veya diğer nesnelerin temsillerini de üretebilir. Psikolojide "zihin teorisi", insanların, hayvanların ve çevredeki diğer varlıkların düşünce ve duygularıyla nasıl hareket edebileceklerini anlama yetisini ifade eder. Toplamların nasıl işlediğini anlamak, sosyal etkileşimi mümkün kılar. Eğer yapay zeka sistemleri toplumumuzun bir parçası olacaksa, her birimizin fikirleri, duyguları ve beklentileri olduğunu anlamaları gerekecektir. Buna bağlı olarak, bireylerin davranışlarını nasıl ayarlamaları gerektiği ortaya çıkacaktır (Lavet, Henderson & Pineau, 2018).

Son kategori ise 'kişisel farkındalık'tır. Kendi temsillerini oluşturabilen sistemler, yapay zeka gelişiminin en son aşamasını temsil eder ve bu, 'zihin teorisi'nin bir genişlemesi olarak kabul edilebilir. Bilinç, sık sık 'öz-farkındalık' olarak adlandırılır. "Bu eşyayı istiyorum" ile "Bu eşyayı istediğimi biliyorum" arasındaki fark büyük önem taşır. İnsanlar, başkalarının duygularını tahmin edebilir, kendi iç dünyalarının farkında olabilir ve bilinçli hareket edebilirler. Örneğin, birisi trafikte bize korna çaldığında, otomatik olarak o kişinin kızgın veya rahatsız olduğunu düşünürüz. Zihin teorisi olmadan, bu tür sonuçlara varmamız zorlaşır. Kendi bilincinin farkında olan makinelerin inşası henüz uzak bir hedef olmakla birlikte, araştırmalarımızı hafıza, öğrenme ve deneyime dayalı karar verme üzerine odaklamamız gerekmektedir (Lavet, Henderson & Pineau, 2018).

1.2. YAPAY ZEKA UYGULAMA ALANLARI

1.2.1. Makine Öğrenimi

Makine öğrenimi, yapay zekanın bir dalıdır ve bu alan, algoritmalara dayalı bir yöntem kullanarak girdi verilerini alıp, bu verileri istatistiksel analizlerle işleyerek çıktıları tahmin eder ve sonuçları güncelleyerek yeni girdiler oluşturur. İnsanlar, sürekli olarak bir ekran karşısında oturup veri girmek yerine, bu işlemleri kendiliğinden

gerçekleştirebilen makineler geliştirmeyi tercih etmişlerdir. Bu makineler, veritabanlarını tarayabilir, erişebilir, yükleyebilir, kaydedebilir ve veri oluşturabilir (Davenport, 2018: 22).

İnsanlar, internetin sağladığı imkanlar sayesinde makinelerin kendi kendine öğrenme yeteneğini de değerlendirirler. Makine, ihtiyaç duyulan her türlü veriyi sağlayabilir ve internet erişiminin ücretsiz olması sayesinde, makine öğrenimi teknolojisini kullanarak çok zaman kazanabilirsiniz. Yapay zeka, teknik gelişimin bir yansıması olarak makinelerin kontrol edilmesini sağlar. Makine öğrenimi, bilgisayar destekli analiz ve kendi kendini eğitme, gözlem ve bilgi yoluyla yeni bilgileri işlemeyi kolaylaştırır. Makineler, yeni senaryolara maruz kalarak, test edilerek ve uyum sağlayarak sürekli olarak gelişir. Desen tanıma ve görünüm tanıma, gelecekteki seçenekleri iyileştirmek için de kullanılır. Makine öğrenimi, veri işleme ve bilgi keşfi ile sıkça karıştırılır, ancak genellikle aynı metodolojiyi kullanır. İnsan sezgisini bir makinede taklit etmek zordur çünkü insanlar genellikle seçimlerini bilinçsizce öğrenir ve uygular (Barbell, Banker & Pihlajamaa, 2018: 170).

Çocuklar gibi makineler de, kapsamlı algoritmalar geliştirildikten sonra belirli eğitim tekniklerine ihtiyaç duyar. Bu teknikler arasında geriye yayılma, takviyeli öğrenme, genetik algoritmalar, segmentasyon, açıklamaya dayalı öğrenme, kümeleme, hata düzeltme, durum kaydı ve makro operatörler bulunur. Yapay zeka hesapları genellikle denetimli veya denetimsiz olarak sınıflandırılır; denetimli hesaplar, hem girdileri hem de çıktıları kontrol edebilmek için yapay zeka yeteneklerine sahip bir bilgi araştırmacısı veya bilgi uzmanı gerektirir (Dignum, 2019: 40).

Bilgi araştırmacıları, modelin işlemesi gereken faktörleri tanımlar ve bu faktörleri tahminler oluşturmak üzere kullanırlar. Hazırlık süreci tamamlandıktan sonra, algoritma bulguları güncellenmiş veriler üzerine uygular. Denetimsiz hesapların hazırlığında, sonuçlar önceden tanımlanmaz. Bunun yerine, derin öğrenme adı verilen yinelemeli bir süreçle veriler analiz edilir ve sonuçlar çıkarılır. Denetimli öğrenme sistemleriyle kıyaslandığında, genellikle sinir ağları olarak adlandırılan denetimsiz öğrenme modelleri, içerik doğrulama ve karakteristik dil yaşamı gibi daha karmaşık görevlerde kullanılır. Bu sinir ağları, çeşitli bilgi işleme adımlarından geçer ve sonuçta farklı parametreler arasındaki basit ilişkileri belirler. Hazır olduklarında, bu ağlar ilişkisel veritabanlarını kullanarak yeni verileri işleyebilir. Büyük veri işleme yöntemleri gerektirdiğinden, bu tür

hesaplamaların büyük veri çağında mümkün olduğu anlaşılmıştır (Prat, Wattiau & Akoka, 2014).

Yapay zeka günümüzde örneğin Facebook'un haber kaynağı gibi çok çeşitli alanlarda kullanılmaktadır. Yapay zeka, kullanıcıların etkileşimlerine bağlı olarak haber kaynağını kişiselleştirmek için kullanılır. Eğer bir kullanıcı belirli bir arkadaşının gönderileriyle etkileşimde bulunmayı durdurursa, haber kaynağı bu arkadaşın daha önceki aktivitelerini daha sık göstermeye başlar. Bu süreç, ölçülebilir analizler ve ileri araştırmalar kullanılarak gerçekleştirilir. Eğer kullanıcı arkadaşlarının gönderileriyle etkileşimi sürdürürse, bu yeni bilgiler sisteme dahil edilir ve haber kaynağı buna göre güncellenir (Dignum, 2019).

Ayrıca, yapay zeka iş dünyasında da geniş bir uygulama alanına sahiptir. CRM sistemleri, e-postaları sıralamak ve çalışanların önemli mesajlara öncelik vermesini sağlamak için yapay zeka modellerini kullanır. Gelişmiş sistemler, başarılı yanıtları öngörebilir. İş zekası (BI) ve pazar araştırmacıları, müşterilerin dikkatlerini önemli bilgilere yönlendirmelerine yardımcı olmak için yapay zeka kullanır. İnsan kaynakları (HR) sistemleri, adayların niteliklerini değerlendirip en uygun işe yerleştirme için bu modellerden yararlanır. Sürücüsüz araçlar da büyük ölçüde yapay zekaya bağımlıdır.

Derin öğrenme sinir ağları, aracın yol üzerinde güvenle yönlendirilmesi için en uygun hareket planını seçmek amacıyla kullanılır (Grover, Kar & Dwivedi, 2022). Makine öğrenimi, sistemlerin deneyimlerden bağımsız olarak otomatik olarak öğrenmelerini mümkün kılan yapay zekanın bir uygulamasıdır. Makine öğrenimi, bilgisayar programlarının yeni bilgilere erişip öğrenmesi sürecidir. Öğrenme, verilerde kalıplar arayarak ve gözlemler veya bilgilerle başlayıp, sağlanan örneklerle pekiştirilerek ileriye dönük kararlar almayı amaçlar. Esas hedef, bilgisayarların insan müdahalesi olmadan öğrenmelerini sağlamaktır (Brynjolfsson, Rock ve Syverson, 2019: 32). Denetimli makine öğrenimi, etiketlenmiş örneklerden öğrendiklerini kullanarak gelecekteki olayları tahmin etmek için tasarlanmıştır. Algoritma, bilinen bir veri setini inceleyerek ve tahminler yaparak başlar ve modeldeki hataları belirleyip düzeltebilir. Öte yandan, denetimsiz makine öğrenimi, etiketlenmemiş verilerden çıkarımlar yaparak gizli yapıları keşfeder ve bu verileri kullanarak çıktılar üretir, hatta uygun çıktıyı bulamasa bile (Acemoğlu & Restrepo, 2018).

Yarı denetimli makine öğrenimi, az sayıda etiketli veri ve çok miktarda etiketlenmemiş veri kullanarak, denetimli ve denetimsiz öğrenme arasında bir yerde konumlanır. Bu sistemler, öğrenme doğruluğunu büyük ölçüde artırabilir. Genellikle, eğitimde önemli miktarda etiketlenmemiş bilgi gerektiğinde ve nitelikli personel ve kaynaklar gerektirildiğinde kullanılır. Takviyeli öğrenme, davranışları üreterek ve çevreyle etkileşimde bulunarak hatalardan veya ödüllerden yararlanan bir öğrenme sürecidir. Bu teknik, makinelerin ve programlarının seçilen ortamda en iyi davranışları bulmalarını sağlayarak performansı maksimize eder (Crandall, 2019: 12).

1.2.2. Derin Öğrenme

Derin öğrenme, makine öğreniminin daha ileri bir seviyesi olup, özellikle gelecek tahminlerinde kullanılır ve daha önce makine öğrenimi tarafından oluşturulan veri tabanlarından yararlanır. Bu süreç, derin sinir ağları kullanılarak daha karmaşık veri analizleri yapar (Joshi, 2019: 58). Derin öğrenme, verileri gruplayarak ve bu gruplardan tahminler yaparak çalışır. Örneğin, Facebook yorumları veya Google fotoğraf kitaplıkları gibi büyük veri kümelerini işleyebilir. Derin sinir ağları, daha karmaşık sınıflandırma, analiz ve karşılaştırma görevlerini yerine getirir, ve bu süreçler, sürekli güncellenen veri girişlerini kullanarak tahminler yapmayı gerektirir. Sinir ağları, insan beyninin kalıp tanıma kapasitesini taklit edecek şekilde geliştirilmiştir ve duysal verileri işleyerek etiketler veya ham veriler aracılığıyla algılar. Bu kalıplar dijital formatta olduğundan, dönüştürülmesi gereken bilgiler arasında resimler, sesler, metinler ve diğer veri türleri bulunur (Lane & Martin, 2021: 16).

Yapay sinir ağları, bilgileri işleyip sınıflandırabilir. Bunlar, saklanan bilgileri bir koleksiyon ve sınıflama sınıfı gibi yönetir. Sinir ağları, benzer özellikler taşıyan etiketsiz bilgileri gruplar ve etiketli veri setlerine sahip olduğunda sınıflandırma yapar. Bu ağlar, farklı sınıflama ve yığılma algoritmalarını destekleyerek, öğrenme ve regresyon tekniklerini içeren geniş kapsamlı makine öğrenimi uygulamalarının bir parçası olabilir. Tüm sınıflama işlemleri, ayrıştırılmış veri kümeleri üzerine kuruludur; bu, sinir ağının veri ve etiketler arasındaki ilişkileri anlamasını sağlar. Benzerlik bulma, birleştirme işlemi olarak adlandırılır. Derin öğrenme, etiket gerektirmeyen alanlarda da benzerlikleri keşfeder. Denetimsiz öğrenme, sınıflandırma içermeyen bir öğrenme türüdür ve dünyadaki verilerin çoğu etiketlenmemiştir. Bir algoritmanın kullanabileceği ek bilgiler her zaman daha fazla gerçeklik sunar, bu yüzden denetimsiz öğrenme doğru modeller sağlama potansiyeline sahiptir (Petropoulos, 2018).

Benzerlikleri görmenin tersi anormallikleri veya tuhaf davranışları tespit etmektir. Olağandışı davranışların bazen dolandırıcılık gibi dikkat etmeniz ve uzak durmanız gereken şeylerle güçlü bir ilişkisi vardır.

Yapay sinir ağları aynı zamanda insan beyni olarak da düşünülebilir. Bir sinir ağının, kendi deposundaki veri kümeleri ile dış ortamdan aldığı yeni veriler arasındaki bağlantıya dayalı olarak çalıştığını hayal edebilirsiniz. Sinir ağı, iki farklı veri türünü birbirine bağlayarak geçmiş ve bugünü karşılaştırarak ve karşılaştırarak tavsiyelerde bulunabilir, bir karara varabilir veya geleceği tahmin edebilir.

Derin öğrenme insan beyninin yaratılıştaki işleyişini simüle eden insan yapımı bir zeka uygulamasıdır. Derin öğrenme, yapılandırılmamış veya sınıflandırılmamış bilgilerden denetimsiz öğrenebilen ağlara sahip yapay zekada (AI) bir dizi makine öğrenimi olabilir. Birlikte derin nöral öğrenme veya derin nöral ağ olarak adlandırılırlar. Derin öğrenme, dijital çağla birlikte gelişmiş ve gezegenin her bölgesinden ve her türden bilgi patlamasına neden olmuştur. Yalnızca "büyük bilgi" olarak sınıflandırılan bu bilgiler, sosyal medya, çevrimiçi arama motorları, e-ticaret platformları ve çevrimiçi sinema salonları dahil olmak üzere çeşitli yerlerden toplanmaktadır. Bulut bilişim gibi sistemler bu büyük miktardaki veriye kolay erişim ve bu verinin paylaşılmasını sağlar (Kober, Bagnell & Peters, 2013).

Bununla birlikte, genellikle yapılandırılmamış veriler devasa boyuttadır ve insanların bunları anlaması ve ilgili verileri çıkarması onlarca yıl alabilir. Şirketler, makine destekli destek için bu zengin verileri ve yapay zekayı açığa çıkarmanın getireceği inanılmaz potansiyeli fark etmeye başlamıştır ve yavaş yavaş kendi sistemlerine adapte etmektedirler. Makinelerin karmaşık sorunları insan müdahalesi olmadan çözmeyi öğrendiği bir çağdayız ve özellikle uğraştığımız sorunlar neler? Bugün derin öğrenmenin desteklediği görevlerden şunlardır:

- Sanal asistanlar

İster Alexa, Siri veya Cortana olsun, çevrimiçi hizmet sağlayıcıların sanal asistanları, konuşmanızı algılamaya yardımcı olmak için derin öğrenmeyi ve ayrıca insanların kendileriyle birlikte taşındıktan sonra kullandıkları dili kullanır.

- Çeviriler

Derin öğrenme algoritmaları aynı zamanda diller arasında mekanik olarak çeviri yapmaktadır. Gezinler, iş adamları ve kamu çalışanları bunu faydalı bulabilir. Otonom

araba bağlantılı otomobiller, yolun gerçeklerini ve ister dur işareti ister sokak lambası veya başka bir araç olsun, bunlara nasıl tepki verileceğini anlamak için derin öğrenme algoritmalarını kullanır. Karla kaplı bir dur işareti hala bir durma sinyali olarak hizmet vermektedir, bu nedenle algoritmalar ne kadar fazla bilgiye sahipse onu anlamak için o kadar çok çalışabilirler (Kober, Bagnell & Peters, 2013).

- Chatbot'lar ve bot'lar reformu

Çok sayıda kuruluşa müşteri hizmeti sağlayan chatbot'lar ve onarım yazılımları, derin öğrenme nedeniyle artan dış algı ve metin sorguları sayesinde akıllı ve yardımcı bir şekilde yanıt verebilir.

- Resmi renklendirme

Siyah beyaz görüntüleri renkliye dönüştürmek bir zamanlar insan elinin hassas bir işiydi. Günümüzde derin öğrenme algoritmaları, siyah beyaz bir görüntüyü renkli olarak yeniden oluşturmak için görüntülerdeki bağlamı ve nesneleri kullanabiliyor. Sonuçlar şaşırtıcı ve doğrudur.

- Kimlik doğrulama

Derin öğrenme, yalnızca güvenlik işlevleri için değil, Facebook gönderilerinde insanları etiketlemek dışında kimlik doğrulama için de kullanılabilir ve yakın gelecekte yüzümüzü feda ederek bir mağazada bir şeyler satın alabiliriz. Kimlik doğrulama için derin öğrenme algoritmalarının zorlukları, saç stilini değiştirdikten, sakal veya tıraş olduktan sonra veya çekilen fotoğrafın tehlikeli aydınlatma veya ilgili engeller nedeniyle kötü olup olmadığını bile eşdeğer bir kişinin olduğunu bilmektir (Reynoso, 2019).

- İlaçlar ve reçeteli ilaçlar

Hastalıkla ilişkili tümörlerin teşhis edilmesinden, insanları yönetmek için özel olarak oluşturulmuş özel ilaçlara kadar, tıp alanındaki derin öğrenme, en önemli ilaç ve tıp şirketlerinin çoğunun dikkatini çekmektedir.

1.3. YAPAY ZEKA ÜZERİNE ETİK TARTIŞMALAR

Yapay zekanın gelişmesinin beraberinde getirdiği çok sayıda tartışmalı konu bulunmaktadır. Yapay zekaya ilişkin hem kamusal hem de profesyonel söylemde bazı felsefi fikirler tartışmanın ön saflarında yer almaktadır. Yapay zeka oluşturulurken ve kullanılırken dikkat edilmesi gereken fikirler üzerine çok sayıda analiz yapılmıştır (Kober, Bagnell & Peters, 2013).

Avrupa Komisyonu, yapay zekanın oluşturulması ve uygulanması konusunda rehberlik sağlamak üzere akademi, iş dünyası ve sivil toplumdan uzmanları Yapay Zeka Kıdemli Uzman Grubuna davet etmiştir. Uzmanlar bireysel özerkliğe saygı duymayı, hasarı önlemeyi, adaleti takip etmeyi ve yoruma yer bırakmayı temel etik ilkeleri olarak kabul etmiştir. Bu ilkeler aynı zamanda yapay zeka sistemleri için yedi temel standarda da dönüştürülmüştür. Bunlar: insan temsili ve gözlemi; teknik emniyet ve güvenlik, gizlilik ve veri yönetimi, şeffaflık, çeşitlilik, ayrımcılık yapmama ve adalet; sosyal ve çevresel refah; ve sorumluluktur (Reynoso, 2019). Değer bilincine sahip bir tasarım metodolojisi kullanarak makine öğrenimi algoritmalarını adil ve güvenli bir şekilde uygularken dikkate alınması gereken değerleri incelendiğinde dürüstlük, bağımsızlık, mahremiyet, adalet, eşitlik, sahiplenme, sorumluluk ve açıklık ilkelerini seçilir. Bu ilkelerin çoğu, makine öğrenimi algoritmalarının güvenlik, adalet ve genel yapay zeka alanlarının dışında uygulanmasıyla ilgilidir.

1.3.1. Bağımsızlık ve Sorumluluk

AI düzenli olarak bağımsızlığa sahip olarak tanımlanır, ancak bununla genellikle ne kastedildiği açık değildir. Bilimsel literatürde ve popüler medyada yazarlar, yapay zekanın tamamen otonom hale geleceği ve insanları önemsiz kılacağı yönündeki korkularını dile getirmişlerdir. Yapay zekanın insanlığın büyük bir bölümünü öldürerek varoluşsal bir tehdit haline gelme olasılığı da tartışılmaktadır. "Yapay zeka" terimi ilk ortaya çıktığında, makinelerin insan zekasına benzer bir zeka türü kazanması bekleniyordu. Beklenti, bir bilgisayar sisteminin birçok farklı görevde insanlardan daha iyi performans gösterebilmesiydi. Bu tür genel bir AI sisteminde çalışmak yerine, çoğu araştırma şu anda belirli bir görevi insan uzmanlardan daha yüksek hız, verimlilik ve doğrulukla gerçekleştirebilen AI sistemleri geliştirmeye odaklanmaktadır (Reynoso, 2019).

Yapay zeka tamamen otonom hale gelmese ve insan kontrolünden çıkmasa bile, şu anda konuşlandırılmış ve geliştirilmiş yapay zeka sistemleri aslında insanların uygulayabileceği özerklik düzeyini etkileyebilir. Avrupa yapay zeka üst düzey uzman grubu, insan özerkliğine saygı ilkesini dahil ederek bu tehdide odaklanmıştır. Bu ilkeye ilişkin yorumlarında, insanların yapay zeka sistemleriyle etkileşim kurarken tam ve etkili bir şekilde kendi kaderini tayin edebilmeleri ve demokratik sürece katılabilmeleri gerektiğini savunmuşlardır. Bunun, AI sistemlerinin insanları haksız yere manipüle

etmemesi, zorlamaması, aldatmaması veya baskı altına almaması gerektiği anlamına geldiğini eklemiştir (Lavet, Henderson & Pineau, 2018).

Hayes ve arkadaşlarına göre (2020: 540) bağımsızlık, bireylerin özgür yaşamlar sürdürmek için bilinçli olarak hareket etme ve düşünme kapasitesidir. Hem kararı veren hem de karar verilen kişinin bağımsızlığının, açıkça yargı sistemini hedef alan karar verme algoritmaları tarafından nasıl tehdit edilebileceğini vurgulamışlardır. Karar vericiler körü körüne veya içgüdüsel olarak kendi kararlarından çok algoritmanın yargılarına güvenme riskiyle karşı karşıyadır. Bu, karmaşıklığı ve açıklığı nedeniyle algoritmanın sonucunu eleştirel olarak değerlendiremeyebilecekleri için karar vericinin bağımsızlığını engelleyebilir.

Makine öğrenimi algoritmalarının (veya yardımıyla) alınan kararların özneliği için, özerkliklerinin çeşitli şekillerde kısıtlanması riski vardır. Adalet ve güvenlik alanında, algoritmalar masumiyet karinesini düşürerek ve öznenin gelecekteki fırsatlarını ve özgürlüklerini engelleyerek insanları şüpheli gösterebilir (Hayes, Poel & Steen, 2020). Gelecekteki fırsatların engellenmesi, kredi tahsisi veya iş fırsatları için çalışanların seçilmesi gibi diğer durumlarda kullanılan algoritmaları da riske atabilir (Johnson ve Verdicchio, 2019: 640).

Yapay zeka teknolojilerinin kullanılması sonucunda bir şeyler ters giderse kimin sorumlu olduğunu belirlemek bazen zor olabilir. Bir yapay zekanın bir dereceye kadar özerk olduğunu düşünürsek, onun kendi faaliyetlerinden en azından bir dereceye kadar sorumlu olduğu sonucuna varabiliriz. Yapay zeka uygulamalarından kimin sorumlu olduğunu açıklığa kavuşturmak için Johnson ve Verdicchio (2019), üç farklı aracı türü arasında bir ayrım yapmıştır. Birinci tür faillik, bir şeyin veya birisinin meydana gelmesine katkıda bulunan nedensel failliktir. Bir şeyin nasıl meydana geldiğini etkileyen herhangi bir teknolojinin nedensel faillığe sahip olduğu söylenebilir. İkinci tür faillik, nedensel faillikte de mevcut olan failin amacı ile nedensel zinciri başlatan kasıtlı failliktir.

Niyetlerin zihinsel durumlar olduğuna inanıldığından, kasıtlı eylemlilik çoğunlukla yalnızca insanlarla sınırlıdır. Niyetler önemlidir çünkü bir kişinin etik ve yasal durumlarda eylemlerinden sorumlu tutulup tutulamayacağını etkiler (Johnson & Verdicchio, 2019: 641).

Sorumluluk, bir tür negatif sorumluluk olarak tanımlanabilir, eğer failer ahlaki yetkiye sahipse, olanlarla nedensel bir bağlantıya sahipse ve bir tür suçla suçlanıyorsa,

sorumlu tutulabileceklerini ve hatta herhangi bir şey için suçlanabilecekleri öne sürülebilir. Hayes ve arkadaşlarına göre (2020: 540), bir kişiyi veya grubu sorumlu tutmak için olaya veya sonuca ve ayrıca olaya dahil olan kişilere ve şeylere ilişkin bilgi sahibi olunması gerekir. Bu nedenle örtük yapay zeka sistemlerinin yapay zeka ile ilgili tüm yönlerde şeffaf olması gerektiğini söylemek mümkündür.

1.3.2. Adalet, Önyargı ve Yorumlanabilirlik

Adil olmak, karar verme algoritmaları hakkında konuşurken dikkate alınması gereken çok önemli bir fikirdir. Yapay zekanın hangi adalet anlayışını desteklemediğini keşfetmek için Saxena ve diğerleri (2020), karar verme algoritmaları için açıkça oluşturulan üç farklı adalet tanımını analiz etmiştir. Adalet kavramlarında dağıtıcı adaleti ve adil sonuçları vurgulanmıştır. "İnsanlara aynı şekilde davranmak", "daha kötü bir insanı daha iyi bir insana tercih etmemek" ve "hesaplanmış adalet", karşılaştırdıkları üç adalet tanımıdır (Saxena vd., 2020).

Dwork ve diğerleri tarafından (2012) önerilen ilk tanım, benzer özelliklere sahip bireyleri benzer şekillerde değerlendirerek bilgilendirici kararlar almayı sağlayan algoritmaların geliştirilmesine yöneliktir. İkinci tanım ise, adil bir algoritmanın bir grup içinden en uygun adayı seçmesi gerektiğini savunarak, bir grup insan arasından aday seçmek için adil bir algoritma oluşturulmasını önermiştir. Liu ve diğerleri ise (2002), standartlaştırılmış adalet tanımlarını önceki iki tanımın bir kombinasyonuna dayandırmaktadır.

Saxena ve diğerleri tarafından yapılan tartışma (2020) şu anda bilgisayar bilimcileri tarafından adil algoritmalar oluşturmak için kullanıldığı için, öncelikle halkın adalet tanımları algısıyla ilgilidir. AI ile ilgili olarak daha felsefi odaklı adalet tartışmaları da yayınlanmıştır. Binns (2018) makine öğrenimi adaletine siyaset felsefesi açısından bakmıştır. Ona göre, makine öğrenimi teknikleri muhtemelen önyargı olarak altta yatan ayrımcılık kalıplarını tespit edecek ve bu da bazı gruplara ve bireylere adaletsiz davranılmasına neden olabilecek sonuçlara yol açabilecektir. Binns (2018: 150), adil makine öğrenimini geliştirmeye yönelik mevcut yöntemlerin, bu grupların ekstra korumaya ihtiyaç duymasının nedenlerini dikkate almadan, belirli, yasal olarak korunan sınıf gruplarına çok fazla odaklandığını belirtmiştir. Belirli durumlardaki kritik endişelerin, çeşitli adalet ve ayrımcılık felsefeleri üzerine düşünülerek felsefi olarak ele alınabileceğini belirtmiştir. Güvenilir Adalet, Uluslararası Af Örgütü'nün temel

değerlerine entegre ettiği bir değerdir. Yapay zeka oluştururken her ikisinin de dikkate alınması gereken prosedürel ve objektif adalet arasında bir ayrım yapılmıştır. Nesnel olarak adil olabilmek için yapay zeka sistemleri, avantajları ve harcamaları adil bir şekilde dağıtmalı ve herhangi bir kişi veya grubu haksız önyargıdan, ayrımcılıktan veya damgalamaktan kaçınmalıdır. Yapay zeka sistemleri ve bunları kullanan kişiler tarafından alınan kararlara, prosedürel adalet sayesinde başarılı bir şekilde itiraz edilebilir ve değiştirilebilir. Bu durum, karar alma prosedürünün şeffaflığının yanı sıra, açıkça tanımlanmış, hesap verebilir tutulabilecek bir kurumun varlığını da gerektirmektedir. Adalet, özerklik ve sorumluluk yapay zekanın açıklama kapasitesinden etkilenir.

Ferrara (2020), AI HLEG'in (2019) tarafından tanımlanan "fiziksel adalet" kavramına dayanarak, yapay zeka sistemlerinin geliştiricileri ve kullanılan eğitim verileri kaynaklı önyargıların, adil olmayan sonuçlar üretilip toplumsal eşitsizlikleri pekiştirebileceğini savunur. Bu bağlamda, AI modellerinin, özellikle azınlık grupları üzerinde olumsuz ve orantısız etkiler yaratabileceği, ayrımcılığı ve dar bakış açılarını sürdürebileceği belirtilir. Özellikle, veri setlerindeki eksiklikler ve yanlılıklar nedeniyle, bu sistemlerin haksız ve yanlı davranış biçimlerini devam ettirebileceği üzerinde durulur.

Önyargı ve adalet konusunda, yapay zeka uygulamalarının özellikle karar verme algoritmaları aracılığıyla haksız sonuçlar üretebileceği sıkça tartışılır. Önyargı, yapay zekadaki adaletsizliklerin potansiyel bir kaynağı olarak görülür. Ferrara (2020) çalışmasında, önyargının sadece gerçeklik ve adaletle ilişkili olmadığını, aynı zamanda algoritmaların tasarım seçimlerinden, verilerin temsiliyet eksikliklerinden ve hatalı veri toplama pratiklerinden de kaynaklanabileceğini belirtir. Özellikle, önyargıların algoritmalar aracılığıyla nasıl haksız sonuçlara yol açabileceğini anlamak, karar verme sistemlerinde önemli bir yer tutar.

Ayrıca, önyargı kavramının yapay zeka uygulamalarında hem zararlı hem de yararlı olabileceği, kasıtlı veya kasıtsız şekilde ortaya çıkabileceği anlaşılmıştır. Önyargılar yalnızca insanların öznel değerlendirmelerinden veya insanlar tarafından toplanan verilerden kaynaklanmaz; yapay zeka sistemlerinin kullanıldığı bağlamlar, çevrimiçi etkileşimler ve sürekli öğrenme süreçleri de bu önyargıları şekillendirebilir. Bu nedenle, yapay zeka geliştiricileri ve uygulayıcıları, sistemlerin tasarımında ve uygulanmasında önyargıları minimize etmek için dikkatli olmalıdır.

Makine öğrenimi algoritmalarında birçok önyargı türü ve bunların nasıl ortaya çıkabileceği Dobbe ve diğerleri (2018) tarafından daha ayrıntılı olarak açıklanmaktadır. Yapay zekadaki adalet literatürünün, önyargıların eğitim verilerinden makine öğrenimi sistemlerine nasıl taşınabileceğini aşırı vurguladığını iddia etmişlerdir. Önceden var olan önyargılara ek olarak, teknik önyargıların ve sonuçta ortaya çıkan önyargıların makine öğrenimi algoritmalarında doğal olarak meydana geldiği belirtilmiştir.

Yapay zekanın açıklanabilirliği ve şeffaflığı, yapay zekanın adaleti hakkındaki tartışmada yinelenen önemli bir temadır. Bir algoritma tarafından verilen bir kararın adil olup olmadığını belirlerken, insanlar genellikle algoritmanın bu karara nasıl vardığına dair bir açıklama ister. Makine öğrenimi algoritmaları söz konusu olduğunda bu karmaşıktır çünkü bu algoritmalar genellikle opaktır. Açıklama ilkesi, yapay zeka geliştirme süreçlerinin şeffaf olması, yapay zeka sistemlerinin yetenekleri ve hedeflerinin açıkça iletilmesi ve yapay zeka sistemleri tarafından alınan kararların, bunlardan etkilenenlere mümkün olduğunca açıklanabilir olması gerektiği anlamına gelir (Prat, Comyn-Wattiau & Akoka, 2014)

Daha önce belirtildiği gibi, Hayes ve ark. (2020, 15), genel olarak algoritmaların ve yapay zekanın şeffaflığının hesap verebilirlik için gerekli olduğunu savundu. Ayrıca özerklik, adalet ve mahremiyet dahil olmak üzere tartıştıkları diğer birçok değer için şeffaflığın önemli olduğundan bahsetmişlerdir. Algoritma bilgisi, algoritmaların karar vericilerin ve karar öznelerinin bağımsızlığını sınırlayabileceği yollara karşı yardımcı olabilir (Hayes vd., 2020: 540).

Bir nesne veya olay hakkında "açıklanabilirlik" bilgi edinme olasılığı olarak kullanılır. İlgili bilgiyi bulma görevi, yapay zeka sistemleri, özellikle de makinelerdeki ilerlemeler nedeniyle sıklıkla daha zor hale gelmektedir. Burrell (2016: 13) gösterge olarak makine ilerlemelerine odaklanmıştır. Algoritmanın seçiminden etkilenenlerin, öğrencilerinin sağladığı girdinin bu sonuca nasıl ve neden katkıda bulunduğunu her zaman anlayamamaları nedeniyle bu patlamaların sıklıkla şeffaf olmadığını açıklamıştır. Bu sistemlerde ve genel olarak yapay zekada sıklıkla görülen üç temel opaklık biçimi Burrell (2016: 13) tarafından ayırt edilmiştir. İlk kırmızı bayrak "kasıtlı kurumsal veya hükümet şeffaflığıdır". Bu tür bir gizlemede, büyümeyi yaratan kuruluş veya hükümet, örneğin rekabet avantajı elde etmek için gizli sitedeki kodu yayınlatabilir, genişlemenin kullanımını yasaklayabilir veya büyümenin gizli konumlarını gizleyebilir. Algoritmik şaşırtmanın ikinci türü "teknik bilgisizlik olarak şaşırtmacadır". Bu şeffaflığın nedeni,

nispeten az sayıda kişinin makine öğrenme algoritmalarını oluşturmak ve kavramak için gereken uzmanlaşmış yeteneklere ve bilgilere sahip olmasıdır. Son ve en temel şaşırtma türü ise "uygulama kapsamında algoritmaların çalışma şekli olan şaşırtmacadır." Bu tür bir şaşırtmaca, makine öğrenimi algoritmalarının tasarımından ve işleyişinden kaynaklanır. Birincisi, bir kişinin bir makine öğrenimi algoritmasının tamamını kavraması son derece zor olabilir çünkü çoğu zaman birçok yazar tarafından oluşturulan çeşitli farklı bileşenlerden oluşur. İkincisi, makine öğrenimi algoritmalarının etkili olabilmesi için çok büyük miktarda verinin algoritmanın koduyla karmaşık bir şekilde etkileşime girmesi gerekir. Son olarak Burrell (2016: 16), bir makine öğrenme algoritmasının kodu ve verileri birbirinden ayrı olarak anlaşılabilirse bile, bilgisayarların bilgiyi çok farklı şekilde işlemesi nedeniyle aralarındaki etkileşimin insanlar için anlaşılabilir olduğunu savunmaktadır.

1.3.3. Tehlikeler ve Riskler

İlk olarak, yapay zekanın mahremiyeti tehdit etme riski vardır. Bu nedenle yapay zeka için geliştirilen etik yönergelerde büyük ilgi görmüştür. Hayes vd., (2020) yapay zekanın değere duyarlı tasarımında dikkate alınması gereken yedi temel değerden biri olarak gizlilikten bahseder. Fiziksel alanımıza ve özel bilgilerimize erişim sağlamanın ve düzenlemenin nasıl gizlilik kategorisine girdiğini özetler. Nissenbaum'un (2009) bilginin bağlamsal entegrasyonu olarak tanımladığı özgüllük kavramını uygulamışlardır ve bu, kişisel bilgilerimizin, kendimizi içinde bulduğumuz bağlamın özelliklerine bağlı olarak, uygun ilkeler çerçevesinde uygun aktörler tarafından iletilmesi durumunda gizliliğimize saygı gösterilmesini gerektirir. Hayes vd., (2020: 541), AI'nın adli bir sistemde kullanımının, kişisel verilerin bağlamlar arasında aktarımı ve grupların oluşturulması ve sınıflandırılması yoluyla bilgilerin bağlamsal entegrasyonu olarak mahremiyeti nasıl tehdit edebileceğini daha ayrıntılı olarak tartışır. Kıdemli AI uzmanlarından oluşan grup, bireysel özgürlüğün etik ilkesinin bir parçası olarak mahremiyet ve mahremiyet hakkını tartışmıştır. Ayrıca, güvenilir yapay zeka için yedi gereksinimden biri olarak gizlilik ve veri yönetimini de içermektedir. Bu gizlilik ve veri yönetimi ilkesi, zararı önleme ilkesiyle yakından ilgilidir ve özel hayata, veri kalitesine ve bütünlüğüne saygıyı ve verilere erişimi içerir. yapay zeka sistemlerinin sistemle etkileşimi yoluyla zaman içinde kullanıcı hakkında üretebileceği bilgilere ek olarak çevresel riskler bulunmaktadır.

Yapay zekanın geliştirilmesi ve kullanılması, bilgisayarlar ve çok fazla bilgi işlem gücü gerektirir. Ensmenger (2018), bilgisayar ve internet kullanımının fiziksel yönlerine

odaklanarak bilgi işlemin çevresel geçmişini analiz etmiştir. 2003 yılında masaüstü bilgisayar üretiminin insan emeği hariç 240 kilo fosil yakıt, 22 kilo kimyasal ve 1500 kilo suya mal olduğuna dikkat çekmiştir. Ayrıca internet üzerinden veri depolamak ve iletmek için çok fazla kaynağa ihtiyaç vardır. Ensmenger (2018: 9), yalnızca Google'ın veri merkezlerinin 2011'de 2,3 milyar kilovat saatten fazla elektrik kullandığını belirtmiştir.

Strubell, Ganesh ve McCallum (2020), doğal dil işleme (NLP) modellerini eğitmek için gereken enerjiyi hesaplayarak makine öğrenimi algoritmalarının eğitiminin çevresel maliyetine bakmıştır. NLP yeteneklerine sahip makine öğrenimi algoritmaları, konuşulan veya yazılı dili tanıyabilir ve kavrayabilir. Son birkaç yılda işlem gücündeki artışa bağlı olarak bu algoritmaların doğruluğu önemli ölçüde arttı. En yaygın kullanılan NLP modellerinin tahmini CO2 emisyonları, donanım, tüketilen enerjiye ve algoritma eğitiminin uzunluğuna bağlı olarak 192 ila 626.166 psi arasında değişmektedir. Buna karşılık, ömrü boyunca tipik bir araç, yakıt dahil 126.000 pound karbondioksit emisyonu üretmektedir.

Son olarak, yapay zekanın toplumu nasıl etkileyeceğine ilişkin tartışmalarda sıklıkla dile getirilen ancak bu makalede ele alınmayan ek tehlikeler de vardır. Yapay zeka uygulamalarının teknolojik güvenilirliği, güvenliği ve hassasiyetiyle bağlantılı risklerin, yapay zekanın kullanımında ve geliştirilmesinde çok önemli bir faktör olduğu sıklıkla dile getirilmektedir.

Yapay zekanın mülkiyet üzerindeki olası etkisi de Hayes ve diğerleri tarafından dikkate alınması gereken bir tehlike olarak tanımlanmıştır. Felsefi literatürde referans alınsa da bu iki tehdit çoğu zaman derinlemesine felsefi tartışmaların odağı olmamıştır.

2. SENTETİK SES ÜRETİMİ

Sentetik ses üretimi, son yıllarda özellikle medya endüstrisinde çeşitli kullanım alanlarıyla dikkat çekmektedir. Bu süreçte, belirli bir konuşmacının sesini taklit edebilen, yani ses klonlama teknolojisi önemli bir yer tutmaktadır. Ses klonlama, derin öğrenme tekniklerinin yardımıyla, gerçek bir bireyin ses özelliklerine yakın sentetik ses üretimini amaçlar. Örneğin, bu sesler metin okuma, sesli kitap anlatımı veya dublaj gibi alanlarda kullanılabilir (Docasal & Alvarez, 2023).

Ses sentezi konusunda önemli gelişmeler kaydeden NVIDIA gibi firmalar, RAD-TTS gibi modeller geliştirerek, kullanıcıların metinleri belirli bir ses tonunda ve duygusal ifadeyle sentezleyebilmelerini sağlamaktadır. Bu tür modeller, sesin tonlamasını, süresini

ve enerjisini kullanıcıların kontrol etmesine olanak tanır, böylece üretilen sesin daha doğal ve ifade gücü yüksek olmasını sağlar (Shih vd., 2021).

Öte yandan, sentetik seslerin geliştirilmesi, büyük veri setleri ve karmaşık algoritmalar gerektirir. Bu süreçte, ses kalitesini değerlendirmek için kullanılan Ortalama Görüş Puanı (Mean Opinion Score-MOS) gibi metrikler, sentetik sesin kalitesini belirlemek için önemlidir. Bu, ses dosyalarının akustik koşullarının optimal olmasını ve eğitim materyalinin ses kalitesinin yüksek olmasını gerektirir (Ning vd., 2019).

Bu teknolojiler, özellikle çeşitli dillerde ve aksanlarda yüksek kaliteli ses üretimi yapabilen firmalar tarafından sürekli olarak geliştirilmektedir. Örneğin, Amazon ve Google gibi büyük teknoloji firmaları, özelleştirilmiş ses hizmetleri sunarak kullanıcıların kendi benzersiz seslerini oluşturmalarına olanak tanır. Bu hizmetler, kullanıcının ses verilerini modellemek ve daha yüksek kalitede bir ses üretimi yapmak için gerekli destekleri içerir (Dale, 2023).

Bu gelişmeler, sentetik ses teknolojisinin sadece teknolojik bir başarı olmakla kalmayıp, aynı zamanda çeşitli pratik uygulamalarda gerçek bir fayda sağlayabilecek potansiyele sahip olduğunu göstermektedir.

2.1. KONUŞMA SENTEZİ (TEXT-TO-SPEECH-TTS)

Konuşma sentezi (Text-to-Speech, TTS), metin tabanlı bilgileri otomatik olarak konuşmaya dönüştüren teknolojidir. Bu dönüşüm, metni okuyan sesli asistanlar, konuşan kitaplar gibi birçok uygulamada kullanılır, özellikle de bilgisayar destekli dil öğreniminde (Computer-Aided Language Learning, CALL) yaygındır (Handley, 2009). TTS teknolojisi, bilgisayar konuşma sentezinin üç ana yöntemini içerir: birleştirmeli, artikülatif ve formant sentezi. Her bir teknik, bilgisayarın konuşma üretme sürecinde farklı yaklaşımlar ve teknikler sunar.

Birleştirmeli Konuşma Sentezi: Bu yöntem, önceden kaydedilmiş ses örneklerini kullanarak konuşma üretir. Konuşmacıların ses kayıtları toplanır, bu kayıtlar istenen konuşmayı oluşturacak şekilde birleştirilir. Bu yöntemin avantajı, son derece doğal ve gerçekçi bir ses üretimi sağlamasıdır. Ancak, önceden kaydedilmiş ses örneklerine dayanması nedeniyle, üretilen seslerin esnekliği sınırlıdır.

Artikülatif Konuşma Sentezi: Bu teknik, konuşmanın fiziksel üretim sürecini taklit eder ve böylece konuşma organlarının hareketlerini simüle eder. Yeni fikirler

üretebilme ve duyguları daha iyi ifade edebilme gibi avantajlara sahiptir, ancak bu yöntem hesaplama yoğunluğu gerektirir ve karmaşık modellemeleri içerir.

Formant Konuşma Sentezi: Bu yaklaşım, konuşma sırasında sesin belirli frekanslarındaki formantları oluşturarak çalışır. Formantlar, konuşmanın akıcılığını ve anlamını önemli ölçüde etkiler. Bu yöntem, daha az hesaplama gerektirir ve gerçekçi bir konuşma sağlayabilir; ancak, bazı durumlarda diğer yöntemlere kıyasla ses doğallığı ve duygusal ifade daha sınırlı olabilir.

TTS teknolojisinin gelişimi önemli ölçüde ilerlemiştir ve bu ilerleme, teknolojiyi sürekli olarak daha verimli hale getirmiştir (O'Cinneide, Dorran & Gainza, 2007). Sharma'ya göre (2007), bir TTS sistemi genellikle iki ana kısımdan oluşur: ön uç ve arka uç. Ön uç, metni işlemek ve sesli bir şekle dönüştürmek için gerekli dil işleme ve özellik çıkarma işlemlerini gerçekleştirir. Arka uç ise, bu bilgileri kullanarak sentezlenmiş sesleri üretir ve düzenler.

TTS teknolojisi, kullanıcıların metin tabanlı içerikleri sesli olarak deneyimlemesini sağlar, bu da özellikle görme engelli bireyler ve çoklu görevler arasında geçiş yaparken bilgiye erişmeyi kolaylaştıran kullanıcılar için değerlidir. Bu teknolojinin sunduğu esneklik ve erişilebilirlik, onu günümüzde ve gelecekteki dijital iletişim araçları arasında vazgeçilmez kılar.

Text-to-Speech (TTS) sistemleri, metni sesli dile çevirme sürecini, metnin temel analizi ve ses sentezi aşamalarından geçirerek gerçekleştirir. Her iki aşama da metni anlaşılır ve doğal bir sesle dönüştürmek için kritik öneme sahiptir.

Temel Analiz Aşaması: Bu aşamada, TTS sistemleri gelen metni derinlemesine analiz eder. Öncelikle, metnin dilbilgisi yapısını, cümle yapısını ve vurgularını dikkate alır. Bu işlem, Doğal Dil İşleme (NLP) tekniklerini kullanarak gerçekleştirilir. Metnin anlamını ve bağlamını doğru bir şekilde kavrayabilmek için, NLP algoritmaları metindeki kelimelerin ve ifadelerin fonetik ve semantik özelliklerini çözümler. Daha sonra, fonetik analiz devreye girer; burada sistem, kelimelerin nasıl telaffuz edileceğini, vurgu ve tonlamaları nasıl uygulanacağını belirler. Bu analiz sonucunda elde edilen bilgiler, sonraki aşama olan ses sentezine hazır hale getirilir.

Ses Sentezi Aşaması: Ses sentezi, TTS'nin metni sesli dile çevirme sürecinin en önemli kısmıdır. Bu aşamada, metin analizinden elde edilen fonetik ve dilbilimsel özellikler kullanılarak son ses çıktısı üretilir. Modern TTS sistemleri, ses birimi tabanlı,

alana özgü ve birim seçimi gibi çeşitli yöntemlerle çalışır (Htun, Zin & Tun, 2015). Ses birimi tabanlı yöntemde, konuşma, temel ses birimlerine ayrılarak oluşturulur; bu birimler genellikle fonemlerdir. Alana özgü yöntem, belirli bir dil veya lehçenin karakteristik özelliklerini yansıtan ses modelleri kullanır, bu da TTS'nin o dili konuşan kullanıcılar için daha doğal ve tanıdık hale gelmesini sağlar. Birim seçimi yöntemi ise, metindeki her kelime veya ifade için en uygun ses birimlerinin seçilmesini içerir; bu sayede metnin doğru tonlama ve vurguyla seslendirilmesi sağlanır.

Ses Değişimi ve Çıktı: Ses sentezi sırasında, sesin tonlaması, vurgusu ve hızı gibi özellikleri ayarlanabilir. Ses değişimi, metindeki duygusal tonları ve anlamları iletme için tonlamada yapılan ayarlamaları içerir. Ses çıktısı ise, son ses dalgalarının kullanıcılara iletilmesi aşamasıdır. Bu aşamada, artikülatif, formant veya birleştirmeli yöntemler kullanılarak ses çıktısı üretilir ve kullanıcının duyabileceği ses dalgaları şeklinde sunulur.

Bu süreçlerin tamamı, TTS sistemlerinin metin tabanlı içerikleri kullanıcılara doğal, anlaşılır ve erişilebilir bir biçimde sunmasını sağlar. Kullanım alanları genişleyen TTS teknolojisi, özellikle eğitim, erişilebilirlik ve multimedya uygulamalarında büyük bir fayda sağlamaktadır. Bu teknoloji, bilgiye erişimi kolaylaştırarak özellikle engelli kullanıcılar için dijital içerikleri daha ulaşılabilir kılar.

2.2. MÜZİK VE SES ÜRETİMİ

Üretken yapay zeka teknolojileri, müzik ve ses üretimi alanında önemli ilerlemeler sağlamıştır. Bu teknolojiler, özellikle müzik üretiminde yaratıcılığı ve yeniliği desteklemek için giderek daha fazla kullanılmaktadır. Yapay zeka destekli müzik üreticileri, müzik yapımcılarının yeni müzikal fikirler üretmesine olanak tanıırken, aynı zamanda farklı müzik türleri ve stilleri arasında geçiş yapma imkanı sunar. Bu sistemler, bestecilerin ve yapımcıların yeni müzikal yapılar ve stiller oluşturmalarına yardımcı olurken, aynı zamanda özgün müzik kompozisyonları üretmelerine olanak tanıyan çeşitli araçlar sunmaktadır (Civit vd., 2022).

Müzik üretiminde kullanılan üretken yapay zeka sistemleri, var olan müzik verilerindeki desenleri analiz ederek benzer yeni müzikler üretebilir. Bu tür sistemler, belirli müzisyenlerin stillerine göre eğitilebilir ve bu sayede müzisyenlerin tercih ettiği tarzlarda müzik üretimi yapabilir. Yapay zeka müzik üreticileri, klasikten çağdaş türlere kadar geniş (Anantrasirichai & Bull, 2022).

Bu teknolojinin sunduğu başlıca avantajlar arasında, müzisyenlerin yeni müzikal fikirler üretmelerine yardımcı olması ve yazar bloğu gibi durumları aşmalarına destek olması yer alır. Ayrıca, müzik üretim sürecinde zaman ve kaynak tasarrufu sağlar. Yapay zeka kullanarak müzisyenler, sıfırdan orijinal kompozisyonlar oluşturmak için saatler harcamak yerine, hızlı ve etkili bir şekilde yeni müzikler üretebilir.¹

Üretken yapay zeka araçlarının müzik sanayi üzerindeki etkisi zamanla artacak ve bu araçları kullananlar, kendi yeteneklerini genişletme ve farklılaşma fırsatı bulacaklardır. Müzik üretiminde yapay zeka kullanımı, müzik endüstrisinde önemli bir yenilik olarak kabul edilmekte ve bu alanda kullanılmaya başlanması müzik yapımcıları için yeni kapılar açmaktadır (Civit vd., 2022).

Yapay zeka tabanlı müzik üretimi, belirli bir tarzda veya duygu durumunda müzik parçaları oluşturmak için gelişmiş algoritmalar ve teknikler kullanır. Bu yöntemin amacı, ritim, melodi, harmoni, enstrümantasyon ve diğer müzikal bileşenleri sentetik bir şekilde oluşturmaktır.²

2.2.1. Veri Analizi, Ritim ve Melodi Oluşturma

Yapay zeka tabanlı müzik üretimi, veri analizi, ritim ve melodi oluşturma süreçlerini kapsamlı bir şekilde içerir. Yapay zeka sistemleri, öğrenme algoritmalarını kullanarak büyük veri kümelerini analiz eder ve bu verilerden müzik parçalarının kalıplarını ve ilişkilerini belirler. Bu veri setleri, çeşitli müzik tarzlarından ve sanatçılardan alınan örnekleri içerebilir, bu da sistemlerin geniş bir yelpazede müzik üretmesine olanak tanır (Zhu vd., 2023).

Ritim ve tempoya dayalı oluşturma konusunda, yapay zeka sistemleri, belirlenen ritim ve tempo dikkate alınarak müzik parçaları üretir. Bu süreç, davul vuruşları ve zamanlamalar gibi temel bileşenleri kullanır. Ayrıca, müzik yapılarını ritim ve melodiye göre şekillendirebilmek için gelişmiş kodlama teknikleri kullanılır.³

Melodi oluşturma sürecinde ise, yapay zeka, belirli bir melodi veya tema etrafında dönen müzikal yapılar oluşturabilir. Bu, çeşitli notaların, aralıkların ve melodi

¹ Üretken yapay zeka teknolojilerinin müzik üretimi üzerindeki etkileri hakkında daha fazla bilgi için Bkz. <https://soundful.com/ultimate-guide-to-generative-ai-music-production/> Erişim Tarihi: 02 Mayıs 2024

² Üretken yapay zeka teknolojilerinin müzik üretimi üzerindeki etkileri hakkında daha fazla bilgi için Bkz. <https://www.stldigital.tech/blog/generative-ai-in-music-production-a-new-era-of-sound-and-composition/> Erişim Tarihi: 01 Mayıs 2024.

³ Yapay zeka tabanlı müzik üretimi, veri analizi, ritim ve melodi oluşturma süreçlerini hakkında daha fazla bilgi için Bkz. <https://openai.com/index/musenet/> Erişim Tarihi: 28 Nisan 2024.

kalıplarının bir araya getirilmesiyle gerçekleştirilir. Yapay zeka, bu süreçte, müziğin yapısal unsurlarını analiz ederek, uyumlu ve estetik açıdan zengin melodiler yaratma yeteneğine sahiptir (Li & Sung, 2023).

Bu teknolojiler, müzik üretiminde yaratıcılığı desteklemek ve yeni müzikal fikirlerin hızla geliştirilmesine olanak tanımak için son derece değerlidir. Yapay zeka tabanlı müzik üretimi, aynı zamanda farklı müzik türlerini keşfetme ve yeni müzikal formlar yaratma potansiyeline sahiptir. Bu süreçler, müzik endüstrisinde inovasyonun önünü açmakta ve sanatçılara yeni ifade biçimleri sunmaktadır.

2.2.2. Armoni, Enstrümantasyon ve Duygusal İçerik

Yapay zeka tabanlı müzik üretiminde harmoni, enstrümantasyon ve duygusal içerik oluşturma, algoritmaların müzikal yapıları derinlemesine anlamasını ve bu bilgileri yaratıcı süreçlerde kullanmasını gerektirir. AI sistemleri, farklı notaların bir araya gelmesiyle oluşan akor ilerlemelerini analiz edebilir ve bu bilgileri yeni ve özgün akor kombinasyonları oluşturmak için kullanabilir. Örneğin, MuseNet gibi sistemler, MIDI dosyalarındaki akor yapılarını analiz ederek yeni müzikal yapılar üretebilir ve bu süreçte belirli bir müzik tarzını veya bestecinin üslubunu taklit edebilir (Zhu vd., 2023).

Enstrümantasyon ve ses rengi konusunda ise, yapay zeka sistemleri çeşitli enstrüman seslerini sentetik olarak oluşturabilir ve bu sesleri bir araya getirerek zengin bir orkestral düzenleme sunabilir. Bu sistemler, her bir enstrümanın müzikal rolünü ve ses rengini, eserin bütünü içinde nasıl işlev göreceğini anlayabilir. Özellikle, OpenAI'nin MuseNet'i gibi araçlar, farklı enstrümanların kombinasyonlarını kullanarak dinleyicilere geniş bir yelpazede müzikal deneyimler sunar (Li & Sung, 2023).

Duygusal ve tematik içerik üretimi, AI'nin belki de en çarpıcı yeteneklerinden biridir. AI sistemleri, belirli duygusal tonları yansıtan müzik parçaları oluşturabilir. Bu, algoritmaların müzikteki duygusal ifadeleri tanıma ve bu ifadeleri yeni kompozisyonlarda kullanma yeteneği ile mümkündür. Örneğin, bir AI sistemi, hüzünlü veya neşeli duyguları uyandıran müzik parçaları üretebilir ve bu, hem melodi seçimleriyle hem de tempo ve harmoni kullanımıyla desteklenebilir (Novelli & Proksch, 2022).

Son olarak, kullanıcı tercihlerine dayalı uyarlamalar sayesinde, yapay zeka, bireysel zevklere uygun özelleştirilmiş müzik parçaları oluşturabilir. Bu, kullanıcının geçmiş dinleme alışkanlıklarından elde edilen verileri analiz ederek ve bu bilgileri yeni müzik önerileri oluşturmak için kullanarak gerçekleştirilir. Bu süreç, müzik öneri

sistemlerinin yanı sıra kişisel müzik deneyimlerini zenginleştirmede de kullanılabilir (Ciu vd., 2022).

Bu geniş kapsamlı yetenekler, yapay zekanın müzik üretiminde yalnızca bir araç olmadığını, aynı zamanda yaratıcı süreçlerde aktif bir katılımcı olabileceğini göstermektedir. Bu teknolojiler, müzik endüstrisinde devrim yaratma potansiyeline sahip olup, sanatçıların yaratıcı ifadelerini genişletmelerine olanak tanır.

2.3. SES EFEKTLERİ VE TASARIMI

Yapay zeka tabanlı ses efektleri ve tasarımı, özellikle sentetik seslerin gerçek seslerle karşılaştırılması bağlamında geniş bir yelpazede kullanım bulmaktadır. Bu alanda, üretken yapay zeka teknolojilerinin gelişimi, film, oyun ve interaktif medya gibi çeşitli alanlarda devrim yaratabilecek potansiyeli taşır. Sentetik ses üretimi ve tasarımı, yalnızca teknik bir beceri olmanın ötesinde, sanatsal bir ifade aracı olarak da giderek daha fazla kabul görmektedir.

AI tabanlı ses tasarım süreci, genellikle büyük veri setlerinden karmaşık ses desenleri çıkararak başlar. Yapay zeka teknolojileri, bu verileri kullanarak gerçekçi ses ortamları yaratma veya var olan sesleri geliştirme yeteneğine sahiptir. Örneğin, yapay zeka, bir şehir ortamındaki karmaşık sesleri modelleyebilir veya bir ormanın sakin seslerini yeniden yaratabilir. Bu tür simülasyonlar, kullanıcıların deneyimlerini zenginleştiren detaylar sunar (Novelli & Proksch, 2022).

Ses efektleri ve tasarımındaki yapay zeka uygulamaları, ayrıca, ses kalitesini artırma, arka plan gürültüsünü azaltma ve genel ses berraklığını iyileştirme gibi teknik iyileştirmeler sunar. AI, bu süreçlerde otomatik olarak gürültüyü azaltabilir veya ses seviyelerini dinamik olarak ayarlayabilir, böylece üretilen son sesin daha temiz ve daha net olmasını sağlar (Novelli & Proksch, 2022).

AI'nin ses tasarımında kullanımı, aynı zamanda zaman ve maliyet tasarrufu sağlar. Geleneksel ses tasarım süreçleri zaman alıcı ve maliyetli olabilirken, AI teknolojileri bu süreçleri hızlandırabilir ve daha az insan müdahalesi ile daha yüksek kaliteli sonuçlar elde edebilir. Bu, özellikle büyük ölçekli prodüksiyonlar için önemli bir avantajdır, çünkü AI, kısa sürede geniş kapsamlı ses tasarım projelerini yönetebilir (Ciu vd., 2022).

Yapay zeka, ayrıca, kullanıcı deneyimini kişiselleştirme potansiyeline sahiptir. Kullanıcıların tercihlerini ve geçmiş etkileşimlerini analiz ederek, kişiye özel ses ayarları sunabilir. Bu, özellikle interaktif uygulamalar ve oyunlar için değerlidir, çünkü

kullanıcıların eylemlerine ve tercihlerine bağı olarak ses ortamını dinamik olarak değıştirebilir (Novelli & Proksch, 2022).

Son olarak, yapay zeka destekli ses efektleri ve tasarımı, sanatçıların ve teknisyenlerin yaratıcılıklarını sergilemelerine olanak tanır. AI, sanatçılara daha önce erişilemeyen ses paletleri ve efektler sunarak, onların yaratıcı ifadelerini genişletmelerine yardımcı olur. Bu, özellikle yeni ve yenilikçi medya biçimlerinin keşfedilmesinde kritik bir rol oynar (Li & Sung, 2023).

Bu geniş kapsamlı yetenekler, yapay zekanın ses efektleri ve tasarım alanında devrim yaratmaya devam edeceğini göstermektedir. Teknolojinin ilerlemesiyle, AI, medya üretimi ve tüketimi süreçlerinde giderek daha etkin bir rol oynamaya başlamıştır ve bu trendin gelecekte de devam etmesi beklenmektedir.

2.3.1. Doğal Ses Taklidi ve Hayali Ses Efektleri

Yapay zeka (AI), doğal ses taklidi ve hayali ses efektleri yaratma konusunda devrim niteliğinde yetenekler sunmaktadır, bu da onu medya ve eğlence sektörlerinde önemli bir araç haline getirmiştir. AI, doğal dünya seslerini taklit edebilir ve tamamen yeni ses efektleri oluşturabilir, bu da yaratıcılığı artırır ve medya projelerinin daha immersif hale gelmesini sağlar (Stickdorn, 2019).

Yapay zeka teknolojileri, çeşitli doğal sesleri, örneğin rüzgarın esintisi, kuş cıvıltıları veya okyanus dalgalarının seslerini detaylı olarak taklit edebilme yeteneğine sahiptir. Bu teknoloji, gerçek ses kayıtlarından öğrenerek bu tür sesleri yaratır ve bu sesler, özellikle doğa belgeselleri veya rahatlama uygulamaları gibi uygulamalar için mükemmeldir. AI, bu sesleri üretirken, öğrenme algoritmaları aracılığıyla elde edilen verilerden yararlanır, bu da sonuçların doğruluğunu ve doğallığını artırır (Baltazar, 2018).

Öte yandan, AI, bilim kurgu ve fantastik türlerde yaygın olarak kullanılan hayali ses efektleri oluşturma kapasitesine de sahiptir. Bu tür sesler, geleneksel yöntemlerle üretilemeyen veya kaydedilemeyen sesleri içerir ve AI, bu sesleri özgün ve yenilikçi bir şekilde üretebilir. Örneğin, bir AI sistemi, fantastik bir yaratığın çığlıklarını veya bir uzay gemisinin motor seslerini tasarlayabilir. Bu ses efektleri, kullanıcıların ve dinleyicilerin deneyimlerini derinleştirir ve dijital içeriklere daha fazla katılım sağlar (Baltazar vd., 2019).

AI'nin ses tasarımındaki bu yetenekleri, medya üreticilerine ve sanatçılara, projelerinde kullanacakları ses paletlerini genişletme olanağı sunar. AI, ses efektleri ve

atmosfer yaratma süreçlerini dönüştürmekte ve bu süreçleri daha erişilebilir hale getirmekte, böylece profesyoneller kadar amatör içerik üreticileri için de yeni olanaklar sunmaktadır (Stickdorn, 2019).

Bu teknoloji, özellikle film, oyun ve interaktif medya gibi alanlarda, ses efektleri ve atmosfer yaratma süreçlerini dönüştürmekte ve bu süreçleri daha erişilebilir hale getirmektedir. AI destekli araçlar, ses efektleri oluşturmayı daha demokratikleştirerek, profesyoneller kadar amatör içerik üreticileri için de yeni olanaklar sunmaktadır. AI'nin ses tasarımındaki bu uygulamaları, medya üreticilerine ve sanatçılara sesler üzerinde daha derin kontrol sağlar ve dinleyiciler için daha zengin, daha çekici deneyimler yaratmalarına yardımcı olur (Baltazar, 2018).

2.3.2. Ses Ortamı Oluşturma ve Ses İyileştirme

Yapay zeka (AI), medya prodüksiyonlarında ses ortamı oluşturma ve ses iyileştirme konularında önemli yenilikler sunmaktadır. AI, film, televizyon, video oyunları ve interaktif medya gibi çeşitli alanlarda ses tasarımı yeniden şekillendirerek, ses atmosferleri yaratma ve ses kalitesini artırma yeteneğiyle dikkat çekmektedir.

AI teknolojileri, belirli bir sahne veya ortam için uygun ses atmosferini yaratmada büyük bir yetenek sergiler. Örneğin, bir orman sahnesi için AI, rüzgar sesleri, kuş cıvıltıları ve yaprak hışırtıları gibi doğal sesleri gerçekçi bir şekilde taklit edebilir. Şehir ortamları için ise trafik sesleri, kalabalık gürültüleri gibi kentsel ses efektleri üretilebilir (Anantrasirichai & Bull, 2022).

AI aynı zamanda ses iyileştirme ve restorasyon alanında da etkilidir. Eski veya hasar görmüş ses kayıtlarının restorasyonu, gürültü azaltma ve ses kalitesinin genel olarak iyileştirilmesi AI araçları ile mümkündür. Özellikle tarihi ses kayıtları veya eski film seslerinin restorasyonu gibi durumlarda, AI sesleri daha temiz ve net hale getirebilir, bu da arşiv materyallerinin korunması ve yeniden kullanılması için büyük önem taşır (Xu vd., 2013).

AI destekli ses tasarım araçları, geleneksel ses düzenleme süreçlerini daha hızlı ve verimli hale getirirken, aynı zamanda prodüksiyon ekiplerine zaman ve maliyet tasarrufu sağlar. Bu araçlar, genellikle gelişmiş algoritmalar ve derin öğrenme teknikleri kullanarak, büyük ölçekli projelerde bile dakikalar içinde sonuçlar üretebilir. Bu, özellikle zaman baskısı altında çalışan ekipler için büyük bir avantajdır (Xu vd., 2021).

Sonuç olarak, yapay zeka, ses ortamı oluşturma ve ses iyileştirme alanlarında, medya prodüksiyon süreçlerini dönüştürmektedir. AI destekli ses tasarımı, yalnızca profesyonel ses mühendislerine değil, aynı zamanda amatör içerik üreticilerine de yeni yaratıcı olanaklar sunmaktadır. Bu teknoloji, ses efektleri ve atmosfer yaratma süreçlerini daha erişilebilir ve esnek hale getirerek, medya ve eğlence sektörlerindeki etkisini artırmaya devam edecektir (Zheng & Zhang, 2018).

3. GELİŞMİŞ SES ÜRETİM MODELLERİ

Gelişmiş ses üretim modelleri, yapay zeka teknolojilerinin ses üretiminde sağladığı devrimsel yenilikleri temsil etmektedir. Bu modeller, sentetik seslerin gerçek seslerle karşılaştırılması ve analiz edilmesi konusunda önemli araçlar haline gelmiştir. Özellikle, WaveNet ve MelNet gibi modeller, doğal konuşma seslerini ve müzikal kompozisyonları üretebilmeleri ile bilinirler.

Bu modeller, derin öğrenme ve büyük veri setlerinden elde edilen bilgilerle beslenir. Örneğin, WaveNet, gerçek zamanlı konuşma sentezi için geliştirilmiş bir modeldir ve ses dalgalarını doğrudan üretebilme yeteneğiyle dikkat çeker. Bu model, ses dalgalarının her bir örneğini üretmek için önceki ses örneklerinden öğrenir ve son derece gerçekçi sesler üretebilir (Natsiou & O’Leary, 2021).

Öte yandan, MelNet, zaman ve frekans boyutlarına göre sesleri modelleyebilen daha karmaşık bir yapıya sahiptir. Bu model, çok katmanlı ses dokularını ve tonlamaları yakalayabilir, bu da onu hem konuşma hem de müzikal ses üretimi için uygun kılar. MelNet, ses üretiminde detay ve derinlik sağlama konusunda önemli bir yetenek sunar (Natsiou & O’Leary, 2021).

Bu teknolojiler, özellikle eğlence ve medya endüstrilerinde, gerçekçilik ve etkileşim açısından kullanıcı deneyimini büyük ölçüde iyileştirmektedir. Gelişmiş ses üretim modelleri, yapay zekanın ses ve müzik üretimindeki potansiyelini gözler önüne sermekte ve bu alandaki çalışmalara yön vermektedir. Ayrıca, bu modeller, robotik ve interaktif sistemlerde de kullanılarak, makinelerin çevreleriyle daha doğal ve anlamlı bir şekilde etkileşimde bulunmalarını sağlamaktadır (Simons, 2016).

Bu yenilikler, yapay zeka uygulamalarının ses ve müzik üretimi alanındaki ilerlemelerini ve bu teknolojilerin nasıl gerçek dünyadaki uygulamalara entegre edilebileceğini daha iyi anlamamıza yardımcı olmaktadır. Bu, yapay zekanın sanatsal

ifade ve yaratıcı süreçler üzerindeki etkisini artırarak, medya ve eğlence sektörlerinde yeni yaratıcı olanaklar açmaktadır.

3.1. WAVENET

WaveNet, Google DeepMind tarafından geliştirilen derin öğrenme tabanlı bir generatif modeldir ve doğal sesin gerçekçi bir şekilde sentezlenmesi konusunda çığır açmıştır. Bu model, özellikle konuşma sentezi ve müzik prodüksiyonu gibi alanlarda kullanılmak üzere tasarlanmıştır. WaveNet, dilatasyonlu konvolüsyonlar gibi özel mimari özellikler içerir, bu da onun zaman içinde geniş bir ses verisi aralığını kapsayacak şekilde öğrenmesine olanak tanır (Akiyama, 2019).

WaveNet'in temel özelliği, ses dalgalarını doğrudan modelleyebilmesidir. Bu model probabilistik bir yapıya sahiptir ve ses dalgalarını olasılık dağılımları aracılığıyla üretir. Bu yaklaşım, modelin gerçekçi ve doğal ses sentezleme yeteneğini artırır. Özellikle, insan konuşmasını taklit etme konusunda büyük başarılar göstermiştir ve bu alanda önceki tekniklere göre önemli bir ilerleme olarak kabul edilir (Shen vd., 2018).

Eğitim süreci, büyük miktarda etiketlenmiş ses verisi üzerinde gerçekleştirilir. WaveNet, bu verileri kullanarak ses örneklerini analiz eder ve zaman içinde ses dalgalarının nasıl değişebileceğini öğrenir. Bu süreç, modelin yeni ve benzer ses örnekleri üretebilmesi için kritik öneme sahiptir. WaveNet, eğitim sırasında elde ettiği bilgileri kullanarak, çok çeşitli ses türlerini sentezleyebilir (Oord vd., 2018).

WaveNet ayrıca, ses dalgalarını farklı seviyelerdeki detaylarıyla modelleyebilen hiyerarşik bir yapıya sahiptir. Bu özellik, modelin karmaşık ses yapılarını ve tonlamaları daha iyi anlamasını ve yorumlamasını sağlar, böylece üretilen sesler daha zengin ve detaylı hale gelir. Yüksek kalite ve gerçekçilik, WaveNet'in en belirgin avantajlarından biridir. Model, özellikle insan benzeri konuşma sentezinde ve müzik üretiminde kullanıldığında, önceki ses sentezi yöntemlerine göre daha doğal ve anlaşılır sonuçlar üretir. Bu, kullanıcı deneyimini önemli ölçüde iyileştiren bir faktördür (Akiyama, 2019).

WaveNet'in bir türevi olan Parallel WaveNet, ses sentezini daha hızlı ve verimli hale getiren paralel işleme tekniklerini kullanır. Bu model, özellikle gerçek zamanlı uygulamalar için geliştirilmiştir ve WaveNet'in sunduğu yüksek kaliteli ses üretimini daha hızlı bir şekilde sağlar. Parallel WaveNet, dijital asistanlar ve interaktif uygulamalar gibi anlık ses üretimi gerektiren durumlar için idealdir (Shen vd., 2018).

Sonuç olarak, WaveNet ve onun türevleri, yapay zeka destekli ses teknolojilerinin sınırlarını zorlamakta ve ses üretimi alanında yeni standartlar belirlemektedir. Bu modeller, sentetik seslerin gerçek seslerle karşılaştırılması ve analiz edilmesi konusunda önemli araçlar haline gelmiştir, bu da onların medya, eğlence ve iletişim teknolojileri sektörlerindeki etkisini artırmaktadır.

3.2. MELNET

MelNet, tanıtılan "MelNet: A Generative Model for Audio in the Frequency Domain" başlıklı makalede ele alınan, sesin frekans alanında modellenmesi için geliştirilmiş bir yapay zeka tabanlı generatif modeldir. Bu model, özellikle mel spektrogramları kullanılarak ses dalgalarını doğrudan modellemek amacıyla tasarlanmıştır, bu da onun ses sentezi konusundaki önceki yöntemlerden farklı olarak öne çıkmasını sağlar (Vasquez & Lewis, 2019).

MelNet, sesin frekans özelliklerini detaylı bir şekilde ele alır ve bu frekanslar üzerinden ses dalgalarını üretir. Model, özellik çıkarımı yaparak ses verisinin frekans domainindeki özellik vektörlerini çıkarır ve bu vektörler, sesin frekans bileşenlerini zaman içinde gösteren temsillerdir. Bu süreç, modelin ses özelliklerini daha iyi kavramasını ve yeni sesler üretmesini sağlar (Ji, Luo & Yang, 2020).

Eğitim programı, geniş bir ses veri seti üzerinden yürütülür. MelNet, bu verilerle eğitilirken, ses örnekleri arasındaki ilişkileri öğrenir ve bu bilgileri yeni ses örnekleri üretmek için kullanır. Bu karmaşık ilişkiler, modelin gerçekçi ses dalgaları üretebilmesi için kritik öneme sahiptir. Ses analizi sürecinde MelNet, girdi olarak alınan özellik vektörlerini kullanarak gerçekçi ses dalgaları üretir. Bu süreç, modelin ses sentezi kapasitesini artırarak, kullanıcıların doğal ve akıcı bir ses deneyimi yaşamasını sağlar. MelNet'in bu yetenekleri, onu film, video oyunları, müzik prodüksiyonu ve diğer medya projeleri için değerli bir araç haline getirir (Vasquez & Lewis, 2019).

Son olarak, MelNet, ses sentezi konusunda güçlü ve esnek sonuçlar üretebilen bir yapay zeka modelidir. Frekans domaininde çalışması, modelin sesin daha spesifik özelliklerini daha etkili bir şekilde işlemesine olanak tanır. Bu özellikler, konuşma sentezi, müzik üretimi ve ses efektleri üretimi gibi birçok alanda kullanılabilir, bu da modelin geniş bir uygulama yelpazesine sahip olmasını sağlar.

4. SES KLONLAMA TEKNOLOJİLERİ

Ses klonlama teknolojileri, yapay zeka (AI) ve makine öğrenimi algoritmalarını kullanarak bir kişinin özgün sesini sentetik olarak yeniden oluşturma sürecidir. Bu teknolojiler, özellikle derin öğrenme tekniklerinden yararlanarak, hedeflenen sesin karakteristik özelliklerini öğrenir ve bu özellikleri kullanarak doğrulukla sesleri yeniden üretebilir (Bateman, 2022).

Ses klonlama süreci genellikle birkaç adımda gerçekleşir. İlk adım, hedef sesin çeşitli örneklerinin toplanmasıdır. Bu örnekler, sesin farklı duygusal durumlarını ve tonlarını kapsayacak şekilde çeşitlendirilir. Toplanan bu veriler, sesin temel özelliklerini—pitch, ton, kadans ve timbre gibi—belirlemek için analiz edilir. Bu özellikler, sesin benzersiz niteliklerini yansıtan önemli faktörlerdir (Lee & Qiufan, 2021).

Eğitim süreci, çeşitli derin sinir ağları kullanılarak gerçekleştirilir. Bu ağlar, ses özelliklerini modelleyerek ve tahmin ederek sentetik ses üretimi yapar. Gelişmiş metinden-konuşmaya (TTS) sistemleri ve Üretici Karşıt Ağlar (GAN'lar) bu sürecin bir parçasıdır. GAN'lar, üretilen ses örneklerinin gerçekçiliğini artırmak için kullanılır; bir üreteç ses örnekleri oluştururken, bir ayırt edici bu örneklerin gerçek ses kayıtlarıyla ne kadar benzer olduğunu değerlendirir ve geri bildirim sağlar (Bateman, 2022).

Ses klonlama teknolojisinin kullanım alanları oldukça geniştir. Erişilebilirlik için konuşma yetisini kaybetmiş bireyler için iletişim cihazlarında kendi seslerinin yeniden oluşturulmasını sağlayabilir. Eğlence ve medya sektörlerinde, seslendirme sanatçılarının devamlılığını sağlamak veya tarihsel figürlerin seslerini yeniden canlandırmak için kullanılır. Ayrıca, kişiselleştirilmiş pazarlama ve müşteri hizmetleri gibi alanlarda da etkili bir araçtır (Lee & Qiufan, 2021). Ancak, ses klonlama teknolojilerinin etik ve yasal sorunları da bulunmaktadır. Özellikle, bu teknolojinin kötüye kullanılma potansiyeli, kişisel izinsiz kullanım ve kimlik hırsızlığı gibi ciddi riskler taşır. Bu nedenle, ses klonlama işleminde şeffaflık ve rıza oldukça önemlidir; bireylerin seslerinin nasıl kullanılacağı konusunda tam kontrol sahibi olmaları gerekmektedir.

Sonuç olarak, ses klonlama teknolojileri, doğru kullanıldığında ve etik kurallara uygun şekilde yönetildiğinde, birçok faydalı uygulamayı mümkün kılar. Ancak bu teknolojilerin geleceği hem teknolojik gelişmelerin hem de etik standartların dengeli bir şekilde ilerlemesine bağlıdır.

4.1. VERİ TOPLAMA VE YÜKSEK KALİTE ÜRETİMİ

Ses klonlama teknolojilerinde veri toplama süreci, modelin hedeflenen kişinin sesini doğru bir şekilde öğrenmesi ve taklit etmesi için kritik öneme sahiptir. Bu süreç, geniş bir ses verisi setinin toplanmasını ve analiz edilmesini gerektirir. Ses klonlama için kullanılan veriler genellikle çeşitli tonlamalar, duygusal ifadeler ve dilbilgisel yapılar içeren ses kayıtlarından oluşur. Bu aşamada, konuşmacının ses özelliklerini yansıtabilecek metinler dikkatlice seçilir ve konuşmacı bu metinleri seslendirir. Kaydedilen bu ses örnekleri, modelin eğitimi için temel teşkil eder. Veri toplama süreci, sesin farklı yönlerini kapsayacak şekilde çeşitlendirilmelidir ki model, sesin tonundan duygusal tonlamasına kadar geniş bir yelpazede öğrenebilsin (Zhu vd., 2023).

Kayıt kalitesi, ses klonlama teknolojisinin başarısı için hayati önem taşır. Yüksek kaliteli mikrofonlar ve ses yalıtımlı stüdyolar, arka plan gürültüsünün en aza indirgenmesi ve kaydedilen sesin doğruluk derecesinin artırılması açısından önemlidir. Bu, özellikle profesyonel ses klonlama uygulamalarında, sesin doğallığını ve gerçekçiliğini korumak için gereklidir (Docasal & Alvarez, 2023).

Verilerin etiketlenmesi süreci, ses örneklerinin hangi metni içerdiğini açıkça belirtir. Bu etiketleme, modelin ses verilerini daha iyi anlamasına ve öğrenmesine yardımcı olur. Doğru etiketlenmiş veriler, modelin eğitim sürecinin verimliliğini ve sonuçlarının doğruluğunu artırır (Docasal & Alvarez, 2023).

Ses klonlama teknolojisinde kullanılan programlama dilleri ve kütüphaneler, modelin geliştirilmesi ve uygulanmasında önemli bir rol oynar. Python gibi diller ve TensorFlow, PyTorch gibi kütüphaneler, ses klonlama modellerinin etkin bir şekilde geliştirilmesini ve uygulanmasını sağlar. Bu araçlar, modelin ses verilerini işlemesine ve öğrenmesine olanak tanır, böylece daha doğal ve etkileyici ses sentezleri gerçekleştirilebilir (Zhu vd., 2023).

Son olarak, ses klonlama teknolojisinin gelişimi, ses verilerinin temizlenmesi ve analizi ile devam eder. Gürültüler, bozukluklar veya istenmeyen ses unsurlarının temizlenmesi, modelin daha temiz ve net sesler üretmesini sağlar. Bu süreç, ses klonlama teknolojisinin kalitesini ve etkinliğini doğrudan etkiler ve modelin eğitimi sırasında önemli bir adımdır.

4.2. HOPARLÖR KODLAMASI VE UYARLAMASI

Hoparlör kodlaması ve uyarlaması, ses klonlama sürecinde hayati öneme sahip iki tekniktir. Bu teknikler, ses klonlama modellerinin klonlamak istedikleri kişinin ses özelliklerini daha etkili bir şekilde öğrenmelerine ve taklit etmelerine yardımcı olur. Özellikle, hoparlör kodlaması (speaker encoding) ve hoparlör uyarlaması (speaker adaptation) olarak iki ana süreçten bahsedilmektedir.

Hoparlör Kodlaması sürecinde, model bir konuşmacının ses özelliklerini öğrenmek ve temsil etmek için kullanılır. Bu aşamada, konuşmacının sesinden elde edilen özellikler bir vektör halinde kodlanır. Bu kodlama işlemi, konuşmacının benzersiz ses niteliklerini bir dizi sayısal değer olarak temsil eder, böylece model bu bilgileri yeni ses sentezleme işlemleri için kullanabilir (Zhu vd., 2023).

Hoparlör Uyarlaması ise, öğrenilen bu hoparlör kodlarının farklı konuşma örneklerine veya farklı konuşmacılara uyarlanmasını sağlar. Bu süreç, modelin mevcut öğrenilmiş bilgileri genelleyerek yeni ve az miktarda veri ile eğitim görmüş konuşmacıların seslerini taklit etmesine olanak tanır. Bu, özellikle veri kısıtlı durumlar için idealdir, çünkü modelin etkinliğini ve esnekliğini artırır (Li & Sung, 2023).

Bu iki süreç, ses klonlama teknolojisinde önemli bir yer tutar çünkü sesin doğruluğunu ve gerçekçiliğini artırırken, aynı zamanda sistemin verimliliğini ve uygulama alanlarını genişletir. Örneğin, az miktarda ses örneğiyle dahi yüksek kalitede ses klonlama yapılabilmesi, bu tekniklerin etkinliğini gösterir.

Ses klonlama teknolojisindeki son gelişmeler, minimum girdi verisi ile maksimum çıktı kalitesi sağlama üzerine odaklanmıştır. Neekhara vd., (2021) tarafından sunulan sentezlenmiş konuşmanın ifade gücünü kontrol etmek için geliştirilen yöntemler, bu alandaki yeniliklere örnek teşkil eder. Aynı zamanda, Zhao & Chen (2020) tarafından önerilen gerçek zamanlı ses klonlama yaklaşımı, az sayıda örnek ile çalışabilen sistemlerin potansiyelini ortaya koymaktadır.

Arik vd. (2018) ve Blaauw vd. (2019) gibi araştırmacılar da hoparlör kodlaması ve uyarlaması tekniklerini kullanarak, şarkı söyleme sentezi gibi alanlarda veri verimliliği sağlayan çalışmalar yapmışlardır. Bu çalışmalar, ses klonlama teknolojilerinin veri kullanımını en üst düzeye çıkararak, geniş bir uygulama yelpazesinde yüksek kaliteli ses üretimi sağlayabileceğini göstermektedir.

Sonuç olarak, hoparlör kodlaması ve uyarlaması, ses klonlama teknolojilerindeki veri verimliliğini ve performansı artırmada kritik rol oynamaktadır. Bu teknikler, daha az veriyle daha doğru ve gerçekçi ses üretimi yapılmasını mümkün kılarak, ses klonlama uygulamalarının daha geniş bir kontekste kullanılmasına olanak tanır.

5. MAKİNE ÖĞRENİMİ VE SES ÜRETİMİ

Makine öğrenimi ve yapay zeka (AI), ses üretimi ve klonlama süreçlerinde temel teknolojiler olarak ön plana çıkmaktadır. Bu teknolojiler, sesin zaman içindeki bağlamını ve nüanslarını anlamak için derin öğrenme metodolojilerini kullanır, bu da onları ses klonlama ve ses sentezleme için ideal hale getirir.

Ses klonlama uygulamalarında, özellikle rekürrensiz ağırları (RNN) ve uzun kısa vadeli bellek (LSTM) ağırları gibi yapay sinir ağırları kullanılır. Bu ağırlar, sesin zaman içerisindeki değişimlerini modelleyebilir ve bu sayede ses özelliklerini daha gerçekçi bir şekilde yeniden üretebilirler. RNN'ler, geçmiş bilgileri saklayabilir ve bu bilgileri yeni veri ile birleştirerek sürekli bir öğrenme süreci sağlar (Kane, Johnstone & Szewczyk, 2024).

Özellikle, LSTM ağırları uzun vadeli bağlantıları sürdürebilir ve bu sayede sesin tonlama, vurgu ve ritim gibi uzun vadeli özelliklerini koruyabilir. Bu özellikler, ses klonlama kalitesini önemli ölçüde artırır ve konuşmacının sesinin doğal tonlarını ve duygusal ifadelerini daha iyi kavramayı sağlar (Mendoza vd., 2022).

Makine öğrenimi, ses klonlama modellerinin geliştirilmesinde de önemli bir rol oynar. Bu modeller, toplanan ses verilerini analiz eder ve bu verilerden öğrenirler. Eğitim süreci sırasında, bu modeller çeşitli ses örnekleri üzerinde çalışarak, sesin nasıl üretileceğini öğrenirler. Bu süreç, ses özelliklerinin doğru bir şekilde kavranmasını ve modelin bu özellikleri kullanarak yeni ve benzer sesler üretmesini sağlar (Chandra, Tata & Anand, 2022).

Eğitimden sonra, modeller belirli metinleri seslendirmek veya belirli bir tonlama ile konuşmak gibi görevlerde kullanılabilir. Bu aşamada, öğrenilen ses özelliklerine dayanarak sentetik ses üretimi gerçekleştirilir. Bu sentetik sesler, gerçek insan sesine oldukça benzer olabilir ve çeşitli uygulamalarda kullanılır, örneğin metin tabanlı haber okuma, sesli kitaplar ve sesli asistanlar gösterilebilir (Kane, Johnstone & Szewczyk, 2024).

Modelin sürekli olarak iyileştirilmesi ve optimize edilmesi, ses klonlama teknolojisinde önemli bir yere sahiptir. Bu süreç, modelin ses kalitesini artırmak, kullanıcı deneyimini iyileştirmek ve daha doğal bir konuşma akışı sağlamak için gereklidir. Modelin performansını değerlendirmek için kullanılan objektif metrikler, üretilen sesin gerçeklik derecesini ve kalitesini ölçer (Mendoza vd., 2022).

Makine öğrenimi ve yapay zeka modelleri, ses klonlama ve ses üretimi alanlarında devrim yaratmaktadır. Bu teknolojiler sayesinde, ses üretimi daha verimli, doğru ve etkili bir şekilde gerçekleştirilebilmekte, ses klonlama uygulamaları daha geniş bir kullanım alanına sahip olmaktadır.

5.1. REKÜRRENSİNİR AĞLARI (RNN) VE UZUN-KISA VADELİ BELLEK (LSTM)

Rekürrensiner ağları (RNN) ve Uzun-Kısa Vadeli Bellek (LSTM), ses klonlama ve ses üretimi teknolojilerinde yaygın olarak kullanılan iki temel yapay sinir ağı tipidir. Bu teknolojiler, özellikle zaman serileri verileri ve sekansların işlenmesinde etkilidir, bu da onları ses gibi sürekli değişen veri akışlarını işlemek için ideal kılar (Mendoza vd., 2022).

RNN'ler, geçmişte alınan bilgileri hafızada tutabilir ve bu bilgileri yeni verilerle birleştirerek sürekli bir öğrenme süreci sağlar. Bu özellik, RNN'leri zaman içinde değişen ses özelliklerini modellemek için son derece uygun kılar. Örneğin, bir konuşmacının sesindeki ton değişiklikleri ve vurguları, RNN tarafından zaman içinde izlenebilir ve modelin bu öğrenilen bilgileri kullanarak gerçek zamanlı olarak ses üretmesi veya klonlaması mümkün olur. Ancak, RNN'ler uzun vadeli bağlantıları korumada zorlanabilirler. Bu sorunu "tutarsızlık sorunu" veya "uzun vadeli bağımlılıkların kaybı" olarak bilinir ve RNN'lerin veri serisinin başındaki bilgileri zamanın ilerleyen kısımlarında hatırlamasını zorlaştırır (Chandra, Tata & Anand, 2022).

LSTM ağları ise, RNN'nin bu temel sorununa çözüm sunar. LSTM'ler, RNN'nin temel yapısına ek olarak, bilgiyi uzun süreler boyunca saklayabilen ve gerektiğinde bu bilgiyi geri çağırabilen "hücre durumu" adı verilen bir bileşen içerir. Bu yapıyla, LSTM'ler sesin tonlama, vurgu ve ritmini uzun vadeli bağlamlarda koruyabilir ve böylece daha doğru ve doğal ses klonlaması sağlayabilir. Ses klonlama kalitesinde bu önemli artış, özellikle uzun konuşmalar veya duygusal ifade gerektiren senaryolar için kritik öneme sahiptir (Lai vd., 2016).

LSTM ve RNN'lerin uygulandığı ses klonlama sistemleri, konuşmacının sesinin doğru bir şekilde yeniden üretilmesi için gerekli olan dinamik özellikleri öğrenebilir. Bu sistemler, konuşmacının ses örneklerinden sesin nasıl üretileceğini öğrenir ve bu bilgiyi kullanarak yeni metinleri konuşmacının benzersiz ses tonu ve stilinde seslendirebilir. Bu teknolojilerin ses klonlama ve üretimine uygulanması, özellikle medya, eğitim ve müşteri hizmetleri gibi alanlarda çeşitli uygulamalar açar. Sesli asistanlar, animasyon karakterleri için ses üretimi, sesli kitap okuma ve hatta interaktif oyuncaklar gibi ürünler, bu teknolojiler sayesinde daha etkileşimli ve erişilebilir hale gelmiştir (Mendoza vd., 2022).

Ses klonlama ve sentezlemede kullanılan bu gelişmiş ağlar, sürekli geliştirilmekte ve optimize edilmektedir. Kullanıcı deneyimini iyileştirmek, ses kalitesini artırmak ve daha doğal bir konuşma akışı sağlamak için sürekli olarak yenilikler ve iyileştirmeler yapılmaktadır. Sonuç olarak, RNN ve LSTM teknolojileri, yapay zeka destekli ses klonlama ve üretimi süreçlerinde devrim yaratmakta ve bu alanlardaki yeniliklerin önünü açmaktadır.

5.2. EĞİTİM SÜREÇLERİ VE OPTİMİZASYON

Ses klonlama ve ses üretimi teknolojilerinde eğitim süreçleri ve optimizasyon, yapay sinir ağlarının etkin şekilde kullanılabilmesi için kritik öneme sahiptir. Bu süreçler, ses modellerinin, sesin doğasını ve dinamiklerini nasıl kavrayıp taklit edeceğini öğrenmesini sağlar. Eğitim süreci, modelin ses özelliklerini anlamasını ve bu bilgileri kullanarak yeni ve benzer sesler üretmesini mümkün kılar. Model eğitimi, genellikle büyük ve karmaşık ses veri setlerini kullanır. Bu veriler, modelin çeşitli ses örnekleri üzerinde çalışmasını ve bu örneklerden öğrenmesini sağlar. Ses klonlama ve üretimi için kullanılan yapay sinir ağları, sesin tonunu, ritmini ve diğer akustik özelliklerini öğrenir. Bu öğrenme süreci, sesin zaman içindeki bağlamını ve nüanslarını anlamada hayati rol oynar (Hu & Zhu, 2023).

Eğitim süreci sırasında, ses verileri üzerinde çeşitli algoritmalar uygulanarak modelin parametreleri ayarlanır. Bu parametreler, modelin ses verilerini nasıl işleyeceğini ve sesi nasıl üreteceğini belirler. Modelin başarılı olabilmesi için, veri setlerinin doğru şekilde etiketlenmiş ve düzenlenmiş olması gerekmektedir. Etiketleme, modelin hangi ses örneklerinin hangi çıktıları üretmesi gerektiğini anlamasını sağlar (Mendoza vd., 2022).

Optimizasyon süreci ise modelin performansını artırmak ve genellebilirliğini sağlamak için gerçekleştirilir. Bu süreç, modelin farklı koşullar altında nasıl performans

gösterdiğini test etmeyi ve gerekirse parametre ayarlarını yeniden yapılandırmayı içerir. Optimizasyon, modelin daha hızlı ve verimli bir şekilde eğitilmesini sağlayarak, gerçek zamanlı uygulamalarda kullanılabilirliğini artırır (Chandra, Tata & Anand, 2022).

Modelin sürekli iyileştirilmesi, ses üretimi teknolojilerinin gelişiminde büyük bir rol oynar. Bu iyileştirmeler, modelin daha doğru ve doğal sesler üretmesine olanak tanır. Kullanıcı deneyimini iyileştirmek, ses kalitesini artırmak ve daha doğal bir konuşma akışı sağlamak için yapılan optimizasyonlar, teknolojinin endüstriyel ve ticari uygulamalarda daha geniş bir kabul görmesini sağlar (Hu & Zhu, 2023).

Eğitim süreçleri ve optimizasyon, makine öğrenimi ve yapay zeka temelli ses klonlama ve üretimi teknolojilerinin temel taşlarından biridir. Bu süreçler sayesinde, ses teknolojileri sürekli olarak gelişmekte ve kullanıcıların ihtiyaçlarına daha iyi cevap verebilmektedir. Eğitim ve optimizasyon, bu teknolojilerin potansiyelini maksimize eder ve geniş bir uygulama yelpazesinde etkili sonuçlar elde edilmesini sağlar.

5.3. VERİ SETİ HAZIRLIĞI VE ETİKETLEME

Veri seti hazırlığı ve etiketleme, ses klonlama gibi yapay zeka uygulamaları için hayati öneme sahiptir. Etkili bir ses klonlama modeli oluşturmak için, modelin eğitileceği ve test edileceği verilerin doğru bir şekilde toplanmış, düzenlenmiş ve etiketlenmiş olması gerekir.

Ses klonlama teknolojisinde, farklı konuşmacıların ses örneklerini içeren geniş ve çeşitli bir veri seti gereklidir. Bu veriler, modelin çeşitli ses tonları, duygusal ifadeler ve dilbilgisel özellikleri öğrenmesine olanak tanır. Her ses örneğinin, hangi metni içerdiği ve hangi duygusal veya tonal özellikleri barındırdığı konusunda açık ve net bir şekilde etiketlenmesi önemlidir (Docasal & Alvarez, 2023).

Etiketleme süreci, ses örneğinin içeriğini, konuşmacının kimliğini, dilini ve konuşma hızını gibi bilgileri içerir. Bu bilgiler, ses örneklerinin model tarafından doğru bir şekilde tanınmasını ve işlenmesini sağlar. Modelin eğitimi sırasında, bu etiketler modelin öğrenme sürecini yönlendirir ve modelin doğru çıktıları üretmesine yardımcı olur (Hu & Zhu, 2023).

Optimal sonuçlar elde etmek için, veri setinin temizlenmesi ve normalleştirilmesi de önemlidir. Ses kayıtlarındaki arka plan gürültüsü, yankılanma ve diğer istenmeyen seslerin minimize edilmesi gerekir. Bu işlem, modelin yalnızca konuşma ile ilgili

özellikleri öğrenmesini ve sonuç olarak daha doğru ve anlaşılır ses sentezi yapmasını sağlar (Kheria, & Karani, 2023).

Veri seti hazırlığı ve etiketleme süreci, ses klonlama modellerinin geliştirilmesindeki başarının temelidir. Bu süreçler, modelin eğitilmesi, test edilmesi ve optimizasyonunda kritik rol oynar. İyi hazırlanmış ve etiketlenmiş bir veri seti, modelin performansını önemli ölçüde artırabilir ve ses klonlama teknolojilerinin daha geniş uygulama alanlarına adaptasyonunu kolaylaştırabilir.

5.4. OBJEKTİF METRİKLER VE SES KLONLAMANIN UYGULAMA ALANLARI

Objektif metrikler, ses klonlama teknolojisinde son derece önemlidir. Bu metrikler, ses klonlama modellerinin performansını değerlendirmek ve modelin ürettiği seslerin kalitesini, doğruluğunu ve doğallığını ölçmek için kullanılır. Bu ölçümler genellikle, gerçek ses örnekleriyle sentetik sesler arasındaki benzerlikleri ve farkları kapsamlı bir şekilde analiz eder (Docasal & Alvarez, 2023).

- Benzerlik Skoru: Ses klonlama teknolojisinde kullanılan objektif metriklerden biri, üretilen sesin gerçek sesle ne kadar benzer olduğunu ölçen benzerlik skorlarıdır. Bu skorlar, frekans, ton ve diğer ses özellikleri açısından analizler yapar.

- Anlaşılabilirlik: Klonlanmış sesin anlaşılabilirliği, sesin netliğini ve dinleyicilerin konuşmayı ne kadar iyi anladığını değerlendiren bir metriktir. Bu, özellikle eğitim ve asistan uygulamalarında kritik bir öneme sahiptir.

- Doğallık Derecesi: Sentezlenen seslerin doğallığı, klonlanmış sesin gerçek bir insan sesine ne kadar benzediğini ölçer. Bu metrik, sesin robotik ya da yapay olup olmadığını belirlemek için kullanılır.

- Duyusal Uygunluk: Ses klonlama uygulamalarında duygusal ifadenin doğruluğu, sentetik sesin duygusal durumu doğru bir şekilde yansıtıp yansıtmadığını değerlendirir. Bu, özellikle oyun ve eğlence sektörü için önemlidir.

Ses klonlama teknolojilerinin uygulama alanları oldukça geniştir. Metin tabanlı haber okuma, sesli asistanlar, sesli kitaplar ve eğitim uygulamaları bu teknolojinin kullanıldığı başlıca alanlardır. Ayrıca, dil öğrenme uygulamaları ve engellilere yönelik erişilebilirlik çözümleri de ses klonlama teknolojileri sayesinde gelişmektedir. Bu teknolojiler, kullanıcıların çeşitli dillerde ve diyalektlerde doğal ve anlaşılır seslerle etkileşim kurmasını sağlayarak, global iletişimde önemli bir rol oynar. Ses klonlama,

teknolojinin temel prensipleri ve gelişmiş uygulama alanlarıyla sürekli olarak evrim geçirmekte ve çeşitli sektörlerde yenilikçi çözümler sunmaktadır (Hu & Zhu, 2023).

6. YAPAY ZEKA VE SES İŞLEME

Yapay zeka (AI), insana benzer algılama, öğrenme, fikir yürütme, karar verme, sorun çözme ve iletişim kurma yetenekleri sergileyen teknolojileri içerir. Ses işleme, bu teknolojilerin etkili uygulama alanlarından biridir, özellikle de ses ve müzik üretimi, konuşma tanıma ve interaktif dijital asistanlar gibi alanlarda önemli rol oynamaktadır.

Ses işlemedeki yapay zeka uygulamaları, algoritmalar, modeller ve teknikler kullanarak karmaşık veri yapılarını analiz eder ve desenleri tanır. Bu süreç, bilgisayar sistemlerinin verilerden öğrenmesine ve kendi kendine gelişmesine olanak tanır. Özellikle derin öğrenme, yapay sinir ağlarından ilham alarak geliştirilmiş ve ses dalgalarının analiz edilmesi gibi karmaşık görevlerde başarılı olmuştur (Mah, Skalna & Muzam, 2022).

Yapay zeka, kişisel verileri kullanarak işlemlerini gerçekleştirdiği için güvenlik konusunda bazı endişeler doğurabilir. Özellikle, kötü amaçlı üçüncü kişiler tarafından veri ihlalleri yapay zekanın önemli bir dezavantajı olarak görülebilir. Bununla birlikte, yapay zeka, müzik alanında inovasyon sağlamış ve çeşitli enstrümanlar için yeni müzikal eserlerin yaratılmasına olanak tanımıştır. Örneğin, algoritmik bestecilik yapay zekanın müzikteki ilk uygulamalarından biridir (Parker & Dockray, 2023).

Ses işlemeyle ilgili temel kavramlar arasında, ses dalgalarının genlik, frekans ve faz gibi özellikleri yer alır. Bu özelliklerin analizi, sesin nasıl üretildiğini ve algılandığını anlamada kritik rol oynar. Yapay zeka tabanlı sistemler, bu özelliklerin daha detaylı analizini sağlayarak, örneğin spektrogramlar ve Hızlı Fourier Dönüşümü (FFT) kullanılarak ses sinyallerini daha doğru bir şekilde işleyebilir (Liu & Bao, 2014).

Yapay zeka ve ses işleme teknolojileri, hem günlük kullanım hem de akademik araştırmalar için zengin potansiyeller sunmaktadır. İleriye dönük olarak, bu teknolojilerin etik kullanımı ve güvenlik gibi alanlarda daha fazla gelişme beklenmektedir.

6.1. SES DALGALARININ TEMEL PRENSİPLERİ VE İŞLEME TERİMLERİ

Ses dalgaları, bir ortamda meydana gelen basınç değişiklikleri olarak tanımlanır ve genellikle havada mekanik dalgalar şeklinde yayılır. Bu dalgalar, kaynaktan çevreye enerji taşıırken ortamın partiküllerinin titreşimiyle ilerler. Ses dalgalarının temel

özellikleri arasında frekans, genlik ve dalga boyu bulunur. Frekans, bir saniyede gerçekleşen dalga sayısını ifade eder ve Hertz (Hz) birimiyle ölçülür. İnsan kulağı genellikle 20 Hz ile 20.000 Hz arasındaki frekansları duyabilir (Christensen, 2019).

Genlik, sesin gücünü veya şiddetini belirler ve genellikle desibel (dB) cinsinden ölçülür. Yüksek genlik, daha güçlü ses dalgalarına işaret eder. Faz, dalga formunun bir noktadaki özel durumunu tanımlar ve ses dalgalarının birbiriyle olan ilişkisini ve senkronizasyonunu anlamak için kullanılır (Kefauver & Patschke, 2007).

Ses işlemede kullanılan temel terimlerden biri olan spektrogram, ses sinyalinin zaman, frekans ve genlik bilgilerini görselleştirir. Bu görsel temsil, sesin nasıl değiştiğini ve farklı bileşenlerini anlamayı kolaylaştırır. Hızlı Fourier Dönüşümü (FFT), ses sinyallerini zaman domeninden frekans domenine dönüştürmek için kullanılan matematiksel bir işlemidir ve genellikle ses analizinde frekans bileşenlerini belirlemek için kullanılır (Lerch, 2012).

Bu temel prensipler ve terimler, yapay zeka ve ses işleme teknolojilerinde, özellikle müzik üretimi, ses tanıma ve interaktif sistemlerde kritik rol oynar. Yapay zeka tarafından yönlendirilen ses işleme algoritmaları, bu temel ses özelliklerini kullanarak daha doğal ve etkileşimli kullanıcı deneyimleri sunmayı amaçlar.

6.2. GENLİK, FREKANS VE FAZ ANALİZLERİ

Frekans, ses dalgalarının bir saniyede oluşturduğu titreşim sayısını ifade eder ve hertz (Hz) birimi ile ölçülür. İnsan kulağı genellikle 20 Hz ile 20.000 Hz arasındaki frekansları duyabilir. Frekans analizi, sesin ton yüksekliğini ve algılanan yüksekliği anlamak için hayati öneme sahiptir (Christensen, 2019).

Genlik, ses dalgasının enerji seviyesini yansıtır ve desibel (dB) cinsinden ölçülür. Genlik, sesin şiddetini veya ses basıncı seviyesini belirler. Sesin algılanan ses seviyesi genlikle doğrudan ilişkilidir (Kefauver & Patschke, 2007).

Faz analizi, ses dalgalarının zamana göre nasıl değiştiğini inceler. Faz, bir dalgaanın belirli bir noktadaki anlık durumunu ve dalgalar arasındaki ilişkiyi belirlemek için kullanılır. Faz farklılıkları, ses dalgalarının birleşiminden kaynaklanan girişim ve rezonans fenomenlerini açıklamada kullanılır (Kefauver & Patschke, 2007).

Bu üç parametre, akustik analizde temel rol oynar. Mühendislik, müzik teorisi, akustik tasarım ve ses tanıma sistemleri gibi birçok alanda uygulanır. Örneğin, konser

salonlarının akustik tasarımında bu analizler, sesin nasıl yayılacağını ve dinleyici tarafından nasıl algılanacağını belirlemede kritik öneme sahiptir.

Son çalışmalar, bu parametrelerin daha detaylı analizini sağlayarak, ses işleme teknolojilerindeki inovasyonları desteklemekte ve geliştirmekte önemli rol oynamaktadır. Bu, özellikle yüksek çözünürlüklü ses formatlarının geliştirilmesi ve 3D ses sistemlerinin optimizasyonu gibi alanlarda ses kalitesinin artırılmasına yardımcı olmaktadır.

6.3. NYQUIST TEOREMİ VE UYGULAMALARI

Bir ses sinyalinin örneklenmesi için gereken en düşük örnekleme frekansı. Teoreme göre, örnekleme frekansının en az iki katı olması gerekir. Bir dizi çalışma, ses işlemede spektrogramların ve FFT'nin kullanımını araştırdı. Umapathy, Ghoraani & Krishnan (2010), özellikle sıkıştırma, sınıflandırma ve koruma gibi uygulamalarda ses sinyallerini etkili bir şekilde analiz etmek için ortak bir zaman-frekans yaklaşımının gerekliliğini vurgulamaktadır. Henderson (2004), SmaartLive ölçüm platformundaki uygulamalarına odaklanarak FFT tabanlı analiz tekniklerine pratik bir genel bakış sunmaktadır. Pucik, Kubinec & Ondracek (2014), FFT kullanarak ses sinyali analizi için frekans bileşenlerinin daha doğru bir temsilini sağlayabilen değiştirilmiş bir frekans ölçeği sunar. Brigham (1988), FFT'nin çeşitli alanlardaki geniş uygulamasını tartışarak sinyal işlemedeki çok yönlülüğünü vurgulamaktadır. Bu çalışmalar topluca, ses sinyallerinin anlaşılmasında ve işlenmesinde spektrogramların ve FFT'nin önemini altını çizmektedir.

Yapay zekâ daki son gelişmeler, sesleri yeniden üretebilen ve doğallığı koruyabilen modellerle sentetik ses üretiminin kalitesini önemli ölçüde artırmıştır (Mahajan vd., 2021). Bu modeller, doğal ses dalgalarını yeniden yapılandırabilen ve daha gerçekçi sentetik seslerle sonuçlanan radyal tabanlı fonksiyon ağlarının kullanımıyla daha da geliştirilmiştir (Naniwa vd., 2012). Bununla birlikte, sentetik konuşmanın bilişsel bir perspektiften değerlendirilmesi, mesaj uzunluğu, anlama ve dinleyici eğitimi gibi faktörlerin insan-makine etkileşimini etkilemesi nedeniyle hala bir zorluk olmaya devam etmektedir (Delogu & Sementina, 1995). Yapay zekâ tarafından üretilen sahte sesin artan karmaşıklığı, dinleyicileri aldatma ve ses kanıtlarına olan güveni zayıflatma potansiyeliyle bilgi güvenliğine yönelik yeni bir tehdit oluşturmaktadır (Galyashina & Nikishin, 2023). Yapay zekâ daki son gelişmeler, sesleri yeniden üretebilen ve doğallığı koruyabilen modellerle sentetik ses üretiminin kalitesini önemli

ölçüde artırmıştır (Mahajan vd., 2021). Bu modeller, doğal ses dalgalanmalarını yeniden yapılandırabilen ve daha gerçekçi sentetik seslerle sonuçlanan radyal tabanlı fonksiyon ağlarının kullanımıyla daha da geliştirilmiştir (Naniwa vd., 2012). Bununla birlikte, sentetik konuşmanın bilişsel bir perspektiften değerlendirilmesi, mesaj uzunluğu, anlama ve dinleyici eğitimi gibi faktörlerin insan-makine etkileşimini etkilemesi nedeniyle hala bir zorluk olmaya devam etmektedir (Delogu & Sementina, 1995). Yapay zekâ tarafından üretilen sahte sesin artan karmaşıklığı, dinleyicileri aldatma ve ses kanıtlarına olan güveni zayıflatma potansiyeliyle bilgi güvenliğine yönelik yeni bir tehdit oluşturmaktadır (Galyashina & Nikishin, 2023).

Yapay sinir ağları, özellikle yazı birimlerinin fonemlere dönüştürülmesinde, ses sentezine başarıyla uygulanmıştır (Weijters & Thole, 1993). Daha yeni çalışmalar, müzik notaları ile akustik özellikler arasındaki karmaşık ilişkileri modellemek için evrişimli sinir ağlarının (CNN'ler) kullanıldığı şarkı söyleme sesi sentezine odaklandı. Bu CNN tabanlı sistemlerin, önerilen bir hesaplama karmaşıklığı azaltma tekniği ile doğal sese sahip şarkı sesleri ürettiği ve verimliliklerini daha da artırdığı gösterilmiştir (Nakamura vd., 2019).

Yapay sinir ağları, özellikle yazılı metinlerin fonemlere dönüştürülmesi ve ses sentezi alanlarında başarıyla kullanılmıştır (Weijters & Thole, 1993). Bu tekniklerin evrimi, daha yeni araştırmalarda öne çıkmaktadır. Örneğin, Nakamura vd. (2019) yılında yaptıkları çalışmalarında, müzik notaları ile akustik özellikler arasındaki karmaşık ilişkileri modellemek amacıyla evrişimli sinir ağlarının (CNN'ler) şarkı söyleme sesi sentezi için nasıl kullanıldığına odaklanılmıştır.

CNN tabanlı sistemler, önerilen bir hesaplama karmaşıklığı azaltma tekniği kullanılarak doğal sese sahip şarkı sesleri üretebilir (Nakamura vd., 2019). Bu gelişmeler, yapay sinir ağlarının ses sentezi alanındaki uygulamalarının verimliliğini artırmaya yönelik önemli adımlar olarak kabul edilebilir. Bu bağlamda, yapay sinir ağları, ses sentezi teknolojisinin gelişmesine ve daha doğal ve çekici seslerin üretilmesine yardımcı oldu.

Yapay sinir ağları, özellikle karmaşık doğrusal olmayan dönüşümler ve sınırlı alana özgü bilgi gerektiren görevlerde, konuşma sentezinde çok önemli bir rol oynamaktadır (Weijters & Thole, 1993). Bu ağlar, metinden algısal katsayıları ve formları tahmin etmek için kullanılmış, konuşma sentezinin uyarlanabilirliğini ve boyutunu

arttırmıştır (Raghavendra, Vijayaditya & Prahallad, 2010). Daha yeni gelişmeler, konuşma sentezinde hem karmaşık değerli hem de gerçek değerli verileri modelleyebilen karmaşık değerli sinir ağlarının kullanımını içermektedir. Sinir ağı tabanlı yaklaşımların konuşma sentezinde uygulanması, alanı önemli ölçüde geliştirmiş ve daha gelişmiş sistemlerin geliştirilmesine yol açmıştır (Perera & Ganegoda, 2023). Özellikle karmaşık doğrusal olmayan dönüşümler ve sınırlı alana özgü bilgi gerektiren görevlerde, yapay sinir ağları konuşma sentezinde çok önemlidir. Metinden algısal katsayıları ve formları tahmin etmek için bu ağlar kullanılmıştır. Bu, konuşma sentezinin boyutunu ve uyarlanabilirliğini artırmıştır (Raghavendra, Vijayaditya & Prahallad, 2010).

Karmaşık değerli sinir ağları, konuşma sentezinde hem gerçek değerli hem de karmaşık değerli verileri doğru bir şekilde modelleyebilen daha sonraki ilerlemelerdir. Sinir ağı tabanlı tekniklerin konuşma sentezi alanında kullanılması, bu gelişmelerin bir sonucu olarak daha gelişmiş sistemlerin oluşturulmasına izin verdi (Perera & Ganegoda, 2023). Bu bağlamda, yapay sinir ağlarının kullanımı konuşma sentezi teknolojisinde önemli ilerlemelere yol açtı ve daha karmaşık, uyarlanabilir ve doğal sesler üretmek mümkün oldu. Bu değişiklikler, yapay sinir ağlarının ses sentezi uygulamalarındaki etkinliğini ve uyarlanabilirliğini vurgulamaktadır.

Yapay Sinir Ağları (YSA), karmaşık doğrusal olmayan ilişkileri yönetebilen ve hata toleransı sunabilen, ilaç keşif modellemesinde önemli bir araçtır (Zou, Han & So, 2009). İnsan sinir sistemini taklit eden ve sınıflandırma, tahmin, karar verme gibi görevleri yerine getiren bir tür yapay zekâ ve makine öğrenmesidir. Makine öğrenmesinde ileri beslemeli, tekrarlı ve geri beslemeli ağlar dahil olmak üzere farklı YSA türleri kullanılmaktadır (Gurjar & Patel, 2021). YSA'nın gelişimi, bulanık mantık ve dalgacık analizi gibi ileri yöntemlerin entegrasyonunun yanı sıra yüksek performanslı bir alternatif olarak destek vektör makinelerinin ortaya çıkmasına da tanık olmuştur (Huang, 2009).

Yapay Sinir Ağları (YSA), karmaşık doğrusal olmayan ilişkileri yönetme ve hata toleransı sağlama yeteneğine sahiptir, özellikle ilaç keşif modelleme (Zou, Han & So, 2009). Makine öğrenmesi ve yapay zekâ, tahmin, sınıflandırma ve karar verme gibi çeşitli görevleri başarıyla yerine getirebilir. İleri beslemeli, tekrarlı ve geri beslemeli ağlar gibi çeşitli YSA türleri, yapay sinir ağlarının geliştirilmesinde kullanılmaktadır (Gurjar & Patel, 2021).

Makine öğreniminde YSA ve bulanık mantık ve dalgacık analizi gibi ileri yöntemler de kullanılmaktadır. YSA'nın başarılı bir alternatif olarak ortaya çıkmasıyla birlikte, destek vektör makineleri gibi etkili yöntemler de gelişti (Huang, 2009). Bu yenilikler, Yapay Sinir Ağları'nı sadece ilaç keşif modellemesi gibi bilim ve teknoloji alanlarında değil, aynı zamanda genel makine öğreniminde de kullanıyor.

Ses sentezindeki son gelişmeler yapay sinir ağı tabanlı teknolojiler tarafından yönlendirilmektedir. Arik vd. (2017), fonem sınırı tespiti ve ses sentezi için derin sinir ağlarını kullanan ve gerçek zamanlı performansa ulaşan bir metinden konuşmaya sistemi olan Deep Voice'u tanıttı. Raitio vd. (2014), otomatik olarak stile özgü uyarı dalga formları üreten derin bir sinir ağı kullanarak eğitilebilir bir ses kaynağı modeli geliştirerek ses sentezini daha da geliştirdi. Ning vd. (2019), etkili özellik temsillerinin öğrenilmesinde büyük ölçekli verilerin rolünü vurgulayarak, derin öğrenmeye dayalı konuşma sentezine ilişkin kapsamlı bir inceleme sunmuştur. Gully vd. (2017), konuşma özelliklerinin kontrolüne olanak tanıyan, derin bir sinir ağı tarafından yönlendirilen bir dijital dalga kılavuzu ağını kullanan eklemleyici metin-konuşma sentezi için bir yöntem önermiştir. Bu çalışmalar toplu olarak yapay sinir ağı tabanlı teknolojilerin çeşitli uygulamalar için ses sentezini geliştirmedeki potansiyelini ortaya koyuyor.

Yapay sinir ağı tabanlı teknolojiler, ses sentezindeki son gelişmeleri hızlandırıyor. Örneğin, Arik vd. (2017) fonem sınırı tespiti ve ses sentezi için derin sinir ağlarını kullanan bir metinden konuşmaya sistemi olan Deep Voice'u tanıtarak bu alanda önemli bir ilerleme kaydetti. Raitio vd. (2014), derin bir sinir ağı kullanarak eğitilebilir bir ses kaynağı modeli geliştirerek ses sentezini daha da ileriye taşıdı. Bu model, otomatik olarak stile özgü uyarı dalga formları oluşturur. Ning vd. (2019), derin öğrenmeye dayalı konuşma sentezine ilişkin kapsamlı bir inceleme sunarak, etkili özellik temsillerinin öğrenilmesinde büyük ölçekli verilerin önemini vurguladı. Gully vd. (2017), derin bir sinir ağı tarafından yönlendirilen bir dijital dalga kılavuzu ağını kullanarak metin-konuşma sentezi için eklemleyici bir yöntem önermiştir. Bu yöntem, konuşma özelliklerinin kontrolünü sağlar.

Yapay sinir ağı sistemlerinde seslerin sentezi, geleneksel sentezleyicilerin karmaşıklığı ve parametre tanımlamanın zorluğu dahil olmak üzere çeşitli zorluklar sunar (Bresin & Vedovetto, 1994). Bu zorluklara rağmen, sinir ağı tabanlı yaklaşımlar, sinir ağlarının kullanımının yönlendirdiği alanda kaydedilen ilerlemelerle konuşma sentezinde umut vaat etmektedir. Bu alanda gelecekteki çalışmalar, sinir ağı tabanlı teknikleri diğer

yöntemlerle birleştiren hibrit bir yaklaşımın geliştirilmesini içerebilir (Perera & Ganegoda, 2023). Ek olarak, soyut açıklamalarla kompozisyona olanak tanıyan ses sentezi için yapay zekâ tabanlı sistemlerin kullanılması, gelecekteki potansiyel araştırmaların bir alanıdır (Miranda, 2019).

Yapay sinir ağı sistemlerinde seslerin sentezi, geleneksel sentezleyicilerin karmaşıklığı ve parametre belirlemenin zorluğu dahil olmak üzere bir dizi sorunla karşı karşıyadır (Bresin & Vedovetto, 1994). Sinir ağı tabanlı yaklaşımlar, konuşma sentezi alanında, özellikle sinir ağlarının kullanımının yönlendirdiği alanda, bu zorluklara rağmen ilerleme sağlayacaktır (Perera & Ganegoda, 2023). Gelecekteki araştırmalar, sinir ağı tabanlı yaklaşımları diğer tekniklerle birleştiren hibrit bir yaklaşımın geliştirilmesini içerebilir. Bu, ses sentezi sistemlerini daha verimli ve çok yönlü hale getirmek için çeşitli teknikleri birleştirmeyi amaçlar.

Yapay sinir ağları, özellikle karmaşık doğrusal olmayan dönüşümler ve sınırlı alana özgü bilgi gerektiren görevlerde, konuşma sentezinde önemli bir potansiyel göstermiştir (Weijters & Thole, 1993). Perera & Ganegoda (2023), konuşma sentezi sistemlerinin geliştirilmesinde sinir ağı tabanlı yaklaşımların rolünü daha da vurgulayarak, bunların mevcut tekniklerin sınırlamalarını ele alma yeteneklerini vurgulamaktadır. Malcangi, & Frontini (2007), sinir ağlarının konuşma sentezindeki pratik uygulamalarını araştırıyor; Malcangi, & Frontini (2007), sentetik konuşma doğallığı için hiyerarşik bir öngörücünün geliştirilmesine ve dilden bağımsız bir konuşma sentezleyicisinin oluşturulmasına odaklanıyor. Bu çalışmalar toplu olarak yapay sinir ağlarının konuşma sentezi sistemlerinin kalitesini ve çok yönlülüğünü artırmadaki umut verici rolünün altını çiziyor.

İKİNCİ BÖLÜM

İLGİLİ ARAŞTIRMALAR

Üretken yapay zeka teknolojileri tarafından üretilen sentetik seslerin gerçek seslerle karşılaştırılması, akademik dünyada oldukça yeni bir araştırma alanıdır. Bu konuda yapılmış olan çalışmalar henüz sınırlıdır; ancak, teknolojik ilerlemelerle birlikte bu alanda artan bir ilgi görülmektedir. Bu nedenle, tezimiz kapsamındaki konulara katkıda bulunabilecek, yapay zeka ve ses işleme teknolojileri ile ilgili yapılmış olan tezler, makaleler ve bildirilere bu bölümde yer verilmiştir. İncelediğimiz çalışmalar, özellikle yapay zekanın ses üretimi ve işleme kapasitesini teorik ve pratik yönleriyle ele alan çalışmalardır. Bu çalışmalar, konunun teorik derinliğini ve uygulama alanlarındaki çeşitliliğini göstermektedir. Konuyla daha doğrudan ilgili olan çalışmalar öncelikli olarak ele alınmış, ancak genel yapay zeka ve ses teknolojileri üzerine yapılan daha geniş kapsamlı araştırmalar da ikincil olarak incelenmiştir. Bu sayede, sentetik seslerin gerçek seslerle karşılaştırılması konusundaki mevcut literatür, geniş bir perspektiften değerlendirilmiş ve tezimizin kapsamı daha da netleştirilmiştir.

1. MAKALELER

Zou, Han, & So (2009), tarafından yapılan "Overview of Artificial Neural Networks" başlıklı makalede, insan beynini taklit etme fikrinden evrilen bir makine öğrenme yöntemi olan yapay sinir ağı (ANN) veya basitçe sinir ağı incelenmiştir. Modern ilaç keşfi araştırmalarında veri patlaması, tek veya çoklu yanıtlar ile geniş bir özellik seti arasındaki gizli neden-sonuç ilişkilerini ortaya çıkarmak için sofistike analiz yöntemleri gerektirir. ANN, ilaç keşfi modellemesinde talebi karşılamak için kullanılan birçok yöntemden biridir. Geleneksel regresyon yaklaşımına kıyasla, ANN karmaşık doğrusal olmayan ilişkileri modelleme yeteneğine sahiptir. Ayrıca, ANN mükemmel hata toleransına sahip olup, hızlı ve paralel işleme ile yüksek ölçeklenebilirliğe sahiptir. Bu bölüm, ANN gelişiminin arka planını tanıtmakta ve daha sofistike ANN'leri anlamak için hayati önem taşıyan temel kavramları özetlemektedir. Bölümün sonunda, sıklıkla kullanılan öğrenme yöntemleri ve ağ yapıları kısaca tartışılmaktadır.

Huang (2009), tarafından yapılan "Advances in artificial neural networks—methodological development and application" başlıklı makalede, son üç on yılda yoğun bir şekilde incelenip uygulanmış olan yapay sinir ağları ve bu teknolojinin metodolojik gelişimleri ile uygulamaları üzerine yapılan çalışmalar ele alınmıştır. Geri yayılım eğitim

algoritmalarına odaklanarak çok katmanlı algılayıcı ağlar için yapılan araştırmalar, radyal temel fonksiyon, geri beslemeli ağ, tekrarlayan ağ ve denetimsiz Kohonen kendiliğinden örgütlenen ağlar gibi diğer ağ türleri için eğitim algoritmalarının geliştirilmesine öncülük etmiştir. Bu ağlar, özellikle geri yayılım eğitim algoritması kullanılan çok katmanlı algılayıcı ağ, çeşitli bilimsel ve mühendislik alanlarında tanınır hale gelmiştir. Eğitim sürecinin hızlandırılması ve veri aşırı uyumunun önlenmesi amacıyla geri yayılım algoritmasının iyileştirilmesi üzerine araştırmalar gerçekleştirilmiştir. Yapay sinir ağlarının, veri yorumlama ve modelleme yeteneklerini artırma ve eğitim algoritmasının işleyişindeki öznelliği azaltma amacıyla bulanık mantık ve dalgacık analizi gibi diğer ileri yöntemlerle entegrasyonu sağlanmıştır. Destek vektör makineleri, son yıllarda yapay sinir ağlarına paralel olarak yüksek performanslı denetimli genelleştirilmiş doğrusal sınıflandırıcılar olarak ortaya çıkmıştır. Yapay sinir ağlarının gelişim tarihine dair bir inceleme sunulmuş ve yapay sinir ağlarının standart mimarileri ve algoritmaları tanımlanmıştır. İlerleyen bölümlerde, destek vektör makineleri ile entegre edilmiş gelişmiş yapay sinir ağları tanıtılmış ve ANNs'nin sınırlamaları belirlenmiştir. Yapay sinir ağlarının ve destek vektör makinelerinin gelişimi, gıda bilimi ve mühendisliği, tarım ve biyoloji mühendisliği alanlarında daha fazla uygulama imkanları sağlayacak şekilde tartışılmıştır.

Baltazar (2018) tarafından yapılan "Where mind and music meet: Affect self-regulation through music" başlıklı makalede, müzikle duygusal durumların nasıl düzenlenebileceği incelenmiştir. Müzik aracılığıyla duygusal düzenlemenin kavramsal zayıflıkları ve gelecek çalışmalar için öneriler sunulmuştur. Müziksel faaliyetler, düzenleme stratejileri ve müziksel mekanizmaların duygusal hedefler doğrultusunda nasıl seçildiği ve iyilik halini nasıl belirlediği üzerine bir model geliştirilmiştir. Duygusal sonuçlar üzerine müziğin ve düzenleme stratejilerinin etkileri bir deneysel çalışma ile değerlendirilmiş, dinlenen müziğin stres azaltma üzerindeki etkisinin stratejilerden daha belirleyici olduğu bulgulanmıştır. Bulgular, müzik kullanımının günlük hayat ve sağlık üzerine etkilerini anlamak için önemli katkılar sunmaktadır.

Zheng & Zhang (2018) tarafından yapılan "Phase-aware speech enhancement based on deep neural networks" başlıklı makalede, konuşma işlemede temel bir yöntem olan kısa zamanlı frekans dönüşümü (STFT) üzerine yoğunlaşmıştır. STFT fazının yüksek derecede yapılandırılmamış olması nedeniyle, çoğu konuşma işleme algoritması genellikle yalnızca STFT büyüklüğü ile işlem yapmakta ve fazı büyük ölçüde

keşfedilmemiş bırakılmaktadır. Ancak, derin sinir ağı (DNN) tabanlı konuşma işlemleri, örneğin konuşma iyileştirme ve tanıma gibi alanlarda, faz işleminin önemi DNN tabanlı yöntemlerin yeni bir gelişim noktası olarak artırılmaktadır. Bu makalede, DNN temelli bir faz-farkında konuşma iyileştirme algoritması önerilmiştir. Özellikle eğitim aşamasında, hedef olarak fazı dahil etmek amacıyla yapılandırılmamış bir faz spektrogramının zaman eksenini boyunca türevine, yani anlık frekans sapmasına (IFD) dönüştürülmesi önerilmiştir; bu, karşılık gelen büyüklük spektrogramı ile benzer bir yapıya sahiptir. Hem IFD hem de büyüklüğün çok hedefli bir öğrenme çerçevesinde birlikte optimize edilmesi önerilmiştir. Test aşamasında, tahmin edilen IFD'den faz spektrogramını geri kazanacak bir son işlem yöntemi önerilmiştir. Deneysel sonuçlar, önerilen yöntemin etkinliğini göstermiştir.

Akiyama (2019), tarafından yapılan "Nothing Connects Us but Imagined Sound" başlıklı makalede, alan kayıtlarının zaman ve mekân algımız üzerindeki etkileyici rolü ele alınmıştır. İyi icra edilmiş bir alan kaydı dinleyen kişi, şimdiki zamanda ve burada yer alırken başka bir zamana ve mekâna tuhaf bir şekilde taşınmış hissedebilir. Alan kayıtlarının, başka bir yerde ve zamanda olan bir şeylerin kanıtı, ispatı veya belgesi olarak durma gücüne sahip gibi görünmesi, bu kayıtların bir pencere açma kapasitesine sahip olduğunu gösterir. Bu bölüm, alan kayıtlarının medya uygulamalarında topluluklar oluşturmak ve sürdürmek amacıyla oynadığı rolü ele almakta ve özellikle 1970'lerde Vancouver merkezli Dünya Sesi Peyzajı Projesi tarafından yapılan çalışmalara ve bu projenin "ses peyzajı kompozisyonu" olarak adlandırdığı forma odaklanmaktadır.

Baltazar vd. (2019), tarafından yapılan "Is it me or the music? Stress reduction and the role of regulation strategies and music." başlıklı makalede, müziksel etki düzenlemesinin stres yönetimi üzerindeki etkileri incelenmiştir. Çalışma, müziğin ve düzenleme stratejilerinin, stres azaltmada nasıl bir rol oynadığını anlamayı amaçlamaktadır. Araştırma, 34 katılımcı üzerinde, laboratuvar ortamında yapılan faktöriyel bir iç-birey tasarımıyla yürütülmüştür. Katılımcılar, stresi azaltma amacıyla, kendileri için “yeterli” ve “yetersiz” olarak nitelendirdikleri müzik örneklerini ve stratejileri çevrimiçi bir anket aracılığıyla seçmişlerdir. Daha sonra laboratuvarında, bir stres indüksiyon prosedürü uygulandıktan sonra, bu müzikleri dinlemeleri ve stratejileri kullanmaları istenmiştir. Ana sonuç ölçütleri arasında kendileri tarafından bildirilen gerginlik, enerji, değerlilik, cilt iletkenliği seviyeleri, başlangıç çarpma genlikleri ve riskten kaçınma yer almaktadır. Bulgular hem müziğin hem de stratejinin, katılımcıların

bildirdiği gerginlik üzerinde anlamlı etkilere sahip olduğunu göstermiştir. Müzik ayrıca enerji, değerlilik, cilt iletkenliği ve riskten kaçınma üzerinde de belirgin etkiler sergilemiştir. İkili karşılaştırmalar, “yeterli müzik ve strateji” koşulunun, “yetersiz müzik ve strateji” koşuluna göre daha etkili olduğunu belirtmiş, ancak hiçbir koşul diğerlerinden açık bir şekilde üstün veya altın olmadığını ortaya koymuştur. Ayrıca, müziğin stratejiye kıyasla kısa vadeli sonuçlar üzerinde daha büyük bir etkisi olduğu tespit edilmiştir. Sonuçlar, duygusal düzenlemenin başarısının seçilen stratejiler ve müzik yeterliliğine bağlı olduğunu, ancak müziğin kısa vadeli duygusal sonuçlar için daha belirleyici olduğunu vurgulamaktadır. Bu bulgular, müziksel etki düzenlemesi üzerine yapılan önceki çalışmalarla birlikte tartışılarak, konu hakkında daha derin bir anlayış sağlanmaya çalışılmıştır.

Crandall (2019) tarafından yapılan "Artificial Intelligence And Manufacturing" başlıklı makalede, yapay zeka teknolojisinin araştırma laboratuvarlarından çıkarak ürünlere entegre edilmesi ve iş dünyası ile günlük yaşamın pek çok yönünü köklü bir şekilde dönüştürme potansiyeline sahip olduğu ele alınmıştır. Makale, yapay zekanın ne olduğu ve ne olmadığı, ne zaman iyi çalıştığı ve ne zaman başarısız olduğu hakkında kısa bir genel bakış sağlar. Ardından, yapay zekanın özellikle imalat sektörü üzerindeki potansiyel etkileri incelenmiştir. Yapay zeka teknolojilerinin, imalat süreçlerini ve bu sektördeki iş akışlarını nasıl değiştirebileceği üzerinde durulmuştur.

Ning vd. (2019), tarafından yapılan "A review of deep learning based speech synthesis. Applied Sciences" başlıklı makalede, konuşma sentezlemesi veya metin-konuşma (TTS) alanında derin öğrenme ve hatta uçtan uca tekniklerin kullanımıyla sağlanan son gelişmeler ele alınmaktadır. Bu teknikler, akıllı konuşma etkileşimi, sohbet robotu veya yapay zeka (AI) ile yapılan konuşmalarda kullanılmak üzere geniş bir uygulama yelpazesini iyileştirmek için kullanılmaktadır. Konuşma sentezi için derin öğrenme tabanlı teknikler, büyük ölçekli <metin, konuşma> çiftlerini kullanarak metin ve konuşma arasındaki boşluğu köprülemek için etkili özellik temsilleri öğrenmekte ve böylece olayların özelliklerini daha iyi karakterize etmektedir. Makale, ilk olarak geleneksel konuşma sentezi yöntemlerini tanıtarak istatistiksel parametrik konuşma sentezi (SPSS) sisteminin bileşimindeki akustik modellemenin önemini vurgulamaktadır. Ardından, son yıllarda sanatın zirvesine ulaşmış olan uçtan uca yaklaşımlar da dahil olmak üzere derin öğrenme tabanlı konuşma sentezindeki ilerlemelere genel bir bakış sunmaktadır. Son olarak, konuşma sentezi için derin öğrenme yöntemlerinin sorunlarını

tartışmakta ve konuşma sentezi araştırmalarını yeni bir sınıra taşıyabilecek çekici araştırma yönlerini işaret etmektedir.

Reynoso (2019), tarafından yapılan "Main Types Of Artificial Intelligence" başlıklı makalede, yapay zekanın çeşitli türleri ve bu türlerin insan benzeri zeka gösterme kapasiteleri ele alınmıştır. Yapay zekanın basit otomasyondan karmaşık karar verme mekanizmalarına kadar nasıl sınıflandırıldığı incelenmiş ve her bir türün özellikleri, güçlü yönleri ve sınırlılıkları tartışılmıştır. Ayrıca, bu türlerin farkları detaylandırılarak yapay zekanın gelecekteki potansiyel uygulamalarına dair önemli içgörüler sunulmuştur.

Docasal & Alvarez (2023), tarafından yapılan "Enhancing Voice Cloning Quality through Data Selection and Alignment-Based Metrics" başlıklı makalede, ses işleme alanında yükselen bir saha olan ses klonlamasının kalitesini artırmak amacıyla çeşitli tekniklerin etkisi incelenmiştir. Bu çalışmada, özellikle düşük kaliteli bir veri setine odaklanılmıştır. Bulgularımızı karşılaştırmak için, iki yüksek kaliteli korpus da karşılaştırmalı analiz için kullanılmıştır. Toplanan korpusların kalitesinin kapsamlı değerlendirmeleri yapılmış ve ses klonlama sistemlerinin eğitimi için en uygun verilerin seçilmesi amaçlanmıştır. Bu ölçümleri takiben, korpuslardaki heterojenliği azaltmak için sinyal-gürültü oranı düşük ve ifade hızı değişkenliği yüksek ses dosyaları çıkarılarak bir dizi ablasyon yapılmıştır. Ayrıca, Tacotron 2 metinden-konuşmaya sistemlerinin dikkat matrisini kullanarak girdi karakterlerinin hizalı bölüm oranını hesaplayan yeni bir algoritma tanıtılmıştır. Bu algoritma, ses klonlama sürecinde hizalama kalitesinin değerlendirilmesi için değerli bir metrik sağlamaktadır. Deneylerimizin sonuçları, yapılan ablasyonların düşük kaliteli korpus için sentezlenen sesin kalitesini önemli ölçüde artırdığını göstermiştir. Özellikle, bulgularımız, eğitilmiş modellerin 3 saatlik bir korpus kullanarak hazır bir modelden eğitildiğinde, çok daha büyük veri miktarlarını baştan eğitilmiş modellerle karşılaştırılabilir ses kalitesine sahip olduğunu göstermiştir.

2. BİLDİRİLER

Weijters & Thole (1993), tarafından yapılan "Speech synthesis with artificial neural networks" başlıklı bildiride, yapay sinir ağlarının konuşma sentezi uygulamalarında nasıl kullanıldığı ele alınmıştır. Çalışmada, semboller (grafemler) başka sembollere (fonemlere) dönüştürülmüştür. Yapay sinir ağları, karmaşık doğrusal olmayan dönüşümlerin gerektiği ve yeterli alan özgü bilginin bulunmadığı görevler için özellikle rekabetçidir. Metinlerin fonetik transkripsiyonunun konuşma parametrelerine dönüşümü,

böyle bir görev gibi görünmektedir. Yazarların bu dönüşümü gerçekleştirmek için bir yapay sinir ağı eğitime yaklaşımına dair bazı önemli sonuçlar gösterilmiştir.

Bresin & Vedovetto, (1994). "Neural networks for a simpler control of synthesis algorithm of musical tones and for their compression" başlıklı bildiride, müzik tonlarının sentez algoritmalarının kontrolünü basitleştiren ve veri sıkıştırma için kullanılan bir hibrit sistem sunulmaktadır. Araştırmanın amacı, bir sese ilişkin parametrelerin sayısını azaltmaktır. Elektronik teknolojinin gelişimiyle birlikte, mevcut sentezleyiciler, kullanıcının zorlanmadan hareket edemediği çok karmaşık sistemler haline gelmiştir ve bu durum zaman kaybına neden olmaktadır. Temel fikirler, nöral ağ tarafından yapılan parametrelerin azaltılması ve ses sentezi için herhangi bir algoritmanın seçilebilmesi özgürlüğüdür. Örneğin, FM sentezi temelli bir makine kullanılırken, FM parametrelerinin uygun bir alt kümesi değiştirilebilir; örnek tabanlı bir klavyede ise kazanç değeri, bazı zarf parametreleri vb. değiştirilebilir. Bu sistem kullanımı çok kolay, esnek ve elektronik bir sentezleyicide gerçekleştirilmesi için çok fazla donanım gerektirmeyen bir yapı sunmaktadır. Ana çalışma geliştirici tarafından yapılır, ancak bu sadece bir kez gerçekleştirilir. Kullanıcı sadece bazı kontrol cihazlarını, yani kaydırıcıları hareket ettirir; nöral ağ, sentezleyici için parametrelerin setini hesaplayarak çıktıyı üretir.

Raghavendra, Vijayaditya & Prahallad (2010), "Speech synthesis using artificial neural networks" başlıklı bildiride, bilgisayar bilimi, yapay zeka ve dil işleme teknolojileri alanlarında yer alan konuşma sentezi ele alınmıştır. Özellikle istatistiksel parametrik sentez yöntemleri üzerinde durulmuştur, bu yöntem son yıllarda adaptasyon kabiliyeti ve sentez büyüklüğü nedeniyle popülerlik kazanmıştır. Çalışma, mel sepestral katsayıları, temel frekans (f_0) ve süre gibi konuşma sentezi için temel bileşenlere odaklanır. Araştırma, metinden yapay sinir ağları kullanarak mel sepestral katsayılarının nasıl tahmin edileceği ve metinden formantların nasıl tahmin edileceği gibi iki katmanlı problemleri çözmeyi amaçlamaktadır.

Raitio vd. (2014), tarafından yapılan "Deep neural network based trainable voice source model for synthesis of speech with varying vocal effort" başlıklı bildiride, sesin çeşitli vokal çabalarla sentezlenmesi için bir derin sinir ağı (DNN) tabanlı ses kaynağı modelleme yöntemi incelenmiştir. Bu yeni eğitilebilir ses kaynağı modeli, akustik özellikler ile zaman alanında ses döngüsüne senkronize glottal akış dalgası arasında bir eşleme öğrenmek için bir DNN kullanır. Ses kaynağı modeli, nefesli, normal ve Lombard konuşmasını içeren çeşitli konuşma materyalleri ile eğitilmiştir. Sentez sırasında, normal

bir ses önce istenen bir tarza uyarlanır ve esnek DNN tabanlı ses kaynağı modeli kullanılarak, uyarlanmış akustik özelliklere dayalı olarak tarza özgü bir uyarılma dalgası otomatik olarak üretilir. Önerilen ses kaynağı modeli, her vokal çaba seviyesi için manuel olarak seçilmiş ortalama glottal akış darbelerine dayanan ve istenen tarza ses kaynağı spektrumunu doğru bir şekilde eşleştirmek için bir spektral eşleştirme filtresi kullanan sağlam ve yüksek kaliteli bir uyarılma modelleme yöntemi ile karşılaştırılmıştır. Subjektif değerlendirmeler, önerilen DNN tabanlı yöntemin, manuel darbe seçiminden kaçınan ve spektral eşleştirme filtresi kullanan bir sistemden daha hızlı olduğu ve temel yöntemle karşılaştırılabilir olarak değerlendirildiğini göstermektedir.

Gully vd. (2017), tarafından yapılan "Articulatory text-to-speech synthesis using the digital waveguide mesh driven by a deep neural network" başlıklı bildiride, konuşma dalgası modellemesinde derin sinir ağlarının kullanımına dayanan yeni bir yöntem önerilmektedir. Bu yöntem, manyetik rezonans görüntüleme verileri yerine konuşma dalgasından vokal kanalının fiziksel modelini doğrudan tahmin etmeyi amaçlar. Bu, modelin vokal kanalının boyutu ve şekli ile net bir ilişki kurulmasını sağlar ve yaş ile cinsiyet gibi konuşma özellikleri açısından önemli bir esneklik sunar. İlk testler, oldukça basitleştirilmiş bir fiziksel modele rağmen, anlaşılır sentezlenmiş konuşma elde edildiğini göstermektedir. Bu, fiziksel modellerin genel kontrolü için birleşik tekniğin potansiyelini ve dolayısıyla daha doğal sounding sentetik konuşma üretimini vurgulamaktadır.

Arık vd. (2017), tarafından yapılan "Deep Voice: Real-time Neural Text-to-Speech" başlıklı bildiride, tamamen derin sinir ağlarından oluşturulmuş üretim kalitesinde bir metin-konuşma sistemi sunulmuştur. Deep Voice, gerçekten uçtan uca nöral konuşma sentezi için temel oluşturur. Sistem, beş ana yapı taşından oluşmaktadır: fonem sınırlarını belirleyen bir segmentasyon modeli, grafemden foneme dönüşüm modeli, fonem süresi tahmin modeli, temel frekans tahmini modeli ve ses sentezi modeli. Segmentasyon modeli için, fonem sınırı tespiti yapmak üzere derin sinir ağları kullanılarak bağlantısalcı zaman sınıflandırması (CTC) kaybı ile bir yöntem önerilmiştir. Ses sentezi modelinde, orijinalinden daha az parametre gerektiren ve daha hızlı eğitilen WaveNet'in bir varyantı uygulanmıştır. Her bileşen için bir nöral ağ kullanılarak, sistem geleneksel metin-konuşma sistemlerinden daha basit ve esnek hale getirilmiştir; bu geleneksel sistemler her bileşen için yoğun özellik mühendisliği ve geniş alan uzmanlığı gerektirir. Son olarak, sistemle çıkarımın gerçek zamanın üzerinde hızlı bir şekilde yapılabildiği gösterilmiş ve hem CPU hem de GPU üzerinde optimize edilmiş WaveNet

ıkarım ekirdekleri ile mevcut uygulamalara gre 400 kata kadar hızlanma saėlandıėı belirtilmiřtir.

Perera & Ganegoda (2023), tarafından yapılan "A Comprehensive Review on Speech Synthesis Using Neural-Network Based Approaches." bařlıklı bildiride, insanlar arasındaki temel iletiřim biimi olan konuřma zerine odaklanılmıřtır. 1980'lerin bařından itibaren insanlıėın yapay konuřma yaratma hayali, gemiř on yıllar boyunca birok konuřma sentezi sisteminin geliřtirilmesine ve mevcut sistemlerin ilerletilmesine yol amıřtır. Ancak, Yapay Zeka aėının bařlamasıyla bu ilerlemeler, nral aėların yardımıyla hızla geliřtirilmiřtir. Bu yeni bulunan ilerleme teknikleri ve yntemleri, "Nral aė tabanlı yaklařımlar kullanılarak konuřma sentezi" arařtırma alanının yeniden ortaya ıkmasına nclk etmiřtir. Bu makale, řu anda kullanılmakta olan konuřma sentezi tekniklerini, nral aė tabanlı yaklařımlar kullanan teknikleri, avantajları, dezavantajları ve sınırlamaları gzden geirmektedir. Ayrıca, bu inceleme makalesi, yeni bir hibrit yaklařım nermekte ve gelecekte yapılabilecek deėiřiklikleri tartıřmaktadır.

3. KİTAPLAR

Davenport (2018) tarafından yapılan "The AI Advantage: How To Put The Artificial Intelligence Revolution To Work" bařlıklı kitapta, yapay zeka teknolojilerinin iř dnyasında etkin kullanımı ve rekabet avantajı saėlamak iin bu teknolojilerden nasıl yararlanılacaėı zerine pratik bir rehber sunulmuřtur. Davenport tarafından, yapay zekanın abartılı algısının tesine geilerek, IBM'in Watson'ının kanseri iyileřtirebileceėi zamanların ne kadar inandırıcı grndėu hatırlatılmıř ve iřletmelerin yapay zekayı gerek dnyada řimdi nasıl kullanabileceėi aıklanmıřtır. "Ay yolculuėu" yerine, řirketi daha verimli kılmak iin "dřk asılı meyvelere" odaklanması nerilmiřtir. Yapay zeka teknolojilerinin insan iřileri yerine koymayacaėı, ancak onların yeteneklerini artıracaaėı belirtilmiřtir; akıllı makinelerin akıllı insanlarla yan yana alıřtıėı ifade edilmiřtir. Yapay zeka, yapılandırılmıř ve tekrarlanan iřleri otomatikleřtirmekte, makine ėrenimi aracılıėıyla veriler zerinde geniř aplı analizler saėlamakta ve chatbotlar ile akıllı ajanlar aracılıėıyla mřteriler ve alıřanlarla etkileřim kurmaktadır. řirketlerin bu teknolojilerle deneyler yapması ve kendi uzmanlıklarını geliřtirmesi gerektiėi vurgulanmıřtır. Davenport tarafından, nemli yapay zeka teknolojileri tanımlanmıř ve bu teknolojilerin kullanımı, Amazon ve Google gibi byk ticari iřletmeler tarafından yapılan yapay zeka alıřmaları rapor edilmiř ve biliřsel bir kuruluřa dnřme adımları ve stratejileri zetlenmiřtir.

Lee & Qiufan (2021), tarafından yapılan "AI 2041: Ten visions for our future" başlıklı kitapta, yapay zekanın (AI) önümüzdeki yirmi yıl içinde dünyamızı nasıl değiştireceği sorgulanmaktadır. Yenilikçi bir teknolog ve ünlü bir yazarın işbirliği ile, "etkileyici" (The New York Times) bir gelecek tasviri sunulmuş ve "ilgi çekici içgörülerle dolu" (Financial Times) olarak değerlendirilmiştir. Bu baskı, Kai-Fu Lee tarafından yazılan yeni bir önsözü içermektedir. Kitap, ChatGPT'nin ortaya çıkışından çok önce, Kai-Fu Lee ve Chen Qiufan'ın yapay zekanın günlük yaşamımızı dönüştürme potansiyelini anladıkları bir dönemi ele almaktadır. Dünya yapay zekanın gücüne uyanırken, birçok kişi hala büyük resmi görememektedir. Chatbotlar ve büyük dil modelleri sadece başlangıçtır. Bu "ilham verici işbirliği" (The Wall Street Journal) kapsamında, Lee ve Chen, 2041 yılında dünyamızın yapay zeka tarafından nasıl şekillendirileceğini hayal etmektedirler. On sürükleyici, dünya çapında kısa hikaye ve eşlik eden yorumlarıyla kitap, okuyucuları derin öğrenme, karma gerçeklik, robotik, yapay genel zeka ve otonom silahlar gibi AI teknolojilerinin yeni bolluğu ve potansiyel zararlarıyla boğuşan çeşitli ayarlar ve karakterlerle tanıştıır.

4. KİTAP İÇİ BÖLÜMLER

Malcangi & Frontini (2007), tarafından yapılan "A Language-Independent Neural Network-Based Speech Synthesizer" başlıklı kitap içi bölümde, metin-ses dönüşümü uygulamalarında yumuşak hesaplama tekniklerinin kullanımı ele alınmıştır. Çalışma, fonetik seviyede metin-ses dönüşümü için yapay sinir ağlarının kullanılmasının sert hesaplama yöntemlerine göre sağladığı avantajları vurgulamaktadır. Yumuşak hesaplamanın genelleme kapasitesi sayesinde, veritabanında bulunmayan kelimelerin haritalanması mümkün hale gelmiş ve aynı kelimenin farklı telaffuzları arasındaki çelişkiler azaltılmıştır. Yapay sinir ağları, uygulamalı desen eşleştirme problemlerinin geniş bir sınıfını optimal olarak çözme yeteneği göstermiştir. Ancak makine-insan etkileşimi için gerekli desen oluşturma konusunda yapılan araştırmalar yetersiz kalmıştır. Dil bağımsız konuşma komutları için tasarlanan bu yapay konuşma sentezleyici, eller serbest arayüzlerde kullanılmak üzere derinlemesine gömülü sistemlere uygulanmaktadır. Bu sentezleyici, özel olarak geliştirilmiş düzenli ifade tabanlı bir metin-ses dönüşüm motoru kullanılarak eğitilen ileri beslemeli, geri yayımlı bir yapay sinir ağı üzerine kurulmuştur. İlk deneysel sonuçlar, dil bağımsızlığı ve sistem içi öğrenme gibi beklenen özellikleri başarıyla göstermiştir.

Miranda (2019), tarafından yapılan "From symbols to sound: AI-based investigation of sound synthesis" başlıklı kitap içi bölümde, ses sentezi için bir zeka sistem olanaklarının araştırılması ele alınmıştır. Bu çalışma, düşük seviye bilgisayar programlaması yerine soyut tanımlar ve işlemlerle ses kompozisyonuna olanak tanıyacak bir zeka sistemi olasılığını incelemektedir. Araştırma, bir bestecinin ses yapısına dair sezgisel kavrayışını elektronik ses sentezinin parametrik kontrolüne dönüştürmek için olası yönergeler ve stratejiler aramaya yöneliktir. Çalışmada, yapay zeka (AI) araştırmalarında kullanılan birkaç tanınmış uzman sistem tasarım tekniğinin derlenmesi yoluyla bu probleme nasıl yaklaşıldığı gösterilmektedir. Probleme müzisyen perspektifinden başlayarak, AI/Müzik alanındaki iki ana konu olan bilgi temsili ve makine öğrenimi üzerinde yoğunlaşılır. Ses sentezi için uygun bir temsil şeması önerilmekte ve bir örnek prototip sentezleyiciye yönelik indüktif bir öğrenme deneyi gerçekleştirilmektedir. Ayrıca, prototipin işlevselliği tanıtılarak ileri çalışmalar için önerilerde bulunmaktadır.

Gurjar & Patel (2021), tarafından yapılan "Fundamental Categories of Artificial Neural Networks" başlıklı kitap içi bölümde, yapay zeka (AI) ve makine öğrenimi kullanımının yeni dönemi ve bu iki alanın kombinasyonunun yapay sinir ağları (ANN) olarak adlandırıldığı ele alınmıştır. Yapay sinir ağları, donanım veya yazılım bazlı bileşenler olarak kullanılabilir. Yapay sinir ağlarında farklı topolojiler ve öğrenme algoritmaları kullanılır. Yapay sinir ağı, insan sinir sisteminin işlevselliğine benzer şekilde çalışır. ANN, insan beyninin gerçekleştirdiği sınıflandırma, tahmin, karar verme, görselleştirme gibi aktiviteleri önceki deneyimleri göz önünde bulundurarak gerçekleştiren bir doğrusal olmayan hesaplama modeli olarak işlev görür. ANN, çevre hakkında bilgi toplayarak karmaşık ve yönetilmesi zor sorunları çözmek için kullanılır. Makine öğreniminde mevcut olan yapay sinir ağlarının farklı türleri bulunmaktadır. Tüm yapay sinir ağ türleri, matematiksel işlemlere dayanarak çalışır ve sonuç elde etmek için bir dizi parametreye ihtiyaç duyar. Bu bölüm, ileri beslemeli, geri beslemeli, tekrarlayan, sınıflandırma-tahmin gibi çeşitli sinir ağ türleri hakkında genel bir bakış sunar.

ÜÇÜNCÜ BÖLÜM

ÜRETKEN YAPAY ZEKA TEKNOLOJİSİNDE ÜRETİLEN SENTETİK SESLERİN GERÇEK SESLER İLE KARŞILAŞTIRILMASI

1. ARAŞTIRMANIN PROBLEM CÜMLESİ

Bu araştırmanın ana problem cümlesi, "Üretken yapay zeka teknolojisi ile üretilen sentetik seslerin gerçek insan sesleriyle olan benzerlikleri ve farklılıkları nelerdir?" sorusudur. Bu soru, yapay zeka teknolojileri kullanılarak üretilen seslerin gerçek seslerle karşılaştırılmasını ve değerlendirilmesini merkezine alarak, seslerin algılanış biçimini ve insan kulağına olan etkilerini detaylı bir şekilde incelemeyi amaçlamaktadır.

Birinci Alt Problem: Katılımcılar sentetik ve gerçek sesleri ayırt ederken hangi özellikleri veya faktörleri temel almaktadır?

İkinci Alt Problem: Yapay zeka teknolojisi kullanılarak üretilen sentetik seslerin ifade yeteneği ve doğallığı, insanlar tarafından nasıl değerlendirilmektedir?

Üçüncü Alt Problem: Katılımcıların sentetik sesler ile gerçek sesler arasındaki tercihleri neye dayanmaktadır ve bu tercihler kullanıcı deneyimini nasıl şekillendirmektedir?

Dördüncü Alt Problem: Gerçek ve sentetik seslerin güvenilirlik algıları arasındaki karşılaştırmalı etkileri nelerdir?

Beşinci Alt Problem: Sentezlenmiş seslerin sosyal etkileşime etkisi ve toplumdaki kabul düzeyi nasıldır?

2. ARAŞTIRMANIN AMACI

Bu araştırmanın temel amacı, üretken yapay zeka teknolojisi kullanılarak üretilen sentetik seslerin gerçek insan sesleriyle olan benzerliklerini ve farklılıklarını detaylı bir şekilde incelemektir. Araştırma, sentetik seslerin doğallığını, ifade yeteneğini ve güvenilirliğini değerlendirerek, bu seslerin insan kulağına olan etkisini ve algılanış biçimlerini ortaya koymayı hedeflemektedir. Ayrıca, katılımcıların bu sesler arasında yaptıkları tercihlerin altında yatan nedenleri ve bu tercihlerin kullanıcı deneyimini nasıl etkilediğini anlamaya çalışır. Bu çalışma, yapay zeka ile üretilen seslerin toplumdaki kabul düzeyini ve sosyal etkileşim üzerindeki etkilerini de araştırarak, sentetik ses teknolojilerinin gelecekteki gelişimine yön vermek için bilimsel veriler sunmayı amaçlar. Ayrıca, sentetik seslerin sosyal kabul düzeyini ve bu seslerin etkileşimde bulunduğu

alanlardaki performansını deęerlendirmek, arařtırmanın dięer bir hedefidir. Bu arařtırma, katılımcıların sentetik seslerle etkileřimleri sonucunda yařadıkları deneyimleri belirleyerek, bu teknolojinin potansiyel avantajları ve sınırlamalarını anlamayı amalar.

3. ARAřTIRMANIN NEMİ

Bu arařtırma, retken yapay zeka teknolojileri tarafından retilen sentetik seslerin gerek insan sesleriyle karřılařtırmalı analizini yaparak, bu alandaki mevcut bilgi bořluęunu doldurmayı hedeflemektedir. Ses teknolojilerindeki hızlı geliřmeler, gnlk hayattan sanayi uygulamalarına kadar geniř bir yelpazede etkiler yaratmaktadır. Ancak, bu seslerin gereklik hissi, algılanıř biimleri ve insanlar zerindeki etkileri yeterince anlařılmamıřtır. Arařtırma, yapay zeka ile retilen seslerin doęallıęını, ifade yeteneęini ve gvenilirlięini derinlemesine inceleyerek, bu seslerin gerek seslerle olan benzerliklerini ve farklarını ortaya koyar. Bu alıřma, ses teknolojilerinin geliřtirilmesi ve optimizasyonu iin kritik veriler sunarak, yapay zeka tabanlı ses retimi alanında teknik ilerlemelerin yanı sıra etik ve sosyal boyutlarına da ıřık tutmaktadır. Ayrıca, sentetik seslerin kullanıcı deneyimi ve sosyal kabul zerine etkilerini deęerlendirerek, bu teknolojinin toplum tarafından nasıl algılandıęını ve kabul edildięini belirlemeye alıřır. Arařtırmanın sonuları, sentetik ses teknolojisinin potansiyelini ve kısıtlamalarını anlamada ve bu teknolojinin daha geniř bir kabul ve uygulama alanı bulmasında yardımcı olabilir. Bu nedenle, arařtırma, teknolojik geliřmelerin yanı sıra sosyal ve kltrel dinamikleri de dikkate alarak, ses teknolojilerinin gelecekteki ynn řekillendirmede nemli bir rol oynar.

4. ARAřTIRMANIN SAYILTILARI

retken yapay zeka teknolojisi ile retilen sentetik seslerin gerek insan sesleriyle karřılařtırılacak dzeyde farklılıklar olduęu,

Yapay zeka tabanlı sentetik seslerin kullanıcı deneyimi zerinde etki oluřturabileceęi,

Arařtırmaya ynelik belirlenen veri toplama yntemlerinin arařtırma amacına uygun olduęu,

Arařtırmaya ynelik yapılan ilgili yayın taramasının arařtırma bulgularını destekleyeceęi varsayılmaktadır.

5. ARAŞTIRMANIN SINIRLILIKLARI

Bu araştırma, Afyon Kocatepe Üniversitesi Konservatuvarında öğrenim gören yalnızca 20 katılımcı ile,

Yalnızca üretken yapay zeka teknolojileri kullanılarak oluşturulan sentetik seslerin ve bu seslerin karşılaştırıldığı gerçek ses kayıtlarının incelenmesi ile,

Dinleme testleri ve yarı yapılandırılmış görüşmeler sonucunda elde edilen veriler ile sınırlıdır.

6. YÖNTEM

Bu bölümde araştırmada izlenen bilimsel yöntemin ayrıntıları, araştırmanın yöntemi, çalışma grubu, verilerin toplanması, verilerin çözümlenmesi ve yorumlanması başlıkları altında incelenmiştir.

6.1. ARAŞTIRMA YÖNTEMİ

Bu çalışmada, üretken yapay zeka teknolojisi ile üretilen sentetik seslerin, gerçek seslerle karşılaştırılması ve değerlendirilmesi amaçlanmaktadır. Bu kapsamda, yapay zeka teknolojileri kullanılarak üretilen seslerin, gerçek insan sesleriyle olan benzerlikleri ve farklılıkları detaylı bir şekilde incelenmektedir. Betimsel araştırmalar, mevcut olayları ve koşulları önceki deneyimlerle ilişkilendirerek olaylar arasındaki etkileşimi açıklamayı amaçlar (Kaptan, 1995, s. 37). Bu bağlamda, var olan durumun, kullanılan yöntem çerçevesinde, olduğu haliyle saptanarak betimlendiği bu araştırma betimsel bir araştırma niteliğindedir.

Bu çalışmada nitel araştırma yöntemlerinden biri olarak kabul edilen betimsel araştırmalar, incelenen durumu aydınlatmak, araştırılan olgular arasındaki ilişkileri saptamak ve belirlenen çerçeveler içinde değerlendirmeler yapmak amacıyla yürütülmektedir (Çepni, 2018, s. 89). Betimsel araştırma yöntemleri, mevcut durumları ve koşulları gözlemleyerek, olaylar arasındaki ilişkileri ve etkileşimleri açıklamayı amaçlar (Karasar, 2023, s. 121). Bu çalışma, sentetik seslerin gerçek seslere ne derece yakın olduğunu veya bu sesler arasındaki farklılıkları belirlemek için betimsel bir metodoloji izlemektedir. Araştırmada kullanılan içerik analizi yöntemiyle dinleme testi ve yarı yapılandırılmış görüşmelerde elde edilen veriler düzenlenmiş ve içerik analizi yöntemiyle derinlemesine çözümlenmiştir. Elde edilen veriler, araştırmanın alt

problemlerine uygun şekilde sınıflandırılarak görsel ve tablolar şeklinde sunulmuş, tanımlanmış ve yorumlanmıştır.

6.2. ÇALIŞMA GRUBU

Bu araştırmada çalışma grubunu, Afyon Kocatepe Üniversitesi Konservatuvarında öğrenim gören yirmi öğrenci oluşturmaktadır. Katılımcılar, amaçlı örneklem yöntemi kullanılarak, araştırmanın amaçlarına uygun olarak önceden belirlenmiş müzikal eğitim ve deneyim kategorilerine göre seçilmiştir (Neuman & Robson, 2014). Katılımcılar, yapay zeka teknolojisi ile üretilen sentetik sesleri değerlendirecek ve bu seslerin gerçek insan sesleriyle olan benzerliklerini ve farklılıklarını belirlemek için dinleme testlerine katılmışlardır. Katılımcıların bu alandaki önceden sahip oldukları müzikal eğitim ve deneyimleri, ses değerlendirmeleri sırasında kritik bir öneme sahiptir. Katılımcılarla yarı yapılandırılmış görüşmelerde elde edilen veriler, sentetik seslerin insan kulağına olan etkisini daha geniş bir perspektiften analiz etmeye olanak tanıyacağı değerlendirilmektedir.

Tablo 1. Katılımcıların Demografik Yapısı

Katılımcı Sayısı	Konservatuvar Sınıfı	Cinsiyetler	Katılımcılar	(f)
1	1. Sınıf	1 Kadın	K1	1
3	2. Sınıf	2 Erkek, 1 Kadın	K2, K16, K3	3
7	3. Sınıf	4 Erkek, 3 Kadın	K8, K9, K7, K14, K12, K19, K20	7
9	4. Sınıf	5 Erkek, 4 Kadın	K4, K5, K6, K10, K11, K13, K15, K17, K18	9

Bu araştırma, Afyon Kocatepe Üniversitesi Konservatuvarı'nda öğrenim gören yirmi öğrenci üzerinden yürütülmüştür. Çalışma grubu, amaçlı örneklem yöntemi ile müzikal eğitim ve deneyim kriterlerine göre özenle seçilmiştir, böylece her bir katılımcının araştırmanın gerekliliklerini karşılayacak düzeyde bilgi ve tecrübeye sahip olması sağlanmıştır. Grup, müzik eğitimi alanında farklı akademik yıllardan öğrencileri barındırmakta olup; birinci sınıftan bir kadın, ikinci sınıftan iki erkek ve bir kadın, üçüncü sınıftan dört erkek ve üç kadın, ve dördüncü sınıftan beş erkek ve dört kadın öğrenci bulunmaktadır. Bu çeşitlilik, yapay zeka teknolojisi ile üretilen sentetik seslerin gerçek seslerle karşılaştırılması konusunda geniş bir perspektif sunmakta ve farklı müzikal algıları ortaya koymaktadır. Katılımcılar, seslerin benzerliklerini ve farklılıklarını

belirlemek için dinleme testlerine katılmış ve yarı yapılandırılmış görüşmelerde değerli bilgiler sağlamışlardır. Müzik eğitimi almış olmaları, ses değerlendirmelerinde önemli bir faktör olarak öne çıkmakta ve bu da araştırmanın derinliğini artırmaktadır. Katılımcıların bu deneyimleri, sentetik seslerin insan kulağına olan etkisini analiz etme konusunda kritik bir öneme sahiptir.

6.3. VERİLERİN TOPLAMA ARAÇLARI

Bu araştırmada kullanılan ana veri toplama araçları, dinleme testleri ve yarı yapılandırılmış görüşmelerdir. Dinleme testleri, katılımcılara ardışık olarak sunulan gerçek ve sentetik ses kayıtlarını değerlendirmeleri amacıyla düzenlenmiştir. Bu testler sırasında, katılımcılardan hangi ses kaydının gerçek olduğunu belirlemeleri istenmiştir. Testler, katılımcıların sesler arasındaki farkları nasıl algıladıklarını anlamak için tasarlanmıştır. Görüşme, sözlü iletişim yoluyla veri toplama tekniğidir ve bu teknik, yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış görüşme olarak üçe ayrılmaktadır (Çepni, 2018). Yarı yapılandırılmış görüşmeler, araştırmacı tarafından önceden hazırlanan ancak görüşme sürecine göre esneklik gösteren sorular içermektedir. Bu görüşmeler, katılımcıların sentetik seslerle ilgili deneyimlerini, tercihlerini ve bu seslerin doğallığına dair görüşlerini derinlemesine anlamak için kullanılmıştır. Görüşmeler dinleme testi sonrasında yüz yüze gerçekleştirilmiş ve yazılı olarak kaydedilmiştir. Bu veri toplama yöntemlerinin kombinasyonu, sentetik seslerin gerçek seslere karşı olan performansını çeşitli açılardan değerlendirme imkanı sunmuştur. Araştırmadan elde edilen verilerin güvenilirliğini tespit etmek için iki alan uzmanının değerlendirmeleri doğrultusunda bir tutarlılık incelemesi yapılmıştır. Elde edilen değer %70'in üzerinde olduğu için araştırmanın güvenilirliği kabul edilmiştir.

6.4. VERİLERİN TOPLANMASI VE ÇÖZÜMLENMESİ

Verilerin toplanması, katılımcılar ile yapılan dinleme testi ve yüz yüze yarı yapılandırılmış görüşmelerle sağlanmıştır. Veriler, içerik ve betimsel analiz yöntemleri ile analiz edilmiştir. Verilerin toplanmasında, yaratıcı yapay zeka teknolojisi kullanılarak üretilen sentetik sesler ve bunların karşılaştırılması için kullanılan gerçek insan ses kayıtları öncelikle detaylı bir süreçle işlenmiştir. Ses kayıtları, yüksek kalitede bir kayıt cihazı (Zoom H6) kullanılarak alınmış ve daha sonra bu kayıtlar bilgisayar ortamına aktarılmıştır. Bilgisayar ortamına aktarılan bu sesler, yapay zeka teknolojisi kullanan web

tabanlı elevenlabs.io sisteminde ses klonlama teknikleri kullanılarak işlenmiş ve sentetik sesler üretilmiştir.

Eleven Labs, Anında Ses Klonlama (IVC) ve Profesyonel Ses Klonlama (PVC) olmak üzere iki ana ses klonlama teknolojisi sunar. IVC, kısa örneklerden hızlı ses klonlamayı sağlar ve sıfır atış öğrenme adı verilen bir yöntem kullanır. Bu yöntem, yapay zekanın kısa bir ses klibini dinleyerek sesi anında çoğaltmasını sağlar ve geniş veri setlerine ihtiyaç duymaz, ancak doğruluğu daha düşüktür. PVC ise, yüksek doğrulukta bir ses modeli oluşturmak için büyük bir ses verisi seti üzerinde eğitim gerektirir ve genellikle en az 30 dakikalık temiz ses verisi gerektirir. Bu model, orijinal sesin nüanslarını daha doğru bir şekilde yakalar ve çeşitli doğrulama adımları içerir. Eleven Labs ayrıca, yaş, cinsiyet ve aksan gibi parametrelere dayalı olarak tamamen yeni sentetik sesler oluşturabilen bir Ses Tasarımı teknolojisi sunar. Bu sesler benzersizdir ve belirli bir kişiyi taklit etmez. Bu özellik, çeşitli endüstrilerin ihtiyaçlarını karşılamak için yaratıcı özgürlük ve maliyet avantajı sağlar. Eleven Labs'ın Metinden Sese (TTS) teknolojisi, doğal insan konuşma modellerini taklit etmek için tasarlanmış sofistike algoritmalar kullanır. Bu teknoloji, yazılı metni hece, kelime ve cümlelere ayırır ve fonem adı verilen sesleri bağlayarak anlaşılır konuşma üretir. Bağlam farkındalığı ve yüksek sıkıştırma gibi gelişmiş özellikler de içeren bu teknoloji, konuşmanın doğal ve metnin bağlamına göre ayarlanmış olmasını sağlar. Bu araştırma bağlamında TTS teknolojisi kullanılmıştır.

Dinleme testleri ve yarı yapılandırılmış görüşmeler, bu sentetik seslerin değerlendirilmesi için kullanılan başlıca veri toplama araçlarıdır. Dinleme testlerinde, katılımcılara ardışık olarak sunulan gerçek ve sentetik ses kayıtları dinletilmiş ve hangi sesin gerçek olduğunu belirlemeleri istenmiştir. Bu testler, katılımcıların sesler arasındaki farkları algılama yeteneklerini ölçmek için kritik öneme sahiptir. Yarı yapılandırılmış görüşmelerde ise, katılımcılara seslerin doğallığı, ifade biçimi ve genel algıları üzerine sorular sorulmuştur. Bu görüşmeler, katılımcıların detaylı yorumlarını ve tepkilerini elde etmek için yüz yüze gerçekleştirilmiştir. Bu araştırma çerçevesinde veriler, içerik analizi ve betimsel analiz yöntemleri kullanılarak derinlemesine analiz edilmiştir. İçerik analizi, verilerin kodlanması, kategorize edilmesi ve temaların belirlenmesi işlemlerini içerir. Bu yöntem, verilerden önceden belirlenmemiş boyutların ve temaların ortaya çıkarılmasını sağlar. Betimsel analiz ise, katılımcıların yanıtlarını ve görüşlerini doğrudan kullanarak, mevcut durumu ve olayları detaylı bir şekilde açıklamaya odaklanır. Araştırmadan elde edilen bulgular, sentetik seslerin insan sesine ne derecede benzediğini veya farklılaştığını

belirlemede kullanılmıştır. Ayrıca, bu bulgular katılımcıların sentetik sesleri nasıl algıladıklarını ve bu seslerin gerçeklik hissini ne derecede yansıttığını göstermiştir. Bu yöntemlerle elde edilen veriler, sentetik ses teknolojisinin potansiyelini ve sınırlamalarının derinlemesine incelenmesine olanak tanımaktadır.

Bu araştırmada verilerin toplanması ve çözümlenmesi beş basamaktan oluşmaktadır. Bu işlem basamakları şunlardır:

- Hazırlık aşaması
- Veri toplama işlemi
- Verilerin analizi
- Bulguların raporlanması
- Tartışma, sonuç ve öneriler.

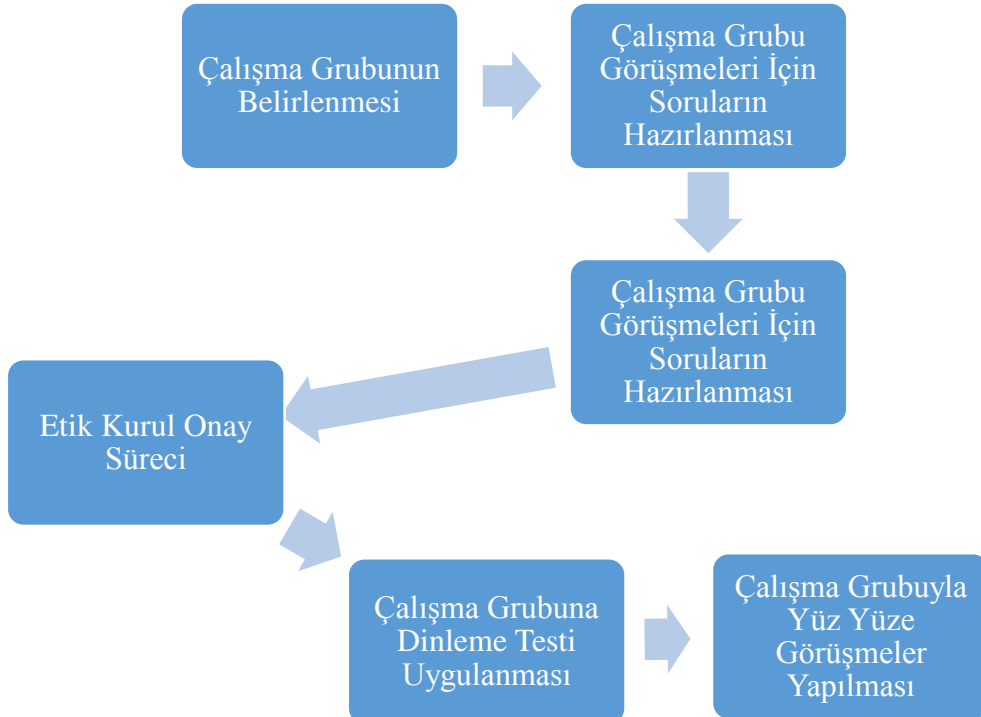
Bu kapsamda Şekil 1’de hazırlık aşamasına ilişkin alt basamaklar gösterilmiştir.

Şekil 1. Hazırlık Aşaması Alt Basamakları



Hazırlık aşamasından sonra veri toplama işlemi gelmektedir. Veri toplama işlemine ilişkin alt basamaklar Şekil 2’de gösterilmiştir.

Şekil 2. Veri Toplama İşlemi Alt Basamakları



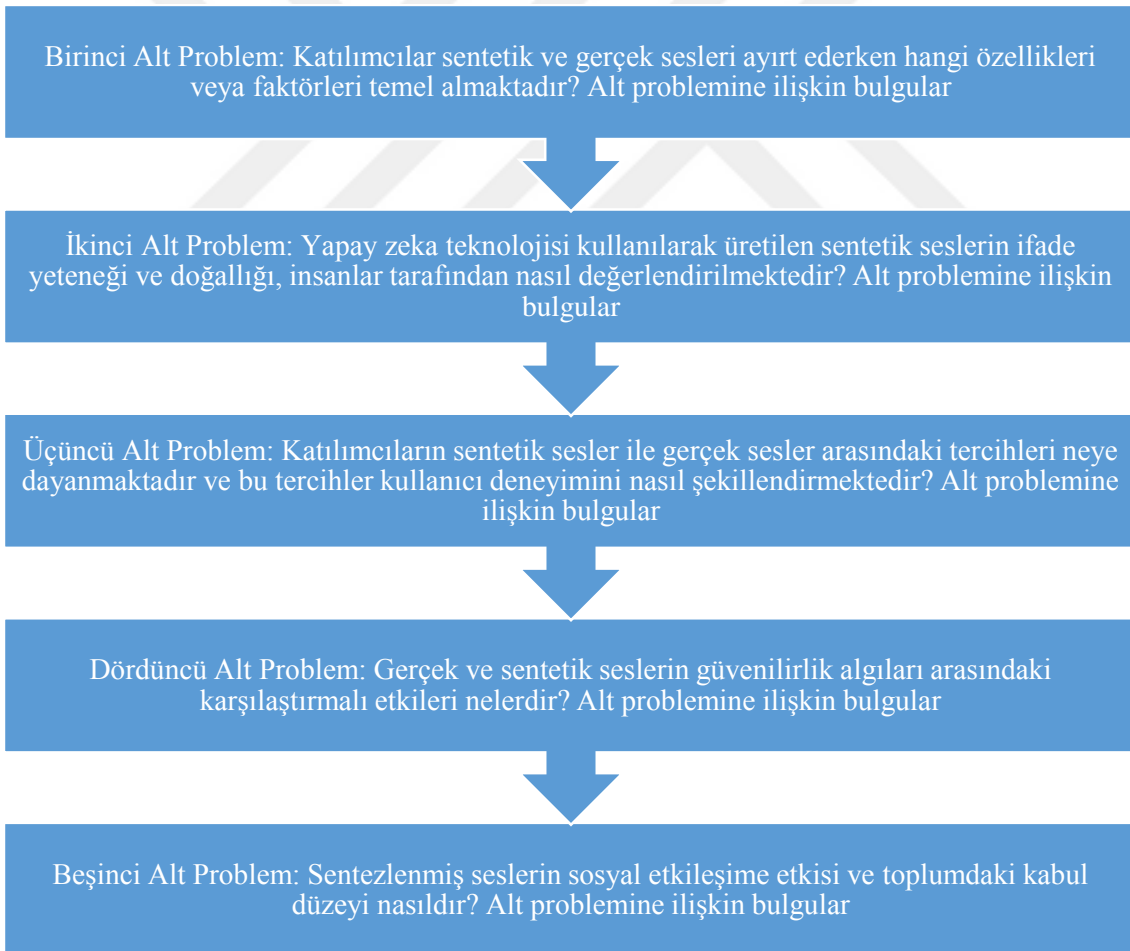
Veri toplama işlemi basamaklarından sonra verilerin analizi işlem basamakları gelmektedir. Verilerin analizi aşamasındaki işlem basamakları Şekil 3'te gösterilmektedir.

Şekil 3. Veri Analizi İşlemi Alt Basamakları



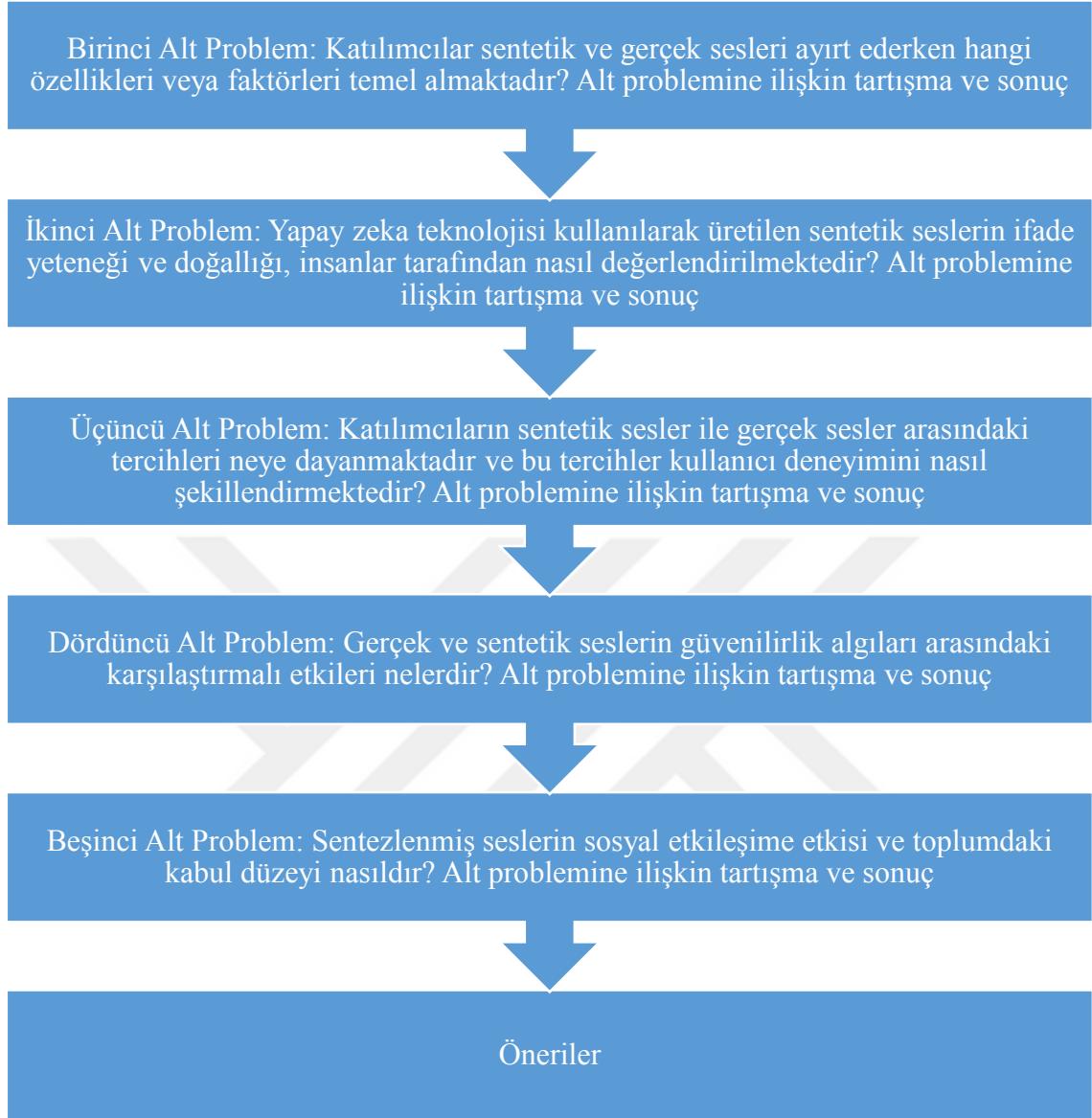
Verilerin analizi işlem basamaklarından sonra bulguların raporlanması işlem basamakları gelmektedir. Bulguların raporlanması işlem basamaklarına ilişkin alt basamaklar Şekil 4'te gösterilmektedir.

Şekil 4. Bulguların Raporlanması İşlemi Alt Basamakları



Bulguların raporlanması sonrası tartışma, sonuç ve öneriler işlem basamakları gelmektedir. Tartışma, sonuç ve öneriler işlem basamaklarına ilişkin alt basamaklar Şekil 5'te gösterilmiştir.

Şekil 5. Tartışma, Sonuç ve Öneriler İşlemi Alt Basamakları



7. ARAŞTIRMANIN BULGULARI VE YORUMLAR

7.1. BİRİNCİ ALT PROBLEME YÖNELİK BULGULAR

Bu alt başlıkta “Katılımcılar sentetik ve gerçek sesleri ayırt ederken hangi özellikleri veya faktörleri temel almaktadır?” alt problemine ilişkin bulgular ve yorumlar yer almaktadır. Katılımcıların sentetik ve gerçek sesleri ayırt ederken hangi özellikleri veya faktörleri temel aldıklarını anlamak amacıyla üç yarı yapılandırılmış araştırma sorusu yöneltilmiştir. Araştırma soruları:

- Dinlediğiniz seslerden biri gerçek ses diğeri ise sentetik ses, gerçek sesi sentezlenmiş sestten ayırırken hangi etmenlerden etkilenerек sonuca ulaştınız?
- Sizce gerçek ses ve sentezlenmiş ses arasındaki fark nedir?

- Bu çalışmanın sizin sonraki yaşamınızda dinlediğiniz bir konuşma, şarkı ya da ses içerisinde gerçek ve sentetik ses ayırımını yapmanızda farkındalık yaratabileceğini düşünüyor musunuz?

Bu görüşme soruları, katılımcıların dinleme testinde ardışık sunulan ses kayıtları arasından gerçek sesi nasıl algıladıklarını ve ayırt ettiklerini ortaya koymak amacıyla tasarlanmıştır. Araştırma sorularının amacı, katılımcıların ses algısı üzerindeki etmenleri belirlemek ve değerlendirmektedir. Katılımcıların bu sorulara yönelik görüşlerinin içerik analizi sonucu elde edilen içerik analizine ilişkin veriler Tablo 2.'de gösterilmektedir.

Tablo 2. Katılımcıların "Algı" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	f
Algı	Ayırt Edici Özellik	Canlılık ve Arka Plan Gürültüsü	K1, K5, K7	3
		Ses Tonunun Yapaylığı	K2, K8, K14, K17	4
	Gerçeklik	Doğal Titreşim ve Monotonluk	K3, K12	2
		Duygusal İfade ve Ses Rengi	K4, K6, K11, K16, K18, K19	6
	Farkındalık Düzeyi	Ses Kalitesi ve Netliği	K10, K15, K20	5
		Medya ve Teknoloji Farkındalığı	K2, K5, K6, K9, K13, K17	6

Tablo 2'den elde edilen bulgular sonucunda; 1 tema, 3 kategori, 6 kod elde edilmiştir. Katılımcıların sentetik seslerin canlılığına, arka plan gürültüsüne, ses tonunun yapaylığına, doğal titreşim ve monotonluk farklarına, duygusal ifade ve ses rengi zenginliğine, ses kalitesi ve netliğine, medya ve teknoloji farkındalığına cevaplar verdiği belirlenmiştir.

Bu cevaplar, katılımcıların gerçek ve sentetik sesler arasındaki ayırt edici özellikleri, gerçeklik algılarını ve teknoloji kullanımı sırasında farkındalık düzeylerini yansıtmaktadır. Her bir kategoride iki kod yer almış ve bu kodlara yönelik belirtilen katılımcıların detaylı cevapları incelenmiştir.

Tablo 2'ye göre, sentetik ve gerçek sesler arasındaki algı farklılıkları üç ana kategori altında incelenmiştir: "Ayırt Edici Özellik", "Gerçeklik" ve "Farkındalık Düzeyi". Her bir kategori, belirli kodlar oluşturularak, bu kodlara katılımcılar ve frekanslar atılmıştır.

Ayrt Edici Özellik:

- Canlılık ve Arka Plan: Bu koddaki üç katılımcı (K1, K5, K7) sentetik seslerdeki canlılık eksikliği ve arka plan seslerinin farklılıklarını vurgulamış, bu özelliklerin ayırt edici olduğunu belirtmişlerdir.

- Yapaylık: Dört katılımcı (K2, K8, K14, K17) sentetik seslerin yapay tonunu ve gerçek seslere kıyasla doğallık eksikliğini ifade etmişlerdir. Bu da, sesin yapaylığının önemli bir ayırt edici özellik olduğunu göstermektedir.

Gerçeklik:

- Doğal Titreşim ve Monotonluk: İki katılımcı (K3, K12) gerçek seslerin doğal titreşimine ve sentetik seslerin monotonluğuna dikkat çekmişlerdir, bu özelliklerin seslerin gerçekliğini algılama üzerinde etkili olduğunu vurgulamışlardır.

- Ses Rengi: Altı katılımcı (K4, K6, K11, K16, K18, K19) ses renginin zenginliğini ve duygusal ifadeleri nasıl taşıdığını belirtmişlerdir. Bu, seslerin duygusal derinliğinin ve tonal nüanslarının gerçeklik algısını güçlendirdiğini ortaya koymaktadır.

Farkındalık Düzeyi:

- Ses Kalitesi: Üç katılımcı (K10, K15, K20) sentetik seslerin netliği ve kalitesi hakkında görüş bildirmiş, bu özelliklerin kullanıcı deneyimini ve sesler arası farkındalığı artırdığını ifade etmişlerdir.

- Medya ve Teknoloji: Altı katılımcı (K2, K5, K6, K9, K13, K17) medya ve teknolojiye kullanılan sentetik seslerin farkındalığını artırıcı etkisine dikkat çekmişlerdir. Bu, teknolojik seslerin kullanımının kullanıcıların ses algısı üzerindeki etkisini göstermektedir.

Birinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Dinlediğiniz seslerden biri gerçek ses diğeri ise sentetik ses, gerçek sesi sentezlenmiş sestten ayırırken hangi etmenlerden etkilenerек sonuca ulaştınız?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Gerçek ses daha canlıydı, nefes alıp verme sesleri ve hafif arka plan gürültüsü sentetik olandan ayırt ediciydi."*

K2: *"Sentezlenmiş sesin tonu biraz yapay geliyordu, gerçek ses ise daha sıcak ve doğal tınılara sahipti. Medya da kullanılan seslerle benzerlik vardı."*

K3: "Gerçek sesin ses dalgalarında doğal bir titreşim var, sentetik ses ise daha düz ve monoton."

K4: "Gerçek ses duygusal ifadelerde daha yoğun, sentetik ses ise duyguları aynı derecede yansıtamıyordu."

K5: "Sentezlenmiş seste arka plandaki sesler kesilmiş gibi netti, gerçek ses daha karmaşık ve zengindi. Teknolojik bir dokunuş olduğu hissediliyordu."

K6: "Gerçek sesin ses rengi daha zengin ve dolgun, sentetik ses ise daha sönük ve tek boyutlu. Sosyal medyada gördüğü çalışmalarda denk geldiğim bir ses gibiydi."

K7: "Gerçek sesin doğal hataları, nefes sesleri ve tonlamaları sentetik sesinkinden farklıydı. Arka plan farklıydı."

K8: "Sentezlenmiş sesin vurguları yapay ve zamanlaması dengesizdi, gerçek ses daha akıcı ve doğaldı."

K9: "Gerçek sesin yüksekliği ve alçaklığı daha doğal değişkenlik gösterirken, sentetik ses daha sabitti. Teknolojik sesler çok belirgindi."

K10: "Sentezlenmiş sesteki titremeler ve sert geçişler, gerçek sesin yumuşak ve doğal geçişlerine kıyasla ses kalitesi dikkat çekiciydi."

K11: "Gerçek sesin duygusal derinliği ve tonal nüansları, sentetik sesteki ayırt edilmesini sağladı."

K12: "Gerçek sesin ses dalgalanmalarında organik bir akıcılık varken, sentetik ses daha mekanikti. Monotonluk dikkat çekiciydi."

K13: "Gerçek sesin insani unsurları, öksürük veya hafif anlam bozuklukları gibi, sentetik seste bulunmuyordu. Sentetik seste teknolojik bazı belirtiler vardı."

K14: "Sentezlenmiş ses daha keskin ve soğuktu, gerçek ses ise sıcaklık ve doğallık içeriyordu. Yapaylık hissettim."

K15: "Gerçek sesin ses kalitesindeki doğal ton farklılıkları, sentetik sesin genellikle düz tonundan ayırt ediciydi."

K16: "Gerçek ses daha çok duygusal ifade ve sıcaklık içerirken, sentetik ses bu boyutlardan yoksundu."

K17: "Sentezlenmiş sesin monoton yapısı ve yapay tonlamaları, gerçek sesin doğallığı ile tezat oluşturmuyordu. Dijital medyada kullanılan sese benziyordu."

K18: "Gerçek sesin varlığı ve sesin duygusal derinliği, sentetik sesin yapay ve yüzeysel yapısından kolaylıkla ayırt edilebiliyordu."

K19: "Sentezlenmiş seste vurgular yapayken, gerçek ses doğal vurgular ve duygusal ifadelerle doluydu."

K20: "Gerçek sesin doğal vibrasyonu ve akustiği, sentetik sesin katı ve düzenli yapısından farklıydı. Ses kalitesi dikkat çekiciydi."

Birinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Sizce gerçek ses ve sentezlenmiş ses arasındaki fark nedir?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Bana göre, sentezlenmiş ses gerçek sesin doğallığından yoksun. Ama ton kalitesi ve netliği açısından daha üstün diyebilirim. Gerçek sesin canlılığı ve arka plan gürültüsü ise oldukça belirgin."

K2: "Gerçek ses daha ham ve düzensiz. Sentetik ses ise bu açıdan mükemmel ve keskin hatlara sahip. Ama bence sentetik sesin tonunda bir yapaylık var."

K3: "Gerçek sesin arka plan gürültüsü var ve bu ona doğal titreşimler katıyor. Sentetik ses ise tamamen temiz ve kesintisiz, bu da onu biraz monoton yapıyor."

K4: "Sentezlenmiş ses, gerçek sesin sağlayamadığı bir netlik ve keskinlik sunuyor. Ama gerçek sesin duygusal ifadesi ve ses rengi bambaşka."

K5: "Gerçek sesin doğal dalgalanmalar içerdiğini düşünüyorum, bu da onu daha canlı kılıyor. Öte yandan, sentetik ses çok daha düzgün ve tutarlı oluyor. Ancak, gerçek sesin canlılığı ve arka plan gürültüsü onu farklı kılıyor."

K6: "Sentezlenmiş ses, gerçek sesin kaçırdığı detayları bile yakalayıp sunuyor. Ancak gerçek sesin duygusal ifadesi ve ses rengi sentetik seste eksik."

K7: "Bence, gerçek sesin organik bir dokusu var, bu da ona bir canlılık katıyor. Sentetik ses ise çok daha düzenli ve pürüzsüz. Gerçek sesin canlılığı ve arka plan gürültüsü ise hemen fark ediliyor."

K8: "Sentezlenmiş sesin vurguları daha belirgin ve net, gerçek ses daha spontane. Sentetik sesin tonunda bir yapaylık hissettiriyor."

K9: "Gerçek sesin sıcaklığı var ancak sentetik ses daha geniş bir spektrum sunuyor."

K10: "Gerçek ses çevresel seslerle dolu, bu da onu daha doğal yapıyor. Ancak sentetik ses daha odaklanmış ve net bir ses kalitesine sahip."

K11: "Gerçek ses bazen düzeltme gerektirebilir, sentetik ses baştan sona kusursuz. Fakat gerçek sesin duygusal ifadesi ve ses rengi sentetik sesle kıyaslanamaz."

K12: "Sentezlenmiş ses, vurgu ve tonlama konusunda gerçek sesi geride bırakıyor. Ancak, sentetik sesin monotonluğu gerçek sesin doğal titreşimlerini aratıyor."

K13: "Gerçek sesin doğallığı hoş fakat sentetik ses çok daha teknik ve detaylı."

K14: "Sentezlenmiş ses, teknoloji sayesinde gerçek sesin erişemediği bir düzeyde. Ancak, sentetik sesin tonunun yapaylığı dikkat çekiyor."

K15: "Gerçek sesin insan dokunuşu hissedilir ve bu onu özel kılar. Sentetik ses ise bu yönü optimize ederek ses kalitesi ve netliği konusunda üstün."

K16: "Gerçek ses dalgalı ve değişkenken, sentetik ses her zaman dengeli ve tutarlı. Yine de, gerçek sesin duygusal ifadesi ve ses rengi daha zengin."

K17: "Sentezlenmiş ses, teknolojik avantajlarıyla gerçek sesin ötesine geçiyor. Yine de, ses tonunun yapaylığı belirgin oluyor."

K18: "Gerçek ses duygusal, sentetik ses ise bu duyguları daha geniş bir yelpazede yansıtıyor. Ama sentetik sesin duygusal ifadesi ve ses rengi, gerçek sesin yerini tutmuyor."

K19: "Gerçek ses doğal ama sentetik sesin tutarlılığı ve doğruluğu daha belirgin. Fakat gerçek sesin duygusal ifadesi ve ses rengi üstün."

K20: "Sentezlenmiş sesin sunumu profesyonel ve net, ses kalitesi yüksek. Gerçek ses ise daha rastgele ve spontane, fakat yine de kendine has bir çekiciliği var."

Birinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Bu çalışmanın sizin sonraki yaşamınızda dinlediğiniz bir konuşma, şarkı ya da ses içerisinde gerçek ve sentetik ses ayırımını yapmanızda farkındalık yaratabileceğini düşünüyor musunuz?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Evet, bu çalışma bana sesler arasındaki farkları daha net görmemi sağladı, özellikle seslerin canlılığı konusunda."

K2: "Bu deneyim, medya tüketimim sırasında ses tonunun yapaylığını fark etme konusunda beni daha bilinçli hale getirdi."

K3: "Artık filmlerde veya reklamlarda doğal titreşimleri olmayan sentetik sesleri ayırt edebiliyorum, bu çalışma farkındalığımı artırdı."

K4: "Daha önce fark etmediğim ses detaylarını şimdi ayırt edebiliyorum, özellikle seslerin duygusal ifadesine ve ses rengine dikkat ediyorum."

K5: "Bu çalışma, teknolojik seslerin varlığını daha çok fark etmeme yardımcı oldu. Artık sesler arasındaki arka plan gürültüsünü daha iyi ayırt ediyorum."

K6: "Ses teknolojisine olan bakış açım değişti, sentetik seslerin duygusal ifadesini ve ses rengini daha kolay fark ediyorum."

K7: "Çalışma, günlük yaşamdaki sesleri daha bilinçli dinlememi sağladı. Özellikle müzik ve podcastlerdeki seslerin canlılığı konusunda farkındalığım arttı."

K8: "Evet, bu çalışma sayesinde sesleri daha dikkatli dinler oldum ve ses tonunun yapaylığını ayırt etmekte daha iyiyim."

K9: "Sentezlenmiş sesler konusunda artık daha duyarlıyım, reklam ve müziklerde farkını anlayabiliyorum."

K10: "Bu deneyim, sentetik seslerin ne kadar gelişmiş olduğunu gösterdi ve artık ses kalitesi ve netliği açısından onları daha rahat tespit edebiliyorum."

K11: "Artık her sesi duyduğumda, gerçek mi yoksa üretilmiş mi diye düşünüyorum, bu çalışmanın duygusal ifade ve ses rengi açısından etkisi büyük oldu."

K12: "Sentezlenmiş ve gerçek sesler arasındaki farkları daha net görebiliyorum, özellikle ses oyunlarında monotonluk dikkatimi çekiyor."

K13: "Bu çalışma sayesinde sesleri daha kritik bir şekilde değerlendirme eğilimindeyim."

K14: "Çalışma, seslerin nasıl üretildiğine dair bilincimi artırdı ve şimdi ses tonunun yapaylığını daha ayrıntılı dinliyorum."

K15: "Evet, bu çalışma bana sesleri analiz etme konusunda yeni bir perspektif sundu, özellikle ses kalitesi ve netliği açısından."

K16: "Şimdi sesleri dinlerken, hangisinin gerçek hangisinin yapay olduğunu ve duygusal ifadeyi nasıl etkilediğini düşünmekten kendimi alamıyorum."

K17: "Bu deneyim, özellikle dijital ortamlarda ses tonunun yapaylığını ayırt etme becerimi geliştirdi."

K18: *"Bu çalışma, sentetik sesleri daha bilinçli fark etmeme olanak sağladı, özellikle animasyon ve video oyunlarında duygusal ifade ve ses rengi açısından."*

K19: *"Seslerin yapay olup olmadığını ve ses rengini düşünerek dinlemek, bu çalışmadan sonra alışkanlık haline geldi."*

K20: *"Bu çalışma sonrasında, özellikle müzik dinlerken ses kalitesi ve netliği açısından sentetik sesleri fark etmeye daha fazla dikkat ediyorum."*

Bu veriler, katılımcıların sentetik ve gerçek sesleri ayırt ederken temel aldıkları özelliklerin geniş bir yelpazeye yayıldığını göstermektedir. Katılımcılar, sesin canlılığı, yapaylık derecesi, doğal titreşimler, ses rengi zenginliği, netlik ve medya ile teknolojik farkındalık gibi faktörleri özellikle önemsemekte ve bu faktörlerin ses algısı üzerindeki doğrudan etkilerini belirtmektedirler.

Bu özellikler, gerçek ve sentetik sesler arasındaki farkları belirginleştirmede kritik rol oynamakta ve katılımcıların bu sesleri nasıl algıladıklarını derinlemesine yansıtmaktadır.

7.2. İKİNCİ ALT PROBLEME YÖNELİK BULGULAR

Bu alt başlıkta “Yapay zeka teknolojisinin ses üretimindeki doğallığına ve sentetik seslerin ifade yeteneğine katkıda bulunup bulunmadığı” alt problemine ilişkin bulgular ve yorumlar yer almaktadır. Katılımcıların yapay zeka tarafından üretilen seslerin doğallığını ve ifade yeteneğini değerlendirme sürecini anlamak amacıyla yarı yapılandırılmış araştırma soruları yöneltilmiştir.

Araştırma soruları:

- Ses sentezi sürecinde yapay zeka teknolojisinin sesin doğallığına ne kadar katkıda bulunduğunu düşünüyorsunuz?
- Sentezlenmiş sesin anlamını ve ifadeyi doğru bir şekilde aktarma yeteneğine nasıl bakıyorsunuz?

Bu görüşme soruları, katılımcıların yapay zeka teknolojisinin sentetik ses üretimindeki doğallığa ve ifade yeteneğine olan etkisini belirlemek ve değerlendirmek amacıyla tasarlanmıştır.

Sorular, katılımcıların bu teknolojinin sesler üzerindeki etkilerini nasıl algıladıklarını ve sentetik seslerin insan sesleriyle kıyaslandığında ne kadar doğal ve ifade

yeteneğine sahip olduğunu ortaya koymak için hazırlanmıştır. Katılımcıların bu sorulara yönelik görüşlerinin içerik analizi sonucu elde edilen içerik analizine ilişkin veriler Tablo 3'te gösterilmektedir.

Tablo 3. Katılımcıların "Doğallık" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	<i>f</i>
Doğallık	Netlik ve Düzen	Netlik Artışı	K1, K3, K6, K10, K12, K17	6
		Mekanik His	K3, K8, K9, K12, K18	5
		Doğal Sesin Sıcaklığı	K4, K7, K14, K16, K18	5
	İfade Yeteneği	Duygusal İfade Eksikliği	K1, K2, K4, K5, K7, K8	6
		Anlamanın Doğru Aktarımı	K1, K6, K11, K16, K19	5
		İfade Gücü ve Otantiklik	K4, K5, K8, K14, K20	5

Tablo 3'ten elde edilen bulgular sonucunda; 1 tema, 2 kategori, 6 kod elde edilmiştir. Katılımcıların yapay zeka ile üretilen seslerin netliğine, mekanik hissine, doğal sesin sıcaklığına, duygusal ifade eksikliğine, anlamanın doğru aktarımına ve ifade gücü ile otantikliğe ilişkin cevapları belirlenmiştir. Bu cevaplar, katılımcıların sentetik seslerin doğallığı ve ifade yeteneği üzerindeki algılarını yansıtmaktadır. Her bir kategoride üçer kod yer almış ve bu kodlara yönelik belirtilen katılımcıların detaylı cevapları incelenmiştir.

Tablo 3'e göre, yapay zeka ile üretilen sentetik seslerin doğallığı ve ifade yeteneği iki ana kategori altında incelenmiştir: "Doğallık" ve "İfade Yeteneği". Her bir kategori, belirli kodlar oluşturularak, bu kodlara katılımcılar ve frekanslar atılmıştır.

Doğallık:

- **Netlik:** Bu koddaki altı katılımcı (K1, K3, K6, K10, K12, K17) yapay zekanın sesin netliğini artırdığını vurgulamış, ancak sesin tamamen doğal hissetmediğini belirtmişlerdir.
- **Mekanik His:** Beş katılımcı (K3, K8, K9, K12, K18) teknolojinin sesi temiz ve net yapmasına rağmen bazen çok mekanik ve yapay hissettirdiğini ifade etmişlerdir.
- **Doğal Sesin Sıcaklığı:** Beş katılımcı (K4, K7, K14, K16, K18) yapay zekanın sesin ifadesini artırmasına rağmen, doğal ses kadar sıcaklık sunmadığını dile getirmişlerdir.

İfade Yeteneği:

- Duygusal İfade Eksikliği: Altı katılımcı (K1, K2, K4, K5, K7, K8) sentetik seslerin anlamı iyi aktardığını, fakat duygusal ifadeyi tam olarak yansıtamadığını belirtmişlerdir.

- Anlamın Doğru Aktarımı: Beş katılımcı (K1, K6, K11, K16, K19) sentetik seslerin anlamı net ve doğru bir şekilde aktardığını, fakat ifadede bazı eksiklikler olduğunu vurgulamışlardır.

- İfade Gücü ve Otantiklik: Beş katılımcı (K4, K5, K8, K14, K20) teknolojinin ifade gücünü artırdığını, ancak seslerin bazen hesaplı ve yapay geldiğini ifade etmişlerdir.

İkinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Ses sentezi sürecinde yapay zeka teknolojisinin sesin doğallığına ne kadar katkıda bulunduğunu düşünüyorsunuz?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Yapay zeka, sesin netliğini artırmakla kalmıyor, aynı zamanda ifade yeteneğini de yükseltiyor; ancak tamamen doğal hissettirmiyor."*

K2: *"Teknoloji, sesin doğallığını belirli bir düzeyde sağlıyor, ama yine de yapay bir hissi var."*

K3: *"Yapay zeka sesi temiz ve net yapabiliyor, fakat bazen çok mekanik ve yapay hissediliyor."*

K4: *"Ses sentezi süreci, ifade yeteneğini artırıyor, ancak doğallık konusunda gerçek sesin yerini tam anlamıyla alamıyor."*

K5: *"Yapay zeka, sesin düzgünlüğü ve vurguları konusunda büyük katkı sağlıyor, ama bu sesler hala yapay duyulabiliyor."*

K6: *"Teknolojinin ifade yeteneği ve doğallığı artırma konusunda olumlu etkileri var, ama gerçek sesin doğallığını tam olarak taklit edemiyor."*

K7: *"Yapay zeka, sesin akışını ve tonunu iyileştiriyor, ancak doğal ses kadar sıcaklık sunmuyor."*

K8: *"Teknoloji, sesin ifadesini geliştiriyor, ama bazen bu sesler çok hesaplı ve yapay geliyor."*

K9: "Ses sentezi, sesin kullanım alanını genişletiyor, fakat doğallığını tam anlamıyla yansıtamıyor."

K10: "Yapay zeka sesleri daha anlaşılır ve net kılıyor, ama doğallıktan bir miktar ödün veriliyor."

K11: "Teknolojinin ses üretimindeki etkisi büyük, ses daha berrak ve net, ancak yine de bazı yapay tonlamalar hissediliyor."

K12: "Yapay zeka, sesin temizliği ve düzeni konusunda mükemmel iş çıkarıyor, ancak doğallık azalıyor."

K13: "Teknoloji sesin ifadesini ve netliğini artırıyor; fakat doğallık hissi tam olarak sağlanamıyor."

K14: "Ses sentezinde yapay zeka, performansı yükseltiyor, doğal sesin sıcaklığını veremese de."

K15: "Yapay zeka, sesin ifadesini güçlendiriyor, ancak bu sesler zaman zaman yapay hissedebiliyor."

K16: "Teknoloji sesin netliği ve düzenini artırıyor, ama doğal sesin yerini tam olarak alamıyor."

K17: "Yapay zeka ile üretilen sesler, netlik ve ifade yeteneği açısından iyi seviyede, ancak doğallıkta eksikler var."

K18: "Sesin ifadesi ve anlaşılabilirliği yapay zeka ile artırılıyor, fakat bu seslerin doğallığı sorgulanabilir."

K19: "Yapay zeka sesi optimize ediyor, daha net ve düzgün hale getiriyor, ama bazen doğallıktan uzaklaşıyor."

K20: "Teknoloji sesin kalitesini ve kullanımını iyileştiriyor, ama doğal sesin duygusal derinliğini tam anlamıyla kopyalayamıyor."

İkinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Sentezlenmiş sesin anlamını ve ifadeyi doğru bir şekilde aktarma yeteneğine nasıl bakıyorsunuz?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Sentezlenmiş sesler anlamı iyi aktarıyor, fakat duygusal ifadeyi tam olarak yansıtamıyor."

- K2: *"Anlam açısından net, fakat ifade yeteneği bazen yapay hissettiriyor."*
- K3: *"Sentezlenmiş sesler genellikle anlaşılır, ancak gerçek sesin duygusal derinliğine ulaşamıyor."*
- K4: *"Anlamı doğru aktarıyorlar, ancak ifadedeki sıcaklık eksik."*
- K5: *"Çoğunlukla doğru anlamı iletiyorlar, ama duygusal ifadede gerçek ses kadar etkileyici değiller."*
- K6: *"Sentez sesler teknik olarak kusursuz anlam aktarımı yapıyor, ama ifadede biraz mekanik kalıyorlar."*
- K7: *"Bu sesler anlamı iyi aktarıyor ancak duygusal ifadeyi tam olarak veremiyor."*
- K8: *"Anlam transferi mükemmel; fakat ifadede gerçek ses kadar otantik değil."*
- K9: *"Anlam açısından güçlü, ifade yönünden ise biraz zayıf."*
- K10: *"İfade yeteneği iyileşiyor, ancak duygusal yankı eksik kalabiliyor."*
- K11: *"Sentezlenmiş sesler anlamı net bir şekilde aktarıyor, ancak ifadede daha fazla gelişime ihtiyaç var."*
- K12: *"İfade gücü gelişmiş, ancak gerçek ses kadar dokunaklı değil."*
- K13: *"Anlam doğru iletiliyor, ama ifadenin duygusal derinliği eksik."*
- K14: *"Seslerin anlam aktarımı keskin, fakat ifade gücü gerçek sesler kadar zengin değil."*
- K15: *"Anlam iyi, ifade ise yapay zeka tarafından geliştirilmeli."*
- K16: *"Sentezlenmiş sesler, anlamı doğru bir şekilde aktarıyor, ifade konusunda ise daha yumuşak ve doğal olabilir."*
- K17: *"Anlam açık ve net, ifade ise gerçek ses kadar etkili değil."*
- K18: *"Teknoloji anlamın doğru aktarılmasında etkili, fakat duygusal ifade sınırlı."*
- K19: *"Anlam aktarımı mükemmel; ifadede ise gerçek ses kadar doğallık yok."*
- K20: *"Anlam yönünden güçlü, ifade yönünden ise geliştirilebilir."*

Bu veriler, katılımcıların yapay zeka ile üretilen sentetik seslerin doğallığını ve ifade yeteneğini nasıl değerlendirdiklerini göstermektedir. Katılımcılar, sesin netliği, mekanik hissi, doğal sesin sıcaklığı, duygusal ifade eksikliği, anlamın doğru aktarımı ve ifade gücü ve otantiklik gibi faktörleri özellikle önemsemekte ve bu faktörlerin ses algısı üzerindeki doğrudan etkilerini belirtmektedirler. Bu özellikler, sentetik ve gerçek sesler arasındaki farkları belirginleştirmede kritik rol oynamakta ve katılımcıların bu sesleri nasıl algıladıklarını derinlemesine yansıtmaktadır.

7.3. ÜÇÜNCÜ ALT PROBLEME YÖNELİK BULGULAR

Bu alt başlıkta "Katılımcıların sentetik sesler ve gerçek sesler arasındaki tercihlerini ve bu seslerin kullanıcı deneyimi açısından nasıl değerlendirildiğini" incelenmektedir. Katılımcıların tercihlerini ve sentetik ile gerçek seslerin kullanıcı deneyimi üzerindeki etkilerini anlamak amacıyla yarı yapılandırılmış araştırma soruları yöneltilmiştir. Araştırma soruları:

- Sentezlenmiş ses veya gerçek ses arasında tercih ettiğiniz hangisidir?
- Gerçek ses veya sentezlenmiş ses arasında kullanıcı deneyimi açısından hangisini tercih ediyorsunuz ve sentezlenmiş sesi günlük yaşamınızda ne kadar kullanışlı buluyorsunuz?
- Dilinin anlaşılabilirliği açısından gerçek ses ile sentezlenmiş sesi nasıl karşılaştırıyorsunuz?

Bu görüşme soruları, katılımcıların sentetik ve gerçek sesler arasında nasıl bir tercih yaptıklarını ve bu tercihlerin kullanıcı deneyimine nasıl yansıdığını belirlemek ve değerlendirmek amacıyla tasarlanmıştır. Sorular, katılımcıların bu seslerin kullanışlılığını, anlaşılabilirliğini ve genel deneyimlerini nasıl algıladıklarını ve sentetik seslerin insan sesleriyle kıyaslandığında hangi özellikler açısından avantaj veya dezavantaj sunduğunu ortaya koymak için hazırlanmıştır. Katılımcıların bu sorulara yönelik görüşlerinin içerik analizi sonucu elde edilen içerik analizine ilişkin veriler Tablo 4'te gösterilmektedir.

Tablo 4. Katılımcıların "Doğallık" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	<i>f</i>
Kullanıcı Deneyimi	Ses Tercihi	Gerçek Sesin Doğallığı	K1, K3, K5, K7, K9, K11, K13, K15, K17, K19	10

Tablo 4 (Devam). Katılımcıların "Doğallık" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	f
Kullanıcı Deneyimi		Sentez Sesin Netliği	K2, K4, K6, K8, K10, K12, K14, K16, K18, K20	10
		Teknolojik Üstünlük	K6, K8, K10, K14, K16, K18, K20	7
	Kullanıcı Deneyimi Faktörleri	Duygusal Bağ	K1, K3, K5, K7, K9, K11, K13, K15, K17, K19	10
		Anlaşılabilirlik	K1, K3, K4, K5, K6, K7, K8, K9, K12, K13, K18, K19, K20	13
		Günlük Kullanışlılık	K2, K4, K6, K8, K10, K12, K14, K16, K18, K20	10
	Sesin Etkileşim Kalitesi	Duygusal İfade	K1, K3, K5, K7, K9, K11, K13, K15, K17, K19	10
		Vurgu ve Tonlama Kontrolü	K2, K4, K6, K8, K10, K12, K14, K16, K18, K20	10
		Doğal Dalgalanmalar	K1, K3, K5, K7, K9, K11, K13, K15, K17, K19	10

Tablo 4'ten elde edilen bulgular sonucunda; 1 tema, 3 kategori, 9 kod elde edilmiştir. Katılımcıların gerçek ve sentetik sesler arasındaki tercihlerine, kullanıcı deneyimi faktörlerine ve sesin etkileşim kalitesine yönelik cevaplar verdiği belirlenmiştir. Bu cevaplar, katılımcıların ses tercihlerinin nedenlerini, kullanıcı deneyimi üzerindeki etkilerini ve sesin etkileşim kalitesini nasıl algıladıklarını yansıtmaktadır. Her bir kategoride üç kod yer almış ve bu kodlara yönelik belirtilen katılımcıların detaylı cevapları incelenmiştir.

Tablo 4'e göre, katılımcıların gerçek ve sentetik sesler arasındaki tercihleri ve bu tercihlerin kullanıcı deneyimi üzerindeki etkileri üç ana kategori altında incelenmiştir: "Tercih", "Kullanıcı Deneyimi Faktörleri", ve "Sesin Etkileşim Kalitesi". Her bir kategori, belirli kodlar oluşturularak, bu kodlara katılımcılar ve frekanslar atılmıştır.

Tercih:

- Gerçek Sesin Doğallığı: Bu koddaki on katılımcı (K1, K3, K5, K7, K9, K11, K13, K15, K17, K19) gerçek sesi daha doğal ve sıcak bulduklarını belirtmiş, bu özelliklerin tercihlerinde belirleyici olduğunu vurgulamışlardır.

- Sentez Sesin Netliği: On katılımcı (K2, K4, K6, K8, K10, K12, K14, K16, K18, K20) sentetik sesin netliğini ve kusursuzluğunu tercih sebebi olarak göstermişlerdir, bu da sesin teknik üstünlüğünün önemli bir faktör olduğunu göstermektedir.

- Teknolojik Üstünlük: Yedi katılımcı (K6, K8, K10, K14, K16, K18, K20) sentetik sesin teknolojik avantajlarını, özellikle de kullanıcı deneyimi açısından sağladığı faydaları tercih nedeni olarak belirtmiştir.

Kullanıcı Deneyimi Faktörleri:

- Duygusal Bağ: On katılımcı (K1, K3, K5, K7, K9, K11, K13, K15, K17, K19) gerçek sesin duygusal bağ kurma kapasitesini kullanıcı deneyimi için önemli bulmuşlardır.

- Anlaşılabilirlik: On üç katılımcı (K1, K3, K4, K5, K6, K7, K8, K9, K12, K13, K18, K19, K20) sesin anlaşılabilirliğinin kullanıcı deneyimini nasıl etkilediğine dikkat çekmişlerdir, özellikle gerçek sesin tonal zenginliği ve sentetik sesin netliği bu kategoride öne çıkmaktadır.

- Günlük Kullanışlılık: On katılımcı (K2, K4, K6, K8, K10, K12, K14, K16, K18, K20) sentetik sesin günlük yaşamdaki kullanılabilirliğini, özellikle de bilgi erişimi ve talimatları izlemede sağladığı kolaylıkları vurgulamışlardır.

Sesin Etkileşim Kalitesi:

- Duygusal İfade: On katılımcı (K1, K3, K5, K7, K9, K11, K13, K15, K17, K19) gerçek sesin duygusal ifade yeteneğini, özellikle insanlar arası etkileşimlerde sağladığı derinlik ve sıcaklık açısından değerlendirmişlerdir.

- Vurgu ve Tonlama Kontrolü: On katılımcı (K2, K4, K6, K8, K10, K12, K14, K16, K18, K20) sentetik sesin vurgu ve tonlama kontrolündeki üstünlüğünü, kullanıcı deneyimindeki etkinliğini ve profesyonellik sunma kapasitesini öne çıkarmışlardır.

- Doğal Dalgalanmalar: On katılımcı (K1, K3, K5, K7, K9, K11, K13, K15, K17, K19) gerçek sesin doğal dalgalanmalarının, konuşmanın akıcılığını ve anlaşılabilirliğini artırdığını, bu nedenle kullanıcı deneyimini iyileştirdiğini belirtmiştir.

Üçüncü alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Sentezlenmiş ses veya gerçek ses arasında tercih ettiğiniz hangisidir?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Gerçek sesi tercih ederim çünkü daha sıcak ve doğal hissettiriyor."*

K2: *"Sentezlenmiş sesin netliği ve kusursuzluğu nedeniyle onu tercih ediyorum."*

K3: *"Duygusal derinlik için gerçek sesi tercih ediyorum."*

K4: "Sentezlenmiş sesin temizliği ve tutarlılığı benim için daha çekici."

K5: "Gerçek sesin doğallığını ve canlılığını tercih ediyorum."

K6: "Teknolojik üstünlükleri nedeniyle sentezlenmiş sesi tercih ederim."

K7: "Gerçek ses, insani duyguları daha iyi yansıttığı için tercihimdir."

K8: "Sentezlenmiş sesin düzenliliği ve belirgin vurguları nedeniyle onu tercih ediyorum."

K9: "Gerçek sesin sıcaklığı ve ifade gücü tercih sebebim."

K10: "Sentezlenmiş sesin arka plan seslerden arınmış olması onu tercih etmemde etkili."

K11: "Gerçek sesin doğal kusurları onu daha gerçekçi ve tercih edilir kılıyor."

K12: "Sentezlenmiş sesin ifade yeteneğinin genişliği onu tercih etmemi sağlıyor."

K13: "Gerçek sesin doğal tonlamaları benim için daha çekici."

K14: "Sentezlenmiş sesin teknik üstünlüğü, onu tercih etmemde belirleyici."

K15: "Gerçek sesin insani yönleri onu tercih etmemde önemli bir faktör."

K16: "Sentezlenmiş sesin istikrarlı performansı benim için daha avantajlı."

K17: "Gerçek sesin duygusal ifadesi ve doğallığı, onu tercih etmemdeki nedenler."

K18: "Sentezlenmiş sesin daha profesyonel ve kesin sunumu, tercihimde etkili."

K19: "Gerçek sesin organikliği ve canlılığı benim için daha önemli."

K20: "Sentezlenmiş sesin temizliği ve düzeni, onu tercih etme sebebim."

Üçüncü alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Gerçek ses veya sentezlenmiş ses arasında kullanıcı deneyimi açısından hangisini tercih ediyorsunuz ve sentezlenmiş sesi günlük yaşamınızda ne kadar kullanışlı buluyorsunuz?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Kullanıcı deneyimi açısından gerçek sesi tercih ediyorum; daha doğal ve samimi. Sentezlenmiş sesi günlük yaşamda kullanmadım, ancak işlevsellik açısından potansiyel taşıdığını düşünüyorum, özellikle bilgi sorgulamada."

K2: "Sentezlenmiş sesin düzgünlüğü kullanıcı deneyimini iyileştiriyor, bu yüzden tercih ederim. Eğitim materyallerinde sentezlenmiş sesin kullanılmasının zaman kazandırabileceğini görebiliyorum."

K3: "Gerçek sesin doğal olması kullanıcı deneyimini zenginleştiriyor, onu tercih ediyorum. Telefon görüşmelerinde ve sesli komutlarda gerçek sesin daha pratik olduğunu düşünüyorum."

K4: "Sentezlenmiş sesin netliği ve kusursuz sunumu, kullanıcı deneyimi için ideal. Sentezlenmiş sesin netliği ve belirginliği günlük işlerde yardımcı olabilir diye tahmin ediyorum."

K5: "Gerçek ses, kullanıcı deneyiminde doğallık ve sıcaklık sağlıyor. Gerçek sesle kurulan duygusal bağ, onu daha kullanışlı kılar."

K6: "Kullanıcı deneyiminde teknolojik açıdan üstün olan sentezlenmiş sesi tercih ederim. Sentezlenmiş sesin yol tarifleri ve bilgi sorgulamada kullanışlı olabileceğini düşünüyorum."

K7: "Gerçek sesin insani unsurları kullanıcı deneyimini daha anlamlı kılıyor. Gerçek sesin doğallığı ve samimiyeti, günlük yaşamda benim için daha değerli."

K8: "Sentezlenmiş sesin vurgu ve tonlama kontrolü, kullanıcı deneyimini optimize ediyor. Sentezlenmiş sesin, multimedya ve eğitim araçlarında işlevsel olabileceğini düşünüyorum."

K9: "Gerçek sesin sıcaklığı ve dalgalanmaları kullanıcı deneyimini iyileştiriyor. Gerçek sesin verdiği huzur ve rahatlık, günlük kullanımda onu tercih etmeme neden oluyor."

K10: "Sentezlenmiş sesin temizliği, kullanıcı deneyimi açısından tercih sebebim. Sentezlenmiş ses, akıllı ev cihazlarında kullanışlı olabilir, fakat ben gerçek sesi tercih ederim."

K11: "Gerçek sesin doğal kusurları, kullanıcı deneyimine gerçeklik katıyor. Gerçek sesle etkileşim, günlük yaşamda benim için daha önemli."

K12: "Sentezlenmiş sesin ifade yeteneği genişliği kullanıcı deneyimi için tercih nedenim. Sentezlenmiş sesin, alışveriş listesi oluşturma gibi pratik kullanımlarını görebiliyorum."

K13: "Gerçek sesin doğal tonlamaları, kullanıcı deneyimini daha zengin kılıyor. Gerçek ses, günlük etkileşimlerimde daha anlamlı ve kullanışlı."

K14: "Sentezlenmiş sesin teknik üstünlüğü kullanıcı deneyimi için daha avantajlı. Sentezlenmiş ses teknolojisinin teknik destek ve müşteri hizmetlerinde etkili olabileceğini düşünüyorum."

K15: "Gerçek sesin insani yönleri, kullanıcı deneyiminde daha tatmin edici. Gerçek sesin tonu ve duygusallığı, günlük etkileşimlerde benim için daha tatmin edici."

K16: "Sentezlenmiş sesin dengeli performansı, kullanıcı deneyimi için daha istikrarlı. Sentezlenmiş sesin hızı ve doğruluğu, bilgi erişiminde kullanışlı olabilir."

K17: "Gerçek sesin duygusal ifadesi ve doğallığı, kullanıcı deneyimini zenginleştiriyor. Gerçek insan sesini, doğallığı ve rahatlatıcı etkisi nedeniyle günlük kullanımda tercih ederim."

K18: "Sentezlenmiş sesin kesin ve profesyonel sunumu kullanıcı deneyimi için tercih ediyorum. Sentezlenmiş sesin, hızlı yanıt gerektiren durumlarda işlevsel olabileceğini düşünüyorum."

K19: "Gerçek sesin doğallığı ve canlılığı kullanıcı deneyimini doğrudan etkiliyor. Gerçek sesle sohbet etmenin günümü daha keyifli kıldığını düşünüyorum, bu yüzden onu tercih ediyorum."

K20: "Sentezlenmiş sesin temizliği ve düzeni, kullanıcı deneyimini iyileştiriyor. Sentezlenmiş sesin, günlük yaşamda hızlı bilgi akışı sağlamada ve anında yanıt gerektiren uygulamalarda kullanışlı olabileceğini görüyorum."

Üçüncü alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Dilinin anlaşılabilirliği açısından gerçek ses ile sentezlenmiş sesi nasıl karşılaştırıyorsunuz?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Sentezlenmiş ses oldukça net ve anlaşılır, fakat gerçek sesin tonal zenginliği bambaşka bir boyut katıyor."

K2: "Gerçek ses bana daha canlı ve anlaşılır geliyor, sentetik sesler bazen mekanik kalabiliyor."

K3: "Sentezlenmiş sesin dili oldukça düzgün, ama gerçek sesin doğal vurguları bana daha anlaşılır geliyor."

K4: "Gerçek sesin anlaşılabilirliği doğal intonasyon sayesinde bence daha üstün."

K5: "Teknik terimlerde ve net ifadelerde sentezlenmiş ses gerçekten etkileyici, ama doğal sesin sıcaklığını arıyorum."

K6: "Gerçek sesin insani unsurları onu daha samimi kılıyor ve bu yüzden daha anlaşılır buluyorum."

K7: "Her kelimeyi açık ve net duyduğum için sentezlenmiş sesi tercih ediyorum."

K8: "Gerçek ses daha duygusal ve bu da bazen anlamı daha derin hale getiriyor."

K9: "Sentezlenmiş ses bilgi aktarımı için her zaman net ve tutarlı, bu yüzden onu kullanmayı seviyorum."

K10: "Gerçek sesin doğal dalgalanmaları onu daha rahat anlaşılır kılıyor, sentetik sesler ise biraz monoton olabiliyor."

K11: "Sentezlenmiş sesin dili belirgin ve net, ama gerçek sesin duygusal ifadesi gerçekten etkileyici."

K12: "Konuşma dilinin doğal akışını yansıtan gerçek ses, bana daha anlaşılır geliyor."

K13: "Sentezlenmiş sesin mükemmel diksiyonu var ve her şeyi net bir şekilde ifade ediyor."

K14: "Gerçek sesin duygusal tonu, mesajın daha kolay anlaşılmasını sağlıyor."

K15: "Net talimatlar ve bilgi verme konusunda sentezlenmiş ses gerçekten üstün."

K16: "Gerçek ses, sözcükler arasındaki bağlantıyı doğal bir şekilde kuruyor ve bu da anlaşılabilirliği artırıyor."

K17: "Sentezlenmiş sesin temiz ve düz dili, teknik açıklamalar için ideal."

K18: "Gerçek ses, anadil gibi akıyor ve daha otantik bir anlaşılabilirlik sunuyor."

K19: "Sentezlenmiş sesin kesinliği ve netliği, özellikle resmi ve bilimsel metinlerde çok işe yarıyor."

K20: "Gerçek ses, konuşma dilindeki ince nüansları yakalayarak daha anlaşılır hale getiriyor."

Bu bulgular, katılımcıların sentetik ve gerçek sesler arasında nasıl bir tercih yaptıklarını ve bu tercihlerin kullanıcı deneyimi üzerinde nasıl etkiler bıraktığını göstermektedir. Katılımcılar, gerçek sesin doğallığı ve duygusal ifadesini önemserken, sentetik sesin netliği ve teknolojik üstünlüklerini de değerlendirmişlerdir. Kullanıcı deneyimi faktörleri olarak; duygusal bağ, anlaşılabilirlik ve günlük kullanılabilirlik ön plana çıkmaktadır. Sesin etkileşim kalitesi kategorisinde ise, duygusal ifade, vurgu ve tonlama kontrolü ile doğal dalgalanmalar katılımcıların ses tercihlerini nasıl şekillendirdiğini ve kullanıcı deneyimine nasıl katkı sağladığını belirlemiştir. Bu çeşitli faktörler, gerçek ve sentetik sesler arasındaki farkları belirginleştirmede kritik rol oynamakta ve katılımcıların bu sesleri nasıl algıladıklarını derinlemesine yansıtmaktadır.

7.4. DÖRDÜNCÜ ALT PROBLEME YÖNELİK BULGULAR

Bu alt başlıkta "Sentezlenmiş seslerin ve gerçek seslerin güvenilirlik algısını ve bilgi aktarımındaki etkisini" alt problemine ilişkin bulgular ve yorumlar yer almaktadır. Katılımcıların sentetik ve gerçek seslerin güvenilirliğini değerlendirme sürecini anlamak amacıyla yarı yapılandırılmış araştırma soruları yöneltilmiştir. Araştırma soruları:

- Gerçek ses ve sentezlenmiş ses arasında hangisi daha güvenilir bir izlenim bıraktı?
- Sentezlenmiş sesin, size bilgi aktarırken güvenilirlik hissiyatınız nasıl etkilendi ve bu sesleri güvenilir olarak nasıl görüyorsunuz?

Bu görüşme soruları, katılımcıların sentetik ve gerçek seslerin güvenilirlik algılarını belirlemek ve değerlendirmek amacıyla tasarlanmıştır. Sorular, katılımcıların bu seslerin bilgi aktarımındaki etkinliklerini nasıl algıladıklarını ve sentetik seslerin insan sesleriyle kıyaslandığında hangi özellikler açısından daha güvenilir veya güvensiz olduklarını ortaya koymak için hazırlanmıştır. Katılımcıların bu sorulara yönelik görüşlerinin içerik analizi sonucu elde edilen içerik analizine ilişkin veriler Tablo 5'te gösterilmektedir.

Tablo 5. Katılımcıların "Güvenilirlik" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	f
Güvenilirlik	İlk İzlenim	Gerçek Sesin Tanıdıklığı	K1, K3, K5, K7, K9, K11, K13, K17, K19	9
		Sentez Sesin Teknik Üstünlüğü	K2, K4, K6, K8, K10, K12, K14, K16, K18, K20	10

Tablo 5 (Devam). Katılımcıların "Güvenilirlik" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	f
		Duygusal İfade	K3, K5, K7, K9, K11, K13, K15, K17, K19	9
	Bilgi Aktarımı Etkisi	Başlangıçta Şüphe	K1, K2, K3, K4, K8, K14, K17, K18	8
		Zamanla Artan Güven	K1, K2, K3, K4, K9, K11, K12, K13, K14, K19, K20	11
		Teknik Detaylar için Güvenilirlik	K10, K12, K14, K16, K18, K20	6

Tablo 5'ten elde edilen bulgular sonucunda; 1 tema, 2 kategori, 6 kod elde edilmiştir. Katılımcıların gerçek ve sentetik seslerin güvenilirlik algısına, bilgi aktarımındaki etkisine ve duygusal ifadelerine verdiği cevaplar belirlenmiştir. Bu cevaplar, katılımcıların gerçek ve sentetik seslerin ilk izlenimlerini, zamanla nasıl bir güven değişimi yaşadıklarını ve teknik detaylar açısından hangi sesin daha güvenilir olduğunu yansıtmaktadır. Her bir kategoride üçer kod yer almış ve bu kodlara yönelik belirtilen katılımcıların detaylı cevapları incelenmiştir. Tablo 5'e göre, gerçek ve sentetik sesler arasındaki güvenilirlik algıları iki ana kategori altında incelenmiştir: "İlk İzlenim" ve "Bilgi Aktarımı Etkisi". Her bir kategori, belirli kodlar oluşturularak, bu kodlara katılımcılar ve frekanslar atılmıştır.

İlk İzlenim:

- Gerçek Sesin Tanıdıklığı: Bu koddaki dokuz katılımcı (K1, K3, K5, K7, K9, K11, K13, K17, K19), gerçek sesin doğallığı ve tanıdıklığından dolayı daha güvenilir bir izlenim bıraktığını belirtmişlerdir.

- Sentez Sesin Teknik Üstünlüğü: On katılımcı (K2, K4, K6, K8, K10, K12, K14, K16, K18, K20), sentetik sesin teknik açıdan üstünlüğünün güvenilir bir izlenim bıraktığını ifade etmişlerdir.

- Duygusal İfade: Bu koddaki dokuz katılımcı (K3, K5, K7, K9, K11, K13, K15, K17, K19), gerçek sesin duygusal ifadesinin onları daha fazla güven duymalarını sağladığını belirtmişlerdir.

Bilgi Aktarımı Etkisi:

- Başlangıçta Şüphe: Sekiz katılımcı (K1, K2, K3, K4, K8, K14, K17, K18), başlangıçta sentetik sesin güvenilirliği konusunda şüphe duyduklarını, ancak zamanla bu durumun değiştiğini açıklamışlardır.

• Zamanla Artan Güven: On bir katılımcı (K1, K2, K3, K4, K9, K11, K12, K13, K14, K19, K20), sentetik sesle ilgili başlangıçtaki şüphelerinin, verilen doğru bilgilerle zamanla azaldığını ve güvenlerinin arttığını ifade etmişlerdir.

• Teknik Detaylar için Güvenilirlik: Altı katılımcı (K10, K12, K14, K16, K18, K20), özellikle teknik detaylar için sentetik sesin sunduğu bilgilerin güvenilir olduğunu belirtmişlerdir.

Dördüncü alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Gerçek ses ve sentezlenmiş ses arasında hangisi daha güvenilir bir izlenim bıraktı?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Gerçek ses daha güvenilir bir izlenim bıraktı çünkü doğallığı ve insanca unsurları daha tanıdık hissettirdi."*

K2: *"Sentezlenmiş ses teknik açıdan kusursuz olduğundan daha güvenilir bir izlenim bıraktı."*

K3: *"Gerçek sesin duygusal ifadesi ve tonal dalgalanmaları daha güvenilir geldi."*

K4: *"Sentezlenmiş sesin netliği ve tutarlılığı onu daha güvenilir kıldı."*

K5: *"Gerçek sesin doğal kusurları onu güvenilir yaparken, sentetik ses çok yapay geldi."*

K6: *"Sentezlenmiş sesin dengeli tonu ve düzgün sunumu daha güvenilir bir his uyandırdı."*

K7: *"Gerçek ses, içerdiği insani unsurlarla daha güvenilir bir izlenim bıraktı."*

K8: *"Sentezlenmiş sesin temizliği ve kusursuzluğu onu daha güvenilir kıldı."*

K9: *"Gerçek sesin sıcaklığı ve duygusallığı onu daha güvenilir yaptı."*

K10: *"Sentezlenmiş ses, teknolojik açıdan mükemmel olduğu için daha güvenilir bir izlenim bıraktı."*

K11: *"Gerçek sesin doğallığı ve akıcılığı, güvenilirliği artırıyor."*

K12: *"Sentezlenmiş sesin belirgin ve net konuşması onu güvenilir kıldı."*

K13: *"Gerçek sesin canlılığı ve doğal ton değişiklikleri daha güven verdi."*

K14: "Sentezlenmiş sesin keskinliği ve netliği, onu gerçek sestten daha güvenilir yaptı."

K15: "Gerçek ses daha samimi ve güvenilir geldi, sentetik ses ise bazen çok mekanik."

K16: "Sentezlenmiş sesin hızı ve doğruluğu onu özellikle bilgi verme konusunda daha güvenilir kıldı."

K17: "Gerçek sesin organik yapısı onu daha güvenilir yaptı."

K18: "Sentezlenmiş sesin düzenli tonlamaları, onu bilgi aktarımında daha güvenilir kıldı."

K19: "Gerçek sesin doğal vibrasyonu ve ton değişiklikleri onu daha güvenilir kıldı."

K20: "Sentezlenmiş ses, özellikle teknik detaylar için daha güvenilir bir seçenek sundu."

Dördüncü alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Sentezlenmiş sesin, size bilgi aktarırken güvenilirlik hissiyatınız nasıl etkilendi ve bu sesleri güvenilir olarak nasıl görüyorsunuz?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: "Başta, sentezlenmiş sesin bilgi aktarması biraz korkutucu geldi, ama zamanla alıştım ve güvenilir buldum."

K2: "Sentezlenmiş ses başlangıçta insan gibi gelmediği için güven vermedi, fakat bilgilerin doğruluğu güvenimi artırdı."

K3: "Sentezlenmiş sesi duymak başta garip olsa da, verdiği bilgiler doğru çıktıkça güven duymaya başladım."

K4: "Sentezlenmiş sesin doğru ve net bilgi vermesi, başta şüpheci olsam da güvenimi kazandı."

K5: "Sentezlenmiş ses, insan sesine çok benzediği için ilk başta fark edemedim ve bu durum güvenimi sarsmadı."

K6: "Sentezlenmiş ses, bilgi aktarımında başarılı olduğundan, güvenilir olduğunu düşünüyorum."

K7: *"Başta sentezlenmiş sesin bilgi vermesi kafa karıştırıcıydı, güvenmek zorunda kaldım çünkü doğru bilgiydi."*

K8: *"Sentezlenmiş sesi duyunca, bir robotun bu kadar doğru bilgi verebilmesi şaşırtıcı ve biraz ürkütücüydü."*

K9: *"İlk başta sentezlenmiş sesin güvenilir olup olmadığını sorguladım, ancak doğru bilgiler verdiğini gördükçe rahatlardım."*

K10: *"Sentezlenmiş ses, insan sesine o kadar benziyordu ki, güvenilir olduğunu hissetmem zaman aldı."*

K11: *"Sentezlenmiş sesin verdiği bilgiler başlangıçta güvenilir gelmedi, ama zamanla doğruluğunu kanıtladı."*

K12: *"Sentezlenmiş sesin bilgileri başta kuşkuyla karşıladım, ama doğruluk oranı yüksek çıktıkça güvendim."*

K13: *"Sentezlenmiş sesin bilgileri doğru çıktıkça, başta hissettiğim güvensizlik azaldı."*

K14: *"Sentezlenmiş sesin bilgi verme şekli ilk başta rahatsız ediciydi, ama doğru bilgiler verince güven duymaya başladım."*

K15: *"Sentezlenmiş sesin doğru bilgi vermesi güven vericiydi, fakat sesin yapaylığı başta ürkütücüydü."*

K16: *"Sentezlenmiş sesi duymak alışılmadık bir deneyimdi, doğru bilgi vermesi güvenimi artırdı."*

K17: *"Sentezlenmiş ses, başlangıçta güven verici gelmedi çünkü çok yapaydı, fakat bilgileri doğru çıktı."*

K18: *"Başta sentezlenmiş sesin bilgi aktarması rahatsız ediciydi, fakat zamanla doğru ve güvenilir buldum."*

K19: *"Sentezlenmiş sesin bilgileri doğrudu, bu da zamanla güven duymamı sağladı."*

K20: *"Sentezlenmiş sesin verdiği bilgiler, başta şüpheyle yaklaşmama rağmen, güven verici çıktı."*

Bu veriler, katılımcıların gerçek ve sentetik seslerin güvenilirliğini değerlendirirken dikkate aldıkları çeşitli özelliklerin geniş bir yelpazeye yayıldığını

göstermektedir. Katılımcılar, gerçek sesin tanındığı, sentetik sesin teknik üstünlüğü ve duygusal ifadeler gibi faktörleri özellikle önemsemekte ve bu faktörlerin güvenilirlik algısı üzerindeki doğrudan etkilerini belirtmektedirler. Bu özellikler, başlangıçtaki şüphelerden, zamanla artan güvene ve teknik detaylarda gözlemlenen güvenilirliğe kadar güvenilirlik algılarında kritik rol oynamakta ve katılımcıların bu sesleri nasıl algıladıklarını derinlemesine yansıtmaktadır.

7.5. BEŞİNCİ ALT PROBLEME YÖNELİK BULGULAR

Bu alt başlıkta "Sentezlenmiş seslerin sosyal etkileşime etkisi ve toplumdaki kabul düzeyi nasıldır?" alt problemine ilişkin bulgular ve yorumlar yer almaktadır. Katılımcıların sentetik seslerin toplumdaki kabul düzeyini ve sosyal etkileşimlerdeki etkinliklerini değerlendirme sürecini anlamak amacıyla yarı yapılandırılmış araştırma soruları yöneltilmiştir. Araştırma soruları:

- Endüstriyel uygulamalarda gerçek ses veya sentezlenmiş ses hangisi daha etkiliydi?

- Sentezlenmiş ses toplumda ne kadar yaygın bir şekilde kabul görüyor?

Bu görüşme soruları, katılımcıların sentetik seslerin sosyal kabulünü ve etkileşimdeki etkinliklerini belirlemek ve değerlendirmek amacıyla tasarlanmıştır. Sorular, katılımcıların bu seslerin toplum içindeki yaygınlığını ve endüstriyel uygulamalardaki etkililiklerini nasıl algıladıklarını ve sentetik seslerin insan sesleriyle kıyaslandığında sosyal kabul seviyelerinin ne olduğunu ortaya koymak için hazırlanmıştır. Katılımcıların bu sorulara yönelik görüşlerinin içerik analizi sonucu elde edilen içerik analizine ilişkin veriler Tablo 6'da gösterilmektedir.

Tablo 6. Katılımcıların "Sosyal Kabul ve Etkileşim" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	<i>f</i>
Sosyal Kabul ve Etkileşim	Endüstriyel Etkililik	Teknik Üstünlük ve Güncellenebilirlik	K1, K4, K6, K9, K18	5
		Doğrudan İnsan Etkileşimi	K3, K7, K10, K13, K15, K19	6
		Bilgi Tabanlı Uygulamalar	K2, K8, K11, K12, K14, K16, K20	7
	Toplumsal Kabul	Artan Kabul ve Yaygınlık	K1, K4, K6, K11, K12, K18	6

Tablo 6 (Devam). Katılımcıların "Sosyal Kabul ve Etkileşim" Temasına İlişkin Görüşleri

Tema	Kategori	Kod	Katılımcılar	f
		Direnç ve Çekinceler	K2, K3, K5, K7, K13, K15, K17, K19	8
		Genç Nesil ve Teknolojiye Açıklık	K4, K8, K14, K16, K20	5

Tablo 6'dan elde edilen bulgular sonucunda; 1 tema, 2 kategori, 6 kod elde edilmiştir. Katılımcıların sentetik ve gerçek seslerin endüstriyel etkililiği ve toplumsal kabul düzeyine yönelik cevapları belirlenmiştir. Bu cevaplar, katılımcıların sentetik ve gerçek seslerin endüstriyel uygulamalardaki etkililiğini, toplumdaki kabul düzeyini ve teknolojiye olan açıklıklarını yansıtmaktadır. Her bir kategoride üçer kod yer almış ve bu kodlara yönelik belirtilen katılımcıların detaylı cevapları incelenmiştir.

Tablo 6'ya göre, sentetik seslerin sosyal kabulü ve etkileşimi iki ana kategori altında incelenmiştir: "Endüstriyel Etkililik" ve "Toplumsal Kabul". Her bir kategori, belirli kodlar oluşturularak, bu kodlara katılımcılar ve frekanslar atılmıştır.

Endüstriyel Etkililik:

- Teknik Üstünlük ve Güncellenebilirlik: Bu koddaki beş katılımcı (K1, K4, K6, K9, K18), sentetik seslerin endüstriyel uygulamalarda teknik üstünlük ve güncellenebilirlik açısından daha etkili olduğunu belirtmişlerdir.

- Doğrudan İnsan Etkileşimi: Altı katılımcı (K3, K7, K10, K13, K15, K19), gerçek sesin doğrudan insan etkileşimi gerektiren durumlarda daha etkili olduğunu ifade etmişlerdir.

- Bilgi Tabanlı Uygulamalar: Yedi katılımcı (K2, K8, K11, K12, K14, K16, K20), sentetik seslerin bilgi tabanlı uygulamalarda daha etkili olduğunu belirtmişlerdir.

Toplumsal Kabul:

- Artan Kabul ve Yaygınlık: Altı katılımcı (K1, K4, K6, K11, K12, K18), sentetik seslerin toplumda giderek daha fazla kabul gördüğünü ifade etmiştir.

- Direnç ve Çekinceler: Sekiz katılımcı (K2, K3, K5, K7, K13, K15, K17, K19), toplumda hala bir direnç ve çekince olduğunu belirtmişlerdir.

- Genç Nesil ve Teknolojiye Açıklık: Beş katılımcı (K4, K8, K14, K16, K20), genç nesillerin sentetik seslere daha açık olduğunu ve teknolojiyi benimsediklerini ifade etmiştir.

Beşinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen “Endüstriyel uygulamalarda gerçek ses veya sentezlenmiş ses hangisi daha etkiliydi?” sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Endüstriyel uygulamalarda, özellikle teknik destek ve bilgi sağlamada sentezlenmiş ses daha etkili olabilir çünkü sürekli güncellenip düzeltilebilir."*

K2: *"Sentezlenmiş ses, bilgi tabanlı uygulamalarda gerçek sesi geçebilir; daha net, anlaşılır ve tutarlı bilgi sağlar."*

K3: *"Gerçek ses, müşteri hizmetleri gibi doğrudan insan etkileşimi gerektiren durumlarda daha etkili olabilir."*

K4: *"Endüstriyel uygulamalarda, sentezlenmiş sesin hızı ve verimliliği, özellikle tekrar eden bilgiler için idealdir."*

K6: *"Sentezlenmiş ses, endüstriyel otomasyon ve uyarı sistemlerinde oldukça etkili çünkü daha dikkat çekici ve sürekli güncellenebilir."*

K7: *"Gerçek sesin duygusal ve kişisel bağ kurma yeteneği, sosyal kabul ve etkileşimde önemli bir avantaj sağlar."*

K8: *"Sentezlenmiş ses, anlam ve tonlama konusunda giderek gelişiyor ve belli uygulamalarda gerçek ses kadar etkili olabilir."*

K9: *"Endüstriyel uygulamalarda, sentezlenmiş sesin kusursuz doğruluğu ve tekrarlanabilirliği, büyük ölçekli operasyonlarda avantaj sağlar."*

K10: *"Gerçek ses, özellikle reklam ve pazarlama gibi alanlarda daha etkili olabilir, insanları daha çok etkileyebilir."*

K11: *"Sentezlenmiş seslerin dil anlaşılabilirliği ve teknik kapasitesi, bazı endüstriyel uygulamalarda gerçek seslerden daha üstün olabilir."*

K12: *"Sentezlenmiş sesin bilgileri başta kuşkuyla karşıladım, ama doğruluk oranı yüksek çıktıkça güvendim."*

K13: *"Gerçek sesin doğallığı ve anlaşılabilirliği, insan etkileşimi gerektiren her alanda daha etkili."*

K14: *"Sentezlenmiş ses, toplumdaki kabul düzeyi arttıkça, özellikle standart bilgilerin verilmesi gereken yerlerde etkili olabilir."*

K15: *"Gerçek ses, hissiyatı ve duygusal bağları güçlendirir, bu da sosyal etkileşimde önemlidir."*

K16: *"Sentezlenmiş ses, sürekli gelişen teknoloji sayesinde, belirli uygulamalarda gerçek ses kadar etkili hale gelebilir."*

K18: *"Sentezlenmiş sesin anlaşılabilirliği ve doğruluğu, özellikle bilimsel ve teknik alanlarda, gerçek sesin sunduğundan daha etkili olabilir."*

K19: *"Gerçek sesin doğrallığı, müşteri hizmetleri ve insanlarla doğrudan etkileşim gerektiren her alanda tercih edilir."*

K20: *"Sentezlenmiş ses, dil anlaşılabilirliğini ve veri sunumunu mükemmelleştirdiği için, bazı endüstriyel uygulamalarda daha etkili olabilir."*

Beşinci alt problemin çözümlenmesine yönelik yarı yapılandırılmış görüşmelerde yöneltilen "Sentezlenmiş ses toplumda ne kadar yaygın bir şekilde kabul görüyor?" sorusuna ilişkin katılımcıların verdikleri cevaplar aşağıda belirtilmiştir.

K1: *"Sentezlenmiş sesler giderek daha fazla kabul görüyor, özellikle müşteri hizmetleri ve asistan uygulamalarında."*

K2: *"Toplumda hâlâ bir çekince var, özellikle duygusal ifade gerektiren durumlarda insan sesi tercih ediliyor."*

K3: *"Sentezlenmiş ses, teknoloji sektöründe oldukça yaygın ancak gündelik kullanımda insan sesi kadar değil."*

K4: *"Yavaş yavaş kabul görüyor, özellikle genç nesiller arasında sentetik seslere büyük bir açıklık var."*

K5: *"Sentezlenmiş sesler bazı alanlarda kabul görse de, toplumun büyük bir kısmı hâlâ gerçek sesi tercih ediyor."*

K6: *"Toplumda kabul düzeyi artıyor, özellikle otomatik sistemlerde ve bilgi sağlama hizmetlerinde kullanımı genişliyor."*

K7: *"Çoğu insan için sentezlenmiş sesler yeterince doğal gelmiyor, bu yüzden kabul görme oranı sınırlı."*

K8: *"Belirli uygulamalar için kabul ediliyor, ancak genel olarak insan sesinin yerini alması zor görünüyor."*

K9: *"Akıllı ev teknolojileri ve asistanlar aracılığıyla kabul görüyor ama duygusal bağ kurulmasını gerektiren durumlarda tercih edilmiyor."*

K10: *"Medya ve eğlence sektörlerinde oldukça kabul görmekte, ancak günlük kişisel kullanımda insan sesi kadar popüler değil."*

K11: *"Sentezlenmiş sesler giderek daha fazla teknolojik cihazda yer alıyor ve bu da toplumsal kabulünü artırıyor."*

K12: *"Özellikle çağrı merkezleri gibi bazı profesyonel alanlarda oldukça yaygın."*

K13: *"Toplumda hala bir direnç var, insanlar genellikle gerçek insan etkileşimini tercih ediyor."*

K14: *"Eğitim araçları ve belirli uygulamalarda kabul görüyor, ancak genel olarak insan sesine karşı bir tercih var."*

K15: *"Kullanımı arttıkça kabul de artıyor, ancak hala insan sesine olan doğal eğilim baskın."*

K16: *"Belli başlı alanlarda kabul görse de, toplumun büyük bir kesimi tarafından hâlâ şüpheyle karşılanıyor."*

K17: *"Sentezlenmiş seslerin kullanımı artıyor, ancak toplumda tam anlamıyla kabul edilmiş değil."*

K18: *"Bazı endüstriyel ve ticari uygulamalarda çok etkili ve kabul görüyor."*

K19: *"Toplumun daha teknolojiye açık kesimlerinde kabul görüyor, ancak daha geniş bir kabul için zaman gerekiyor."*

K20: *"Sentezlenmiş sesler, genellikle bilgilendirme ve eğitim gibi belirli kontekstlerde kabul görüyor ama genel bir kabul için hâlâ yol var."*

Bu veriler, katılımcıların sentetik ve gerçek seslerin sosyal kabulünü değerlendirirken dikkate aldıkları çeşitli özelliklerin geniş bir yelpazeye yayıldığını göstermektedir. Katılımcılar, sentetik seslerin endüstriyel etkililiği, teknik üstünlüğü ve bilgi tabanlı uygulamalardaki başarısı gibi faktörleri özellikle önemsemektedirler. Aynı zamanda, gerçek sesin doğrudan insan etkileşimindeki avantajları ve duygusal bağ kurma kapasitesi de vurgulanmaktadır. Bu faktörlerin sosyal kabul ve etkileşim üzerindeki doğrudan etkilerini belirtmektedirler. Bu özellikler, toplumsal kabul düzeyinde artış,

başlangıçtaki direnç ve çekinceler ve genç nesillerin teknolojiye açıklığı gibi unsurlarla birlikte, katılımcıların bu sesleri nasıl algıladıklarını derinlemesine yansıtmaktadır.



TARTIŞMA, SONUÇ VE ÖNERİLER

Birinci Alt Probleme İlişkin Tartışma Sonuçlar

Bu alt başlıkta, katılımcıların sentetik ve gerçek sesleri ayırt ederken temel aldıkları özellikleri ve faktörleri ele alınmaktadır. Araştırma, katılımcıların sentetik ve gerçek sesleri nasıl algıladıkları ve ayırt ettikleri üzerine odaklanmıştır.

Katılımcılar, sentetik ve gerçek sesler arasındaki farkları belirlerken, canlılık, arka plan gürültüsü, ses tonunun yapaylığı gibi etmenleri ön plana çıkarmışlardır. Örneğin, sentetik seslerdeki canlılık eksikliği ve arka plan seslerinin farklılıkları, gerçek sesleri ayırt etmede önemli rol oynamıştır. Ayrıca, sentetik seslerin yapay tonu ve doğallık eksikliği, dinleyiciler tarafından kolayca fark edilmiştir.

Katılımcılar, gerçek seslerin doğal titreşimlerine ve monoton olmayan yapısına dikkat çekmişlerdir. Gerçek seslerin duygusal ifade kapasitesi ve ses renginin zenginliği, sentetik seslerle karşılaştırıldığında daha belirgin olmuş ve gerçeklik hissini pekiştirmiştir.

Araştırma ayrıca, katılımcıların ses kalitesi ve netliği konusundaki farkındalığını artırmıştır. Medya ve teknoloji kullanımı sırasında sentetik seslerin varlığını daha bilinçli bir şekilde fark etme eğilimi göstermişlerdir. Bu, teknolojik seslerin kullanımının kullanıcıların ses algısı üzerinde etkili olduğunu göstermektedir.

Bu araştırmada elde edilen bulgularla, sentetik ve gerçek sesler arasındaki farkları ve kullanıcıların bu sesleri nasıl algıladıklarını derinlemesine incelenmiştir. Literatürde yer alan çalışmalar, özellikle yapay zekanın ses üretimi kapasitesi üzerine odaklanmış ve ses algısıyla ilgili temel teorik ve pratik yönleri tartışmıştır (Huang, 2009; Ning vd., 2019).

Katılımcı yorumlarına göre, sentetik seslerin genellikle daha temiz ve monoton olduğu, gerçek seslerin ise daha doğal titreşimlere ve arka plan gürültüsüne sahip olduğu görülmüştür. Zheng ve Zhang (2018) tarafından yapılan "Phase-aware speech enhancement based on deep neural networks" çalışmasında ele alınan, sentetik seslerde faz bilgisinin yönetimi ve bu yönetimin sentetik seslerin gerçek sesleri taklit etme kapasitesini nasıl sınırladığına dair bulgular, bu çalışmanın sonuçlarıyla paralellik göstermektedir.

Katılımcılar, sentetik seslerin yapay tonunu ve canlılık eksikliğini önemli ayırt edici özellikler olarak belirtmişlerdir. Crandall (2019) tarafından gerçekleştirilen "Artificial Intelligence And Manufacturing" adlı çalışmada ses klonlama teknolojilerinin duygusal tonlamaları ve ses renklerini tam olarak kopyalayamama sorunlarına dikkat çekilmiştir. Bu, sentetik seslerin gerçek seslerin sağladığı duygusal derinlik ve tonal zenginlikleri henüz tam anlamıyla yerine getiremediğini ortaya koymaktadır.

Katılımcıların artan medya ve teknoloji farkındalığı, teknolojinin ses algısı üzerindeki etkilerini de gözler önüne sermektedir. Weijters ve Thole (1993) tarafından yapılan "Speech synthesis with artificial neural networks" bildirisinde vurgulanan, yapay sinir ağlarının konuşma sentezi uygulamalarında nasıl kullanıldığına dair analizler, kullanıcıların sentetik seslere olan aşinalıklarını artıran teknolojik ilerlemelerle uyumlu bir gelişme olarak değerlendirilebilir.

Bu çalışmada elde edilen sonuçlar, gerçek ve sentetik sesler arasındaki farklılıkların tanımlanmasında, yapay zeka teknolojilerinin ses üretimindeki sınırlılıkları ve potansiyelleri hakkında değerli bilgiler sunmaktadır. Özellikle ses teknolojilerindeki ilerlemeler, bu seslerin daha doğal ve duygusal ifade yeteneklerinin geliştirilmesine odaklanmalıdır.

Bu çalışmanın sonuçları, sentetik seslerin gerçek seslerle karşılaştırılmasında önemli teorik ve pratik katkılar sağlamakta olup, yapay zeka uygulamalarının insan hayatındaki rolünü daha iyi anlamamıza yardımcı olmaktadır. Elde edilen bulgular, yapay zekanın ses üretimindeki geleceği hakkında daha bilinçli kararlar alınmasına olanak tanımaktadır.

İkinci Alt Probleme İlişkin Tartışma Sonuçlar

Bu bölüm, "Yapay zeka teknolojisinin ses üretimindeki doğallığına ve sentetik seslerin ifade yeteneğine katkıda bulunup bulunmadığı" alt problemine ilişkin bulguları ve yorumları içermektedir. Araştırma, yapay zekanın ses üretimi ve işleme kapasitesini değerlendirirken, katılımcıların görüşlerini yarı yapılandırılmış sorular aracılığıyla toplamıştır.

Araştırmada incelenen ikinci alt problem, yapay zeka teknolojisinin ses üretimindeki doğallığına ve sentetik seslerin ifade yeteneğine katkıda bulunup bulunmadığına odaklanmaktadır. Katılımcılar, yapay zekanın ses üretimindeki rolünü ve sentetik seslerin doğallık ile ifade yeteneğini çeşitli açılardan değerlendirmişlerdir. Genel

olarak, yapay zeka teknolojisinin sesin netliğini artırdığı ve ifade yeteneğini geliştirdiği belirtilmiştir; ancak, bu seslerin tamamen doğal hissetmediği ve bazı duygusal ifadeleri yansıtamadığı vurgulanmıştır.

Yapay zeka, seslerin netliğini ve düzenini önemli ölçüde artırmış, ancak bu sesler bazen çok mekanik ve yapay hissedilmiştir. Katılımcılar, teknolojinin sesi temiz ve net yapmasına karşın, bu netliğin bazen sesleri yapay ve mekanik yapabildiğini ifade etmişlerdir. Yapay zekanın sesin sıcaklığını ve doğallığını tam olarak yansıtamadığı belirtilmiştir.

Sentetik seslerin duygusal ifadeyi yansıtmada yetersiz kaldığı, bu seslerin anlamı iyi aktarsa da duygusal derinliği yeterince sunamadığı belirtilmiştir. Sentetik seslerin anlamı doğru ve net bir şekilde aktardığı, ancak ifadedeki bazı eksikliklere dikkat çekilmiştir. Yapay zeka ile üretilen seslerin ifade gücü artmış, ancak bu seslerin otantiklik açısından zaman zaman eksik kaldığı ifade edilmiştir.

Katılımcıların görüşleri, yapay zeka ile üretilen seslerin teknik açıdan ileri düzeyde olduğunu göstermekle birlikte, bu seslerin hala insan seslerinin sahip olduğu duygusal zenginlik ve sıcaklığı tam olarak yansıtamadığını ortaya koymaktadır. Sentetik sesler, anlamı etkili bir şekilde iletebilmekte, ancak duygusal ifadeleri ve insani dokunuşları yeterince sunamamaktadır. Bu, yapay zeka ve ses sentezi teknolojilerinin gelişiminde duygusal ifade ve doğallığın daha fazla vurgulanması gerektiğini göstermektedir.

Bu bağlamda, yapay zeka destekli ses sentezleme teknolojilerinin, insan sesine benzer duygusal ifadeler sunma ve doğallık sağlama kapasitesi değerlendirilmiştir. Özellikle, Docal ve Alvarez (2023) tarafından yapılan "Enhancing Voice Cloning Quality through Data Selection and Alignment-Based Metrics" çalışması, ses klonlama tekniklerinin kalitesini artırmak amacıyla gerçekleştirilen kapsamlı analizlerle bu alanın anlaşılmasına önemli katkılar sunmaktadır. Araştırmacılar, ses klonlama sistemlerinin eğitimi için optimal verilerin seçilmesine yönelik stratejiler geliştirmişlerdir.

Katılımcılar tarafından ses sentezleme sürecinde yapay zeka teknolojisinin sesin doğallığına sağladığı katkılar ve sentetik seslerin anlam ile ifadeyi aktarma yeteneği hakkında çeşitli görüşler sunulmuştur. Görüşler, Weijters ve Thole (1993) tarafından ele alınan "Speech synthesis with artificial neural networks" çalışmasında vurgulanan yapay sinir ağlarının ses sentezleme uygulamalarında nasıl kullanıldığıyla uyumlu bir şekilde,

yapay zeka teknolojilerinin ses sentezi sürecine sağladığı teknik iyileştirmeleri yansıtmaktadır.

Katılımcıların çoğu, yapay zeka teknolojilerinin seslerin netliğini artırdığını ve anlam aktarımını iyileştirdiğini belirtmiş, ancak duygusal ifade ve sesin sıcaklığı konusunda gerçek sesler kadar etkili olamadığı konusunda hemfikir olmuşlardır. Bu durum, Reynoso (2019) tarafından yapılan "Main Types Of Artificial Intelligence" çalışmasında belirtilen yapay zeka teknolojilerinin sınırlılıklarıyla paralellik göstermektedir. Reynoso, yapay zekanın çeşitli uygulamalarının güçlü yönlerini ve sınırlılıklarını ele almıştır.

Özellikle, seslerin duygusal ifade yeteneği ve doğallığını artırma çabaları, Crandall (2019) tarafından gerçekleştirilen "Artificial Intelligence And Manufacturing" makalesinde tartışılan, yapay zekanın çeşitli alanlardaki potansiyelini ve bu teknolojilerin pratik uygulamalarını geliştirme yöntemleriyle uyumludur. Crandall, yapay zeka teknolojilerinin işlevselliğini ve iş dünyasındaki uygulamalarını geniş bir perspektiften incelemiştir.

Bu bağlamda, elde edilen bulgular, yapay zeka destekli ses sentezleme teknolojilerinin, ses üretimindeki doğallığı ve ifade yeteneğini önemli ölçüde iyileştirebildiğini, ancak hâlâ gerçek insan sesinin sağladığı duygusal derinlik ve tonal zenginliği tam olarak replike edemediğini göstermektedir. Araştırma sonuçları, yapay zeka teknolojilerinin bu alandaki potansiyellerinin yanı sıra, duygusal ifade ve doğallık gibi alanlarda gereken ilerlemelere dikkat çekmektedir.

Sonuç olarak, bu çalışma, yapay zeka teknolojilerinin ses sentezi alanındaki etkilerini detaylı bir şekilde ortaya koymakta ve gelecekteki araştırmalar için yol gösterici olabilecek önemli içgörüler sunmaktadır. Bu sonuçlar, teknolojik gelişmelerin insan sesine ne derece yaklaşabildiğini ve bu süreçte karşılaşılan temel zorlukları anlamamıza yardımcı olmaktadır.

Üçüncü Alt Probleme İlişkin Tartışma Sonuçlar

Bu alt problem kapsamında, katılımcıların sentetik ve gerçek sesler arasındaki tercihlerini ve bu seslerin kullanıcı deneyimi açısından nasıl değerlendirildiği incelenmiştir. Araştırma bulguları, katılımcıların ses tercihlerinde doğallık, teknolojik üstünlük ve anlaşılabilirlik gibi faktörleri önemli bulduklarını göstermektedir.

Bu bölüm, "Katılımcıların sentetik sesler ve gerçek sesler arasındaki tercihlerini ve bu seslerin kullanıcı deneyimi açısından nasıl değerlendirildiğini" incelemektedir. Katılımcılardan elde edilen verilere dayanarak, yarı yapılandırılmış sorularla toplanan görüşler değerlendirilmiştir.

Katılımcıların tercihleri ve sentetik ile gerçek seslerin kullanıcı deneyimi üzerindeki etkileri, çeşitli faktörlerle incelenmiştir. Genel olarak, kullanıcılar arasında gerçek sesin doğallığı ve sentetik sesin teknolojik avantajları arasında bir bölünme görülmektedir.

Çoğu katılımcı, gerçek sesin doğal tonunu ve insani dokunuşunu tercih ettiğini belirtmiştir. Gerçek sesin duygusal ifade kapasitesi, katılımcılar için önemli bir faktör olarak öne çıkmıştır. Diğer katılımcılar, sentetik sesin netliği ve hatasız sunumunu tercih ettiklerini ifade etmişlerdir. Bu, özellikle teknik ve bilgi ağırlıklı kontekstlerde kullanışlı bulunmuştur.

Gerçek sesin kullanıcılarla duygusal bir bağ kurma yeteneği, onu günlük ve kişisel kullanım için daha uygun hale getirmiştir. Her iki ses türü de anlaşılabilirlik açısından farklı avantajlar sunmuştur; gerçek ses tonal zenginliği ile, sentetik ses ise netlik ve tutarlılık ile öne çıkmıştır. Sentez ses, özellikle otomatize edilmiş sistemler ve belirgin talimat gerektiren durumlar için tercih edilmiştir.

Gerçek sesin duygusal derinliği ve vurgulama yeteneği, kullanıcı deneyimini zenginleştirdiği belirtilmiştir. Sentez sesin bu alandaki kontrolü ve ayarlanabilirliği, profesyonel ve teknik senaryolarda avantaj sağlamıştır. Gerçek sesin doğal dalgalanmaları, konuşmanın akıcılığını ve anlaşılabilirliğini artırmış, böylece kullanıcı deneyimini iyileştirmiştir.

Araştırma sonuçları, kullanıcı tercihlerinin gerçek sesin duygusal zenginliği ile sentetik sesin teknolojik avantajları arasında bölündüğünü göstermektedir. Bu iki ses türü arasındaki farklar, özellikle kullanım senaryolarına ve kişisel tercihlere bağlı olarak kullanıcı deneyiminde belirgin bir etki yaratmaktadır. Gerçek ses, daha kişisel ve duygusal etkileşimler için tercih edilirken, sentetik ses netlik, hatasız sunum ve teknik kullanım için ideal bulunmuştur.

Katılımcıların yarısı, gerçek sesin doğallığı ve duygusal ifade yeteneğini tercih etmiştir. Gerçek ses, kullanıcılar tarafından sıcaklık ve samimiyet hissi vermesi nedeniyle tercih edilirken, bu seslerin insanlar arası iletişimde daha etkili olduğu belirtilmiştir.

(Baltazar vd., 2019). Diğer yarısı ise, sentetik sesin netliği ve teknolojik avantajlarını tercih sebebi olarak göstermiştir. Sentez sesler, özellikle teknik ve bilgilendirici kullanımlarda, tutarlılık ve hata toleransı avantajlarıyla öne çıkmaktadır (Huang, 2009).

Duygusal ifade ve doğallık açısından gerçek sesler üstünken, teknolojik üstünlük ve pratik kullanım kolaylığı bakımından sentetik sesler daha avantajlıdır. Zheng ve Zhang (2018) tarafından belirtilen, sentetik seslerdeki faz işlemenin önemli gelişmelere işaret etmesi, bu seslerin gerçek seslerle rekabet edebilirliğini artırmaktadır. Ancak, gerçek sesin duygusal bağlama ve tonal zenginlik sağlamadaki üstünlüğü, özellikle terapötik ve kişisel asistan uygulamalarında önemli bir avantaj olarak kalmaya devam etmektedir.

Bu bulgular, yapay zeka ve ses işleme teknolojilerinin, sentetik seslerin doğallığını ve duygusal ifade kapasitesini artırma yönündeki ilerlemelerini vurgulamaktadır. Crandall (2019) tarafından yapılan araştırmalar, yapay zekanın ürünlere entegrasyonu ile birlikte sentetik seslerin, gerçek seslerle daha iyi bir rekabet içine girebileceğini öngörmektedir. Bu doğrultuda, teknolojinin ilerlemesi ile birlikte sentetik seslerin kullanımının daha geniş alanlarda yaygınlaşacağı ve kullanıcı deneyimini daha fazla iyileştireceği beklenmektedir.

Sonuç olarak, sentetik ve gerçek seslerin her birinin kendine has avantajları ve kısıtlamaları bulunmaktadır. Teknolojik gelişmeler ve kullanıcı ihtiyaçları doğrultusunda, bu iki ses tipinin birbirini tamamlayıcı şekilde kullanılması, kullanıcı deneyimini optimize edecek ve ses teknolojilerinin geleceğini şekillendirecek ana faktörlerden biri olacaktır.

Dördüncü Alt Probleme İlişkin Tartışma Sonuçlar

Bu alt başlıkta, sentetik sesler ile gerçek sesler arasındaki güvenilirlik algısının ve bilgi aktarımındaki etkilerinin değerlendirilmesi üzerine elde edilen bulgular irdelenmiştir. Araştırma, bu iki ses türünün kullanıcılar üzerindeki izlenimlerini ve güvenilirlik algılarını anlamak amacıyla tasarlanmış yarı yapılandırılmış görüşme sorularını temel almaktadır.

Katılımcıların ilk izlenimleri üzerinden gerçek ve sentetik seslerin güvenilirlik düzeyleri değerlendirilmiştir. Gerçek sesler, tanıdıklık ve doğallık gibi özellikler nedeniyle dokuz katılımcı tarafından daha güvenilir bulunmuştur. Bu algı, sesin insani özelliklerinin aşinalık duygusu yaratmasıyla desteklenmektedir. Öte yandan, teknik üstünlük açısından on katılımcı, sentetik seslerin daha güvenilir olduğunu belirtmiştir.

Sentez seslerin, bilgi işleme ve sunum konusunda gösterdikleri tutarlılık ve hata toleransı, bu algıyı pekiştiren faktörler arasında yer almaktadır.

Güvenilirlik algısının bilgi aktarımı üzerindeki etkilerine dair bulgular, başlangıçta sentetik seslere karşı şüpheyle yaklaşan sekiz katılımcının zamanla bu seslere olan güvenlerinin arttığını göstermiştir. Katılımcıların, verilen doğru bilgilerin etkisiyle sentetik seslerin güvenilirliğine dair algıları olumlu yönde değişmiştir. Bu değişim, teknik detaylar açısından altı katılımcının sentetik sesleri daha güvenilir bulduğunu gösteren bulgularla desteklenmiştir.

Elde edilen bulgular, mevcut literatürde de yansıtıldığı üzere, yapay zeka destekli ses teknolojilerinin insan algısı üzerinde önemli etkiler bırakabileceğini göstermektedir. Zou, Han, ve So (2009) tarafından incelenen yapay sinir ağlarının karmaşık modelleme yetenekleri, sentetik seslerin insan sesine benzer şekilde güvenilir olabileceğini düşündürmektedir. Ayrıca, Crandall (2019)'ın yapay zeka ve imalat üzerine yaptığı çalışmalar, teknolojinin insan etkileşimleri içinde nasıl bir rol oynayabileceğine dair fikirler sunmaktadır. Bu çalışmalar, tez kapsamında elde edilen bulguların, yapay zeka teknolojilerinin günlük yaşam ve özellikle bilgi aktarımı kontekstlerinde nasıl etkili olabileceğini daha geniş bir perspektiften değerlendirme olanağı sağlamaktadır.

Bu bağlamda, sentetik seslerin teknik detaylar ve tutarlı bilgi sunumu konusundaki üstünlükleri, özellikle Reynoso (2019) tarafından ele alınan yapay zekanın çeşitli türleri ve sınıflandırmalarıyla paralellik göstermektedir. Reynoso'nun çalışması, yapay zeka sistemlerinin ne kadar geniş bir yelpazede işlev görebileceğini ortaya koymaktadır. Bu da sentetik seslerin, belirli alanlarda insan seslerinden daha güvenilir olabileceğini destekleyen bir argüman olarak değerlendirilebilir.

Sonuç olarak, bu çalışma, sentetik ve gerçek sesler arasındaki güvenilirlik algıları üzerine değerli içgörüler sunmakta ve yapay zeka destekli teknolojilerin insan algısı üzerindeki etkilerini daha iyi anlamamızı sağlamaktadır. Elde edilen bulgular, ileri araştırmalar için bir temel oluştururken, yapay zeka uygulamalarının sosyal kabulü ve etkileşim stratejileri üzerine de düşündürmektedir.

Beşinci Alt Probleme İlişkin Tartışma Sonuçları

Bu alt başlıkta, sentetik seslerin sosyal etkileşime etkisi ve toplumdaki kabul düzeyine yönelik değerlendirmeler ele alınmaktadır. Araştırma, katılımcıların sentetik

seslerin endüstriyel etkililiği ve toplumsal kabulü hakkındaki görüşlerini yarı yapılandırılmış görüşme soruları aracılığıyla incelemiştir.

Katılımcıların görüşleri, sentetik seslerin endüstriyel uygulamalarda özellikle teknik üstünlük ve güncellenebilirlik açısından önemli bir etkililik sunduğunu göstermiştir. Beş katılımcı, bu seslerin sürekli güncellenme potansiyeli sayesinde teknik destek ve bilgi sağlama gibi alanlarda daha etkili olduğunu ifade etmiştir. Diğer yandan, gerçek seslerin doğrudan insan etkileşimi gerektiren durumlarda daha etkili olduğu belirtilmiştir, bu durum altı katılımcı tarafından vurgulanmıştır. Bilgi tabanlı uygulamalarda ise yedi katılımcı, sentetik seslerin anlaşılabilirlik ve tutarlılık avantajları sayesinde daha etkili olduğunu belirtmiştir.

Sentezlenmiş seslerin toplumda giderek daha fazla kabul gördüğüne dair kanıtlar bulunmaktadır. Altı katılımcı, bu seslerin özellikle teknoloji entegre hizmetlerde ve müşteri hizmetlerinde kabulünün arttığını rapor etmiştir. Ancak sekiz katılımcı, özellikle duygusal ifade gerektiren durumlarda hala direnç ve çekincelerin bulunduğunu ifade etmiştir. Genç nesillerin bu teknolojiye daha açık olduğu, beş katılımcı tarafından belirtilmiştir.

Katılımcıların görüşleri, Docasal & Alvarez (2023) tarafından yapılan çalışmanın bulgularıyla uyumlu olarak, sentetik seslerin sosyal kabulünün belirli kontekstlerde yüksek olduğunu göstermektedir. Ancak, Crandall (2019) tarafından ele alınan yapay zekanın iş dünyasındaki entegrasyonu gibi, bu kabulün tüm sektörlerde homojen olmadığı da vurgulanmaktadır.

Bu bulgular, Ning vd. (2019) tarafından ele alınan konuşma sentezi üzerine yapılan çalışmaların, sentetik seslerin sosyal ve endüstriyel uygulamadaki yerini daha da pekiştirdiğini göstermektedir. Özellikle, yapay sinir ağları kullanılarak geliştirilen konuşma sentezi teknolojilerinin, gelişmekte olan uygulamalarda nasıl bir etki yaratabileceği önemli bir araştırma alanı olarak karşımıza çıkmaktadır.

Sonuç olarak, bu çalışma sentetik seslerin sosyal kabulü ve etkileşimdeki rolünü derinlemesine inceleyerek, bu teknolojilerin toplumsal ve endüstriyel açıdan nasıl değerlendirildiğini ve gelecekteki potansiyellerini ortaya koymaktadır. Katılımcıların yansıttığı çeşitli görüşler, bu ses teknolojilerinin toplumsal entegrasyon sürecinin karmaşık ve dinamik bir yapıda olduğunu göstermektedir. Bu da gelecekteki araştırmalar için zengin bir zemin sunmaktadır.

Öneriler

Araştırma, yapay zeka tarafından üretilen sentetik seslerin gerçek seslerle karşılaştırılmasını ele almıştır ve katılımcıların sesleri ayırt etme yeteneklerine dair önemli bulgular sunmuştur. Araştırmanın gösterdiği üzere, katılımcılar sentetik ve gerçek sesler arasındaki farkları belirlerken çeşitli özellikleri temel almışlardır; özellikle sesin doğallığı, canlılığı ve arka plan seslerindeki farklılıklar gibi faktörler, seslerin ayırt edilmesinde kritik rol oynamıştır. Buna ek olarak, teknolojik seslerin kullanımı sırasında katılımcıların medya ve teknolojiye olan farkındalıkları da artmıştır.

Bu bulgular ışığında, sentetik seslerin doğallığını ve ifade yeteneğini artırmak amacıyla yapay zeka teknolojilerinde yapılabilecek iyileştirmeler önem taşımaktadır. Araştırmacılar ve teknoloji geliştiriciler, ses sentezi algoritmalarını geliştirerek, insan sesinin duygusal ve tonal zenginliklerini daha iyi taklit edebilir. Bununla birlikte, eğitimciler ve dil bilimciler, ses teknolojilerinin dilsel doğruluk ve duygusal ifade kapasitesini artırmaya yönelik multidisipliner çalışmalar yapabilir. Ayrıca, teknolojik seslerin sosyal medya platformları gibi geniş kullanım alanlarında nasıl algılandığına dair daha kapsamlı çalışmalar yapılabilir.

Önerilen diğer bir çalışma alanı ise, sentetik seslerin eğitim ve öğretim materyallerinde nasıl kullanılabileceğini araştırmaktır. Özellikle dil öğrenimi gibi alanlarda sentetik seslerin kullanılması, dilsel çeşitliliği ve erişimi artırabilir. Bu doğrultuda, sentetik seslerin eğitimsel içeriklerde nasıl daha etkin kullanılabileceği üzerine pilot projeler geliştirilebilir.

Son olarak, sentetik seslerin güvenilirlik algısı ve sosyal kabul düzeylerinin artırılması için kamuoyu bilgilendirme kampanyaları ve eğitici seminerler düzenlenebilir. Bu tür faaliyetler, sentetik ses teknolojilerinin faydalarını ve potansiyel kullanım alanlarını daha geniş kitlelere tanıtmak için önemlidir. Katılımcıların sentetik seslere olan güveninin artırılması, bu teknolojilerin daha geniş bir kabul görmesine yardımcı olabilir, bu da teknolojik ilerlemenin toplumsal etkilerini pozitif yönde destekleyecektir.

KAYNAKÇA

- Abrell, T., Benker, A., & Pihlajamaa, M. (2018). User Knowledge Utilization In Innovation Of Complex Products And Systems: An Absorptive Capacity Perspective. *Creativity and Innovation Management*, 27(2), 169-182.
- Acemoğlu, D., & Restrepo, P. (2018). The Race Between Man And Machine: Implications Of Technology For Growth, Factor Shares, And Employment. *American Economic Review*, 108(6), 1488-1542.
- Akiyama, M. (2019). Nothing Connects Us but Imagined Sound. *Sound, Media, Ecology*, 1(2), 113-129.
- Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. *Artificial intelligence review*, 55(1), 589-656.
- Arik, S. Ö., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., & Shoenybi, M. (2017). Deep voice: Real-time neural text-to-speech. In *International conference on machine learning*, PMLR, July 2017, 195-204.
- Arik, S.Ö., Chen, J., Peng, K., Ping, W., & Zhou, Y. (2018). Neural Voice Cloning with a Few Samples. *ArXiv, abs/1802.06006*.
- Baltazar, M. (2018). Where mind and music meet: Affect self-regulation through music. *Jyväskylä studies in humanities*, 2(345), 20-34.
- Baltazar, M., Västfjäll, D., Asutay, E., Koppel, L., & Saarikallio, S. (2019). Is it me or the music? Stress reduction and the role of regulation strategies and music. *Music & Science*, 2(1), 20-42.
- Bateman, J. (2022). *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. Carnegie Endowment for International Peace.
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* 1(4), 149-159.
- Blaauw, M., Bonada, J., & Daido, R. (2019). Data Efficient Voice Cloning for Neural Singing Synthesis. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2019, 6840-6844.
- Borano, J. (2016). Applications Of Artificial Intelligence & Associated Technologies. *Science*, 5(6), 64-67.
- Bresin, R., & Vedovetto, A. (1994). Neural networks for a simpler control of synthesis algorithm of musical tones and for their compression. In *Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, IEEE, October 1994, 628-631.
- Brigham, E. O. (1988). *The fast Fourier transform and its applications*. Prentice-Hall, Inc.
- Brundage M. (2016). Modeling Progress In AI Analysis Of Relevant Literatures. *The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence*, 1(1), 71–79.
- Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial intelligence and the modern productivity paradox. *The economics of artificial intelligence: An agenda*, 23(2), 23-57.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), 12-30.
- Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, 209(2), 118-190.
- Chandra, G. R., Tata, V., & Anand, D. (2022). Real-Time Voice Cloning System Using Machine Learning Algorithms. In *Conference of Innovative Product Design and*

- Intelligent Manufacturing System*, Singapore: Springer Nature Singapore, November 2022, 525-534.
- Christensen, M. G. (2019). *Introduction to audio processing*. Springer.
- Crandall, D. J. (2019). Artificial Intelligence And Manufacturing. *Smart Factories: Issues of Information Governance*, 1(2), 10-16.
- Cui, X., Wu, Y., Wu, J., You, Z., Xiahou, J., & Ouyang, M. (2022). A review: Music-emotion recognition and analysis based on EEG signals. *Frontiers in Neuroinformatics*, 16(2), 99-109.
- Çepni, S. (2018). *Araştırma ve Proje Çalışmalarına Giriş*, İstanbul: Celepler Matbaacılık Yayın ve Dağıtım.
- Davenport, T. H. (2018). *The AI Advantage: How To Put The Artificial Intelligence Revolution To Work*. Cambridge: MIT Press.
- Dale, R. (2023). NLP startup funding in 2022. *Natural Language Engineering*, 29(1), 162-176.
- Delogu, C., & Sementina, C. (1995). Towards a more realistic evaluation of synthetic speech: a cognitive perspective. In *Speech Recognition and Coding: New Advances and Trends* (337-340). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dignum, V. (2019). *Responsible Artificial Intelligence: How To Develop And Use AI In A Responsible Way*. Cham: Springer.
- Dobbe, R., Dean, S., Gilbert, T., & Kohli, N. (2018). A broader view on bias in automated decision-making: Reflecting on epistemology and dynamics. *arXiv preprint arXiv:1807.00553*.
- Docasal, A.G., & Alvarez, A. (2023). Enhancing Voice Cloning Quality through Data Selection and Alignment-Based Metrics. *Applied Sciences*, 13(14), 49-80.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness Through Awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference 1(2)*, 214-226.
- Ensmenger, N. (2018). The environmental history of computing. *Technology and culture*, 59(4), 7-33.
- Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 3-15.
- Galyashina, E. I., & Nikishin, V. D. (2023). AI Generated Fake Audio as a New Threat to Information Security: Legal and Forensic Aspects.
- Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding Artificial Intelligence Adoption In Operations Management: Insights From The Review Of Academic Literature And Social Media Discussions. *Annals of Operations Research*, 308(1), 177-213.
- Gully, A. J., Yoshimura, T., Murphy, D. T., Hashimoto, K., Nankaku, Y., & Tokuda, K. (2017). Articulatory text-to-speech synthesis using the digital waveguide mesh driven by a deep neural network. In *Interspeech, ISCA-INST SPEECH COMMUNICATION ASSOC*, April 2017, 234-238.
- Gurjar, A. P., & Patel, S. B. (2021). Fundamental Categories of Artificial Neural Networks. In *Applications of Artificial Neural Networks for Nonlinear Data* (30-64). IGI Global.
- Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning?. *Speech Communication*, 51(10), 906-919.
- Hayes, P., Poel, I., & Steen, M. (2020). Algorithms and values in justice and security. *Ai & Society*, 35, 533-555.
- Henderson, P.D. (2004). The Fundamentals of FFT-Based Audio Measurements in SmaartLive.

- Htun, H. M., Zin, T., & Tun, H. M. (2015). Text to speech conversion using different speech synthesis. *Int J Sci Technol Res*, 4(7), 104-108.
- Hu, W., & Zhu, X. (2023). A real-time voice cloning system with multiple algorithms for speech quality improvement. *PLoS One*, 18(4), 28-39.
- Huang, Y. (2009). Advances in artificial neural networks—methodological development and application. *Algorithms*, 2(3), 973-1007.
- Ji, S., Luo, J., & Yang, X. (2020). A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions. *arXiv preprint arXiv:2011.06801*.
- Johnson, D. G., & Verdicchio, M. (2019). AI, Agency And Responsibility: The VW Fraud Case And Beyond. *Ai & Society*, 34(3), 639-647.
- Joshi, N. (2019). Types Of Artificial Intelligence. *Received from: Forbes Cognitive World*, 1(2), 57-64.
- Kane, J., Johnstone, M. N., & Szewczyk, P. (2024). Voice Synthesis Improvement by Machine Learning of Natural Prosody. *Sensors*, 24(5), 16-24.
- Kaptan, S. (1995). *Bilimsel Araştırma ve İstatistik Teknikleri*. Ankara: Rehber Yayınevi.
- Karasar, N. (2023). *Bilimsel İrade Algı Çerçevesi ile Bilimsel Araştırma Yöntemi: Kavramlar İlkeler Teknikler*. Ankara: Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic. Ltd. Şti.
- Kefauver, A. P., & Patschke, D. (2007). *Fundamentals of digital audio*. AR Editions, Inc.
- Kheria, I., & Karani, R. (2023). CloneAI: A Deep Learning-Based Approach for Cloned Voice Detection. In *International Conference on Data Science and Applications*, Singapore: Springer Nature Singapore, July 2023, 267-282.
- Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement Learning In Robotics: A Survey. *The International Journal of Robotics Research*, 32(11), 1238-1274.
- Korinek, A., & Stiglitz, J. E. (2018). Artificial Intelligence And Its Implications For Income Distribution And Unemployment. In *The economics of artificial intelligence: An agenda*, 1(1), 349-390.
- Lai, J., Chen, B., Tan, T., Tong, S., & Yu, K. (2016). Phone-aware LSTM-RNN for voice conversion. In *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, IEEE, November 2016, 177-182.
- Lane, M., & Martin, A. S. (2021). The Impact Of Artificial Intelligence On The Labour Market: What Do We Know So Far?. *Organisation for Economic Co-operation and Development*, 3(4), 12-22.
- Lavet, V. F., Henderson, P., & Pineau, J. (2018). An Introduction To Deep Reinforcement Learning. *Foundations and Trends in Machine Learning*, 11(3-4), 219-354.
- Lee, K. F., & Qiufan, C. (2021). *AI 2041: Ten visions for our future*. Crown Currency.
- Lerch, A. (2012). *An introduction to audio content analysis: Applications in signal processing and music informatics*. Wiley-IEEE Press.
- Li, S., & Sung, Y. (2023). Mrbert: Pre-training of melody and rhythm for automatic music generation. *Mathematics*, 11(4), 798-808.
- Liu, X., & Bao, C. C. (2014). EURASIP Journal on Audio, Speech, and Music Processing. *EURASIP Journal on Audio, Speech, and Music Processing*, 201(41), 41-60.
- Liu, D., Xu, J., Li, R., Dai, R., & Gong, W. (2002). Measurements of sound speed in the water by Brillouin scattering using pulsed Nd: YAG laser. *Optics communications*, 203(6), 335-340.
- Mah, P. M., Skalna, I., & Muzam, J. (2022). Natural language processing and artificial intelligence for enterprise management in the era of industry 4.0. *Applied Sciences*, 12(18), 92-107.

- Mahajan, D., Gapat, A., Moharkar, L., Sawant, P., & Dongardive, K. (2021). Artificial Generation of Realistic Voices. *International Journal of Applied Sciences and Smart Technologies*, 3(1), 11-26.
- Malcangi, M., & Frontini, D. (2007). A Language-Independent Neural Network-Based Speech Synthesizer. In *Proceedings of the 10th International Conference on Engineering Applications of Neural Networks*, (395-402). Publishing Centre Alexander TEI.
- Mendoza, M. F., Albanil, I. M., Rodriguez, C., Lezama, P., Rodriguez, D., & Pomachagua, Y. (2022). Neural Network Strategies and Models for Voice Cloning in a Multi-speaker Mode: An Overview. In *International Conference on Intelligent Technologies*, Singapore: Springer Nature Singapore, December 2022, 229-237.
- Miranda, E. R. (2019). From symbols to sound: AI-based investigation of sound synthesis. In *Timbre Composition in Electroacoustic Music* (211-232). Routledge.
- Nakamura, K., Hashimoto, K., Oura, K., Nankaku, Y., & Tokuda, K. (2019). Singing voice synthesis based on convolutional neural networks. *arXiv preprint arXiv:1904.06868*.
- Naniwa, Y., Kondo, T., Kamiyama, K., & Kamata, H. (2012). Study on the artificial synthesis of human voice using radial basis function networks. In *Advanced Methods, Techniques, and Applications in Modeling and Simulation: Asia Simulation Conference 2011, Seoul, Korea, Proceedings*, Springer, Japan, November 2011, 291-300.
- Natsiou, A., & O'Leary, S. (2021). Audio representations for deep learning in sound synthesis: A review. In *2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA)*, IEEE, April 2021, 1-8.
- Neekhara, P., Hussain, S.S., Dubnov, S., Koushanfar, F., & McAuley, J. (2021). Expressive Neural Voice Cloning. *ArXiv, abs/2102.00151*.
- Neuman, W. L. & Robson, K. (2014). *Basics Of Social Research*. Toronto: Pearson Canada.
- Ning, Y., He, S., Wu, Z., Xing, C., & Zhang, L. J. (2019). A review of deep learning based speech synthesis. *Applied Sciences*, 9(19), 40-50.
- Nissenbaum, H. (2009). Privacy in context: Technology, policy, and the integrity of social life. In *Privacy in Context*. Stanford University Press.
- Novelli, N., & Proksch, S. (2022). Am I (deep) blue? music-making ai and emotional awareness. *Frontiers in Neurorobotics*, 16(2), 89-110.
- O'Cinneide, A., Dorran, D., & Gainza, M. (2007). A brief introduction to speech synthesis and voice modification.
- Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., & Hassabis, D. (2018). Parallel wavenet: Fast high-fidelity speech synthesis. In *International conference on machine learning*, PMLR, July 2018, 3918-3926.
- Parker, J. E., & Dockray, S. (2023). 'All possible sounds': speech, music, and the emergence of machine listening. *Sound Studies*, 9(2), 253-281.
- Paschen, U., Pitt, C., & Kietzmann, J. (2020). Artificial Intelligence: Building Blocks And An Innovation Typology. *Business Horizons*, 63(2), 147-155.
- Perera, N. N., & Ganegoda, G. U. (2023). A Comprehensive Review on Speech Synthesis Using Neural-Network Based Approaches. In *2023 3rd International Conference on Advanced Research in Computing (ICARC)*, IEEE, February 2023, 214-219.
- Petropoulos, G. (2018). The Impact Of Artificial Intelligence On Employment. *Praise for Work in the Digital Age*, 119(4), 121-127.

- Prat, N., Wattiau, I. C., & Akoka, J. (2014). *Artifact Evaluation In Information Systems Design-Science Research-A Holistic View*. Proceedings of the Pacific Asia Conference for Information Systems.
- Pucik, J., Kubinec, P., & Ondracek, O. (2014). FFT with modified frequency scale for audio signal analysis. In *2014 24th International Conference Radioelektronika*, IEEE, April 2014, 1-4.
- Raghavendra, E. V., Vijayaditya, P., & Prahallad, K. (2010). Speech synthesis using artificial neural networks. In *2010 National Conference On Communications (NCC)*, IEEE, January 2010, 1-5.
- Raitio, T., Suni, A., Juvela, L., Vainio, M., & Alku, P. (2014). Deep neural network based trainable voice source model for synthesis of speech with varying vocal effort. In *Interspeech*, September 2014, 1969-1973.
- Reynoso, R. (2019). Main Types Of Artificial Intelligence. *G2 Learning Hub*. 1(2), 15-24.
- Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2020). How Do Fairness Definitions Fare? Testing Public Attitudes Towards Three Algorithmic Definitions Of Fairness In Loan Allocations. *Artificial Intelligence*, 283(2), 103-238.
- Semel, B. M. (2022). Listening Like A Computer: Attentional Tensions And Mechanized Care In Psychiatric Digital Phenotyping. *Science, Technology & Human Values*, 47(2), 266-290.
- Sharma, R., (2007). *Speech Synthesis* (Unpublished PhD Thesis). Patiala, Electrical And Instrumentation Engineering Department, Thapar Institute Of Engineering & Technology (Deemed University).
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., & Wu, Y. (2018). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, April 2018, 4779-4783.
- Shih, K. J., Valle, R., Badlani, R., Lancucki, A., Ping, W., & Catanzaro, B. (2021). RAD-TTS: Parallel flow-based TTS with robust alignment learning and diverse synthesis. In *ICML Workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models*.
- Simons, A. C. (2016). Artificial Intelligence Produces Realistic Sounds That Fool Humans. *MIT News*.
- Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). *This is service design doing*. " O'Reilly Media, Inc.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.
- Umapathy, K., Ghoraani, B., & Krishnan, S. (2010). Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking. *EURASIP Journal on Advances in Signal Processing*, April 2010, 1-28.
- Vasquez, S., & Lewis, M. (2019). Melnet: A generative model for audio in the frequency domain. *arXiv preprint arXiv:1906.01083*.
- Weijters, T., & Thole, J. (1993). Speech synthesis with artificial neural networks. In *IEEE International Conference on Neural Networks*, IEEE, March 1993, 1764-1769.
- Xu, Z., Elshamy, S., Zhao, Z., & Fingscheidt, T. (2021). Components loss for neural networks in mask-based speech enhancement. *EURASIP Journal on Audio, Speech, and Music Processing*, 20(1), 24-36.

- Xu, Y., Du, J., Dai, L. R., & Lee, C. H. (2013). An experimental study on speech enhancement based on deep neural networks. *IEEE Signal processing letters*, 21(1), 65-68.
- Zhao, L., & Chen, F. (2020). Research on Voice Cloning with a Few Samples. *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, April 2020, 323-328.
- Zheng, N., & Zhang, X. L. (2018). Phase-aware speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1), 63-76.
- Zhu, Y., Baca, J., Rekabdar, B., & Rawassizadeh, R. (2023). A Survey of AI Music Generation Tools and Models. *arXiv preprint arXiv:2308.12982*.
- Zou, J., Han, Y., & So, S. (2009). Overview of Artificial Neural Networks. *Methods in molecular biology*, 458(2), 15-23.



EKLER

Ek 1: Etik Kurul Kararı

