YAPAY ZEKÂ İLE SAHTE HABER ÜRETİMİNİN SOSYAL MEDYADAKİ GÖRÜNÜMÜ: TEYİT.ORG ÖRNEĞİ

THE PRESENCE OF AI-GENERATED FAKE NEWS ON SOCIAL MEDIA:THE CASE OF TEYİT.ORG

İbrahim YILDIZ

15

Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi
Yıl: 2025 Cilt: 34 No: Özel Sayı Sayfa: 255-274
https://dergipark.org.tr/tr/pub/cusosbil

DOI: 10.35379/cusosbil.1696022

Geliş Recieved: **09.05.2025**
Kabul Accepted **08.07.2025**

YAPAY ZEKÂ İLE SAHTE HABER ÜRETİMİNİN SOSYAL MEDYADAKİ GÖRÜNÜMÜ: TEYİT.ORG ÖRNEĞİ

THE PRESENCE OF AI-GENERATED FAKE NEWS ON SOCIAL MEDIA:THE CASE OF TEYİT.ORG

İbrahim YILDIZ [1]

Anahtar Kelimeler:
Yapay Zekâ,
Sahte Haber,
Sosyal Medya,
Teyit.Org.

Keywords:
Artificial Intelligence,
Fake News,
Social Media,
Teyit.Org

ÖZ

Modern teknolojik gelişmenin yeni halkası olan yapay zekâ teknolojileri, gerçekçi görüntüler ve videolar üretebilme yetenekleri nedeniyle sahte haber üretiminde kullanılabilmektedir. Sahte haberlerin üretimi ve tespitinde yapay zekâ teknolojilerinin kullanımı, akademik çevrelerde giderek daha sık incelenen yeni bir araştırma alanı olarak öne çıkmaktadır. Mevcut literatür, yapay zekâ ve sahte haberlerle ilgili çalışmaların ağırlıklı olarak bu tür bilgileri tespit etmeye yönelik teknolojilerle ilgilendiğini ortaya koymaktadır. Sahte haberlerin yayıldığı kanallar ve bu tür içeriklerin özellikleriyle ilgili araştırmalar son derece sınırlıdır. Bu çalışma, sahte haber üretiminde yapay zekâ teknolojilerinin rolüne ve sosyal medyadaki görünümüne odaklanmaktadır. Buradan hareketle çalışmanın amacı, yapay zekâ tarafından üretilen sahte haberlerin varlığını, yanıltma potansiyeli, hitap ettiği konular, yayıldığı sosyal medya platformları ve içerdiği bilgi türleri açısından analiz etmektir. Belirlenen amaç doğrultusunda, 1 Ocak 2025-31 Mart 2025 tarihleri arasında teyit.org doğrulama platformunda yapay zekâ anahtar kelimesi ile arama yapılmış ve arama sonucu elde edilen yapay zekâ ile üretilmiş 24 sahte haberin içerik analizi yapılmıştır. Araştırmanın sonuçları, yapay zekâ teknolojilerinin sahte haber üretme amacıyla kullanıldığını desteklemektedir. Araştırmada en çok Instagram ve TikTok'ta görülen ve daha çok video içerikler şeklinde yayılan sahte haber içeriklerinde, yaşam, doğa ve çevreyle ilgili konuların öncelikli olduğu tespit edilmiştir.

ABSTRACT

The advent of artificial intelligence (AI) technologies, which represent a pivotal nexus in contemporary technological evolution, has engendered a newfound capacity to generate highly realisticimages and videos. This capacity, however, is not without its potential implications, particularly with regard to the dissemination of misinformation through the medium of AI-generated content. The employment of artificial intelligence technologies in the production and detection of fake news has emerged as a novel research domain that is garnering increasing attention in academic circles. A review of the extant literature indicates that research on artificial intelligence and fake news has focused predominantly on technologies for detecting such information. Research on the channels through which fake news is disseminated and the characteristics of such content is extremely limited. The present study focuses on the role of artificial intelligence technologies in the production of fake news and its reflection on social media platforms. From this perspective, the objective of the present study is to examine the prevalence of disinformation generated by artificial intelligence, with a focus on its potential to deceive, the subjects it addresses, the social media platforms on which it disseminates, and the types of information it contains. In accordance with the predetermined objective, a keyword search was conducted on the verification platform teyit.org between January 1, 2025 and March 31, 2025 using the keyword “artificial intelligence.” A content analysis of 24 fake news articles produced by artificial intelligence was conducted as a result of the aforementioned search. The findings of the study substantiate the utilization of artificial intelligence technologies in the fabrication of disinformation. The study’s findings indicated that such fake news predominantly appears on Instagram and TikTok, is mostly disseminated in video format, and primarily focuses on topics related to life, nature, and the environment.

[1] Öğr. Gör. Dr., İstanbul Medeniyet Üniversitesi, ibrahim.yildiz@medeniyet.edu.tr, ORCID: 0000-0002-2542-389X

Alıntılamak için/Cite as: Yıldız İ. (2025). Yapay Zekâ İle Sahte Haber Üretiminin Sosyal Medyadaki Görünümü: Teyit.org Örneği, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 34 (Özel Sayı), 255-274

GİRİŞ

İlk olarak 1950'lerde ortaya atılan "yapay zekâ" kavramı, başlangıçta makineler tarafından sergilenen insan zekâsının basit teorisi olarak değerlendirilirken, hızlı teknolojik ilerleme ve son derece büyük hacimli veri kümelerinin yönetiminin olduğu günümüzde teoriden somut bir pratik gerçekliğe dönüşmüştür. Yapay zekâ, büyük hacimli veri kümelerinin değerlendirilmesinden çevrim içi satın alma önerilerinin sağlanmasına, reklamcılığa ve dolandırıcılık tespitine kadar toplumun birçok yönünün ayrılmaz bir parçası haline gelmiştir (Helm vd. 2020). Yapay zekâ alanı, "belirli bir sorunu çözmek veya tanımlanmış bir sonuca ulaşmak için atılan bir dizi adım" olarak tanımlanan algoritmalar tarafından yönlendirilir. Bu süreçler makine öğrenimi veya otomasyon olarak programlanabilir (Beckett 2019, s. 12). Temel amacı, doğası gereği insan beyninin işlevlerini yerine getirmek ve hatta bazı durumlarda bunları aşabilmek olan yapay zekâ, medya sektöründe bilgi üretimi ve yayılımının karmaşık süreçlerinde önemli bir oyuncu olarak ortaya çıkmıştır.

Çeşitli alanlarda kullanımı yaygınlaşan yapay zekâ teknolojileri, iletişim ve haber medyasında sahte haberlerle ilgili yeni bir durum ortaya çıkarmıştır. Üretken yapay zekâ teknolojisindeki gelişmeler sahte haber üretmeyi kolaylaştırırken, sosyal medya gibi çevrimiçi platformlar bu tür sahte haberlerin yayılmasını hızlandırmıştır (Joshi ve Patel, 2024). Yapay zekâ tarafından üretilen içeriklerin gerçekleri ile ayırt edilemeyecek düzeyde benzerliği, kamuoyunu kolayca yanıltabilen ikna edici sahte haberlerin artmasına yol açmıştır. Bu durum, demokratik süreçleri baltalama ve zarar verebilecek yanlış bilgiler yayma potansiyeli de dahil olmak üzere, yapay zekâ tarafından üretilen sahte haberlerin olumsuz etkisine ilişkin endişelere yol açmıştır (Fusco, 2022, s. 19). Nitekim son yıllarda yapay zekâ alanında benzeri görülmemiş bir büyüme ve sahte haberlerin yayılmasında eş zamanlı yaşanan artış, dünya genelinde hukuk sistemleri için önemli zorluklar yaratmıştır (Pantserev, 2020).

Genel bir ifadeyle, kurumsal sorumluluğa sahip medya kuruluşlarının formatı ve içeriği ile benzerlik gösteren bir biçimde, dünyayla ilgili olduğu belirtilen yanlış iddiaların sunulması (Levy 2017, s. 20) olarak tanımlanan sahte haberlerin oluşturulmasında iki motivasyon etkilidir. Birincisi parasal motivasyondur. Bu motivasyon sosyal medyada viral olan haberler aracılığı ile reklam geliri elde edilmesini ifade eder. İkinci motivasyon ise ideolojiktir. Bu motivasyona göre sahte haber üreticileri, genellikle karşıt görüşleri itibarsızlaştırarak, destekledikleri fikirleri, kişileri veya grupları tanıtmak amacıyla anlatılar üretir (Allcott ve Gentzkow, 2017, s. 217). Sahte haberlerin oluşturulmasındaki ilk motivasyon maddi kazancın ön planda olduğu bir faaliyet olarak toplumsal sonuçlar açısından kısmen daha az olumsuzluk doğururken; ideolojik temele dayanan ikinci motivasyon, toplumu yanıltıcı içerikler açısından bir dezenformasyon aracına dönüşebilmektedir.

Literatürde sahte haber, kurumsal yapıya sahip haber kuruluşlarının gerçek haber içerikleri taklit edilerek, yanlış veya yanıltıcı içerik olarak sunulmasını ifade etmesi bakımından bir dezenformasyon türü olarak değerlendirilir. Sahte haber üretimi, bir haberin doğruluğu ve güvenirliğinin sağlanmasında gerekli olan habercilik kurallarından ve süreçlerinden yoksun olarak gerçekleştirilir (Gelfert, 2018, s. 103; Lazer vd., 2018, s. 1094). Dolayısıyla sahte haberler, biçimsel olarak kurumsal yapıya sahip haber medyasının haberleri ile benzerlik gösteren ancak kurumsal süreçler ve amaçlar bakımından bu kapsamda yer almayan gerçek dışı ve yanıltıcı bilgiler olarak dezenformasyona hizmet eder (Guess ve Lyons, 2020, ss. 10-11). Kurumsal süreçlerden yoksun yapısı, sahte haber üretimini kolaylaştırıp hızlandırırken, dezenformasyon sürecinin bir parçası olması sebebiyle tespit edilmesini de zorlaştırmaktadır.

Yapay zekâ teknolojileri ile üretilen sahte haber içeriklerinin hızla yayılmaya başlaması kişisel ve kamusal iletişimin çeşitli alanlarını giderek daha fazla etkilemektedir. Öte yandan bu teknolojilerin potansiyel olarak medyaya olan güveni zedelemesi ve kamusal bilgiye yönelik tehdidi artırması son derece mümkündür (Godulla vd., 2021, s. 78). Dolayısıyla sahte haber ve dezenformasyon bağlamında yapay zekâya dayalı haber üretimlerinin iletişim çalışmaları açısından önemi günden güne artmaktadır. Bu alanda sahte haberlerle ilgili akademik çalışmalar yakın dönemde ortaya çıkan bir gelişme olup, konuya ilişkin ilk literatür 2017 yılında

ortaya çıkmıştır (Nyilasy, 2019, s. 337). Sahte haber üretiminde ve tespitinde yapay zekâ teknolojilerinin kullanılması yeni bir araştırma alanı olarak akademik çevrede daha fazla çalışmaya konu olmaktadır. Mevcut literatürün incelenmesi, konunun bilişim teknolojileri, sahte haber üretiminde yapay zekânın kullanımının hukuki etkileri ve deepfake bağlamında sahte haberlerin tespiti ile ilişkili olarak ele alındığını ortaya koymaktadır (Almasi ve Schiønning, 2023; Fusco, 2022; Godulla vd., 2021). Yapay zekâ ile üretilen sahte haberlerin özellikleri ve yayınlandıkları mecralara ilişkin çalışmalar kısıtlıdır. Başta sosyal medya olmak üzere dijital platformlarda yaygın olarak kullanılan sahte haberlerin olumsuz etkilerine karşı toplumda dijital medya okuryazarlığı konusunda farkındalık oluşturmak giderek önem kazanmaktadır. Bu kapsamda çevrimiçi doğrulama platformları sahte haberlerin tespitinde dijital medya okurayazarlığına yönelik önemli katkılar sunmaktadır.

Çevrimiçi doğrulama yapmaya başlayan ilk site, web tarihinin erken dönemlerinde, 1995 yılında gazetecilik geçmişi olmayan bir çift tarafından kurulan Snopes.com sitesidir. Profesyonel gazeteciler tarafından kurulan ilk doğrulama sitesi 2003 yılında kurulan FactCheck.org sitesidir (Graves, 2016: 28-29). Özellikle çevrimiçi ortamlar ve sosyal medyanın yükselişi sonrası sayıları giderek artan doğrulama platformlarının ülkemizde ilki ise 2009 yılında kurulan yalansavar.org isimli internet sitesidir. Günümüzde en etkin doğrulama platformlarından biri olan teyit.org ise 2016 yılında faaliyete geçmiştir. Teyit.org platformu 2015 yılında kurulan dogrulukpayi.com sitesi ile beraber Türkiye'den Uluslararası Doğruluk Kontrolü Ağına (IFCN) üyeliği bulunan ikinci platform olma özelliğine sahiptir. Doğrulama platformları son yıllarda özellikle sosyal medyada artan sahte haberlerin tespiti ve yayıldıkları mecraların belirlenmesinde önemli bir işleve sahiptir. Sahte haberlerin oluşturulmasında yapay zekâ teknolojilerinin de kullanılması, sosyal medya gibi haber akışının çok sayıda yapıldığı ve son derece hızlı paylaşıldığı mecralarda haberlerin doğruluğu noktasındaki belirsizliğin artmasına neden olmuştur. Dolayısıyla doğrulama platformları, yapay zekâ ile üretilen sahte haberlerin tespitine yönelik dijital medya okuryazarlığı açısından da önemli bir katkı sağlamaktadır. Bu bağlamda bilimsel çalışmalar için geniş veri kümeleri sunan doğrulama platformları bilimsel araştırmalara konu olmaktadır (Aydın, 2020; Çömlekçi, 2019; Erkan ve Ayhan, 2018). Literatürde çevrimiçi doğrulama platfomları üzerine yapılan çalışmalar, dezenformasyonla mücadele kapsamında sahte haber tespitine yönelik olup, yapay zekâ destekli haberleri kapsamamaktadır. Diğer taraftan litaretürde, iletişim çalışmaları perspektifinden yapay zekâ ile üretilmiş sahte haberlerin özellikleri ve kullanıldıkları mecralara yönelik doğrulama platformlarını içeren çalışmalarla ilgili önemli bir boşluk olduğu görülmektedir.

İletişim çalışmaları perspektifinden bakıldığında sahte haberlerin yanıltıcı yönleri, içerdiği konular, sunduğu bilgi türleri ve sosyal medyadaki yansımaları incelenmesi gereken bir konu olarak görünmektedir. Bu çalışma, iletişim çalışmaları perspektifinden sahte haberlerin üretimi, dağıtımı ve tespitinde yapay zekânın rolüne ve sosyal medyadaki yansımasına odaklanarak, yapay zekâ ile üretilen sahte haberleri doğrulama platformları üzerinden incelemektedir. Çalışmanın amacı, yapay zekâ ile üretilmiş sahte haberlerin yanıltıcı yönleri, odaklandıkları konular, yayıldığı sosyal medya platformları ve içerdikleri bilgi türleri bakımından ortaya çıkarılmasıdır. Çalışmanın amacı doğrultusunda, teyit.org doğrulama platformunda 1 Ocak 2025 – 31 Mart 2025 tarihleri arasında yayınlanmış, yapay zekâ ile üretilen 24 sahte haberin içerik analizi yapılmıştır.

Araştırmada şu sorulara yanıt aranmaktadır:

1. Sahte haber içeriklerinin yanıltıcı yönü nedir?
2. Sahte haber içeriklerinin haber kategorilerine göre dağılımı nasıldır?
3. Sahte haber içeriklerinin dolaşıma girmesinde hangi sosyal medya platformları ön plana çıkmaktadır?
4. Sahte haber içeriklerinin bilgi türleri nedir?

Çalışma, yapay zekâ ile üretilen sahte haberlerin özelliklerini analiz ederek, dijital medya okuryazarlığına ilişkin iç görüler sağlamaktadır. Çalışma, sonuçları itibari ile, yapay zekâ aracılığıyla sahte haber üretiminin boyutlarının sosyal medyaya yansımasını ortaya koyarak iletişim çalışmaları perspektifinden literatürün genişlemesine katkı sunmaktadır. Araştırmada elde edilen bulgular, ileride yapılacak çalışmalar için güncel bir veri niteliği taşımaktadır.

Çalışma bu kapsamda dört bölüme ayrılmıştır. İlk bölümde, yapay zekâ tarafından sahte haberlerin üretimi ve tespiti için teorik çerçeve çizilmekte ve sahte haberlerin sosyal medya üzerindeki zararlı etkilerinden bahsedilmektedir. Çalışmanın ikinci bölümü, literatürde yapay zekâ ve sahte haber konusu üzerine mevcut çalışmalara kapsamlı bir genel bakış sunmaktadır. Üçüncü bölümü metodolojiyi ve bulguları sunmaktadır. Son bölümde, araştırma bulguları değerlendirilmekte ve sonuca ilişkin açıklamalar yapılmaktadır.

YAPAY ZEKÂ TEKNOLOJİLERİNİN SAHTE HABER ÜRETİMİ VE TESPİTİNDEKİ ROLÜ

İlk olarak 1950'lerde ve 1960'larda geliştirilen yapay zekânın ilk tanımları insan zekâsına atıfla yapılmıştır. Minsky (1968) yapay zekâyı, makinelerin, insanlar tarafından yapıldığında zekâ gerektirecek şeyleri yapmasını sağlama bilimi olarak tanımlamıştır. Kaplan ve Haenlein (2019, s. 17) teknolojik ilerlemenin geldiği düzeyi dikkate alarak yapay zekâ için daha özellikli bir tanımlama yapar. Buna göre yapay zekâ, bir sistemin dış verileri doğru bir şekilde yorumlama, bu verilerden öğrenme ve bu öğrenmeleri esnek adaptasyon yoluyla belirli hedeflere ve görevlere ulaşmak için kullanma yeteneğidir. Daha basit bir ifadeyle yapay zekâ, öğrenme ve karar alma yeteneğine sahip akıllı makineler olarak tanımlanabilir. Yapay zekâ ile bilgisayar sistemlerini ayıran temel özellik, insan zekâsına benzer biçimde ancak sadece bilgileri özümseyip işlemekle sınırlı kalmayıp, yeni bilgilere dayanarak öğrenebilmeleri ve kararlarını güncelleyebilmeleridir. Dolayısıyla yapay zekâ sistemleri insan beyninin işlevlerini taklit etmek üzere tasarlandıkları için "yapay zekâ" kavramı ile açıklanmaktadır (Vincent, 2021, s. 427). Tanımlara göre yapay zekâ, insan zihninin bilişsel yeteneklerini taklit etmek ve bazı durumlarda onları aşmak amacıyla geliştirilmiş bir teknolojiyi ifade eder.

Yapay zekâ, öğrenme yeteneğine sahip bir teknolojidir. Yapay zekânın bilgi edinme kapasitesi, genetik bir algoritma kullanan simülasyonların yürütülmesiyle kolaylaştırılarak önemli miktarda eğitim verisi sağlanmasına bağlıdır. Yapay zekâ tarafından üretilen sonuçların kalitesi ve doğruluğu, öğrenme süreci sırasında sisteme girilen verilerin niceliği ve niteliği ile yakından ilişkilidir. Bu öğrenme süreci zaman alıcı bir durum olsa da bilgisayar teknolojisi ilerledikçe, kullanılan algoritmayı veya formülleri ayarlayarak onu optimize etme olasılığı her zaman vardır (Nazar ve Bustam, 2020). Yapay zekâ, büyük miktardaki veriyi hızlı şekilde analiz etmenin yanı sıra, oyunun kurallarını değiştiren ve "hem kendi kendine öğrenme yeteneklerini hem de karar kalitesini iyileştirme yarışını" kolaylaştırabilen bir araç olarak görülmektedir (Akhtar vd., 2023, s. 637). Bu açıdan insanların talimatlarına göre hareket etmek yerine, öğrenme ve karar kalitesini iyileştirme yetenekleri, yapay zekâyı teknolojideki bir sonraki sınır konumuna taşımaktadır (Vincent, 2021, s. 425).

Günümüzde çok sayıda alanda kullanılan yapay zekâ teknolojileri özellikle iletişim ve haber medyasında da giderek yaygınlaşmaktadır. Yapay zekâ haber medyasında, arama gibi günlük işlevlerden, metin, fotoğraf veya video oluşturmak amacıyla derin öğrenmeden yararlanan karmaşık algoritmalara kadar uzanan önemli bir teknoloji olarak kullanılmaktadır. Medya ve iletişim alanında hızla gelişen ve yaygınlaşan bir teknoloji olarak yapay zekâ, makine öğrenimi, doğal dil işleme ve bilgisayar görüşü gibi teknikleri kullanarak insan zekâsını ve karar alma süreçlerini taklit etmek için makinelerin kullanılmasını içerir. Bu teknikler, makinelerin dil çevirisi, görüntü ve konuşma tanıma ve öngörücü analiz gibi görevleri gerçekleştirmesini sağlar. Medya sektöründeki iş modelleri ve diğer gelişmelerle beraber hızla gelişen bir teknoloji olarak yapay zekânın haberciliğin nasıl yapıldığı ve haberlerin ne şekilde tüketildiğine yönelik derin bir etki potansiyeli olduğu açıktır. Yapay zekânın bu özellikleri, medya ve iletişim alanında sahte haber üretimi ve haber içeriklerinin manipülasyonu gibi bazı sorunları da gündeme getirmiştir. Yapay zekâ teknolojisinde yaşanan gelişmeler neticesinde, sahte haberlerin büyük ölçekte ve doğruluk oranının daha yüksek şekilde üretilmesinin kolaylaştığı, buna karşın tespit edilmesinin ve karşı konulmasının zorlaştığı dile getirilmiştir (Beckett 2019, s. 12; Fusco, 2022, s. 21).

Sahte haber, gerçek dünyadaki olayları, geleneksel medya haberciliğinin kurallarını taklit ederek, önemli ölçüde yanlış olduğu bilinen ve geniş dolaşıma sokularak hedef kitlesini aldatmak amacıyla iletilen haberlerdir (Rini,

2017, s. 45). Nyilasy (2019) sahte haberlerin üç önemli bileşeni olduğunu belirtir. İlk olarak, sahte haberler nesnel referanslar üzerinden değerlendirildiğinde yanlış bilgiyi temsil eder. İkincisi, sahte haberler gerçek bir içerik gibi sunularak taklit unsuru taşır. Üçüncüsü ise sahte haberler kötü niyetli bir gönderici tarafından dağıtıma çıkarılır ve gizli bir amaca hizmet eden aldatma niyeti taşır. Sahte haberleri oluşturan üç unsur günümüzde yapay zekâ teknolojilerinin kullanımı ile birleşerek farklı bir boyut kazanmaktadır. Sahte haber üretiminin geleneksel yöntemlerden yeni nesil teknolojilere uzanan üretim yöntemleri yapay zekâ ile yeni bir olguya dönüşmüştür. Yanlış ve yanıltıcı bilginin oluşturulması, taklit içerikler ve kamuoyunu aldatmaya yönelik sahte haberler için çeşitli yapay zekâ teknolojileri kullanılmaktadır. Bu teknolojilerden birisi "deepfake" teknolojisidir.

Deepfake Teknolojisi ve Medya Manipülasyonu

Derin öğrenme (deep learning) ve sahte (fake) kelimelerinin birleşiminden oluşan (Chadha vd. 2021, s. 558) deepfake terimi, 2017 yılında Reddit platformunda "deepfakes" isimli bir kullanıcı tarafından ortaya atılmıştır. Bu kullanıcı, ünlü kişileri yetişkinlere yönelik videolara yerleştirerek paylaşmıştır. Paylaşımlarında Emma Watson, Katy Pery, Barack Obama ve Donald Trump gibi sanat ve politika dünyasından ünlü kişileri kullanmış ve bu ünlülerin yüzlerini, izinleri olmadan başka kişilerin yüzlerine aktaran videolar yayınlamıştır. Yapay zekâ ve derin öğrenmenin yeteneklerini sergileyen deepfake'lerin dikkat çekici ilk örneklerinden biri, 2017 tarihli "Obama'yı Sentezlemek" idi. Bu çalışmada, ilgi çekici bir etki yaratmak için mevcut ses kayıtlarından yararlanılarak gelişmiş dudak senkronizasyonu teknolojisi kullanılmıştır (Kietzmann vd. 2020, s. 136; Kirchengast, 2020, s. 310).

Deepfake teknolojisi, yapay zekâ alanındaki gelişmelerin bir sonucu olarak, başlangıçta eğlence amaçlı kullanılan ancak zamanla psikolojik güvenliği ciddi şekilde tehdit eden yeni bir olgudur. Kötü amaçlı kullanım potansiyeli giderek artan yapay zekâ destekli deepfake'ler, görünüm, konuşma ve davranış açısından gerçeğine şaşırtıcı derecede benzeyen klonların yaratılmasına olanak tanıyan yeni bir teknolojik gelişmişlik çağını başlatmıştır. Deepfake, bugün alanında tanınmış kişilerin klonunu oluşturup sözlerini manipüle etme becerisine sahip bir teknoloji olarak sahte haber üretiminde kullanılmaktadır (Pantserev, 2020, ss. 37-38). Deepfake teknolojisinin gerçeğine çok yakın oranda benzer içerikleri üreten bir teknoloji olması, bu teknolojinin kendine özgü özelliklerine işaret eder.

Deepfake görsellerin oluşturulmasındaki temel teknolojik unsur derin öğrenmedir (Kietzmann vd. 2020, s. 138). Deepfake teknolojisi video oluşturmak için Google'ın görsel arama özelliğinden ve sosyal medya verilerinden yararlanır. Sahte videolardaki yüzleri gerçeğine en yakın şekilde oluşturmak için topladığı verileri girer. Bir kişinin yüz ifadelerini, sesini ve tonlamalarını nasıl taklit edeceğini öğrenmek için büyük veri kümelerini işleyen sinir ağlarını kullanır. Deepfake algoritması, iki farklı kişinin videoları kullanılarak eğitilir ve böylece yüz özelliklerinin değiştirilmesine olanak tanır. Daha sonra videodaki kişinin yüzünü öznenin yüzüyle değiştirmek için görüntü eşleştirmesi yapar. Deepfake ile sahte videoları oluşturan teknoloji, videoları daha gerçekçi bir hale getirmek için bir kişinin mimiklerini, jestlerini ve sesini taklit etmeye devam ederek videoyu iyileştirmeye çalışır. Deepfake teknolojisi ilk makine öğrenme sürecinin ardından insan denetimine ihtiyaç duymaksızın otonom şekilde videoyu iyileştirmeye devam eder. Sonuç olarak, söz konusu kişinin yüzü değiştirilmiş olarak aynı olayın yeni bir videosu üretilir. Sahte videonun üretilmesinin yanı sıra deepfake teknolojisi ile aslında kişinin söylemediği şeyler de söylenmiş gibi videoya eklenebilir. Deepfake videoları insanların gerçek görüntüleri üzerinden oluşturulduğu için çıplak gözle bakarak gerçeğinden ayırt edilmesi son derece zordur (Chadha vd., 2021, s. 558; Godulla vd., 2021, s. 74; Maras ve Alexandrou, 2018). Derin öğrenme yeteneği ile çalışan deepfake teknolojisi farklı şekillerde gerçeğe benzer içerikler oluşturulmasına imkân tanır. Dolayısıyla deepfake'ler farklı türlerde oluşturulabilen sahte içeriklerdir. Kietzmann vd. (2020) deepfake türlerini dört ana kategoride sınıflandırmaktadır.

Tablo 1. Deepfake Türleri

Tür	Tanım
Fotoğraf Deepfake'ler	**Yüz ve vücut değiştirme:** Bir yüzde değişiklik yapmak, yüzü (veya vücudu) başka birinin yüzü (veya vücudu) ile değiştirmek veya harmanlamak.
Ses Deepfake'ler	**Ses değiştirme:** Birinin sesini değiştirmek veya başkasının sesini taklit etmek.
	Metinden konuşmaya: Yeni metin yazarak bir kayıttaki sesi değiştirme
Video Deepfake'ler	**Yüz değiştirme:** Bir videodaki birinin yüzünü başka birinin yüzüyle değiştirme.
	Yüz şekillendirme: Bir yüzün kusursuz bir geçişle başka bir yüze dönüşmesi.
	Tüm vücut kuklacılığı: Bir kişinin vücudundaki hareketin bir başkasının vücuduna aktarılması.
Ses ve Video Deepfake'ler	**Dudak eşzamanlama:** Konuşan bir kafanın yer aldığı videoda ağız hareketlerini ve konuşulan kelimeleri değiştirmek.

Kaynak: Kietzmann vd. 2020, s. 142.

Günümüzde teknolojinin geldiği nokta itibari ile sahte haberlerde insanların doğru olduğuna inanabileceği ölçüde ikna edici fotoğraflar ve videolar dijital ortamlarda sıkça yayınlanmaktadır. Fotoğraflar ve videolar ile sunulan bu görsel içeriklerin hazırlanmasını yapay zekâ teknolojileri önemli oranda kolaylaştırmaktadır. Deepfake ile hazırlanan videolar başta sosyal medya olmak üzere diğer dijital ortamlarda yoğun bir şekilde paylaşılmaktadır. Bu videoların tespit edilmesi veya filtrelenmesi ise teknik zorluklar içermektedir (Nazar ve Bustam, 2020). Dijitalleşmenin son derece yaygınlaştığı günümüzde deepfake ile oluşturulan içerikler nedeniyle haber medyası açısından gerçek bilgi ile sahte bilgiyi ayırt edemeyeceğimiz bir bilgi dünyasına doğru ilerlemekteyiz (Fallis, 2021, s. 624). Bu durum aynı zamanda bazı endişeleri de beraberinde getirmektedir. Bu endişelerin başında dezenformasyona yönelik manipülasyon faaliyetleri gelmektedir. Özellikle dijital medya üzerinden yapay zekâ teknolojilerine dayalı sahte haberler dezenformasyon yöntemleri açısından sıkça kullanılmaktadır.

Metin Tabanlı Sahte Haber Üretimi

Sahte haber üretiminde video içeriklerin yanı sıra metin tabanlı içerikler de oluşturulabilmektedir. Son yıllarda, özellikle doğal dil işlemedeki hızlı gelişmeler, çeşitli dil görevlerinde önemli ilerlemelere yol açmıştır. Bu anlamda Open AI tarafından geliştirilmiş GPT aracı gündeme gelmektedir. GPT-3, belge sınıflandırması, metin tamamlama, dil çevirisi ve metin özetleme gibi görevlerde oldukça başarılı bir performans göstermiştir. Yapay zekâlı büyük bir dil modeli olan GPT esasen, sahte içerik üretmek amacıyla geliştirilmemiş olsa da yaratıcıları, bu amaçla kullanılabileceğinden endişe duymaktadır (Almasi ve Schiønning, 2023, s. 54; Özsalih, 2023, s. 539). Nitekim yapay zekâ teknolojilerinin beceri ve başarı potansiyeli, bu teknolojilerin kötü amaçlarla kullanılabileceğine yönelik endişeyi tetiklemektedir.

GPT-3, insanlar tarafından yazılan metinlerle yüksek oranda benzerlik gösteren, ikna edici içerikler ve veri kümeleri üretme yeteneğine sahip derin öğrenmeyi kullanan üçüncü nesil, özbağlanımlı bir dil modelidir. Basitçe ifade etmek gerekirse, istem adı verilen bir kaynak girdisinden başlayarak kelime, kod veya diğer veri dizileri üretmek için tasarlanmış bir hesaplama sistemidir. Örneğin, kelime dizilerini istatistiksel olarak tahmin etmek için makine çevirisinde kullanılır. Dil modeli, öncelikle İngilizce olmak üzere Wikipedia ve birçok web sitesi gibi diğer dillerde de etiketlenmemiş bir metin veri kümesi üzerinde eğitilir. Bu istatistiksel modellerin eğitimi, ilgili sonuçların üretilmesini sağlamak için önemli veri kümelerinin kullanımını gerektirir. GPT-3, talep edilen yüksek kaliteli metni, otonom olarak üretme kapasitesine sahiptir. Örneğin The Guardian, GPT-3 tarafından yazılan ve ilgi gören bir makale yayınladığında bunun gazeteciliğin kötü bir örneği olduğu savunulmuştur (Floridi ve Chiriatti, 2020, s. 684).

Benzer bir endişeyi belirten Biswas (2023), yapay zekâ sohbet robotu olan ChatGPT'nin bilgi savaşları için dezenformasyon aracı olarak kullanılabileceğine dikkat

çeker. GPT dil modelleri ailesinin üyesi olan ChatGPT'nin (Törnberg, 2023) propaganda oluşturma amacıyla bir siyasi figürü veya politikayı olumlu biçimde yansıtan içerikler üretmek ve sosyal medya platformlarında yaymak üzere eğitilebileceğini dile getirir. Yine ChatGPT'nin, gerçekçi deepfake videoları üretecek şekilde programlanabilme kapasitesine sahip olduğunu ve bu şekilde üretilen videoların, siyasi figürleri veya önemli kişileri taklit etme, yanlış bilgi yayma veya dezenformasyon amacıyla kullanılabileceğine işaret eder. Almasi ve Schiønning (2023), yaptıkları çalışmada GPT-3'ün gazeteciler tarafından yazılan gerçek haber içeriklerinden neredeyse ayırt edilemeyen sahte haber makaleleri üretecek şekilde ince ayar yapma potansiyelinin İngilizce dışındaki dillerde de olabildiğini göstermişlerdir.

Yapay Zekâ ve Sahte Haber Tespiti

Gerçek zamanlı önemli miktarda bilgi sağlama kapasiteleri nedeniyle haber tüketimi için birincil ortam haline gelen sosyal medya platformları, sahte haberlerin yayılması, dezenformasyon ve siyasi manipülasyon gibi riskler doğurmaktadır. Sahte haberlerle mücadelede kullanılan manuel gerçek kontrolü ve kaynak doğrulaması gibi geleneksel yöntemler, verilerin çokluğu ve yanlış bilginin viral olma hızı nedeniyle yayılmasını engellemede yetersiz kalmaktadır (Fusco, 2022, s. 22). Yapay zekânın büyük miktarda veriyi gerçek zamanlı olarak işleme kapasitesi, yanlış bilginin yayılmasını insan çabalarından daha verimli bir şekilde tespit etmek ve önlemek için yeni imkânlar sunmaktadır. Yapay zekâ destekli sistemler, sahte haberleri tespit etmek için makine öğrenimi, doğal dil işleme (NLP) ve derin öğrenme algoritmalarını kapsayan bir dizi teknik kullanır. Bu teknolojiler, yapay zekânın çok miktarda çevrim içi içeriği otonom bir şekilde incelemesini, yanlış bilgiye işaret eden kalıpları belirlemesini ve içeriği var olan veri kümelerine göre doğru veya yanlış olarak kategorize etmesini sağlar (Swaroop, 2024, ss. 78-79).

Sosyal medyada metin tabanlı sahte haberlerin tespitinde Doğal Dil tabanlı teknikler yoğun olarak kullanılırken, metin sınıflandırması için Derin Öğrenme ve Makine Öğrenmesinden yararlanılır (Narang ve Sharma, 2021, s. 484). Makine öğrenimi yöntemleri sahte haberlerin tespit edilmesini kolaylaştıran doğru, verimli ve pratik modeller üretmiştir. Doğal Dil İşleme, sahte haberleri gerçek haberlerden ayıran dilsel kalıpları belirlemek için algoritmalar ve makine öğrenimi tekniklerini kullanır. Bu yaklaşım, yanlış bilginin yayılmasıyla ölçeklenebilir ve etkili şekilde mücadelede büyük önem taşır (Paredes, 2023). Böylece insan gerçek denetçilerinin çabalarını tamamlayarak sahte haberlerin tespitinde insan faktörüne destek sağlayan teknolojik bir araç olarak işlev görür. Bu sayede daha hızlı ve etkili sonuç alınmasında önemli bir rol oynar.

Yapay zekânın gerçek ve sahte haber içerikleri arasında ayrım yapma kapasitesi, hacimli veri kümelerini analiz etme becerisine dayanmaktadır. Bu teknolojiler, gerçek haberle sahte haber arasındaki ayrımı kolaylaştırmak için haber başlıklarındaki ve içeriklerdeki duygusal ton, kelime kullanımı ve yapısal özellikler gibi unsurları analiz eder. Doğal dil işleme, makine öğrenimi ve derin öğrenme gibi tekniklerin kullanılmasının sahte haberlerin belirlenmesinde yüksek doğruluk oranlarına ulaştığı gösterilmiştir (Deverajan, 2024).

Shu vd. (2017) yapay zekâ tabanlı sistemler aracılığıyla sahte haberlerin dilsel ve görsel tabanlı özelliklerinin işlenerek yanlış bilgilerin belirlenebileceğini öne sürmüştür. Dil tabanlı yaklaşıma göre sahte haberlerde, nesnel bilgiler yerine ticari veya siyasi fayda elde etme amacıyla üretildiği için genellikle öznel görüş bildiren ifadeler ve düşmanca bir dil kullanılır. Bu nedenle kafa karışıklığı oluşturan veya "tık tuzağı" olarak tasarlanan içerikler hazırlanır. Bu tür haberlerin tespitinde farklı dil kullanımlarını ve sansasyonel başlıkları yakalayan dil özelliklerini kullanmak mantıklıdır. Görsel tabanlı yaklaşıma göre, sahte haberler insanların bireysel zaaflarını istismar ederek, onların öfkesini veya diğer duygusal tepkilerini kışkırtmak amacıyla genellikle sansasyonel veya sahte görüntülere güvenir. Bu yaklaşımda görsel ipuçlarının sahte haberler için önemli bir manipülasyon aracı olduğu varsayımından hareketle görsel tabanlı özellikler sahte haberlerdeki farklı özellikleri yakalamak için görsel öğelerden ayrıştırılır.

Paschen (2020) literatürde sahte haber tespitinin kaynak, bağlam ve içerik tabanlı olmak üzere üç farklı kategoriye ayrıldığını belirtir. Kaynak tabanlı sahte haber tespitinde, bilgiyi paylaşan kaynağın güvenirliği değerlendirilerek

sahteciliğin oranı tespit edilmeye çalışılır. Makine öğrenimi aracılığıyla haber kaynaklarının gerçekliği ve önyargısı tahmin edilir. Bağlam tabanlı sahte haber tespitinde, haberin çevrimiçi ortamlarda yayılma biçimine ve mesajın alıcısına odaklanılır. Bağlam tabanlı sahte haberlerde yanlış haberlerin gerçek haberlere göre daha hızlı ve daha uzağa ulaştığı tespit edilmiştir (Vosoughi vd., 2018). İçerik tabanlı sahte haberlerin tespitinde ise iletişim sürecindeki üçüncü unsur olan mesaja odaklanılarak şüpheli içerik metinsel veya görsel analize tabi tutulur. Analizde şüpheli haber içeriğindeki gerçekler ile bilinen gerçekler karşılaştırılır ve tutarsızlıklar işaretlenir. Bu yaklaşımın temelindeki varsayım, bir gerçeğin doğruluğunun göstergesinin, o gerçeğin belirtilme sıklığı ile doğru orantılı olmasıdır. Potthast vd. (2017) bu yaklaşımı sorunlu bulur ve inandırıcılığının ve güvenilirliğinin sorgulanabileceğini dile getirir.

Sosyal Medyada Sahte Haberlerin Olumsuz Etkileri

Dijital teknolojilerin yaygın olarak kullanıldığı günümüzde sahte haber ve dezenformasyonla ilgili sorunlar, sosyal medya platformlarının yükselişi ile giderek artmaktadır. Bu bağlamda sahte haber olgusu yeni bir gelişme değildir. Ancak, bilişsel önyargıların ve sezgilerin, hedef kitleye özgü manipülasyonunu kolaylaştıran sosyal medya ile entegre edildiğinde, güçlü bir araç haline gelmektedir. Çevrim içi sosyal medyanın ortaya çıkışı, tüketicilere tasarım gereği yanıltıcı olarak üretilen, haber benzeri içerikler sunan yeni bir sistem doğurmuştur. Bilişsel önyargılar bağlamında "steroid gibi çalışmak" olarak adlandırılan bu olgu, sosyal medyanın bilginin hızlı dağıtımı üzerindeki zararlı etkisini örneklemektedir. Çevrim içi ve çevrim dışı haber kaynakları arasındaki geçirgenliğin artması, giderek daha fazla haber kisvesi altında kamuoyuna yayılan, ancak asıl amacı kendi sürekli üretimini sağlamak için bilişsel önyargılarımızı beslemek olan sahte içeriklerle toplum karşı karşıya kalmaktadır (Akhtar vd., 2023, s. 633; Gelfert, 2018, ss. 108-109).

Sahte haberler yaygın bir çekiciliğe sahip olmaları nedeniyle sosyal medyada gerçek haberlerden daha hızlı yayılmaktadır. Sahte haber trafiğinin büyük bir kısmının kötü niyetli aktörler tarafından geliştirilen algoritmalar olan "botlar" tarafından oluşturulması bu yayılmada etkilidir (Nyilasy, 2019, s. 337). Sahte haber üretimine yapay zekâ teknolojilerinin dâhil olması bu sorunla mücadelede yeni zorluklar ortaya çıkarmıştır. Özellikle sosyal medya üzerinden kamuoyunun manipüle edilmesine yönelik sahte içerikler, kamusal yaşam için önemli bir tehdit unsuru haline gelmiştir. Kötü niyetli aktörler, üretimini kolaylaştırması ve yayılmasını hızlandırması bakımından yapay zekâ ile üretilmiş haberlere ilgi duymaktadır. Yapay zekâ ile üretilen haberlerin gerçeğine benzerliği nedeniyle meşru görünme yeteneği, sahte haber ile gerçek haber arasındaki sınırı belirsizleştirmektedir (Fusco, 2022, ss. 21-22). Buna sosyal medyanın içerikleri hızlı şekilde yayma yeteneğinin eklenmesi ile sahte haberler, kamuoyunun algısını yönetme ve toplumu manipüle etme amacıyla dezenformasyona sebep olmaktadır.

Shu vd. (2017) sosyal medyada sahte haberlerin yayılması ile ilgili iki önemli özelliği vurgular. Bunlar sosyal medyada propaganda için kötü amaçlı hesaplar ve yankı odası etkisidir. Buna göre sosyal medyadaki birçok kullanıcı gerçek ve meşru bir kimliğe sahip olsa da bunun dışında çok sayıda kötü niyetli ve insanlar tarafından yönetilmeyen sosyal bot hesapları vardır. Sosyal botlar, insan davranışını taklit eden ve çevrim içi platformlarda etkileşime girerek genellikle yanlış bilgi yayma veya kamuoyunu manipüle etme gibi potansiyel olarak kötü niyetli eylemlerde bulunan bilgisayar algoritmalarıdır. Sosyal botlar, yapay zekâ ve doğal dil işleme teknolojilerinin geliştirilmesi ile çevrim içi iletişim ve siber güvenlik alanlarında yeni zorluklar oluşturmaktadır (Ferrera vd., 2016, s. 96). Sosyal medyanın yükselişiyle birlikte sosyal botların iyi niyetli ve kötü niyetli kullanımları artmıştır. Sosyal botların iyi niyetli kullanımı, haber toplamak, müşteri hizmetlerini otomatikleştirmeye yardımcı olmak gibi bir dizi faydalı eylemleri içerir. Buna karşın kötü niyetli kullanımlarda çevrim içi topluluklara zarar verme potansiyelleri ve kamuoyunu etkileme yetenekleri ön plana çıkmaktadır. Çevrim içi ekosistemlere müdahale eden kötü niyetli sosyal botların gerçekleştirdiği eylemlerin geniş yelpazesi, sosyal medya platformlarında sahte haberlerin yayılması ve fikirlerin manipüle edilmesini amaçlayan bilgi savaşlarında etkin bir rol oynamaktadır (Cresci, 2020, s. 73; Ferrera, 2023).

Sosyal medyada sahte haberlerin yayılmasındaki ikinci önemli özellik yankı odası etkisidir. Sosyal medyanın ortaya çıkışı, kullanıcılar için içerik oluşturma ve tüketme konusunda yeni bir paradigma oluşturmuştur. Bu yeni paradigmada içerik arama ve tüketme süreci, gazetecilerin kolaylaştırdığı gibi aracılı bir formdan daha aracısız bir forma geçiş yapmıştır. Haber akışlarının bireylerin ana sayfalarında ve sosyal medya akışlarında sunulma biçimi, tüketicilerin yalnızca belirli haber türlerine maruz kalmasına yol açmıştır. Bu olgu, sahte haberlerin ortadan kaldırılmasıyla ilişkili zorlukların artmasına neden olmuştur. Dolayısıyla, sosyal medya kullanıcıları genellikle benzer bakış açılarına sahip bireylerden oluşan gruplarda bir araya geldiği için fikirlerin giderek daha fazla kutuplaştığı bir ortam meydana gelmektedir. Yankı odası etkisi olarak bilinen bu olgu, grup üyeleri arasında perspektifin daralmasına ve var olan inançların güçlenmesine yol açmaktadır (Shu vd. 2017). Yapay zekâ ile oluşturulan sahte haberler yankı odası etkisi ile doğrulama yanlılığını istismar ederek haber içeriğinin etkisini artırabilmektedir. Aynı zamanda yapay zekâ ile oluşturulan sahte haberlerin büyük hacmi, geleneksel tespit yöntemlerini aşarak insan moderatörlerin bununla baş edebilmesini zorlaştırmaktadır (Fusco, 2022, s. 22).

LİTERATÜR TARAMASI

Literatürde yapay zekâ ile ilgili çalışmaların bir kısmı bu teknolojinin haber üretimine odaklanmıştır. Yapay zekâ teknolojilerinin sahte haber üretim sürecindeki rolü ve etkisi bir araştırma alanı olarak ilgi görmüştür. Almasi ve Schiønning (2023), yapay zekâ dayalı bir teknoloji olarak GPT-3'ün haber üretim sürecindeki rolünü incelemişlerdir. Sonuçlar, GPT-3 teknolojisinin gerçek haber içeriklerinden ayırt edilemeyecek düzeyde haber üretme potansiyelinin İngilizce dışındaki dillerde de başarılı olabildiğini göstermiştir. Fusco (2022) yanlış bilginin sosyal medya aracılığıyla hızla yayılmasının getirdiği zorlukların yanı sıra, yapay zekâ ile üretilen sahte haberlerin Pakistan ve Suudi Arabistan'daki hukuki yönüne odaklanarak bu haberlerin suç yönlerini araştırmıştır. Araştırmada Pakistan ve Suudi Arabistan'da sahte haberle mücadelede ayrıntılı yasal tanımlara ve kamu kurumları ile sivil toplum kurumlarının iş birliğini içerecek güçlü uygulama mekanizmalarına ihtiyaç olduğunu vurgulamıştır.

Yapay zekâ teknolojileri sadece sahte haber üretiminde değil, bu haberlerin tespitinde de önemli bir araştırma konusu olmuştur (Narang ve Sharma, 2021; Paredes, 2023; Swaroop, 2024; Raza vd., 2025). Nitekim sahte haberle mücadelede yeni bir teknoloji olarak yapay zekânın kullanımı bir araştırma alanı olarak literatürde yerini almıştır. Örneğin Devarajan (2024) derin doğal dil işleme modeliyle yeni bir yapay zekâ destekli sahte haber tespiti modeli önerdiği çalışmada, Buzzface, FakeNewsNet ve Twitter gibi üç veri kümesinde önerilen modelin sahte haber tespitinde %99,72'lik bir ortalama ile doğruluk sağladığını belirlemiştir. Roumeliotis vd. (2025) yakın zamanda yaptığı bir çalışmada, sahte haberlerin sınıflandırılmasında gelişmiş makine öğrenimi modellerinin etkinliğini incelemiştir. Bulgular, ince ayarlı GPT-4 Omni modellerinin %98,6 doğruluk oranı yakaladığını ve bu kritik sorunu ele almak için umut verici olduğunu göstermiştir. Nazar ve Bustam (2020) yapay zekâ ile oluşturulan haberlere odaklanmışlar ve medya tüketicilerinin sahte haberlere yönelik farkındalığını deepfake konusu üzerinden araştırmışlardır. Sonuçlar, deepfake teknolojisi kullanılarak oluşturulan videoların sahte olup olmadığının anlaşılmasının zor olduğunu göstermiştir.

Haber merkezlerinin sahte haberleri kontrol etme çabalarında yapay zekânın rolü, El Gody (2021) tarafından yapılan bir araştırmada ele alınmış ve sahte haber kaynaklarını tespit etmede yapay zekânın sağladığı öngörü teknolojisinin önemli bir rolü olduğunu göstermiştir. Yapay zekâ temeline dayanan bir teknoloji olarak deepfake'ler de bir dizi çalışmaya konu olmuştur. Khodabakhsh vd. (2019) medya tüketicilerinin deepfake'leri tespit etme becerisini araştırırken, Godulla vd., (2021) İngilizce'deki deepfake araştırmalarının sistematik incelemesini yapmış, deepfake araştırmalarının bilgisayar bilimi ve hukuk tarafından yönlendirildiğini, çalışmaların deepfake'lerin tespiti ve düzenlenmesine odaklandığını bulgulamıştır.

Yapay zekâ teknolojileri ile üretilen sahte haberlerin sosyal medyadaki etkinliği de araştırmalara konu olmuştur. Özsalih (2023) bu tür haberlerin sosyal medya gündemini şekillendirmedeki rolüne odaklandığı çalışmada X platformunda Donald Trump'ın tutuklanma anını gösterdiği iddia edilen haberi incelemiştir. Araştırmada

elde edilen sonuçlar, söz konusu sahte haberin X platformunda yayınlanmasının ardından Türkiye'nin önemli haber kuruluşlarının web sitelerinde yayınlandığını doğrulamıştır. Araştırmaya göre bu sonuç aynı zamanda X platformunun, ulusal medya kuruluşlarının gündemini belirlediğini göstermiştir. Ulusoy ve İlhan (2025) ABD seçim kampanyalarında dolaşıma giren deepfake haber içeriklerini, kullanım yoğunluğu, paylaşıldıkları mecra, duygu tonu, içerik türü ve hedef aldığı siyasi aktörler açısından analiz etmiştir. Analiz sonucunda deepfake içeriklerin daha çok X platformunda paylaşıldığını, fotoğraf ve video-ses içeriklerin daha fazla tercih edildiğini ve hedef aldıkları siyasi aktörlerle ilgili içeriklerde çoğunlukla olumsuz duygu tonu oluşturduklarını bulgulamışlardır.

Yapa zekâ konusunda önceki çalışmalara ilişkin literatür, yapay zekânın sahte haber üretiminde kullanılması (Almasi ve Schiønning, 2023), sahte haberlerin tespitindeki rolünün bilişim teknolojileri açısından ele alınması (Devarajan, 2024; Narang ve Sharma, 2021; Paredes, 2023; Swaroop, 2024), sahte haber üretiminde yapay zekâ kullanımının hukuki yönü (Fusco, 2022); deepfake uygulaması özelinde sahte haber tespiti (Godulla vd., 2021; Khodabakhsh vd., 2019; Ulusoy ve İlhan, 2025) çerçevesinde ele alındığını göstermektedir. Yapay zekâ ile üretilen sahte haberlerin özellikleri ve yayınlandığı mecralar üzerine yapılan araştırmalar sınırlıdır. Öte yandan, iletişim çalışmaları açısından bakıldığında, yapay zekâ ile üretilen sahte haberlerin özellikleri ve yayıldığı mecralar hakkında doğrulama platformlarını kapsayan çalışmalarla ilgili belirgin bir boşluk bulunmaktadır.

Doğrulama platformları, özellikle sosyal medya bağlamında son yıllarda belirgin bir artış gösteren bir olgu olan sahte haberlerin tespitinde önemli bir rol oynamaktadır. Bu platformlar, sahte haberlerin tespitinde ve yayıldıkları mecraların belirlenmesinde dijital medya okuryazarlığı bağlamında önemli bir işlev görmektedir. Sahte haberlerin üretiminde yapay zekâ teknolojilerinin kullanılması, haber yayılımının büyük miktarlarda ve hızla gerçekleştiği sosyal medya gibi kanallar aracılığıyla yayılan haberlerin doğruluğuna ilişkin belirsizliğin artmasına neden olmuştur. Bu bağlamda, konunun iletişim çalışmaları perspektifinden ve sahte haberlerin yanıltıcı yönleri, içerdiği konular, bilgi türleri ve sosyal medyaya yansımaları incelenmesi gereken bir konu olarak görünmektedir. Bu çalışma iletişim çalışmaları perspektifinden sahte haberlerin üretimi, dağıtımı ve tespitinde yapay zekânın rolüne ilişkin sosyal medya açısından incelemeye odaklanmaktadır. Bu kapsamda yapay zekâ teknolojileri ile üretilen sahte haberlerin sosyal medyadaki görünümünün incelenmesinde Teyit.org doğrulama platformu örnek olarak seçilmiştir. Buradan hareketle çalışmada, yapay zekâ ile üretilmiş sahte haberlerin yanıltıcı yönleri, odaklandıkları konular, yayıldığı sosyal medya platformları ve içerdikleri bilgi türleri analiz edilmiştir. Bu çalışmada yapay zekâ tarafından üretilen sahte haberlerin özelliklerinin analizi yapılarak, bunların sosyal medyadaki görünümüne ilişkin öngörüler sunulmaktadır.

YÖNTEM

Araştırmanın Amacı

Yapay zekâ teknolojilerinin çeşitli disiplinlerde yaygınlaşması, yeni araştırma alanlarının ortaya çıkmasına neden olmuştur. Mevcut literatürün incelenmesi, yapay zekâ ve sahte haberlerle ilgili çalışmaların ağırlıklı olarak bu tür bilgileri tespit etmeye yönelik teknolojilerle ilgilendiğini ortaya koymaktadır. Sahte haberlerin yayıldığı kanallar ve bu tür içeriklerin özellikleriyle ilgili mevcut araştırmalar son derece sınırlıdır. Dahası, bu çalışmalar akademik araştırmaya özgü sınırlamalarla belirli bir çerçeve içinde yürütülmüştür. Buna karşın, haber medyası alanında yeni bir teknoloji olarak yapay zekâ üzerine yeni araştırmalarla alanın kapsamını genişletme gerekliliği göz ardı edilemeyecek bir gerçekliktir. Literatür, yapay zekâ ile üretilen sahte haberlerin doğrulama platformları aracılığıyla analizlerinin yetersiz olduğunu ortaya koymaktadır. Bu bağlamda çalışma, alandaki boşluğa yönelik olarak doğrulama platformları aracılığıyla yapay zekâ teknolojilerinin sahte haber üretimindeki rolüne ve bu haber içeriklerinin sosyal medyadaki görünümüne odaklanmaktadır. Buradan hareketle çalışmanın amacı, sahte haberlerin yanıltıcı yönlerini, haber kategorilerini, yayıldıkları platformları ve içerdikleri bilgi türlerini tespit etmektir.

Çalışma amacına göre araştırma soruları şunlardır:

1. Sahte haber içeriklerinin yanıltıcı yönü nedir?
2. Sahte haber içeriklerinin haber kategorilerine göre dağılımı nasıldır?
3. Sahte haber içeriklerinin dolaşıma girmesinde hangi sosyal medya platformları ön plana çıkmaktadır?
4. Sahte haber içeriklerinin bilgi türleri nedir?

Mevcut çalışma, yapay zekâ ile sahte haber üretiminin boyutlarını ortaya koyarak araştırmada elde edilen bulgularla iletişim bilimleri açısından alanı genişletmeye yönelik bir katkı sunmaktadır. Ayrıca sonuçları itibari ile gelecekte yapılacak çalışmalar için güncel bir veri niteliği taşımaktadır.

Araştırmanın Yöntemi ve Sınırlılıkları

Bu çalışmada, yapay zekâ ile üretilmiş sahte haberlerle ilgili tespitler yapan ve bunu kamuoyu ile paylaşan teyit.org doğrulama platformunda yer alan yapay zekâ ile üretilmiş sahte haberler içerik analizi tekniğiyle incelenmiştir. İçerik analizi, metin veya görsel içeriklerdeki kavram veya anlamların sistematik şekilde incelenerek yorumlandığı, nicel ve nitel araştırmalarda kullanılan bir veri analizi tekniğidir. İçerik analizinin temelinde, içeriklerin belirli olgular hakkında değerli bilgiler ortaya koyma potansiyeli yüksek, zengin bir veri kaynağı olduğu varsayımı yer almaktadır (Kleinheksel vd., 2020; Neuendorf ve Kumar, 2016). İçeriklerin kullanım bağlamlarından geçerli çıkarımlar yapmaya ilişkin bir teknik olarak içerik analizinde, analize konu olan veriler çalışmanın konusuna ve amacına uygun şekilde oluşturulan kategorilere ve temalara göre sınıflandırılarak yorumlanır (Krippendorff, 2018, s. 24).

Araştırma, 1 Ocak 2025 – 31 Mart 2025 tarihleri arasında teyit.org sitesinde yapay zekâ ile üretilen sahte haberler ile sınırlandırılmıştır. Araştırma kapsamında kullanılacak veriler için belirlenen tarih aralığını kapsayacak şekilde teyit.org sitesinde "yapay zekâ" anahtar kelimesi ile arama yapılmış ve 24 sahte haber içeriğine ulaşılmıştır. Araştırma verileri için teyit.org sitesinin seçilme sebebi, araştırmanın amacına uygun en geniş veri setinin elde edilebileceği bir platform olması ve haber içeriklerinin doğrulanmasına yönelik akademik çalışmalarda sıklıkla kullanılmasıdır. Yapay zekâ ile üretilen haberlere ilişkin elde edilen veri seti, araştırma soruları doğrultusunda sınıflandırılarak, yanlış içeriklerin yanıltıcı yönü, haber kategorileri, yayınlanan sosyal medya platformları ve bilgi türü açısından analiz edilmiştir.

ARAŞTIRMA BULGULARI

Çalışmanın bu bölümünde araştırma kapsamında teyit.org sitesinden yapay zekâ ile üretilmiş 24 sahte haber içeriğine ilişkin belirlenmiş kategorilere göre yapılan içerik analizinde elde edilen bulgulara yer verilmiştir.

Tablo 2. Teyit.org Web Sitesinde Yapay Zekâ İle Üretilen Sahte Haberler Kategorisinde İncelenen İddialar

Sıra	İddia	Tarih	Analiz Sonucu	Yanıltıcı Yön
1	Polisleri ve Pikaçu kostümlü kişiyi gösteren görsel	27.03.2025	Yanlış	Manipülasyon
2	Nadir görülen lotus mantis böceğini gösteren video	12.03.2025	Yanlış	Manipülasyon
3	Stockholm'deki devasa obruğu gösteren video	11.03.2025	Yanlış	Manipülasyon
4	İstanbul Boğazı'nda yakalanan dev ahtapot videosu	10.03.2025	Yanlış	Manipülasyon
5	Buzlardan kurtarılan yavru fok videosu	4.03.2025	Yanlış	Manipülasyon
6	Balinanın yatı yuttuğunu gösteren video	4.03.2025	Yanlış	Manipülasyon
7	Yılın fotoğrafı olduğu iddia edilen görsel	3.03.2025	Yanlış	Manipülasyon
8	İlk elektrikli otomobili Nikola Tesla'nın icat ettiği iddiası	26.02.2025	Yanlış	Uydurma
9	ABD'deki bir mamut fosilini gösteren fotoğraf	18.02.2025	Yanlış	Manipülasyon
10	Fil besiciliği yapan bir adamı gösteren video	12.02.2025	Yanlış	Manipülasyon
11	Fransa Cumhurbaşkanı Macron'un Hindistan Başbakanı Modi'nin elini sıkmadı iddiası	12.02.2025	Yanlış	Çarpıtma-Bağlamdan Koparma
12	Videodaki vücut geliştirme sporcusunun gerçek olduğu iddiası	4.02.2025	Yanlış	Manipülasyon
13	Hava balonuna takılarak havalanan kadını gösteren video Türkiye'den iddiası	31.01.2025	Yanlış	Uydurma - Hatalı İlişkilendirme
14	Sony'nin tanıttığı uçan otomobil videosu	28.01.2025	Yanlış	Manipülasyon
15	Timsah besleyen kişiyi gösteren video	27.01.2025	Yanlış	Manipülasyon
16	Los Angeles'daki yangında hasar almayan evlerin fotoğrafları	17.01.2025	Yanlış	Manipülasyon - Hatalı İlişkilendirme
17	Dubaili milyarderin üçüncü eşini gösteren video	17.01.2025	Yanlış	Parodi
18	Hollywood yazısının güncel yangınlardan etkilendiği videosu	13.01.2025	Yanlış	Manipülasyon
19	Los Angeles yangınında dev alevlerin görüldüğü video	13.01.2025	Yanlış	Manipülasyon
20	Los Angeles'taki yangını gösteren video	13.01.2025	Yanlış	Manipülasyon
21	Çöp kutularında sigara söndürülmemesi için uygulanan yöntemi gösteren video	10.01.2025	Yanlış	Manipülasyon
22	Squid Game dizisinin gerçek olduğunu gösteren görseller	7.01.2025	Yanlış	Manipülasyon
23	Arizona'daki UFO kazasını gösteren video	7.01.2025	Yanlış	Manipülasyon
24	Uçaktan görülen UFO videosu	3.01.2025	Yanlış	Manipülasyon

Tablo 2'de teyit.org web sitesinde yapay zekâ ile üretilen sahte haberlere ilişkin iddialarla ilgili bilgiler yer almaktadır. Buna göre incelenen 24 sahte haber içeriğinin tamamının yapay zekâ ile üretilmiş yanlış içerikler olduğu tespit edilmiştir. Sahte haberlerin tamamında yanıltıcı bilgi üretimi için yapay zekâ araçlarının kullanıldığı ve görsellerde manipülasyon yapıldığı bulgulanmıştır.

Tablo 3. Sahte Haberlerin Yanıltıcı Yönlerine Göre Dağılımı

Yanıltıcı Yön	n	%
Manipülasyon	19	79,17
Parodi	1	4,17
Uydurma	1	4,17
Uydurma - Hatalı İlişkilendirme	1	4,17
Manipülasyon - Hatalı İlişkilendirme	1	4,17
Çarpıtma-Bağlamdan Koparma	1	4,17
Toplam	**24**	**100,00**

Araştırma kapsamında incelenen sahte haberlerin yanıltıcı yönlerine ilişkin verilerin dağılımı Tablo 3'te yer almaktadır. Sahte haberlerin yanıltıcı yönünün analizinde Wardle ve Derakshan (2017) tarafından oluşturulan yedi farklı bilgi türü sınıflandırması kullanılmıştır: Parodi/hiciv, sahte/taklit, uydurma, hatalı ilişkilendirme, bağlamdan koparma, çarpıtma ve manipülasyon. Buna göre incelenen 24 içeriğin 19'unun manipülasyon olduğu tespit edilmiştir. Bu rakam incelenen sahte haberlerin %79,17'sinin manipülasyon içerdiğine işaret etmektedir. Diğer bir ifade ile her beş haberden dördünün manipülasyon olduğunu göstermektedir. Diğer dezenformasyon türleri olarak parodi, uydurma, uydurma-hatalı ilişkilendirme, manipülasyon-hatalı ilişkilendirme ve çarpıtma-bağlamdan koparma içeren birer sahte haber içeriği olduğu belirlenmiştir. Tablo 3'teki verilere göre incelenen sahte haberlerin yapay zekâ kullanılarak manipülasyon amacı ile hazırlandığı anlaşılmaktadır. Dolayısıyla bu içeriklerin toplumu yanıltmaya yönelik kasıtlı olarak oluşturulmuş içerikler olduğu görülmektedir. Sadece bir haberin parodi amacı taşıdığı düşünüldüğünde söz konusu haberin doğrudan kötü niyetli bir girişim olmadığı ancak yine yanlış bilgi içerdiği söylenebilir.

Tablo 4. Sahte Haberlerin Kategorilerine Göre Dağılımı

Kategori	n	%
Yaşam	6	25,00
Kaza-Afet	5	20,83
Doğa-Çevre	3	12,50
Bilim-Tarih	2	8,33
Kültür-Sanat	2	8,33
Siyaset	2	8,33
Teknoloji	2	8,33
Spor	1	4,17
Tarım-Gıda	1	4,17
Toplam	**24**	**100,00**

Tablo 4'te sahte haberlerin hangi konularda yapıldığını gösteren kategorilerine göre dağılımına ilişkin veriler yer almaktadır. Sahte haberlerin en fazla üretildiği üç kategori yaşam, kaza-afet ve doğa-çevre kategorisidir. Buna göre yaşam kategorisinde 6 (%25), kaza-afet kategorisinde 5 (%20,83) ve doğa-çevre kategorisinde 3 (%12,50) haber tespit edilmiştir. Bu bulgu yapay zekâ kullanılarak hazırlanan sahte haberlerin büyük bir kısmının yaşam, doğa ve çevresel durumlarla ilgili konuları içerdiğini göstermektedir. Daha sonrasında bilim-tarih, kültür-sanat, siyaset ve teknoloji kategorilerinde 2'şer haber olduğu elde edilen bulgular arasındadır. En az sahte haber üretimi birer haber ile spor ve tarım-gıda kategorisindedir. Tablo 4'teki verilere göre araştırma kapsamında incelenen sahte haberlerde doğa, yaşam ve çevresel felaketler ön planda iken, siyaset, bilim-tarih, kültür-sanat, teknoloji gibi konuların geri planda kaldığı görülmektedir.

Tablo 5. Sahte Haberlerin Yayınlandığı Sosyal Medya Platformları

Mecra	n	%
Instagram	9	37,50%
TikTok	7	29,17%
X	4	16,67%
Facebook	3	12,50%
Belirsiz	1	4,17%
Toplam	**24**	**100,00%**

Tablo 5'te yapay zekâ ile üretilen sahte haberlerin ilk olarak yayınlandığı sosyal medya platformlarına göre

dağılımı bulunmaktadır. Instagram 9 sahte haber (%37,50) ile en fazla paylaşımın yapıldığı platform olarak ön plana çıkmaktadır. Instagram'dan sonra sırasıyla 7 haber (%29,17) TikTok, 4 haber (%16,67) X, 3 haber (%12,50) Facebook olarak gerçekleşmiştir. Araştırmada incelenen 1 sahte haberin (%4,17) ilk olarak yayınlandığı kaynak belirsizdir. Tablo 5'teki verilere göre tüm haberlerin %66,67'sinin Instagram ve TikTok'ta yayınlandığı belirlenmiştir. Diğer bir ifade ile her üç sahte haberin 2'si bu iki sosyal medya platformundan birinde yayınlanmıştır. Bu iki platform yapay zekâ ile üretilen sahte haberlerin en fazla yayınlandığı platformlar olarak ön plana çıkmaktadır.

Tablo 6. Sahte Haberlerin Bilgi Türü

Bilgi Türü	**n**	**%**
Video	19	79,17%
Fotoğraf	5	20,83%
Toplam	**24**	**100,00%**

Araştırmada incelenen sahte haberlerin bilgi türüne göre dağılımının yer aldığı Tablo 6'ya göre sosyal medya platformlarında dolaşıma giren sahte haberlerin %79,17'si video, %20,83'ü fotoğraf kullanılarak oluşturulmuştur. Sahte haber içeriklerinde video kullanımının yoğunluğu dikkat çekici düzeydedir. Bu bulgu, sahte haber üretiminde daha çok video içeriklerin kullanıldığını göstermektedir. Ayrıca sahte haberlerde yanıltıcı içerik olarak sadece fotoğraf ve videodan yararlanılmış olması görsel manipülasyona yönelik bir çabanın olduğuna işaret etmektedir. Toplumun yanıltıcı bilgi ile manipüle edilmesinde görsel içeriklerin gücünden faydalanıldığı açıkça görülmektedir. Görsel içeriklerin ilgi çekici etkisi düşünüldüğünde bu durum normal bir sonuç olarak değerlendirilebilir.

Tablo 7. Sahte Haberlerin Bilgi Türünün Sosyal Medya Platformlarına Göre Dağılımı

	Instagram	Tiktok	X	Facebook	Belirsiz
Fotoğraf	1	0	2	2	0
Video	8	7	2	1	1

Tablo 7'de sahte haberlerin bilgi türlerinin yayınlandıkları sosyal medya platformlarına göre dağılımı yer almaktadır. Buna göre video içeriklerin yoğunlukla Instagram ve TikTok'ta paylaşıldığı görülürken, X ve Facebook'ta ise hem fotoğraf hem de video içeriklerin dengeli dağıldığı görülmektedir. Sahte haberlerde video içeriklerin Instagram ve TikTok üzerinden dolaşıma girdiği anlaşılmaktadır. Instagram ve TikTok platformlarının hem kullanıcı profili hem de kullanım amacı bir arada düşünüldüğünde normal bir sonuç olarak değerlendirilebilir.

SONUÇ

Yapay zekâ teknolojilerinin kullanımı diğer birçok alanda olduğu gibi medya ve habercilik alanında da giderek artmaktadır. Sahte haberlerin üretiminde yapay zekânın kullanımı, bu tür içeriklerin oluşturulmasını kolaylaştırmış ve sosyal medyanın özelliklerinden yararlanarak yayılmasını hızlandırmıştır. Bu durum yapay zekâ teknolojilerinin kötü amaçlı sahte haberlerin üretimindeki rolüyle ilgili endişelere yol açmıştır. Dolayısıyla sosyal medyada karşılaşılan bir içeriğin gerçek mi sahte mi olduğu konusundaki ayırt edebilme durumu zorlaşmıştır.

Bu çalışma sahte haber üretiminde yapay zekâ teknolojilerinin kullanımı ve bu haberlerin sosyal medyadaki görünümüne odaklanmaktadır. Çalışmada bu kapsamda teyit.org doğrulama platformunda yer alan 24 sahte haber içeriği analiz edilmiştir. Araştırmanın amaç soruları bağlamında ilk olarak sahte haber içeriklerinin yanıltıcı yönü incelenmiştir. Bu doğrultuda analiz edilen 24 adet sahte haber içeriğinin tamamının yanlış olduğu ve yapay zekâ kullanılarak üretildiği belirlenmiştir. Bu sonuç, Almasi ve Schiønning (2023) tarafından sahte haber üretiminde yapay zekâ teknolojisi olarak GPT-3'ün kullanımına ilişkin sonuçla örtüşür niteliktedir. Çalışma, yapay zekâ araçlarının yanıltıcı bilgi üretmek için kullanıldığını ve tüm sahte haber öğelerinde görsel manipülasyon kullanıldığını ortaya koymaktadır. Benzer bir bulgu Özsalih (2023) tarafından yapılan çalışmada, Donald Trump ile ilgili bir haberde yanıltıcı bilgi oluşturmada yapay zekâ destkeli görsel manipülasyonun yapıldığını göstermektedir. Sahte haber içeriklerinin manipülasyon amaçlı kullanımına ilişkin literatürde benzer sonuçlara ulaşan araştırmalar bulunmaktadır (Aydın, 2020; Çömlekçi, 2019). Bu araştırma, yapay zekâ destekli sahte haberlerin de benzer amaçla kullanımını ortaya koymaktadır. İçeriklerin büyük bir kısmının manipülasyon içermesi, sahte haber içeriğini üretenlerin niyetine ilişkin

fikir vermektedir. Bu sonuç, sosyal medyada yer alan haber içeriklerinin doğruluğu ile ilgili ciddi bir sorunu yansıtmaktadır. Haberlerin doğruluğu hakkında medya tüketicilerinin dikkatli ve bilinçli hareket etmesi yanlış bilginin dolaşımını engellemeye yönelik önemli bir adım olacaktır. Bu durum aynı zamanda bir medya okuryazarlığı bilgisini de gerektirmektedir.

Araştırmanın amaç soruları bağlamında elde edilen bir diğer sonuç, sahte haberlerin hangi konularda üretildiğidir. İncelenen haberlerin birçok farklı konuda üretildiği, ancak yaşam, doğa ve çevre felaketlerine ilişkin haberlerin daha fazla olduğu tespit edilmiştir. Burada yapay zekâ teknolojilerinin kullanıldığı sahte haberlerle ilgili siyaset kategorisinin düşük oranda olması dikkat çekici bir sonuçtur. Nitekim yapay zekâ ile üretilen sahte içeriklerin öncelikli hedefleri arasında siyasi aktörler olduğu bilinmektedir (Shu vd., 2017; Törnberg, 2023). Yine benzer şekilde Vosoughi vd. (2018) X platfrormunda sahte haberlerle ilgili yaptıkları araştırmada siyasi konuların terörizm, doğal afetler, bilim, şehir efsaneleri ve finansal konulardan daha fazla olduğunu tespit etmiştir. Araştırmamızda elde edilen bu bulgu diğer araştırmalardan farklı olarak, yapay zekâ ile üretilen haberlerin siyaset dışında farklı konularda da yoğun şekilde üretildiğini göstermektedir.

Araştırma sonuçları, sosyal medyanın incelenen çalışmalarda son derece etkili bir mecra olarak kullanıldığını doğrulamaktadır. 24 sahte içeriğin tamamının Instagram, TikTok, X ve Facebook gibi sosyal medya platformlarında dolaşıma girdiği tespit edilirken, sahte içeriklerin daha çok Instagram ve TikTok üzerinden dolaşıma girdiği belirlenmiştir. Bu sonuçta, gençlerin içerik paylaşım hızı dikkate alındığında söz konusu platformların gençler arasında daha yaygın olmasının etkisi olduğu düşünülmektedir. Ayrıca bu platfomların fotoğraf ve video gibi görsel içeriklerin paylaşmına uygun doğasının da etkili olduğu söylenebilir. X platformunun geri planda kalmasında siyasi içerikli videoların az olmasının etkili olması muhtemeldir. X platformunda politik içeriklerin diğer platformlara göre daha yaygın olması bu sonuçta etkili olabilir. Nitekim Ulusoy ve İlhan (2023) ABD seçim kampanyalarına ilişkin deepfake içeriklerin çoğunlukla X platformunda paylaşıldığını tespit etmiştir.

Araştırma soruları kapsamında incelenen bir diğer sonuç ise, sahte içeriklerin bilgi türüne ilişkindir. Sahte içeriklerin bilgi türü açısından sonuçları incelendiğinde, tamamen fotoğraf ve video içeren görsel içeriklerin tercih edildiği tespit edilirken, içeriklerin çok büyük oranda video olarak üretildiği belirlenmiştir. Video ve fotoğraf içerikli sahte haberlerin sosyal medyada ağırlıklı olarak Instagram ve TikTok üzerinden dolaşıma girdiği belirlenmiştir. X ve Facebook ise video ve fotoğraf içeriklerinin yakın oranlarda paylaşıldığı platformlar olarak işlev görmüştür. Bu sonuç Instagram ve TikTok platformlarının video içerik paylaşımındaki etkin rolünü yansıtması bakımından anlamlıdır. Ayrıca haber içeriklerinde politik gündemden uzak doğa, yaşam ve çevre konularının ağırlıklı olmasının bu sonuçta etkili olduğu söylenebilir.

Araştırma sonuçlarına göre yapay zekâ teknolojilerinin sahte haber üretiminde manipülasyon amacıyla etkin bir şekilde kullanıldığı görülmüştür. Sahte haber içeriklerinde yaşam, doğa ve çevre konularının ön planda olduğu, genellikle Instagram ve TikTok'ta dolaşıma girdiği ve daha çok video olarak üretildiği sonucuna varılmıştır. Araştırma sonuçları, sosyal medyanın dezenformasyon aracı olarak kullanılmasına yönelik kuramsal zeminle örtüşmektedir. Diğer bir ifade ile sosyal medyada dezenformasyonla ilgili literatürü destekler niteliktedir. Araştırma sonuçları ayrıca, yapay zekâ teknolojilerinin kullanımı konusunda deepfake ve görsel içerik manipülasyonu teknikleri açısından literatürdeki kuramsal çerçeveyi desteklemektedir. Bu anlamda medya manipülasyonunda görsel içerik kullanımının yaygınlığını tespit etmektedir. Araştırmada incelenen haberlerin tamamının görsel manipülasyon içermesi, yapay zekâ teknolojilerinde toplumu yanıltıcı haberlerde görsel içeriklerin gücünden yararlanıldığını göstermektedir. Araştırma sonuçları, yapay zekâ destekli sahte haberlerle ilgili dijital medya okuryazarlığı bağlamında literatüre doğrulama platformu üzerinden veriler sunmaktadır. Başta sosyal medya olmak üzere, sahte haberlerin dijital platformlardaki yaygın kullanımının olumsuz etkilerine karşı toplumun dijital medya okuryazarlığı konusunda farkındalık kazanması önemlidir. Nazar ve Bustam (2020) çevrimiçi ortamlardaki sahte haberlere yönelik farkındalık konusunda yapay zekâ ile üretilen haberlere dikkat çekmişlerdir. Bu

açıdan araştırmanın sonuçları, dezenformasyonla sınırlı literatürün yapay zekâ destekli sahte haber içerikleri ile genişletilmesine katkı sağlamaktadır.

Bu çalışma, yapay zekâ tabanlı sahte haberlerin sistematik bir incelemesini yaparak konuya ilişkin kapsamlı ve güncel bir bakış sunmaktadır. Çalışmanın sonuçları, sahte haber üretiminde yapay zekânın kullanımı ve sosyal medyada ne şekilde paylaşıldığına ilişkin iletişim çalışmaları alanındaki yeni çalışmalar için çıkarımlar sağlamaktadır. Araştırma ayrıca sahte haber üretiminde ve dağıtımında yapay zekânın ve sosyal medyanın rolünün anlaşılmasına yardımcı olmaktadır. Yapay zekânın günümüzdeki popülaritesi ve yaygınlaşan kullanımı dikkate alındığında yeni araştırmalar için önemli bir motivasyon kaynağı olacağı açıktır. Bu çalışma teyit.org doğrulama platformunda 1 Ocak 2025 – 31 Mart 2025 tarihleri arasındaki yapay zekâ ile üretilen sahte haberlerin analizi ile sınırlıdır. İleride yapılacak çalışmalarda farklı örneklem gruplarının analiz edilmesi alanın genişletilmesi açısından önemli bir katkı olacaktır.

KAYNAKLAR

Akhtar, P., Ghouri, A. M., Khan, H. U. R., Haq, M. A., Awan, U., Zahoor, N., Khan, Z. & Ashraf, A. (2023). Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Annals of Operations Research*. *327*(2), 633-657. https://doi.org/10.1007/s10479-022-05015-5

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives*. *31*(2), 211–235. https://doi.org/10.1257/jep.31.2.211

Almasi, M. & Schiønning, A. (2023). Fine-tuning gpt-3 for synthetic Danish news generation. *In Proceedings of the 16th International Natural Language Generation Conference*, 54–68, Prague, Czechia.

Aydın, A. F. (2020). Post-truth dönemde sosyal medyada dezenformasyon: Covid-19 (yeni koronavirüs) pandemi süreci. *Asya Studies, 4*(12), 76-90. https://doi.org/10.31455/asya.740420

Beckett, C. (2021). *New powers, new responsibilities a global survey of journalism and artificial intelligence.* The London School of Economics.

Biswas, S. (2023). Prospective role of chat gpt in the military: According to chatgpt. *Qeios*. https://doi.org/doi:10.32388/8WYYOD.

Chadha, A., Kumar, V., Kashyap, S. & Gupta, M. (2021). Deepfake: An overview. P. K. Singh, S. T. Wierzchoń, S. Tanwar, M. Ganzha & J. J. P. C. Rodrigues (Eds.), *Proceedings of second international conference on computing, communications, and cyber-security* (s. 557-566). Springer.

Cresci, S. (2020). A decade of social bot detection. *Communications of the ACM. 63*. 72-83. https://doi.org/10.1145/3409116.

Çömlekçi, M. F. (2019). Sosyal medyada dezenformasyon ve haber doğrulama platformlarının pratikleri. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi, 7*(3), 1549-1563. https://doi.org/10.19145/e-gifder.583825

Devarajan, G.G., Nagarajan, S.M., Amanullah, S.I., Mary, S.A., & Bashir, A.K. (2024). AI-assisted deep NLP-based approach for prediction of fake news from social media users. *IEEE Transactions on Computational Social Systems, 11*(4), 4975-4985,

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics, 26*(1), 69-91. https://doi.org/10.1177/1940161220944364

El Gody, A. (2021). Using artificial intelligence in the Al Jazeera newsroom to control fake news. Al Jazeera Media Institute.

Erkan, G. & Ayhan, A. (2018). Siyasal iletişimde dezenformasyon ve sosyal medya: Bir doğrulama platformu olarak teyit.org. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi (29. Özel Sayısı)*, 202-223. https://doi.org/10.31123/akil.458933

Fallis D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology*, *34*(4), 623–643. https://doi.org/10.1007/s13347-020-00419-2

Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday, 28*(6). https://doi.org/10.5210/fm.v28i6.13185

Ferrara, E., Varol, O., Davis, C. B., Menczer, F. & Flammini, A. (2016). The rise of social bots. *Communications of the ACM 59*(7), 96-104.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science, 30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Fusco, F. (2022). Artificial intelligence and fake news: Criminal aspects in Pakistan and Saudi Arabia. *Pakistan Journal of Criminology. 14* (4), 19-33.

Gelfert, A. (2018). Fake news: A definition. *Informal Logic. 38* (1), 84-117. https://doi.org/10.22329/il.v38i1.5068

Godulla, A., Hoffmann, C. P., & Seibert, D. (2021). Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. *SCM Studies in Communication and Media. 10*(1), 72-96. https://doi.org/10.5771/2192-4007-2021-1-72

Graves, L. (2016). *Deciding what's true: the rise of political fact-checking in American journalism*. Columbia University Press.

Guess, A. M. & Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. N. Persily and J. A. Tucker (Eds.), *Social Media and Democracy - The State of the Field, Prospects for Reform* (s. 10-33). Cambridge University Press.

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and artificial ıntelligence: Definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Joshi, V. & Patel, S. (2024). Unveiling deception: a gan-based unsupervised learning approach for real-time generation and detection of text-based fake news. *Journal of Electrical Systems.* 20, 6189-6195. https://doi.org/10.52783/jes.6586.

Kaplan, A., & Haenlein, M. (2019). Siri, siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*, 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Khodabakhsh, A., Ramachandra, R. & Busch, C. (2019). Subjective evaluation of media consumer vulnerability to fake audiovisual content. *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, Berlin, Germany, 1-6, https://doi.org/doi: 10.1109/QoMEX.2019.8743316.

Kietzmann, J., Lee, L. W., McCarthy, I. P., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146. https://doi.org/10.1016/J.BUSHOR.2019.11.006

Kirchengast, T. (2020) Deepfakes and image manipulation: Criminalisation and control. *Information & Communications Technology Law, 29*(3), 308-323. https://doi.org/10.1080/13600834.2020.1794615

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H. & Wyatt, T. R. (2020). Demystifying content analysis. *American Journal of Pharmaceutical Education. 84*(1). https://doi.org/doi: 10.5688/ajpe7113.

Krippendorff, K. (2018). Content analysis - an introduction to its methodology (4th ed.). Sage Publications.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F and Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094-1096. https://doi.org/10.1126/science.aao2998

Levy, N. (2017). The bad news about fake news. *Social Epistemology Review and Reply Collective 6*, *(8)*, 20-36.

Maras, M. & Alexandrou, A. (2018) Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *The International Journal of Evidence & Proof, 23*, 255-262. https://doi.org/10.1177/1365712718807226

Minsky, M. L. (1968). *Semantic information processing.* The MIT Press.

Narang, P. & Sharma, U. (2021). A study on artificial intelligence techniques for fake news detection. *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, Uzbekistan, 482-487. doi: 10.1109/ICTAI53825.2021.9673252.

Nazar, S. & Bustam, R. (2020). Artificial intelligence and new level of fake news. *IOP Conference Series: Materials Science and Engineering*. 879. Bandung, Indonesia. https://doi.org/10.1088/1757-899X/879/1/012006.

Neuendorf, K. & Kumar, A. (2016). Content analysis. G. Mazzoleni (Eds.), *The International Encyclopedia of Political Communication* (s. 1-10). Wiley-Blackwell. https://doi.org/10.1002/9781118541555.wbiepc065.

Nyilasy, G. (2019) Fake news: When the dark side of persuasion takes over. *International Journal of Advertising, 38*(2), 336-342, https://doi.org/10.1080/02650487.2019.1586210

Özsalih, A. (2023). Yapay zekâ yoluyla oluşturulan sahte haberlerin medya gündemini belirlemesi. *The Turkish Online Journal of Design, Art and Communication,* 13(3), 533-550. https://doi.org/10.7456/tojdac.1285554

Pantserev, K. A. (2020). The malicious use of ai-based deepfake technology as the new threat to psychological security and political stability, H. Jahankhani, S. Kendzierskyj, N. Chelvachandran & J. Ibarra (Eds.), *Cyber defence in theage of ai, smart societies and augmented humanity* (s. 37-56). Springer.

Paredes, D. G. (2023). Strategies based on artificial intelligence for the detection of fake news. *AWARI; 4,* 1-6. https://doi.org/10.47909/awari.58

Paschen, J. (2020), Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management, (29)*2, 223-233. https://doi.org/10.1108/JPBM-12-2018-2179

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv: Computation and Language*. https://arxiv.org/abs/1702.05638

Raza, S., Paulen-Patterson, D. & Ding, C. (2025). Fake news detection: Comparative evaluation of BERT-like models and large language models with generative AI-annotated data. *Knowl Inf Syst. 67*, 3267–3292. https://doi.org/10.1007/s10115-024-02321-1

Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal. 27*(S2), 43-64. https://doi.org/10.1353/ken.2017.0025

Roumeliotis, K.I., Tselikas, N.D., Nasiopoulos, D.K. (2025). Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models. *Future Internet. 17*, 28. https://doi.org/10.3390/fi17010028

Shu, K., Sliva, A.L., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ArXiv,* https://doi.org/10.48550/arXiv.1708.01967

Swaroop, T.S. (2024). Social media and the influence of fake news detection based on artificial intelligence. *ShodhKosh: Jour-*

nal of Visual and Performing Arts, 5(7), 77–87. https://doi.org/ 10.29121/shodhkosh.v5.i7.2024.1955

Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot learning. *ArXiv.* https://doi.org/10.48550/arxiv.2304.06588

Ulusoy, H. ve Kaya İlhan, Ç. (2025). Siyasal iletişimde yapay zekâ etkisi ve deepfake (derin sahte) dezenformasyonu: 2024 ABD başkanlık seçimleri örneği. *TRT Akademi, 10*(23), 42-73. https://doi.org/10.37679/trta.1563828

Vincent, V. U. (2021). Integrating intuition and artificial intelligence in organizational decision-making. *Business Horizons*. (64)3, 425-438. https://doi.org/10.1016/j.bushor.2021.02.008

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science (New York, N.Y.)*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wardle, C. & Derakhshan, H. (2017). *Information disorder: toward an interdisciplinary framework for research and policy making*. Council of Europe. https://rm.coe.int/information-disorder-report-november-2017/1680764666