MAKALE

TRT Akademi Dergisi

ISSN 2149-9446 | Cilt 10 | Sayı 23 | Ocak 2025 | Dezenformasyon

# Yapay Zekâ ve Dezenformasyon: OpenAI Raporu Doğrultusunda Küresel Dezenformasyon Kampanyalarının Analizi

**Bahadır AVŞAR***

**Öz**

Bu çalışma, yapay zekâ teknolojilerinin dezenformasyon kampanyalarında kullanımını ele alarak dijital ortamda ortaya çıkan yeni tehditleri anlamayı amaçlamaktadır. Yapay zekânın hızlı gelişimi ve geniş çapta veri işleme kapasitesi, dezenformasyonun yayılma hızını ve etkisini artırarak toplumsal yapıları ve demokratik süreçleri savunmasız hâle getirmektedir. Çalışma, dezenformasyon stratejilerini yapay zekâ bağlamında inceleyerek, bilginin manipülasyonu ve sahte gerçekliklerin oluşturulmasındaki rolünü araştırmaktadır. Kuramsal olarak yapay zekânın dezenformasyon ortamına olan katkısı ele alınmış ve literatürdeki dezenformasyon stratejileri detaylandırılmıştır. Metodolojik olarak literatür taraması ve elektronik materyal inceleme yöntemleri kullanılmıştır. Bu çalışmada sistematik bir literatür taraması yapılmış ve OpenAI tarafından yayımlanan, yapay zekâ kullanılarak dezenformasyon faaliyetlerini ele alan dört ülkeye ait raporlar doküman analizi yöntemiyle incelenmiştir. Özellikle, OpenAI raporu üzerinden Rusya, Çin, İran ve İsrail gibi ülkelerin yürüttüğü yapay zekâ destekli dezenformasyon kampanyaları vaka çalışması olarak incelenmiştir. Bulgular, bu ülkelerin dezenformasyon kampanyalarının küresel güvenlik ve diplomasi üzerindeki olumsuz etkilerine işaret etmektedir. Çalışma, dezenformasyonun toplumsal etkilerini anlamaya ve bu soruna karşı çözüm geliştirme çabalarına katkı sunmayı hedeflemektedir.

**Anahtar Kelimeler:** Yapay Zekâ, Dezenformasyon, Dijital Ekosistem, OpenAI Raporu, Dezenformasyon Kampanyaları

** Dr., Araştırmacı, Türkiye Radyo Televizyon Kurumu, Ankara, Türkiye*
*E-mail: bahadir.avsar@trt.net.tr*
*ORCID: 0000-0002-6805-537X*

*Avşar, B. (2025). Yapay Zekâ ve Dezenformasyon: OpenAI Raporu Doğrultusunda Küresel Dezenformasyon Kampanyalarının Analizi. TRT Akademi, 10(23), 208-237. https://doi.org/10.37679/trta.1562358*

Derleme Makalesi

**Geliş Tarihi:** 06.10.2024 **Revize Tarihi:** 14.11.2024 **Kabul Tarihi:** 17.01.2025

**DOI:** 10.37679/trta.1562358

ARTICLE

ISSN 2149-9446 | Volume 10 | Issue 23 | Jan 2025 | Disinformation

TRT Akademi Dergisi

# Artificial Intelligence and Disinformation: A Comprehensive Analysis of Global Disinformation Campaigns Underpinned by the OpenAI Report

**Bahadır AVŞAR**

**Abstract**

This study aims to understand the new threats emerging in the digital environment by examining the use of artificial intelligence (AI) technologies in disinformation campaigns. The rapid development of AI and its extensive data processing capabilities have increased the speed and impact of disinformation, rendering societal structures and democratic processes more vulnerable. The study investigates disinformation strategies within the context of AI, exploring its role in manipulating information and constructing false realities. Theoretically, the contribution of AI to the disinformation environment is discussed, and disinformation strategies in the literature are elaborated. Methodologically, the study employs literature review and electronic material analysis methods. In this study, a systematic literature review was conducted, and reports from four countries, published by OpenAI and addressing disinformation activities using artificial intelligence, were analyzed using the document analysis method. Specifically, an OpenAI report serves as a case study, examining AI-supported disinformation campaigns conducted by countries such as Russia, China, Iran, and Israel. The findings highlight the adverse effects of these countries' disinformation campaigns on global security and diplomacy. This study contributes to understanding the societal impacts of disinformation and aims to support efforts to develop solutions to this issue.

**Keywords:** Artificial Intelligence, Disinformation, Digital Ecosystem, OpenAI Report, Disinformation Campaigns

Review Article

**Received:** 06.10.2024 **Revised:** 14.11.2024 **Accepted:** 17.01.2025

## 1. Giriş

Yapay zekâ, ilk kez 1956 yılında Dartmouth Konferansı'nda gündeme gelmiş (Lungarella vd., 2007) ve biz farkına varmadan uzun yıllar boyunca pek çok alanda yaşamımızın önemli bir parçası olarak etkisini sürdürmüştür. Dijital alanda kullanılan harita geliştirmelerinden mobil uygulamalardaki önerilere kadar yapay zekâ; çoğu zaman varlığını fark etmediğimiz ama hayatımızı kolaylaştıran, daha verimli olmasını sağlayan ve bunu verilerimizi kullanarak gerçekleştiren kritik bir teknolojidir. Genel kullanıma sunulan yapay zekâ platformlarının hayatımıza girmesiyle birlikte, bu teknoloji yeni bir aşamaya taşınmıştır. İnternete ulaşabilen herkesin kullanabildiği, toplumsal tabana hızla yayılan yapay zekâ, bilgiye erişim biçimimizi köklü bir şekilde dönüştürmeye başlamıştır. Yapay zekânın gelişim hızı ve öğrenme kapasitesi, mevcut teknolojik paradigmanın sınırlarını zorlayarak insan tahayyülünün ötesine geçmektedir. Bu teknolojinin, sürekli olarak trilyonlarca veri noktasını işleyip kendini güncelleyebilme yeteneği, yalnızca bilgi üretim süreçlerini dönüştürmekle kalmayıp toplumsal, ekonomik ve entelektüel yapılarda da köklü değişimlere yol açma potansiyelini taşımaktadır. Özellikle yapay zekânın öğrenme algoritmaları ve sinir ağları veriyle etkileşim kurdukça daha rafine hâle gelerek gelecekteki keşiflerin ve uygulamaların öngörülemez hız ve derinlikte ilerlemesine neden olmaktadır. Bu durum, sadece teknik bir başarıdan öte, bilginin doğası ve kullanımı üzerine derin bir felsefi sorgulama ihtiyacını da beraberinde getirmektedir. Yapay zekâ, özelleştirilmiş süper bilgisayarlarda saniyeler içinde trilyonlarca işlem yaparak, öğrendiği her yeni veriyle sinir ağlarını yeniden şekillendirip kendini sürekli güncellemektedir. Bu hız, Moore Yasası'nın (Pennekamp vd., 2019) öngördüğü teknolojik ilerlemeyi gölgede bırakacak seviyededir.

Yapay zekâ, gündelik yaşamdan ileri teknolojilere kadar hayatımızın pek çok alanında katkılar sağlamakla birlikte, aynı zamanda toplumsal yapıları tahrip etme ve ülkeleri kaosa sürükleme potansiyeline sahip olan dezenformasyonun etkisini güçlendiren bir unsur olarak karşımıza çıkar. Yapay zekânın dezenformasyon ortamına girişi, yanlış bilgilerin benzeri görülmemiş bir kesinlikle ve ölçekte oluşturulmasını ve yayılmasını mümkün kılmaktadır. Yapay zekâ ile desteklenen dezenformasyon kampanyaları, sadece bireyler arasında değil, uluslararası platformlarda da ciddi krizler oluşturabilmektedir. Dezenformasyonun dijital ekosistemdeki etkisi, yapay zekânın hızlı ve geniş çaplı içerik üretme kapasitesi ve sahte gerçeklikler inşa etme gü-

cüyle birleştiğinde, toplumsal yapının kırılganlığını artırmakta, demokratik süreçleri daha savunmasız hâle getirmektedir. Bu da toplumsal dinamikleri etkileme potansiyelini yükseltirken, aynı zamanda dezenformasyonu tespit etme potansiyelini daha karmaşık bir hâle getirmektedir (Karinshak ve Jin, 2023). Ayrıca, yapay zekâ teknolojilerinin yaygınlaşmasıyla birlikte; kişisel mahremiyetin ihlali, algoritmik ön yargılar, otomasyon kaynaklı iş gücü kayıpları ve toplumsal dejenerasyonu derinleştiren bilgi asimetrisi gibi birçok problem içinden çıkılmaz bir hâle gelmektedir. Sosyal medya ve diğer dijital platformlarda yapay zekâ ile üretilen bilginin ışık hızında yayılması, dezenformasyonun erişimini ve etkisini artırmaktadır. Dezenformasyondaki bu hız politika yapıcılar ve küresel kamuoyu için kritik bir sorundur (Wardle ve Derakhshan, 2017).

Çalışmamız, literatürde yer alan dezenformasyon stratejilerini inceleyerek yapay zekâ teknolojileri ile üretilen dezenformasyon yöntemlerini anlamaya yönelik kapsamlı bir analiz sunmayı hedeflemektedir. Bu bağlamda öncelikle, yapay zekâ teknolojilerinin dezenformasyon kampanyalarında kullanımını ele alan bilimsel literatür kapsamlı bir şekilde incelenmiştir. Çalışmada mevcut çalışmaların incelenmesi ile sonuç ve bulguların sistematik, şeffaf ve yeniden üretilebilecek şekilde sentezlenerek ortaya konmasında etkili bir araştırma metodolojisi olarak kabul edilen sistematik literatür taraması metodolojisinden yararlanılmıştır (Yıldız 2022). Kullanılan elektronik materyal "Yapay Zekâ ve Gizli Nüfuz Operasyonları: Son Trendler" isimli OpenAI raporu (Nimmo, 2024) doküman analizi yöntemi ile incelenmiştir. Bu analiz yöntemi anlamlı bilgiler elde etmek için yazılı materyallerin sistematik olarak incelenmesini içeren nitel bir araştırma yöntemidir. Kitapları, raporları, mektupları ve elektronik dosyaları içerebilen belgeleri anlamak, yorumlamak ve içgörüler elde etmek için kullanılır (Kıral, 2020). OpenAI raporunda Rusya, Çin, İran ve İsrail'in dezenformasyon kampanyaları analiz edilmektedir. Rusya'nın "Bad Grammar" ve "Doppelganger" operasyonları, Çin'in "Spamouflage" kampanyası, İran'ın Uluslararası Sanal Medya Birliği (IUVM) girişimi ve İsrail'in "Zero Zeno" operasyonları detaylandırılmaktadır (Nimmo, 2024). Bu rapor, yapay zekâ destekli dezenformasyonun küresel güvenlik ve diplomasi üzerinde oluşturacağı olumsuz etkileri gözler önüne sermesi açısından kritik bir önemdedir. Çalışmada, bu kampanyalar hem stratejik yöntemler hem de uluslararası güvenlik açısından etkileri bağlamında doküman analizi yöntemiyle değerlendirilmeye çalışılmıştır. Literatür taraması, dezen-

formasyon stratejilerinin teorik çerçevesini sunarken doküman analizi vaka bazlı verilerle araştırmayı desteklemiştir.

## 2. Dijital Çağda Dezenformasyon Biçimleri

Dezenformasyon, bir izleyiciyi aldatma veya etkileme niyetiyle kasıtlı olarak yanlış, yanıltıcı veya manipüle edilmiş bilgilerin yayılmasıdır. Yanlış bilgi paylaşımında hata yapılması veya kasıtsız bilgi paylaşılması anlamından farklı olarak, dezenformasyon belirli stratejik hedeflere ulaşmak için kasıtlı olarak oluşturulur. Bu hedefler, kamuoyunu etkilemeyi, bireyleri veya kuruluşları itibarsızlaştırmayı, siyasi varlıkları istikrarsızlaştırmayı veya toplumsal bölünmeler oluşturmayı içerebilir (Wardle ve Derakhshan, 2017). Tarihsel olarak, dezenformasyon çeşitli bağlamlarda, özellikle savaş zamanı propagandalarında, siyasi karalama kampanyalarında ve ideolojik çatışmalarda kullanılmıştır (Hayden, 2024). Ancak dijital medyanın ortaya çıkması ve çevrim içi platformların çoğalması, dezenformasyonun ulaşımını ve etkisini önemli ölçüde genişletmiş, onu psikolojik ve bilgi savaşı için güçlü bir araç hâline getirmiştir (Warin, 2024).

Sosyal medya, bilginin yayılma hızında devrimsel bir etki oluşturarak içeriğin hızlı ve yaygın olarak paylaşılmasına olanak sağlayan bir yapıya sahiptir. Etkileşim odaklı algoritmalar tarafından yönlendirilen bu ortam, genellikle sansasyonel ve yanlış içeriğe öncelik vererek ve dezenformasyonun yayılmasını şiddetlendirmektedir (Balcioğlu ve Dogan, 2023). Dezenformasyon kampanyalarında kullanılan teknikler, dijital teknolojideki ilerlemelerle birlikte önemli ölçüde dönüşmüş, daha karmaşık ve çok yönlü hâle gelmiştir. Bu teknikler, etkinliklerini ve erişimlerini en üst düzeye çıkarmak için birden fazla strateji kullanan, genellikle iç içe geçmiş yöntemlerdir. Dezenformasyon stratejileri genellikle dört temel kategoriye ayrılır: Sızdırma, yalan söyleme, tohumlama ve karalama (Arce, 2024). Sızdırma, genellikle gizli bilgilerin kamuoyuna ulaşması amacıyla yayılmasını içerir. Bu strateji, bilgilerin doğruluğuna dair bir şüphe uyandırsa da çoğunlukla gerçek bilgilerin bir karışımı olabilir ve hedef kitle üzerinde güvenilirlik kazanmayı amaçlar (Zannettou vd., 2019). Yalan söyleme stratejisi, özellikle sosyal medya platformları ve geleneksel medya aracılığıyla yayılan yalan haberler ile toplumları kutuplaştırmak için kullanılır (Vosoughi, Roy ve Aral, 2018). Tohumlama ve Karalama stratejileri ise dezenformasyonun daha yaygın ve etkili yöntemlerindendir.

Tohumlama, yanlış bilgilerin küçük, görünüşte önemsiz parçalar hâlinde yayılması ve zamanla kitlelerin bilinçaltına yerleşmesini amaçlar (Jing vd., 2023). Bu strateji, özellikle internet trolleri ve bot ağları aracılığıyla yürütülür. Karalama ise hedef kişilere ya da kurumlara yönelik saldırgan ve yanıltıcı bilgilerin yayılmasıdır ve özellikle siyasi liderler ya da kamuoyu önündeki kişiler üzerinde büyük etkiye sahiptir. Karalama kampanyaları genellikle dijital platformlar aracılığıyla yürütülerek hedefin itibarını zedelemeyi ve kamuoyunda güven kaybına yol açmayı hedefler (Howard ve Bradshaw, 2018).

Sızdırma, yalan söyleme, tohumlama ve karalama gibi dezenformasyon yöntemleri; aldatıcı reklamlar, siyasal propaganda, tahrif edilmiş fotoğraflar, sahte belgeler, sahte haritalar, internet dolandırıcılıkları, sahte web siteleri ve manipüle edilmiş Wikipedia maddeleri (Fallis, 2014) ile yeniden biçimlenmiştir. Genellikle yanlış ve gerçek bilgilerin bir karışımı olan sahte haberler, okuyucuların kafasını karıştırmak için hazırlanır ve tık tuzağı (Clickbait) başlıkları genellikle gerçeğin feda edilmesi pahasına dikkat çekmek ve trafiği artırmak için kullanılır (Giachanou vd., 2022). Bozulmuş içerik, hibrit hesaplar, kimliğe bürünme, web tugayları ve hicivli haberleri içerir ve hepsi izleyicileri yanıltmak için bilişsel güvenlik açıklarından ve sosyal dinamiklerden yararlanır (Caled ve Silva, 2022). Kötü düzenlenmiş sosyal medya ortamları, dezenformasyonun hızla yayılmasını kolaylaştırır. Bu platformlar genellikle yanlış bilgilerin yayılmasını önlemek için gerekli kontrollerden yoksundur (Rubin, 2019). Doğrulama yanlılığı gibi bilişsel ön yargıları, bireyleri dezenformasyonu gerçek olarak kabul etmeye daha duyarlı hâle getirir. Bireyler inançlarıyla uyumlu içeriği paylaştıkça dijital platformların algoritmaları bu içeriği güçlendirir ve doğrulama yanlılığını sosyal ağlarda daha da sağlamlaştırır. Bu, yanıltıcı anlatıların ilgi çektiği ve hızla yayıldığı bir yankı odası etkisi oluşturur (Caled ve Silva, 2022). İnternet erişiminin yaygınlığı ve sosyal medya platformlarının hızla büyümesi, dezenformasyon kampanyalarının benzeri görülmemiş bir erişim ve etki kazanmasına olanak sağlamıştır. Bilgi yayılımının genellikle editoryal denetime ve profesyonel doğrulama süreçlerine tabi olduğu geleneksel medyanın aksine, dijital platformlar, genellikle hiçbir denetim veya doğrulama olmaksızın sahte anlatıların hızla yayılabileceği büyük ölçüde düzenlenmemiş bir alan sağlar (Vosoughi, Roy ve Aral, 2018). Sosyal medyada algoritma tabanlı içerik teslim sistemleri genellikle sansasyonel, tartışmalı veya duygusal açıdan yüklü bilgileri önceliklendirir

(McLoughlin ve Brady, 2024). Sosyal medyanın algoritmik yapısı, yapay zekânın gerçekliği çarpıtan ve anlaması son derece güç olan dezenformatik veri üretme kapasitesiyle birleştiğinde, toplumun bilgi ekosistemi için ciddi bir tehdit oluşturmaktadır. Yapay zekâ teknolojileri, yanıltıcı içerik oluşturmak ve halkın algısını manipüle etmek için algoritmalardan yararlanarak dezenformasyonun oluşturulmasını ve yayılmasını benzeri görülmemiş ölçeklerde kolaylaştırır (Cybenko ve Cybenko, 2018). Dezenformasyon, bilgi düzensizliği (information disorder) çerçevesinde incelenen üç ana kategoriden biridir (Wardle ve Derakhshan, 2017). Bu çerçeveye göre bilgi düzensizliği; dezenformasyon, mezenformasyon ve malenformasyon olarak üçe ayrılmaktadır. Dezenformasyon, kasıtlı olarak yanlış bilgi üretilmesi ve yayılması anlamına gelirken mezenformasyon, yanlış bilgilerin farkında olunmadan ve kasıtsız bir şekilde paylaşılmasını ifade eder (Kandel, 2020). Buna karşın malenformasyon, doğru bilgilerin bağlamından koparılıp manipüle edilerek zarar verme amacıyla kullanılmasıdır (Omoregie, 2021). Örneğin, özel bilgilerin ifşa edilmesi veya doğru bir haberin yanlış bir algı oluşturacak şekilde çarpıtılması, malenformasyona örnek gösterilebilir. Bu üç bilgi düzensizliği türü, toplum üzerinde farklı etkiler oluşturur ve demokratik süreçlere zarar verebilir. Yapay zekâ teknolojileri ise bu düzensizlikleri daha karmaşık hâle getirebilir. Özellikle, yapay zekâ destekli algoritmalar, dezenformasyonu daha inandırıcı bir hâle getirerek mezenformasyon gibi yayılma riskini artırabilir.

### 3. Yapay Zekâ Destekli Dezenformasyon Kampanyaları

Yapay zekâ, insan beyninin karmaşık yapısına benzer şekilde hareket eden, bir sonuca varma süreçlerinde makinenin kullanılmasını sağlayan; belirli algoritmalarla bir insanın yapabileceği çeşitli işlevleri taklit etmek üzerine kurulu bir sistemdir (Turğal ve Küçükerdoğan, 2023). Yapay zekâ teknolojileri, özellikle Doğal Dil İşleme (NLP), makine öğrenimi algoritmaları ve Generative Adversarial Networks (GAN - Çekişmeli Üretici Ağlar) gibi gelişmiş yöntemlerle dezenformasyonu yeni bir boyuta taşımıştır. Bu teknolojiler, dijital bilgi ortamında yeni tehditler oluşturarak, daha hedefli, verimli ve büyük ölçekli dezenformasyon operasyonları için zemin hazırlamaktadır. Yapay zekâ destekli dezenformasyon kampanyaları, politik mesajların kişiselleştirilmesi ve kitle segmentasyonu gibi stratejilerle, toplumsal ve siyasi süreçleri manipüle etmek için kullanılmaktadır (Islas, Gutiérrez ve Arribas, 2024). Özellikle belirli bir eğitim verileri kümesinden daha özgün yeni veriler üretmek için iki

sinir ağını birbirleriyle rekabet edecek şekilde eğiten GAN'lar, sosyal medya platformlarında dağıtılabilen son derece gerçekçi ve yanıltıcı içerikler üretmekte, böylece propagandacıların gerçek zamanlı manipülasyon yapmasına olanak tanımaktadır (Na vd., 2024). Dezenformasyon oluşturmada Doğal Dil İşleme (NLP) ve makine öğreniminin kullanılması, özellikle insan benzeri metin üretebilen büyük dil modellerinin (LLM'ler) ortaya çıkmasıyla artan bir endişe kaynağıdır. Bu teknolojiler, dezenformasyonu benzeri görülmemiş bir ölçekte üretme ve yayma potansiyeline sahiptir, bu da bilgi bütünlüğü ve toplumsal güven için önemli zorluklar ortaya çıkarabilir. ChatGPT gibi LLM'ler, dezenformasyon amacıyla silahlandırılabilecek ilgi çekici, bağlama duyarlı içerik oluşturma yeteneğini göstermiştir. Bu modeller metin, görüntü, ses ve video üretebilir ve bu da onları yanlış bilgileri yaymak için çok yönlü araçlar hâline getirebilir (Barman, Guo ve Conlan, 2024).

Yapay zekâ, dezenformasyon kampanyalarında yalnızca yanlış bilgilerin oluşturulması ve yayılmasında değil, aynı zamanda deepfake gibi başlı başına kamuoyunu manipüle etme potansiyeli ile de kilit bir rol oynamaktadır. Yapay zekâ teknolojileri, insan yazısını taklit eden sentetik metninler oluşturarak bireylerin gerçek ve sahte haberleri ayırt etmesini zorlaştırır. Bu yetenek, özellikle dış politika gibi hassas alanlarda kamuoyunu etkilemek için kötü niyetli aktörler tarafından kullanılmaktadır (Kreps, McCain ve Brundage, 2020).

Dezenformasyon kampanyaları hakkındaki literatür incelendiğinde; deepfake (Derin Sahte), botlar (Sahte Hesaplar), astroturfing (Yapay Taban Hareketi), bad grammar (Kötü Dil Bilgisi - Yanıltıcı Dil Kullanımı), doppelganger (İkiz Kimlik - Çift Karakter Oluşturma), spamouflage (Spam Kamuflajı) ve zero zeno (Hiçlik Etkisi) gibi kampanyalarının ön plana çıktığı görülmektedir (Arce, 2024; Bontridder ve Poullet, 2021; Ferrara vd., 2016; Fraga-Lamas ve Fernandez-Carames, 2020; García Serrano, Romero-Rodríguez ve Hernando Gómez, 2019; Weber ve Neumann, 2021). Bu stratejilerin yapay zekâ teknolojileriyle entegrasyonu, dezenformasyonun etkisini daha sofistike ve geniş kapsamlı hâle getirmektedir.

Astroturfing, botlar ve deepfake gibi dezenformasyon yöntemleri, çeşitli kaynaklarda strateji olarak sınıflandırılırken farklı kaynaklarda kampanya olarak sınıflandırılmıştır (Keller vd., 2020; Kovic vd., 2018; Rossetti ve Za-

man, 2023; Ruffin vd., 2024). Araştırmamıza konu olan OPEN AI raporunda kampanya olarak sınıflandırıldığından bizde kampanya olarak bahsedeceğiz (Nimmo, 2024). Bu yöntemler, bilgi manipülasyonu amacıyla kullanılan spesifik araçlar ve tekniklerdir. Hem strateji hem kampanya kategorisine girmelerinin sebebi, bu yöntemlerin tekil olarak uygulanabilmesi (strateji) ve daha geniş çapta bir dezenformasyon kampanyasının parçası olabilmeleridir (Chesney ve Citron, 2018; Ferrara vd., 2016; García Serrano, Romero-Rodríguez ve Hernando Gómez, 2019; Keller vd., 2020). Bu özelliklerinden dolayı Astroturfing, botlar ve deepfake konularını çalışmamızda kampanyalar başlığı altında incelenmiştir.

### 3.1. Bad Grammar

Kötü dil bilgisi ve dezenformasyon arasındaki ilişki, dilsel ve sosyal bağlamlarda karmaşık bir yapıya sahiptir. Dezenformasyonun hedefi, dilsel hatalar da dâhil olmak üzere dili manipüle ederek algıları yanıltmak ve yanlış bilgilerin yayılmasını sağlamaktır. Özellikle dil bilgisi hataları, dezenformasyona özgünlük katıyor gibi görünse de aslında gerçeği gizlemek ve bilgiyi eleştirel bir şekilde değerlendirmeyi zorlaştırmak için kullanılır (Nefedova ve Samkova, 2019). Standart dil bilgisinden sapmalar bilginin doğruluğuna olan güveni aşındırarak okuyucunun eleştirel düşünme yetisini zayıflatabilir (Francis, 2018). Bu durum, çok dilli dezenformasyonlarda da karşımıza çıkar; dil engellerini aşan yanlış bilgiler, dil bilgisi hatalarıyla daha da çarpıtılarak gerçeklerin kontrol edilmesini zorlaştırmaktadır (Quelle vd., 2023). Dezenformasyonda kullanılan dil bilgisi hatalarının arkasında derin bilişsel sınırlamalar vardır. Karmaşık gramer yapılarının anlaşılmasındaki bilişsel sınırlamalar, bu hataların kullanıldığı stratejilerde, yanlış bilgilerin daha kolay yayılmasına neden olabilmektedir (Chambers, 2010). Ayrıca, kötü dil bilgisi ile dezenformasyonun kullanılması, bilginin güvenilirliğini azaltabilir. Ancak, aynı zamanda belirli hedef kitlelerle daha ilişkilendirilebilir hâle gelebilir (Clos vd., 2023). Bu nedenle, kötü dil bilgisi içeren dezenformasyon, özellikle sosyal medya platformlarında geniş kitleler üzerinde etkili olabilmektedir. Rusya'nın sosyal medya platformları üzerinden kasıtlı dil bilgisi hatalarıyla dezenformasyon yayması buna bir örnektir (Nimmo, 2024). Benzer bir şekilde, İsrail merkezli Stoic adlı siyasi firma da ABD'deki öğrenci protestolarını antisemitik olarak göstermek amacıyla sahte sosyal medya hesapları üzerinden dezenformasyon yaymış bu operasyonlarda yapay zekâ destekli

dil bilgisi hatalarını kullanmıştır (Jingnan, 2024). Dezenformasyonu tespit etmek her zaman kolay değildir çünkü dilsel sapmalar her zaman aldatmak için yapılmayabilir. Zayıf dil bilgisi, bazı durumlarda yalnızca yazarın dilsel geçmişini veya bilişsel sınırlarını yansıtabilir. Bu yüzden, dil bilgisi analizi dezenformasyonun tespitinde faydalı olsa da tek başına yeterli değildir ve diğer yöntemlerle desteklenmelidir (Mosleh, Cole ve Rand, 2024). Üretici AI teknolojileri, dezenformasyonu yaymak için seçim süreçlerinde giderek daha fazla kullanılmaktadır. Bu teknolojiler daha ikna edici ve dil bilgisi açısından doğru içerik oluşturabilir ve hâlkın yanlış bilgileri meşru kaynaklardan ayırt etmesini zorlaştırır (Ustinovich, 2024). Yapay zekâ tarafından oluşturulan içeriğin karmaşıklığına rağmen bazı dilsel ipuçları hâlâ dezenformasyonun göstergesi olarak hizmet edebilir. Örneğin, 2019 Hong Kong protestoları bağlamında, potansiyel dezenformasyon kampanyalarını belirlemek için çevrim içi yorumlardaki metinsel özellikler kullanıldı. Bu ipuçları, her zaman kesin olmasa da devam eden dezenformasyon çabalarının durumsal olarak indekslenmesine yardımcı olur (Deschrijver, 2024).

### 3.2. Doppelganger

Doppelganger, genellikle dezenformasyonu yaymak amacıyla tek bir varlığın çok sayıda takma isimli kimlik kullanarak geniş çapta destek veya fikir birliği yanılsaması oluşturma çabası olarak tanımlanabilir (Pennekamp vd., 2019). Doppelganger stratejisi, dezenformasyon kampanyalarında kimlik manipülasyonunu kullanarak kamuoyunu yanıltma amacıyla tasarlanmış karmaşık bir tekniktir. Bu yöntem, takma isimli hesapların tek bir varlık tarafından yönetildiği ve sahte bir destek ya da fikir birliği oluşturulduğu bir senaryo oluşturmayı amaçlar. Doppelganger stratejisi aktörlerin mesajlarını güçlendirmelerine ve yanlış bir fikir birliği veya meşruiyet duygusu oluşturmalarına olanak tanır (Martin, 1982). Doppelganger bilgi bütünlüğünü zayıflatır ve demokratik süreçler için ciddi tehditler oluşturur (Pennekamp vd., 2019). Doppelganger operasyonları, yapay zekânın gerçeğe piksel düzeyinde yakın sonuçlar üreten yapısıyla birlikte, tespiti oldukça zor olan bir hâl almıştır. Çünkü bu sahte kimlikler, verilerin anonimleştirilmesi ve platformlarda çok sayıda sahte hesapla oluşturulan içeriklerin kolayca yayılmasıyla daha karmaşık hâle gelir. Ancak, stilometrik analiz gibi teknikler, dilsel özelliklerin izini sürerek ve yazım tarzlarındaki benzerlikleri analiz ederek bu tür dezenformasyon yöntemlerinden korunmak için kullanabilir görün-

mektedir (Pennekamp vd., 2019). Dezenformasyon yöntemi olarak doppelganger'ı uygulamalarından ilk dikkat çeken Çin'in Hong Kong protestoları sırasında gerçekleştirdiği uygulamalardır. Burada sahte sosyal medya hesapları ile protestocular itibarsızlaştırmaya çalışılmış, Çin yanlısı sahte hesaplar, protestocular hakkında yanlış bilgi yayarak uluslararası kamuoyunda olumsuz bir algı amaçlanmıştır (Howard ve Bradshaw, 2018). Rusya merkezli bir Doppelganger dezenformasyon kampanyası da dikkat çekicidir. Bu operasyonun, meşru medya ve hükûmet sitelerinin kopyalarını oluşturup sahte içeriklerle Ukrayna karşıtı ve Rusya yanlısı propagandası yaptığı iddia edilmektedir (EEAS, 2024). 2016 ABD başkanlık seçimlerinde Rusya merkezli gruplar, sahte sosyal medya hesapları aracılığıyla aynı stratejiyi kullanırlar. Özellikle Facebook ve X (Twitter) platformlarında sahte kimliklerle yönetilen hesaplar, siyasi içerik üreterek ABD halkını yanıltıcı bilgilerle etkileyerek seçim sürecine müdahale etmeye çalışmışlardır (Howard ve Bradshaw, 2018).

### 3.3. Spamouflage

Spamouflage, toplu hesap oluşturma ve düşük kaliteli, yüksek hacimli içerik üretimi de dâhil olmak üzere spam içeren tekniklerin kullanılmasıyla karakterize edilir. Bu işlem, bağımsız araştırmacılar tarafından defalarca ortaya çıkarılarak çeşitli platformlarda içerik denetlemesi üzerindeki etkisini vurgulanmıştır (François ve Douek, 2021). Spamouflage stratejisi, dezenformasyonun geniş çapta yayılmasını sağlamak için spam benzeri taktiklerin dezenformasyon içerikleriyle entegre edilmesini ifade eder. Bu strateji, özellikle dezenformasyon kampanyalarının görünürlüğünü artırmak ve yanlış bilgilerin kökenini gizlemek amacıyla kullanılır. Spamouflage, yanlış bilgi yayılımının kaynağını gizleyerek dezenformasyonun tespit edilmesini zorlaştırır ve kitleler üzerinde güçlü bir etki oluşturabilir (Aboutayeb, 2023). Spamouflage kampanyası, sahte sosyal medya hesapları ve botlar aracılığıyla içeriklerin hızlıca yayılmasını sağlar. Bu dezenformasyon kampanyası, özellikle Reddit, Telegram ve YouTube gibi platformlarda kullanılan bir dezenformasyon yöntemidir (Dias, Lopes ve Borges, 2024). Sahte içerikler, karakter düzeyinde sinir ağları ve transfer öğrenimi teknikleri ile üretilir ve spam benzeri dağıtım taktikleri ile desteklenir. Örneğin, Çinli aktörler bu stratejiyi kullanarak İngilizce, Japonca ve Korece içerikler üretmiş ve çeşitli sosyal medya platformlarında yaymıştır (Dias, Lopes ve Borges, 2024). Özellikle sosyal medya platformlarının dağınık yapısı, bu tür dezenformasyon kampanyaları için

verimli bir zemin sağlar. Spamouflage, bu kaosu artırarak yanlış bilgilerin daha geniş kitlelere yayılmasını sağlar (Renedo, Farpón ve García, 2023). Spamouflage stratejisi, özellikle Çin'in Batı kamuoyunu etkileme amacıyla gerçekleştirdiği dezenformasyon kampanyalarında dikkat çekmiştir. Çinli aktörler, spam benzeri teknikleri kullanarak Batı karşıtı içerikler üretmiş ve bu içerikleri sosyal medya platformlarında geniş kitlelere yaymıştır. Bu strateji, ABD ve Avrupa'da siyasi süreçleri manipüle etmek amacıyla sıkça kullanılmıştır (Arce, 2024). Ayrıca, İran'ın da benzer şekilde dezenformasyon kampanyalarını yönettiği ve ABD karşıtı içerikler ürettiği bilinmektedir (Innes, 2020). Özellikle sosyal medya platformlarında yaygın olarak kullanılan bu strateji, demokratik süreçleri ve kamuoyunu manipüle etme kapasitesine sahiptir. Tespit ve karşı koyma stratejileri ise genellikle sinir ağları ve transfer öğrenimi gibi ileri düzey yapay zekâ tekniklerine dayalıdır (Dhamani vd., 2019).

### 3.4. Zero Zeno

İsrail merkezli Stoic şirketi tarafından yürütülen, OpenAI ve Meta tarafından tespit edilen yapay zekâ destekli dezenformasyon kampanyasıdır. OpenAI'nin bu dezenformasyon kampanyasını "Zero Zeno" olarak adlandırır. Bu isimlendirme ilginç bir felsefi bağlantıya işaret etmektedir. Zenon, Antik Yunan'da Stoacılık felsefesinin kurucusu olarak bilinir. Stoacılık, mantık, etik ve doğa felsefesi üzerine odaklanan bir düşünce okuludur (Mansfeld, 1978). "Zero Zeno" ismi, muhtemelen kampanyanın etik dışı doğasına ve Stoacılık felsefesinin etik prensipleriyle olan çelişkisine bir gönderme olabilir. Ayrıca, "zero" (sıfır) kelimesi, kampanyanın gerçek olmayan, yapay doğasını vurgulayabilir. Bu isimlendirme, dezenformasyon kampanyalarının etik boyutlarına dikkat çekmeyi amaçladığı düşünülmektedir.

Zero Zeno kampanyası, modern dezenformasyon tekniklerinin gelişmiş bir örneğini teşkil etmektedir. İsrail merkezli Stoic şirketi tarafından yürütülen bu operasyon, yapay zekâ teknolojilerini kullanarak çeşitli sosyal ve politik konularda kamuoyunu etkilemeyi amaçlamıştır (OpenAI, 2024). Kampanya, özellikle OpenAI'nin dil modellerini kullanarak metin içeriği üretmiş ve Generative Adversarial Networks (GAN) teknolojisinden faydalanarak sahte profil fotoğrafları oluşturmuştur. Stoic, Zero Zeno operasyonu kapsamında ürettiği içerikleri yaymak için Facebook, Instagram, X (eski adıyla Twitter), YouTube ve Telegram gibi çeşitli sosyal medya platformlarını kullanmıştır

(Yang, 2024). Kampanyanın hedef kitleleri arasında Filistin-İsrail çatışması bağlamında, Hamas karşıtı ve İsrail yanlısı görüşleri benimseyenler, Hindistan seçimleri gibi spesifik politik olaylarla ilgilenenler ve farklı ülkelerdeki yerel vatandaşlar, Yahudi öğrenciler ve Afrikalı Amerikalılar bulunmaktadır (Yang, 2024). OpenAI'nin raporundan bir gün sonra, Meta da STOIC şirketinin yürüttüğü Zero Zeno kampanyası hakkında bir rapor yayınlamış ve bir dizi önlem aldığını açıklamıştır. Meta bu kampanyanın Vietnam'dan satın alınmış takipçi ve beğeniler kullanarak etkileşimini artırmaya çalıştığını da tespit etmiştir. Meta'nın Adversarial Threat Report'unda (Franklin vd., 2024) belirtildiğine göre; 510 Facebook hesabı, 11 sayfa, 1 grup ve 32 Instagram hesabının silinmiş, Stoic şirketinin Meta platformlarından yasaklanmış, Stoic'e, Meta politikalarını ihlal eden faaliyetlerini derhal durdurması için bir "cease-and-desist" (dur ve vazgeç) mektubu gönderilmiştir.

Zero Zeno kampanyası, diğer bilinen dezenformasyon kampanyalarıyla karşılaştırıldığında bazı benzerlikleri ve farklılıkları ortaya koymaktadır. Örneğin, 2016 ABD başkanlık seçimleri sırasında Rus Internet Araştırma Ajansı (IRA) tarafından yürütülen kampanya ile bazı benzerlikler göstermektedir. Her iki kampanya da sosyal medya platformlarını yoğun şekilde kullanmış ve hedef kitleleri manipüle etmeye çalışmıştır (Mueller, 2019). Ancak, Zero Zeno kampanyası, yapay zekâ teknolojilerinin yoğun kullanımıyla öne çıkmaktadır. Bu, kampanyanın daha hızlı ve geniş ölçekte içerik üretmesine olanak sağlamıştır. Ayrıca, GAN teknolojisinin kullanımı, sahte profillerin daha inandırıcı olmasını sağlamıştır. Zero Zeno kampanyası, yapay zekâ destekli dezenformasyon kampanyalarının potansiyel tehlikelerini ve karmaşıklığını gözler önüne sermektedir.

### 3.5. Deepfake

Yapay zekâ bağlamında dezenformasyon açısından en kritik konulardan biri deepfake teknolojileridir. Deepfake, yapay zekâ kullanılarak tamamen sahte ancak son derece gerçekçi video ve ses içerikleri üretmek için kullanılmasıdır. Bu teknolojiler, gerçek insanların hiç yapmadıkları şeyleri söyledikleri veya yaptıkları gibi görünen videolar üretebilir ve böylece son derece ikna edici ve çürütülmesi zor sahte görsel kanıtlar oluşturabilir (Chesney ve Citron, 2018). Deepfake'ler, politikadan ekonomiye, sosyal medyadan kişisel mahremiyete kadar geniş bir yelpazede dezenformasyon aracı olarak kullanılır. Bu nedenle, deepfake teknolojisi, dijital dezenformasyonun psikolojik ve sosyal etkile-

rini derinleştirirken, aynı zamanda halkın bilgiye olan güvenini sarsar.

Deepfake teknolojisi, hibrit savaş stratejilerinin bir parçası olarak kullanılmakta ve dezenformasyon kampanyalarını güçlendirmek amacıyla sahte medya içerikleri üretmektedir. Özellikle politik ve toplumsal olaylarda deepfake kullanılarak toplumu yönlendirme ve yanlış bilgiyi yayma amacı güdülmektedir. Bu içerikler, sosyal medya platformları aracılığıyla hızla yayılır ve geniş kitleler tarafından gerçekmiş gibi algılanabilir (Havlík, 2023). Deepfake'lerin sosyal medya üzerindeki etkileri, dezenformasyonun yayılmasını kolaylaştırmakla kalmaz, aynı zamanda kamuoyunu derinden etkileyen sahte haberlerin ve bilgilerin oluşmasına neden olur. Bu bağlamda deepfake hem siyasi hem de ekonomik hedeflere ulaşmak için güçlü bir araç hâline gelir. Özellikle seçim süreçlerinde kullanılan deepfake videoları, politik liderlerin sahte konuşmalarını üreterek halkı yanıltmayı hedeflediği tespit edilmiştir. 2020 ABD başkanlık seçimlerinde deepfake'ler, politikacıların itibarını sarsmak ve seçim sonuçlarını etkilemek amacıyla kullanılmıştır (Starbird, DiResta ve DeButts, 2023). Yalnızca siyasi alanı değil, ekonomik alanı da etkileyen bu teknolojiler kullanılarak sahte CEO konuşmaları üretilmiş ve bu durum piyasalarda dalgalanmalara yol açmıştır (Montasari, 2024). Deepfake aynı zamanda bireylerin özel hayatlarını hedef alarak mahremiyet ihlallerine neden olabilir (R. Kumar vd., 2024). Bu teknoloji, toplumsal bölünmeleri derinleştirir ve halk arasında yanlış anlatıların yayılmasına yol açar. Sahte içeriklerin, özellikle politik süreçlerde toplumu manipüle etme potansiyeli, toplumsal çıkarımların ne kadar geniş olabileceğini göstermektedir (Shoaib vd., 2023).

Deepfake teknolojisinin dijital dezenformasyon alanındaki etkilerini azaltmak için çeşitli stratejiler geliştirilmiştir. Bu stratejiler arasında gelişmiş AI tabanlı tespit algoritmaları, blockchain teknolojileri ve dijital okuryazarlık programları yer alır. Bu teknolojiler, deepfake içeriklerinin doğrulanmasını sağlamak ve yanlış bilgilerin yayılmasını engellemek amacıyla kritik bir rol oynar (R. Kumar vd., 2024). Deepfake algılamadaki son gelişmeler, gerçek ve sahte video gömmeleri arasındaki özellik alanı mesafesini artıran metrik öğrenmeden yararlanmaktadır. Özellikle üçlü bir ağ mimarisi kullanan bu yaklaşım, sosyal medya platformlarında yaygın olan yüksek sıkıştırma senaryolarında bile deepfake'leri sınıflandırmada yüksek etkinlik gösterir. Yöntem, Celeb-DF veri setinde %99,2'lik son teknoloji bir AUC puanı ve yüksek oranda sıkıştırılmış bir Nöral Doku veri setinde %90,71 doğruluk elde etmiştir (A.

Kumar, Bhavsar ve Verma, 2020). Başka bir yenilikçi yöntem, deepfake videolarda doğru bir şekilde kopyalanması zor olan insan göz kırpma modellerini analiz etmeyi içerir. Bu biyometrik özellik, gerçek ve sentetik medya arasında ayrım yaparak video içeriğindeki anormallikleri tanımlamaya yardımcı olur (V. Baravkar vd., 2023). Ölçek-Değişmeyen Özellik Dönüşümü (SIFT) özelliklerinin deepfake analizinde kullanımı da araştırılmıştır. SIFT anahtar noktaları, durağan görüntülerdeki veya video kliplerdeki görsel bilgileri analiz ederek derin sahteleri belirlemede değerli olabilir (Dordevic, Milivojevic ve Gavrovska, 2019). Aynı zamanda Deepfake araştırmasının bibliyometrik analizi; deepfake'lerin eğilimleri, uygulamaları ve zorlukları hakkında derinlemesine tespitler sağlayarak gelecekteki araştırmalar için değerli rehberlik sunabilir (Garg ve Gill, 2024).

### 3.6. Astroturfing

Diğer bir sofistike dezenformasyon tekniği ise astroturfing'dir. Astroturfing, bağımsız görünen ancak aslında şirketler veya siyasi kuruluşlar tarafından desteklenen sahte taban örgütleri veya kampanyaları oluşturmayı içerir. BBu örgütler, gerçek sosyal hareketleri taklit ederek kamuoyunu etkilemeyi amaçlar (Cho vd., 2011). Bu teknikte, genellikle sahte sosyal medya hesapları, ücretli aktörler veya influencer'lar kullanılarak kamuoyunda bir fikir birliğine varıldığı ya da geniş bir destek sağlandığı algısı oluşturulur. Bu dezenformasyon stratejisi, insanların benzer görüşlü oldukları insanlara ve topluluk temelli hareketlere duyduğu güveni kötüye kullanarak, onları yanlış bilgi sürecine dâhil eder. 2016 ABD başkanlık seçimlerinde de benzer stratejiler kullanılmış ve belirli politik adayların lehine sahte sosyal medya kampanyaları yürütülerek kamuoyunun algısı manipüle edilmiştir (Keller vd., 2020). Astroturfing sürecinde, sahte sosyal medya hesapları, sahte yorumlar ve içerikler kullanılarak halk arasında yaygın bir destek izlenimi oluşturulur (Chan, 2024; Zerback ve Töpfl, 2022). Astroturfing'in en tehlikeli yönlerinden biri, insanların çoğunluk görüşüne uyma eğiliminden yararlanarak dezenformasyonu hızla yaymasıdır. Özellikle, bu uygulama uluslararası siyasi arenada, propaganda ve yabancı müdahaleler için tehlikeli bir araç olabilir (Keller vd., 2020; Zerback ve Töpfl, 2022). Astroturfing'in karmaşık yapısı, onu tespit etmeyi zorlaştırmaktadır. Sosyal medya platformlarında, özellikle koordineli davranışları gizlemek için kitle kaynak kullanımı ve anonim hesaplar gibi yöntemler yaygındır. Bu tür sahtekârlıkları tespit etmek için algoritmik çö-

zümler geliştirirken, bu çözümler genellikle astroturfing'in kökenini tamamen ortaya çıkarmakta yetersiz kalmaktadır (Wu ve Liu, 2017).

### 3.7. Botlar

Yapay zekâ destekli botlar ve otomatik sistemler, dijital çağda ortaya çıkan başka bir dezenformasyon biçimini temsil eder. Botlar, insan davranışını taklit edecek şekilde programlanmıştır ve sosyal medyada içerik paylaşma ve yayınlama yetenekleri, insan yeteneklerinin çok ötesindedir. Bu şekilde, yaygın bir fikir birliği veya muhalefet yanılsaması oluşturabilir, trend konuları manipüle edebilir ve belirli demografik grupları hedefleyen dezenformasyonlarla etkileyebilir (Ferrara vd., 2016). Sosyal botlar, dijital dezenformasyonun yayılmasında önemli bir araç hâline gelmiştir. Bu otomatik programlar, özellikle X (Twitter) gibi sosyal medya platformlarında dezenformasyon içeriklerini hızla yayarak kamuoyunu manipüle eder. Botlar, retweet veya paylaşım gibi tekrarlanan görevleri çok hızlı bir şekilde gerçekleştirebilir ve bu sayede dezenformasyonun hızla geniş kitlelere ulaşmasına olanak tanır. Özellikle ABD Başkanı Donald Trump'ın görevden alınması sırasında, sosyal botlar görevden alma ile ilgili içeriklerin %31'ini oluşturarak dezenformasyonun yayılmasında önemli bir rol oynamıştır (Rossetti ve Zaman, 2023). Sosyal botlar, kamuoyunu şekillendirmede kritik bir öneme sahiptir. Botlar, özellikle yankı odaları içinde yanlış bilgileri yayarak izole topluluklar üzerindeki etkilerini artırır. QAnon gibi hareketlerde, dezenformasyonun hızlı ve etkili bir şekilde yayılması için bu botlar kullanılmıştır. Botlar, ayrıca seçimler ve siyasi krizler gibi hassas dönemlerde organize dezenformasyon kampanyalarının bir parçası olarak insan hesaplarıyla koordineli çalışarak dezenformasyonun daha doğal görünmesini sağlar (Woolley 2022). Sosyal botların giderek daha karmaşık hâle gelmesi, dezenformasyon kampanyalarında yeni bir dönemin habercisi olmuştur. Artık yapay zekâ ve insan etkinliğiyle birleştirilen "yarı organik" kampanyalar daha yaygın hâle gelmektedir. Ancak, bu botların etkisiyle mücadele etmek için dijital okuryazarlık ve yapay zekâ tabanlı tespit araçları geliştirilmiştir. Bu çabalar, dezenformasyonun yayılmasını önlemek için hayati önem taşır (Maathuis, Janssens, ve Rahimi 2024).

## 4. OpenAI Tehdit İstihbarat Raporu Doğrultusunda Yapay Zekâyı Dezenformasyon Yaymak için Kullanan Ülkeler

OpenAI tarafından hazırlanan Tehdit İstihbarat Raporu ülkelerin yapay zekâ

kullanarak gerçekleştirdikleri dezenformasyon kampanyalarını ortaya koymaktadır (Nimmo, 2024). OpenAI raporuna göre Rusya (iki ağ), Çin, İran ve İsrail'deki ticari bir şirketteki operatörlerle bağlantılı kampanyalar gerçekleştirmiştir. Rusya'dan "Kötü Dil bilgisi" olarak adlandırdığımız, esas olarak Telegram'da faaliyet gösteren ve Ukrayna, Moldova, Baltık Devletleri ve ABD'yi hedef alan daha önce bildirilmemiş bir dezenformasyon kampanyası ile İnternette Ukrayna hakkında içerik yayınlayan ve "Doppelganger" olarak bilinen bir kampanya yürütmüşlerdir. Çin kendi politikalarını olumlu olarak ön plana çıkarmak ve eleştirileri baskılamak için spamouflage kampanyası gerçekleştirmiştir. İran ise İran'ı destekleyen ve İsrail ile ABD'yi eleştiren web içeriklerini Uluslararası Sanal Medya Birliği (IUVM) üzerinden yapay zekâ kullanarak gerçekleştirmişlerdir. Aynı dönemde İsrail'de Stoic isimli bir şirket Gazze'de yaşanan katliam ve Hindistan'da yapılan seçimler üzerinde dezenformasyon oluşturacak içerikleri yapay zekâ kullanarak üretmiştir (Nimmo, 2024, s. 6).

Rapor, yapay zekânın dezenformasyonun operasyonel kapasitesini artırma noktasındaki kritik rolünü vurgulamaktadır. Rapor, gizli operasyonların etkililiğini değerlendirmek için bir "Breakout Scale" (Etkililik Ölçeği) sunmuştur. Bu ölçekte operasyonlar, 1'den 6'ya kadar sıralanmış ancak hiçbir operasyon 2'nin üzerine çıkamamıştır. Rapordaki bu bulgu, yapay zekânın içerik üretme kapasitesini önemli ölçüde artırsa da bu içeriklerin toplumsal düzeyde anlamlı bir etkileşim oluşturmakta başarısız olduğudur (Nimmo, 2024, s. 6). Bir yapay zekâ platformunun, yapay zekâ kullanılarak dezenformasyon kampanyası yürütülmesi konusunda yayınladığı raporda sınırlı etki iddiası ironik ve taraflı ancak anlaşılır bulunabilir.

Rapora göre Bad Grammar ve Zero Zeno, modellerimizi daha sonra Telegram, X, Instagram ve diğer sitelerde yayınlanan büyük miktarlarda kısa yorumlar oluşturmak için kullanılmıştır. IUVM adına hareket eden kişiler, İngilizce ve Fransızca daha uzun makaleler oluşturmak ve düzeltmek için OpenAI dil modelini kullanmışlardır. Spamouflage ve Doppelganger kampanyalarında, ChatGPT hem nitelik hem de nicelik açısından kullanılmış; dil bilgisi hatalarını düzeltilmiş ve aynı zamanda çeşitli dillerde düzinelerce kısa yorum oluşturulmuştur (Nimmo, 2024, s. 7). Raporda sahte etkileşim (fake engangament) için yapay zekâ platformunun kullanıldığı aktarılmaktadır. Zero Zeno kampanyasında, Instagram ve X'te Gazze konusu başta ol-

mak üzere belirli temalar hakkında kısa metinler yayınlanmıştır. Bu metinler modeller (ChatGPT) kullanılarak oluşturulmuştur. Bu platformlardaki başka bir hesap grubu daha sonra yine bu işlem tarafından oluşturulan yorumlarla yanıt vermiştir (Nimmo, 2024, ss. 7, 31-33). Benzer şekilde, Spamouflage kampanyası yapay zekâ desteği ile kullanılmış, X'e Çinli muhalif Cai Xia'yı eleştiren kısa yorumlar yayınlanmıştır. Bunlar bir ilk gönderi ve bir dizi yanıt şeklindedir (Nimmo, 2024, s. 25). Her Konuşmadaki yorumlar, ChatGPT kullanılarak yapay oluşturulmuştur. Rapora göre çok sayıda sosyal medya gönderisinin, özellikle de Çince gönderilerin duygularını özetlemek ve analiz etmek için yapay zekâ kullanılmıştır. Analizler sonrası ortaya çıkan sonuçlar doğrultusunda web sitesi etiketleri oluşturmak için yapay zekâ kullanıldığı belirlenmiştir (Nimmo, 2024, ss. 8-9).

### 4.1. Rusya: Bad Grammar ve Doppelganger Kampanyası

Rapora göre Bad Grammar kampanyası, kimliği belirsiz bir Rus tehdit aktörü/aktörleri tarafından yürütülmüştür. Aktör/aktörler, Ukrayna, Amerika Birleşik Devletleri ve Baltık Devletleri gibi çeşitli ülkeleri hedef alarak özellikle Telegram üzerinden politik yorumlar yayınlanmıştır. Operasyon, dil bilgisi hatalarıyla dolu İngilizce kullanılmasıyla karakterize edilmiştir. Bu yolla politik tartışmaları kışkırtmak veya kafa karışıklığı oluşturmak amacıyla yerel aktivistler veya yorumcular gibi görünme çabası olarak yorumlanmıştır (Nimmo, 2024, s. 16). Kampanya aktörleri İngilizce ve Rusça olmak üzere düşük kaliteli ve hacimli içerik üretmek için yapay zekâ araçlarını kullanır. Bu içerik, politik huzursuzluk ekme veya kamuoyunu manipüle etme amacıyla çeşitli Telegram kanallarında sistematik olarak dağıtılmıştır. Rapora göre izleyici etkileşiminde sınırlı başarı elde edildi. İçerik, yolsuzluk, savaş ve ulusal politika gibi politik açıdan yüklü temalar etrafında şekillenmiştir. Kullanıcı etkileşimi veya tartışma oluşturma olasılığı yüksek hassas konulara odaklanma stratejik bir tercih olarak gösterilir. Etkileşimin düşük olması, izleyicilerin içeriğin yapay veya manipüle edilmiş doğasını ayırt edebildiğini veya basitçe kullanıcıların perspektifleri ve ilgi alanlarıyla örtüşmediğini göstermektedir (Nimmo, 2024, s. 17).

Doppelganger Dezenformasyon Kampanyası, çoklu hesap kümeleri kullanılarak daha sofistike bir şekilde yürütülmüştür. İçerik, çeşitli dijital platformlar arasında, özellikle 9GAG ve X'te yayınlanmıştır. Temelde Ukrayna karşıtı

içeriklerin üretildiği bu dezenformasyon kampanyasında yapay zekâ ile üretilen içerik manuel olarak oluşturulan, memler ve yorumlarla harmanlanan çok yönlü bir yaklaşım sergilenmiştir. Bu strateji, içeriğin inandırıcılığını ve farklı sosyal medya manzaralarında etkileşimini artırmayı amaçlar. Kampanya belirgin bir şekilde Ukrayna karşıtı Rusya yanlısı duyguları yükseltmeyi amaçlamıştır. Bu hem metin hem de yanıltıcı veya bağlamsal olarak manipüle edilmiş medya aracılığıyla gerçekleştirildi. "Kötü Dil bilgisi" kampanyası gibi, Doppelganger kampanyası da önemli bir etkileşim veya etki elde edememiştir. Rapora göre bot tarafından yönlendirilen yorumlar ve beğeniler sık sık gerçek kullanıcılar tarafından ifşa edilmesi kampanyanın güvenilirliği azaltmıştır (Nimmo, 2024, ss. 17-22).

### 4.2. Çin: "Spamouflage" Kampanyası

Rapora göre Çin'den kaynaklanan Spamouflage Kampanyası; küresel izleyicileri, özellikle Çin diasporasını ve Çin hükûmetinin eleştirmenlerini hedef almıştır (Nimmo, 2024, s. 23). Operasyon sofistike bir şekilde yürütüldü, içerik birden fazla dilde ve platformda yapay zekâ kullanılarak üretilmiştir. Çin kolluk kuvvetleriyle ilişkili aktörlere atfedilen "Spamouflage" operasyonunda Çince, İngilizce, Japonca ve Korece Batı karşıtı içerikler üreterek bunları sosyal medya platformlarında ve blog sitelerinde hızla yaymıştır. Operasyon ayrıca, Çin hükûmetini eleştiren sosyal medya gönderilerini özetlemek ve duygu analizlerini yapmak için de yapay zekâ araçlarını kullanılmış, böylece mesajlaşma stratejisini geliştirmek hedeflenmiştir. Spamouflage stratejisi, Reddit, Telegram ve YouTube gibi platformlarda etkili olmuştur (Nimmo, 2024, s. 25). Spamouflage, çekici içerik üretmenin yanı sıra, duyarlılık analizi ve içerik yönetimi yapmak için yapay zekâ kullanılmıştır. Bu kampanya stratejilerine teknik entegrasyonun yüksek derecesini işaret etmektedir. İçerik, Çin hükûmetini öven ve karşı olanları eleştiren, Çin hükûmeti ve diplomatik çıkarlarıyla uyumlu temalar üzerine odaklanmıştır. Rapora göre geniş kapsamlı erişimine ve yapay zekâ araçlarının sofistike kullanımına rağmen Spamouflage gerçek izleyicilerle etkili bir şekilde etkileşim kurmada başarısız olmuştur (Nimmo, 2024, ss. 26-27).

### 4.3. İran: Uluslararası Sanal Medya Birliği (IUVM)

Raporda Uluslararası Sanal Medya Birliği (IUVM), 2018'den beri açık kaynak topluluğu tarafından incelenen İran'a bağlı bir kuruluş olarak tanımlanmaktadır. IUVM, özellikle İsrail ve ABD karşıtı mesajlar olmak üzere, İran'ın jeo-

politik anlatılarını destekleyen içerik üretimine odaklanmıştır (Nimmo, 2024, s. 28). Bu kampanya, İngilizce ve Fransızca dâhil olmak üzere birden fazla dilde içerik üretmek için yapay zekâ araçlarını kullanmıştır. IUVM, makaleler, başlıklar ve web sitesi etiketleri oluşturarak çeşitli platformlarda içerik yayınlamıştır. Bu içerikler genellikle bir gün önce hazırlanmış ve IUVM'nin güncel web sitesi iuvmpress.co üzerinde yayınlanmıştır. Bazı web site etiketleri, otomasyon belirtileri göstererek veya kötü düzeltmeler içererek operasyonel sofistikelik eksikliklerini ortaya koymuştur. IUVM tarafından üretilen içerik, genellikle ABD ve İsrail karşıtı olup Filistinlileri, İran'ı ve "Direniş Ekseni"ni övmektedir. İçerikler, politik propaganda yanında, ABD ve İsrail gibi ülkeleri hedef alan yanlış bilgilerle uluslararası kamuoyunu etkilemeyi amaçlamaktadır. IUVM'nin çevrim içi varlığı, sosyal medya platformlarından sürekli kaldırılmalar ve FBI tarafından alan adlarının ele geçirilmesiyle azalmıştır. IUVM'nin web sitesi dışında, 23 Mayıs 2024 itibarıyla TikTok, VKontakte ve Odnoklassniki üzerinde IUVM markalı hesaplar tespit edilmiştir; bu sosyal medya hesaplarının sırasıyla 10, 76 ve 274 takipçisi bulunmaktadır. Operasyon, birden fazla platformda faaliyet göstermesine rağmen, herhangi bir platformda önemli bir kitle etkileşimi veya büyüme sağlayamamıştır. Bu durum, IUVM'nin Kategori 2 operasyon olarak değerlendirilmesine neden olmuştur. Bu bağlamda rapora göre dezenformasyon kampanyasının etkisinin sınırlı kaldığını göstermektedir (Nimmo, 2024).

### 4.4. İsrail: "Zero Zeno" Kampanyası

Zero Zeno kampanyası, İsrail merkezli ve genellikle Hamas ve Katar karşıtı, İsrail yanlısı içerikler üreten bir etki operasyonudur Operasyon, İsrail'de faaliyet gösteren bir siyasi kampanya yönetim firması olan Stoic tarafından yürütüldü. Kampanya, adını Stoacı felsefe okulunun kurucusundan alarak düşük katılım seviyelerini ironik bir şekilde vurgulamaktadır (Nimmo, 2024, s. 31). Bu dezenformasyon kampanyasıyla çok dilli içerikler üreterek geniş bir coğrafi yelpazede etki oluşturmayı amaçladı. Yukarıda açıklanan devlet destekli çabalardan farklı olarak, Zero Zeno belirli jeopolitik anlatıları hedefleyen, örneğin Gazze katliamı, İsrail sendikaları ve Hindistan seçimleri gibi konularda dezenformasyon üreten kiralık bir operasyondu. Kampanya, X, Facebook, Instagram ve YouTube gibi platformlarda izleyicilerle etkileşimde bulunmak için web makaleleri, sosyal medya gönderileri ve kurgusal kişilikler oluşturmak için yapay zekâ kullandı (Nimmo, 2024, ss. 32-34). Kam-

panya, modeli (ChatGPT) kullanarak sosyal medya için kurgusal kişilikler ve biyografiler oluşturdu ve çeşitli sosyal medya yorumları hazırladı. Bu dezenformasyon kampanyasında Filistin karşıtı, İsrail yanlısı içerikler üretilmiştir. Ayrıca, Hindistan seçimlerine yönelik içerikler de üretildi. İçerik, genellikle güncel siyasi olaylarla ilgiliydi ve bu temalar etrafında dönen bir dizi kampanya yürütüldü. Rapora göre Zero Zeno kampanyasının etkinliği sınırlıydı. Sosyal medya gönderileri genellikle az sayıda etkileşim aldı ve çoğunlukla kendi oluşturduğu hesaplardan gelen etkileşimlerdi (Innes, 2020). Kampanya, gerçek kitlelere ulaşmakta zorlandı ve çoğunlukla görsel etkileşimler almadı. Kampanyanın sosyal medya hesaplarındaki etkinliği, genellikle daha önce kullanılan yapay zekâ türleriyle oluşturulan profil resimlerini kullandığı için tanınabiliyordu. Bu profil resimleri, çoğu zaman birçok farklı hesapta kullanılarak etkileşimlerin yapay olduğu izlenimini pekiştirdi (Nimmo, 2024).

## 5. Sonuç

Bu çalışmada Rusya, Çin, İran ve İsrail'in dezenformasyon faaliyetleri doküman analizi yöntemi kullanılarak karşılaştırılmaktadır. Doküman analizi, yazılı ve dijital materyallerin sistematik bir incelemesiyle, bu kampanyaların kullandıkları yöntemler, araçlar ve hedefler arasında ortak desenleri ve özgün yaklaşımları anlamamıza yardımcı olur (Yıldız, 2022). Bu yöntem, dezenformasyonun dinamik doğasını anlamak ve küresel bilgi ekosistemi üzerindeki etkilerini değerlendirmek için ideal bir çerçeve sunmaktadır.

Rusya'nın dezenformasyon kampanyaları, genellikle düşük kaliteli ve hacimli içerik üretimi üzerine yoğunlaşmıştır. Bad Grammar kampanyasında, dil bilgisi hatalarıyla dikkat çeken İngilizce metinler, Telegram gibi platformlarda yayılarak politik tartışmaları kışkırtmayı ve kafa karışıklığı oluşturmayı hedeflemiştir. Doppelganger kampanyasında ise çoklu hesap kümeleri ve yapay zekâ ile üretilen içerikler, memler ve sahte profillerle birleştirilerek Ukrayna karşıtı mesajlar yayılmıştır. Ancak bu kampanyalar, dil hatalarının ve bot hesapların gerçek kullanıcılar tarafından fark edilmesi nedeniyle düşük bir etkileşim oranına sahip olmuştur. Bu durum, Rusya'nın dezenformasyon faaliyetlerinde, teknolojik sofistikasyondan ziyade içerik hacmine dayalı bir yaklaşım izlediğini göstermektedir.

Çin'in dezenformasyon kampanyaları, sofistike teknolojik araçlarla dikkat

çekmektedir. Spamouflage kampanyası, Batı karşıtı ve Çin hükûmetini öven içeriklerin çok dilli olarak üretilip Reddit, Telegram ve YouTube gibi platformlarda yayılmasıyla karakterize edildiği gözlenmiştir. Çin'in stratejisi, yalnızca içerik üretmekle kalmayıp aynı zamanda yapay zekâ destekli duygu analizi ve mesaj optimizasyonu yoluyla bu içeriklerin etkisini artırmayı hedeflediği anlaşılmaktadır. Bununla birlikte, Çin'in kampanyalarının geniş bir erişime sahip olmasına rağmen izleyicilerle anlamlı bir etkileşim sağlayamaması, manipülasyon niyetinin fark edilmesinden kaynaklanmış olabilir. Çin'in teknoloji odaklı yaklaşımı, diğer ülkelerle kıyaslandığında, daha sofistike bir strateji olarak öne çıkmaktadır.

İran, dezenformasyon faaliyetlerinde Uluslararası Sanal Medya Birliği (IUVM) adlı platform üzerinden çok dilli içerikler üreterek İsrail ve ABD karşıtı içerikler üretilmesi şeklinde yürütüldüğü anlaşılmaktadır. IUVMPress adlı web sitesi, bu içerikleri çeşitli sosyal medya platformlarına yayma işlevi görmüştür. İran'ın stratejisi, jeopolitik çıkarlarını destekleyen bir anlatı oluşturmak için yapay zekâ ile otomatik başlık ve makale üretimine dayandığı anlaşılmaktadır. Ancak, IUVM tarafından kullanılan otomatikleştirilmiş etiketleme ve içeriklerin düşük kalitesi, kampanyanın profesyonellik eksikliğini ortaya koymuştur. Sosyal medya hesaplarının sınırlı erişimi de İran'ın dezenformasyon faaliyetlerinin etkisini zayıflattığı görülmektedir.

İsrail'in Zero Zeno kampanyası, Hamas karşıtı ve İsrail yanlısı mesajların kurgusal kişilikler üzerinden yayıldığı bir dezenformasyon stratejisini temsil etmektedir. ChatGPT gibi yapay zekâ araçları kullanılarak oluşturulan sahte biyografiler ve profiller, sosyal medya platformlarında etkileşimi yüksek olacak şekilde yayılması amacı ile üretildiği anlaşılmaktadır. Ancak, sahte profillerde aynı görsellerin tekrar kullanılması gibi hatalar, manipülasyonun fark edilmesine neden olmuş ve İsrail'in yürüttüğü dezenformasyon kampanyasının etkisini sınırlamıştır.

Ülkelerin dezenformasyon stratejileri arasında hem benzerlikler hem de belirgin farklılıklar bulunmaktadır. Tüm ülkeler, dezenformasyon kampanyalarında yapay zekâ teknolojilerini kullanmış ve sosyal medya platformlarını birincil dağıtım kanalı olarak tercih etmiştir. Ancak, bu kampanyaların büyük çoğunluğu düşük kullanıcı etkileşim oranlarıyla sınırlı kalmıştır, bu da izleyicilerin manipülatif içerikleri kolayca fark edebildiğini göstermektedir. Çin'in teknolojik gücü, Rusya ve İran'ın içerik hacmine dayalı yaklaşımların-

dan farklılaşmaktadır. Öte yandan, İsrail'in daha dar bir hedef kitlesi ve spesifik bir mesajla çalışması, onu diğerlerinden ayıran bir stratejik tercih olarak öne çıkmaktadır.

Dezenformasyon kampanyaları, küresel bilgi bütünlüğüne ciddi tehditler oluşturmaktadır. Bu faaliyetler, bilgi ekosisteminde güvensizliği artırmakta ve kamusal tartışmaları manipüle etmektedir. Yapay zekâ sistemlerinin kompleks çalışma biçimi ele alındığında tek bir çözüm önerisinden ziyade çok yönlü ve bütüncül bir yapılanmaya ihtiyaç duyulduğunu söylemek yanlış olmayacaktır (Gül Ünlü ve Küçükşabanoğlu, 2023). Bu doğrultuda hükûmetler, teknoloji şirketleri ve sivil toplum arasındaki işbirlikçi çabalar, dezenformasyona karşı dirençli bir savunma oluşturma adına önemli adımlar olabilir. Ancak, yapay zekâ teknolojilerinin hızlı dönüşümü ve dezenformasyonun küresel doğası, karşı önlemlerde sürekli adaptasyon ve yenilik gerektirdiği de göz ardı edilmemesi gereken önemli bir konudur. Gelecekteki araştırmaların; tespit teknolojilerinin sağlamlığını artırmaya, uluslararası iş birliğini geliştirmeye, medya okuryazarlığı eğitimlerini yaygınlaştırmaya ve dezenformasyonla mücadelede yapay zekâ kullanımının çok boyutlu olarak ele almaya odaklanması yerinde olacaktır.

**Çıkar Çatışması Beyanı**

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan etmiştir.

## Kaynakça

Aboutayeb, Mostafa. 2023. "Démystification de la désinformation en ligne : Une approche analytique". FRANCISOLA 8(2): 113-20. doi:10.17509/francisola.v8i2.63493.

Arce, Daniel. 2024. "Disinformation Strategies". Defence and Peace Economics: 1-14. doi:10.1080/10242694.2024.2302236.

Auezov, M., E.A. Nysanov, Zh.S. Kemelbekova, A.N. Zhidebayeva, A. Kuatbekov University of Peoples' Friendship, S.E. Kozhabaev, M. Auezov South Kazakhstan University, A.U. Korokbaev, ve A. Kuatbekov University of Peoples' Friendship. 2024. "Computer simulation and description of the phenomenon of the development of information technology using moore's law". Bulletin of the National Engineering Academy of the Republic of Kazakhstan 91(1): 93-102. doi:10.47533/2024.1606-146X.10.

Balcioğlu, Yavuz Selim, ve Bülent Dogan. 2023. "Dissecting Disinformation Dynamics: Insights from a Social Media Environment". İletişim ve Diplomasi (11): 107-25. doi:10.54722/iletisimvediplomasi.1374744.

Barman, Dipto, Ziyi Guo, ve Owen Conlan. 2024. "The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination". Machine Learning with Applications 16: 100545. doi:10.1016/j.mlwa.2024.100545.

Bontridder, Noémi, ve Yves Poullet. 2021. "The Role of Artificial Intelligence in Disinformation". Data & Policy 3: e32. doi:10.1017/dap.2021.20.

Caled, Danielle, ve Mário J. Silva. 2022a. "Digital Media and Misinformation: An Outlook on Multidisciplinary Strategies against Manipulation". Journal of Computational Social Science 5(1): 123-59. doi:10.1007/s42001-021-00118-8.

Caled, Danielle, ve Mário J. Silva. 2022b. "Digital Media and Misinformation: An Outlook on Multidisciplinary Strategies against Manipulation". Journal of Computational Social Science 5(1): 123-59. doi:10.1007/s42001-021-00118-8.

Chambers, JK. 2010. "Bad'grammar and the Language Faculty". University of Pennsylvania Working Papers in Linguistics 16(38): 19-25. https://core.ac.uk/download/pdf/76365019.pdf.

Chan, Jovy. 2024. "Online Astroturfing: A Problem beyond Disinformation". Philosophy & Social Criticism 50(3): 507-28. doi:10.1177/01914537221108467.

Chesney, Robert, ve Danielle Citron. 2018. "Deepfakes and the New Disinformation War The Coming Age of Post-Truth Geopolitics". Foreign Affairs. https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war.

Cho, Charles H., Martin L. Martens, Hakkyun Kim, ve Michelle Rodrigue. 2011. "Astroturfing Global Warming: It Isn't Always Greener on the Other Side of the Fence". Journal of Business Ethics 104(4): 571-87. doi:10.1007/s10551-011-0950-6.

Clos, Jeremie, Emma Mcclaughlin, Pepita Barnard, Tino Tom, ve Sudarshan Yajaman. 2023. "A Privacy-Preserving Observatory of Misinformation Using Linguistic Markers - A Work in Progress". Içinde Proceedings of the First International Symposium on Trustworthy Autonomous Systems, Edinburgh United Kingdom: ACM, 1-4. doi:10.1145/3597512.3597530.

Cybenko, Anne K., ve George Cybenko. 2018. "AI and Fake News". IEEE Intelligent Systems 33(5): 1-5. doi:10.1109/MIS.2018.2877280.

Deschrijver, Cedric. 2024. "Assessing Potential Disinformation Campaigns in Anonymous Online Comments: Evaluating Available Textual Cues in Debates on the 2019 Hong Kong Protests". Language & Communication 95: 31-41. doi:10.1016/j.langcom.2024.01.002.

Dhamani, Numa, Paul Azunre, Jeffrey L. Gleason, Craig Corcoran, Garrett Honke, Steve Kramer, ve Jonathon Morgan. 2019. "Using Deep Networks and Transfer Learning to Address Disinformation". http://arxiv.org/abs/1905.10412 (Erişim Tarihi:15 Eylül 2024).

Dias, Emmanuelle, Letícia Lopes, ve Felipe Borges. 2024. "Estratégias de desinformação na produção de videoensaios". Esferas (29). doi:10.31501/esf.v1i29.14892.

Dordevic, Miljan, Milan Milivojevic, ve Ana Gavrovska. 2019. "DeepFake Video Analysis using SIFT Features". İçinde 2019 27th Telecommunications Forum (TELFOR), Belgrade, Serbia: IEEE, 1-4. doi:10.1109/TELFOR48224.2019.8971206.

EEAS. 2024. Doppelganger Operation - EEAS Technical Report. https://euvsdisinfo.eu/uploads/2024/06/EEAS-TechnicalReport-DoppelgangerEE24_June2024.pdf.

Fallis, Don. 2014. "A Functional Analysis of Disinformation". Içinde iConference 2014 Proceedings, iSchools. doi:10.9776/14278.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, ve Alessandro Flammini. 2016. "The Rise of Social Bots". Communications of the ACM 59(7): 96-104. doi:10.1145/2818717.

Fraga-Lamas, Paula, ve Tiago M. Fernandez-Carames. 2020. "Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality". IT Professional 22(2): 53-59. doi:10.1109/MITP.2020.2977589.

Francis, E. 2018. MisInfoWars: A Linguistic Analysis of Deceptive and Credible News. Simon Fraser University. https://books.google.com.tr/books?id=TMcGywEACAAJ.

François, Camille, ve Evelyn Douek. 2021. "The Accidental Origins, Underappreciated Limits, and Enduring Promises of Platform Transparency Reporting about Information Operations". Journal of Online Trust and Safety 1(1). doi:10.54501/jots.v1i1.17.

Franklin, Margarita, Lindsay Hundley, Mike Torrey, David Agronovich, ve Mike Dvilyanski. 2024. Adversarial Threat Report. Meta. https://md.teyit.org/file/meta-threat-report.pdf.

García Serrano, Jesús, Luis M. Romero-Rodríguez, ve Ángel Hernando Gómez. 2019. "Análisis del 'clickbaiting' en los titulares de la prensa española contemporánea / Estudio de caso: Diario 'El País' en Facebook". Estudios sobre el Mensaje Periodístico 25(1): 197-212. doi:10.5209/ESMP.63724.

Garg, Diya, ve Rupali Gill. 2024. "A Bibliometric Analysis of Deepfakes : Trends, Applications and Challenges". ICST Transactions on Scalable Information Systems 11(6). doi:10.4108/eetsis.4883.

Giachanou, Anastasia, Xiuzhen Zhang, Alberto Barrón-Cedeño, Olessia Koltsova, ve Paolo Rosso. 2022. "Online Information Disorder: Fake News, Bots and Trolls". International Journal of Data Science and Analytics 13(4): 265-69. doi:10.1007/s41060-022-00325-0.

Gül Ünlü, Derya, ve Zafer Küçükşabanoğlu. 2023. "Dezenformasyon ve Yapay Zekâ: Dezenformasyonla Mücadele Yollarına Yapay Zekâ Uzmanlarının Gözünden Bakmak". İletişim ve Diplomasi (11): 83-106. doi:10.54722/iletisimvediplomasi.1375478.

Havlík, Martin. 2023. "Deepfake as an Advanced Manipulative Technique for Spreading Propaganda". Vojenské rozhledy 32(1): 3-17. doi:10.3849/2336-2995.32.2023.01.003-017.

Hayden, Joseph R. 2024. A History of Disinformation in the U.S. 1. bs New York: Routledge. doi:10.4324/9781003331551.

Howard, P, ve S Bradshaw. 2018. "The global organization of social media disinformation campaigns". Journal of International Affairs 71(1.5).

Innes, Martin. 2020. "Techniques of Disinformation: Constructing and Communicating 'Soft Facts' after Terrorism". The British Journal of Sociology 71(2): 284-99. doi:10.1111/1468-4446.12735.

Islas, Octavio, Fernando Gutiérrez, ve Amaia Arribas. 2024. "Artificial Intelligence, a Powerful Battering Ram in the Disinformation Industry". New Explorations 4(1): 1111639ar. doi:10.7202/1111639ar.

Jing, Junchang, Fei Li, Bin Song, Zhiyong Zhang, ve Kim-Kwang Raymond Choo. 2023. "Disinformation Propagation Trend Analysis and Identification Based on Social Situation Analytics and Multilevel Attention Network". IEEE Transactions on Computational Social Systems 10(2): 507-22. doi:10.1109/TCSS.2022.3169132.

Jingnan, Hou. 2024. "How Israel tried to use AI to covertly sway Americans about Gaza". Wusf npr. https://www.wusf.org/2024-06-05/how-israel-tried-to-use-ai-to-covertly-sway-americans-about-gaza.

Kandel, Nirmal. 2020. "Information Disorder Syndrome and Its Management". Journal of Nepal Medical Association 58(224). doi:10.31729/jnma.4968.

Karinshak, Elise, ve Yan Jin. 2023. "AI-Driven Disinformation: A Framework for Organizational Preparation and Response". Journal of Communication Management 27(4): 539-62. doi:10.1108/JCOM-09-2022-0113.

Keller, Franziska B., David Schoch, Sebastian Stier, ve JungHwan Yang. 2020. "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign". Political Communication 37(2): 256-80. doi:10.1080/10584609.2019.1661888.

Kıral, Bilgen. 2020. "Nitel bir veri analizi yöntemi olarak doküman analizi." Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (8(15)): 170-89.

Kovic, Marko, Adrian Rauchfleisch, Marc Sele, ve Christian Caspar. 2018. "Digital astroturfing in politics: Definition, typology, and countermeasures". Studies in Communication Sciences 18(1). doi:10.24434/j.scoms.2018.01.005.

Kreps, Sarah E., Miles McCain, ve Miles Brundage. 2020. "All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation". SSRN Electronic Journal. doi:10.2139/ssrn.3525002.

Kumar, Akash, Arnav Bhavsar, ve Rajesh Verma. 2020. "Detecting Deepfakes with Metric Learning". İçinde 2020 8th International Workshop on Biometrics and Forensics (IWBF), Porto, Portugal: IEEE, 1-6. doi:10.1109/IWBF49977.2020.9107962.

Kumar, Rajeev, Suhel Ahmad Khan, Nawaf Alharbe, ve Raees Ahmad Khan. 2024. "Code of Silence: Cyber Security Strategies for Combating Deepfake Disinformation". Computer Fraud & Security 2024(4): S1361-

3723(24)70013-X. doi:10.12968/S1361-3723(24)70013-X.
Lungarella, Max, Fumiya Iida, Josh C. Bongard, ve Rolf Pfeifer. 2007. "AI in the 21st Century – With Historical Reflections". Içinde 50 Years of Artificial Intelligence, Lecture Notes in Computer Science, ed. Max Lungarella, Fumiya Iida, Josh Bongard, ve Rolf Pfeifer. Berlin, Heidelberg: Springer Berlin Heidelberg, 1-8. doi:10.1007/978-3-540-77296-5_1.
Maathuis, Clara, Frederick Janssens, ve Ebrahim Rahimi. 2024. "Design of a Disinformation Awareness Digital Game". European Conference on Social Media 11(1): 127-36. doi:10.34190/ecsm.11.1.2053.
Mansfeld, J. 1978. "Zeno of Citium". Mnemosyne 31(2): 134-78. doi:10.1163/156852578X00337.
Martin, L. John. 1982. "Disinformation: An Instrumentality in the Propaganda Arsenal". Political Communication 2(1): 47-64. doi:10.1080/10584609.1982.9962747.
McLoughlin, Killian L., ve William J. Brady. 2024. "Human-Algorithm Interactions Help Explain the Spread of Misinformation". Current Opinion in Psychology 56: 101770. doi:10.1016/j.copsyc.2023.101770.
Montasari, Reza. 2024. "The Dual Role of Artificial Intelligence in Online Disinformation: A Critical Analysis". Içinde Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution, Advanced Sciences and Technologies for Security Applications, Cham: Springer International Publishing, 229-40. doi:10.1007/978-3-031-50454-9_11.
Mosleh, Mohsen, Rocky Cole, ve David Gertler Rand. 2024. "Misinformation and harmful language are interconnected, rather than distinct, challenges". doi:10.31234/osf.io/y5n4u.
Mueller, Robert S. 2019. "Report on the Investigation Into Russian Interference in the 2016 Presidential Election". https://digital.library.unt.edu/ark:/67531/metadc1933948/.
Na, David, Samuel Nathanson, Yungjun Yoo, Yinzhi Cao, ve Lanier Watkins. 2024. "Showcasing the Threat of Scalable Generative AI Disinformation through Social Media Simulation". İçinde IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, BC, Canada: IEEE, 1-2. doi:10.1109/INFOCOMWKSHPS61880.2024.10620878.
Nefedova, Lilia, ve Maria Samkova*. 2019. "Detection Of Disinformation In A Media Text (Structural And Pragma-Linguistic Approaches)". İçinde , 265-72. doi:10.15405/epsbs.2019.08.02.31.
Nimmo, Ben. 2024. AI and Covert Influence Operations: Latest Trends. OpenAI: OpenAI. https://downloads.ctfassets.net/kftzwdyauwt9/5IMxzTmUclSOAcWUXbkVrK/3cfab518e6b10789ab8843bcca18b633/Threat_Intel_Report.pdf.
Omoregie, Uyiosa. 2021. "The 'Harm Principle' and Information Disorder Online". Academia Letters. doi:10.20935/AL3425.

OpenAI. 2024. "Disrupting deceptive uses of AI by covert influence operations". https://openai.com/index/disrupting-deceptive-uses-of-ai/.

Pennekamp, Jan, Martin Henze, Oliver Hohlfeld, ve Andriy Panchenko. 2019. "Hi Doppelgänger : Towards Detecting Manipulation in News Comments". Içinde Companion Proceedings of The 2019 World Wide Web Conference, San Francisco USA: ACM, 197-205. doi:10.1145/3308560.3316496.

Quelle, Dorian, Calvin Cheng, Alexandre Bovet, ve Scott A. Hale. 2023. "Lost in Translation -- Multilingual Misinformation and its Evolution". doi:10.48550/ARXIV.2310.18089.

Renedo Farpón, Cristina, ve Francisco José García. 2023. "PRÓLOGO. Análisis de la desinformación: estrategias (en) de los desórdenes informativos". Miguel Hernández Communication Journal 14: 15-18. doi:10.21134/mhjournal.v14i.1791.

Rossetti, Michael, ve Tauhid Zaman. 2023. "Bots, Disinformation, and the First Impeachment of U.S. President Donald Trump" ed. Alexandre Bovet. PLOS ONE 18(5): e0283971. doi:10.1371/journal.pone.0283971.

Rubin, Victoria L. 2019. "Disinformation and Misinformation Triangle: A Conceptual Model for 'Fake News' Epidemic, Causal Factors and Interventions". Journal of Documentation 75(5): 1013-34. doi:10.1108/JD-12-2018-0209.

Ruffin, Margie, Haeseung Seo, Aiping Xiong, ve Gang Wang. 2024. "Does It Matter Who Said It? Exploring the Impact of Deepfake-Enabled Profiles on User Perception towards Disinformation". Proceedings of the International AAAI Conference on Web and Social Media 18: 1328-41. doi:10.1609/icwsm.v18i1.31392.

Shoaib, Mohamed R., Zefan Wang, Milad Taleby Ahvanooey, ve Jun Zhao. 2023. "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models". doi:10.48550/ARXIV.2311.17394.

Starbird, Kate, Renée DiResta, ve Matt DeButts. 2023. "Influence and Improvisation: Participatory Disinformation during the 2020 US Election". Social Media + Society 9(2): 20563051231177943. doi:10.1177/20563051231177943.

Turğal, L., ve B.B. Küçükerdoğan. 2023. "Bir Dezenformasyon Aracı Olarak Yapay Zekâ: Bing Arama Motoru Örneğinde İklim Değişikliği Konulu Haber Fotoğraflarının İncelenmesi." İletişim Ve Diplomasi (11): 57-82. doi:https://doi.org/10.54722/iletisimvediplomasi.1376404.

Ustinovich, Elena Stepanovna. 2024. "Generative artificial intelligence in the electoral processes of 2024 in the world: disinformation campaigns and online trolls". Social'naja politika i social'noe partnerstvo (Social Policy and Social Partnership) (3): 197-204. doi:10.33920/pol-01-2403-03.

V. Baravkar, Prof. Pooja, Namita Survase, Sakshi Shrimandale, ve Gaurav Hande. 2023. "Survey On 'Deepvision's Human Eye Blink Pattern Analysis for Deepfake Detection'". INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT 07(10): 1-11. doi:10.55041/IJSREM26418.

Vosoughi, Soroush, Deb Roy, ve Sinan Aral. 2018. "The Spread of True and False News Online". Science 359(6380): 1146-51. doi:10.1126/science.aap9559.

Wardle, Claire, ve Hossein Derakhshan. 2017. INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making. Published by the Council of Europe. Council of Europe report. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

Warin, Thierry. 2024. Disinformation in the Digital Age: Impacts on Democracy and Strategies for Mitigation. CIRANO. doi:10.54932/GQWB1497.

Weber, Derek, ve Frank Neumann. 2021. "Amplifying Influence through Coordinated Behaviour in Social Networks". Social Network Analysis and Mining 11(1): 111. doi:10.1007/s13278-021-00815-2.

Woolley, Samuel C. 2022. "Digital Propaganda: The Power of Influencers". Journal of Democracy 33(3): 115-29. doi:10.1353/jod.2022.0027.

Wu, Liang, ve Huan Liu. 2017. "Detecting Crowdturfing in Social Media". Içinde Encyclopedia of Social Network Analysis and Mining, ed. Reda Alhajj ve Jon Rokne. New York, NY: Springer New York, 1-9. doi:10.1007/978-1-4614-7163-9_110196-1.

Yang, Angela. 2024. "Meta and OpenAI say they disrupted influence operations linked to Israeli company". https://www.nbcnews.com/tech/security/meta-openai-say-disrupted-israeli-companys-influence-campaign-rcna154774.

Yıldız, A. 2022. "Bir araştırma metodolojisi olarak sistematik literatür taramasına genel bakış". Anadolu Üniversitesi Sosyal Bilimler Dergisi (22(Özel Sayı 2)): 367-86.

Zannettou, Savvas, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, ve Jeremy Blackburn. 2019. "Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls". İçinde Proceedings of the 10th ACM Conference on Web Science, Boston Massachusetts USA: ACM, 353-62. doi:10.1145/3292522.3326016.

Zerback, Thomas, ve Florian Töpfl. 2022. "Forged Examples as Disinformation: The Biasing Effects of Political Astroturfing Comments on Public Opinion Perceptions and How to Prevent Them". Political Psychology 43(3): 399-418. doi:10.1111/pops.12767.


Kaynak: AA