iLETİŞİM ve DİPLOMASİ
Communication and Diplomacy
ISSN: 2147-6772 e-ISSN: 2791-7649 Yıl/Year: Aralık/December 2023 Sayı/Issue: 11

Dezenformasyon ve Yapay Zekâ : Dezenformasyonla Mücadele Yollarına Yapay Zekâ Uzmanlarının Gözünden Bakmak

Disinformation and Artificial Intelligence: Looking at Ways to Combat Disinformation through Artificial Intelligence Experts' Eyes

Araştırma Makalesi / Research Article

Derya GÜL ÜNLÜ[1], **Zafer KÜÇÜKŞABANOĞLU**[2]

Sorumlu yazar/ Corresponding author:
Derya Gül Ünlü

ORCID:
0000-0003-3936-7988

Geliş tarihi/Received:
13.10.2023

Son revizyon teslimi/Last revision received:
13.11.2023

Kabul tarihi/Accepted:
14.11.2023

Yayın tarihi/Published:
16.12.2023

Atıf/Citation:
Gül Ünlü, D. & Küçükşabanoğlu, Z. (2023). Dezenformasyon ve yapay zekâ: Dezenformasyonla mücadele yollarına yapay zekâ uzmanlarının gözünden bakmak. *İletişim ve Diplomasi, 11*, 83-106.

doi: 10.54722/iletisimvediplomasi.1375478

öz

İletişim teknolojilerindeki gelişim ve kullanıcı kaynaklı içeriğin yükselişi, her türlü içeriği herhangi bir kontrol mekanizmasına takılmadan kolaylıkla dolaşıma sokulabilir kılmıştır. Bu durum, günümüzde dijital platform kullanıcılarının sınırsız sayıda içeriğe hızlı erişimini sağlamakla birlikte; bireylerin maruz kaldıkları yoğun dezenformasyonu da beraberinde getirmiştir. Çevrimiçi dezenformasyonla mücadele süreci, yapay zekâ tekniklerinin kullanımıyla yakından ilişkilenmekte; söz konusu teknoloji hem dezenformasyonun üretilip yaygınlaştırılmasında hem de sorunlu içeriğin tespiti ve denetiminde önemli bir rol üstlenmektedir. Dezenformasyon ve yapay zekâ ilişkisinin bu iki yönü, yapay zekâ teknolojilerinin sorunlu içeriğin üretimi ve dağıtımı sürecindeki belirleyiciliğinin ve çevrimiçi dezenformasyonun tespit edilip azaltılabilmesi için yapay zekâ sistemlerinden en efektif biçimde nasıl yararlanılabileceğinin anlaşılmasını da gerekli kılmaktadır. Bu odak noktasından hareketle gerçekleştirilen çalışma kapsamında, yapay zekâ sistemlerinin dezenformasyonla mücadele sürecindeki potansiyelinin yapay zekâ uzmanlarının gözünden değerlendirilmesi hedeflenmektedir. Bu hedef doğrultusunda, Yapay Zekâ Politikaları Derneği (AIPA) üyesi ve paydaşı olan yapay zekâ uzmanlarıyla yarı yapılandırılmış görüşme tekniğinin kullanıldığı betimsel nitelikli bir alan araştırması gerçekleştirilmiştir. Çalışma sonucunda, günümüz yapay zekâ sistemlerinin dezenformasyo-

[1] Doç. Dr., İstanbul Üniversitesi İletişim Fakültesi, Halkla İlişkiler ve Tanıtım Bölümü, İstanbul, Türkiye, derya.gul@istanbul.edu.tr

[2] Misafir Öğr. Gör., Gazi Üniversitesi Teknoloji Fakültesi, Ankara, Türkiye, zafer.kucuksabanoglu@aipaturkey.org, ORCID: 0000-0003-2686-4109

nun artırılmasında olduğu kadar azaltılması için de nasıl aktif kullanılabileceği; bunun için dezenformasyon tespit ve filtreleme mekanizmalarının, doğrulama platformlarının yaygınlaştırılmasının gerekliliği, bu amaçla geliştirilecek politikalar kamu-dijital platform iş birliğiyle oluşturulurken kullanıcıya karşı sorumluluğun da öncelenmesine ihtiyaç duyulduğu tespit edilmiştir.

Anahtar Kelimeler: Dezenformasyon, yapay zekâ, algoritma, dijital ekosistem, çevrimiçi platformlar

ABSTRACT

The process of combating online disinformation is closely related to the use of artificial intelligence techniques. The technology in question plays an important role both in producing and disseminating disinformation and in detecting and controlling problematic content. These two aspects of the relationship between disinformation and artificial intelligence also require an understanding of the decisiveness of artificial intelligence technologies in the production and distribution of problematic content, as well as how artificial intelligence systems can be used most effectively to detect and reduce online disinformation. The aim of the study carried out with this focal point is to evaluate the capacity of artificial intelligence systems to combat disinformation as perceived by artificial intelligence experts. In line with this goal, descriptive field research was conducted using the semi-structured in-depth interview technique. The study's findings established that contemporary artificial intelligence systems have the capacity to both amplify and diminish disinformation. It has come to light that disinformation detection and filtering mechanisms and verification platforms should be disseminated, and the public and digital platforms should collaborate in the development of policies for this purpose. At the same time, accountability to the user should be given top priority.

Keywords: Disinformation, artificial intelligence, algorithm, digital ecosystem, online platforms

EXTENDED ABSTRACT

Disinformation has been a problem ever since the pre-digital era. Still, as a result of the new opportunities provided by artificial intelligence techniques found in the digital ecosystem, disinformation has become much more difficult to circulate and spread. Disinformation generated or supported by AI systems is delivered to consumers via a variety of digital channels without any checks or filters, and it is backed up with highly convincing-looking proof. Moreover, the algorithms that grant access to

artificial intelligence systems encircle the user with inaccurate content that is very pertinent to his interests; conversely, his exposure to dissenting perspectives is limited. This circumstance gives rise to issues such as the user believing the false information, encountering difficulties in finding and accessing accurate information, and a decline in the user's trust in the information obtained.

Disinformation impairs the functioning of numerous mechanisms, including political decisions, corporate operations, and public trust in institutions, by causing individual and collective uncertainty. Beyond this, other ethical and human rights issues come into play, including user freedom, privacy, information access, autonomy, the ability to select from viable options, and democratic principles (Bontridder & Poullet, 2021). Disinformation poses a serious and all-encompassing challenge in this regard, one that has the capacity to alter the political, social, and cultural foundations of any society and undermine the fundamental values of democracies (Montoro-Montarosso et al., 2023). Algorithms are gaining power as artificial intelligence's influence over the technology that runs our daily lives increases. These artificial intelligence systems are among the ones that bad actors favour for activities like information theft, privacy invasion, and the skewing of political and social decisions. For all of these reasons, it is essential to consider internet disinformation to be a threat that must be countered using both legal and technical measures, as well as to take into account the part artificial intelligence systems play in its creation and dissemination.

On the other hand, artificial intelligence (AI) techniques are also thought to be useful in the process of battling internet misinformation. In this context, there is growing interest in using artificial intelligence systems to detect and control online disinformation automatically; in fact, several initiatives are being carried out with this objective in mind (Bouziane et al., 2020; Funke, 2019; Graves, 2018; Jackson 2016). Once more, the scalability, reasonably low cost, and ability to be updated against different disinformation components through machine learning allow artificial intelligence to be seen as a technology that produces solutions in the fight against disinformation. It should be noted, though, that in the effort to combat disinformation, the boundaries of what artificial intelligence technologies are capable of are still not clearly defined, and even structural limitations and unfavourable outcomes are encountered in reducing disinformative content (Bontridder & Poullet, 2021; Kertysova, 2018). Because of this multi-dynamic nature of the relationship between disinformation and artificial intelligence, it is important to comprehend how decisive these technologies are in the creation and dissemination of problematic content, as well as the most efficient ways to use them to identify and combat online disinformation. In the context of the research conducted from this vantage point, it is intended to assess the potential of AI systems in the process of battling misinformation through the eyes of AI specialists. By doing so, it aims to show how

specialists in artificial intelligence characterize the features and dangers of online disinformation and to create a roadmap for how to stop it using current artificial intelligence techniques. The results of the research are also expected to contribute to the international literature and serve as a guide from the perspective of Türkiye in the process of battling disinformation, given that there is no study in the pertinent literature examining the relationship between artificial intelligence and disinformation through the opinions of artificial intelligence experts. In this regard, the relationship between artificial intelligence, algorithms and disinformation is discussed below. Afterwards, the study findings obtained through semi-structured in-depth interviews conducted with artificial intelligence experts who are members and stakeholders of the Artificial Intelligence Policy Association (AIPA) are presented. The study's findings established that contemporary artificial intelligence systems have the capacity to both amplify and diminish disinformation. It has come to light that disinformation detection and filtering mechanisms and verification platforms should be disseminated, and the public and digital platforms should collaborate in the development of policies for this purpose. At the same time, accountability to the user should be given top priority.

GİRİŞ

Yapay zekâ sistemleriyle üretilen ya da desteklenen dezenformasyon, herhangi bir denetim mekanizmasıyla karşılaşmadan kullanıcıyla birçok dijital kanal üzerinden buluşturulmakta ve oldukça gerçek görünümlü kanıtlara dayalı bir biçimde sunulmaktadır. Dahası, yapay zekâ sistemlerinin kullanıcıya erişimini sağlayan algoritmalar aracılığıyla, kullanıcı, ilgisiyle yüksek düzeyde ilişkilenen dezenformatif içerikle sarmalanmakta ve kullanıcının alternatif görüşlere erişimi sınırlanmaktadır. Bu durum, kullanıcının karşılaştığı dezenformasyona güvenmesi, gerçek bilgiyi arama yollarının tıkanması, doğru bilgiye erişiminin kısıtlanması, edinilen bilgiye güvenin zedelenmesi sorunlarını da beraberinde getirmektedir.

Dezenformasyonun neden olduğu bireysel ve toplumsal belirsizlik (Akhtar vd., 2022), politik kararlardan iş süreçlerine, aşı olmadan kamu kurumlarına güvene kadar çok sayıda mekanizmanın işleyişini olumsuz etkilemektedir. Bunun da ötesinde kullanıcı özgürlüğü, mahremiyet, bilgi edinme, özerk olabilme, gerçek seçenekler arasından tercihte bulunabilme, demokratik değerlerin korunması gibi çok sayıda etik ve insan hakları kaygısı da ortaya çıkmaktadır (Bontridder & Poullet, 2021). Bu bakımdan, dezenformasyon, herhangi bir toplumun siyasi, ekonomik ve kültürel dokusunu potansiyel olarak dönüştürebilecek, dolayısıyla demokratik ulusların temel ilkelerini aşındırabilecek önemli ve geniş kapsamlı bir tehdit oluşturmaktadır (Montoro-Montarosso et al., 2023). Yapay zekâningünlük hayatımızdaki kullanım alanı genişledikçe, algoritmalar giderek daha fazla nüfuz sahibi olmakta ve söz konusu ya-

pay zekâ sistemleri, kötü niyetli aktörler tarafından bilgi çalma, kişisel mahremiyeti tehlikeye atma, siyasal ve toplumsal kararları çarpıtma gibi amaçlar için tercih edilen araçların başında gelmektedir. Tüm bu nedenler, çevrimiçi dezenformasyonun yasal ve teknik yollarla mücadele edilmesi gereken bir tehdit olarak görülmesini ve yapay zekâ sistemlerinin çevrimiçi dezenformasyonun üretilmesi ve yayılmasında üstlendiği rolün dikkate alınmasını gerekli kılmaktadır.

Diğer yandan, yapay zekâ sistemleri, çevrimiçi dezenformasyonla mücadele sürecinde yararlanılabilecek uygun araçlar olarak da değerlendirilmektedir. Bu çerçevede, yapay zekâ sistemleri aracılığıyla çevrimiçi dezenformasyonu otomatik olarak tespit etmeye ve denetlemeye yönelik ilgi giderek artmakta; hatta doğrudan bu hedefle gerçekleştirilen çok sayıda girişimle karşılaşılmaktadır (Bouziane et al., 2020; Funke, 2019; Graves, 2018; Jackson 2016). Yine yapay zekâ sistemlerinin ölçeklenebilme, görece uygun maliyet, makine öğrenmesiyle dezenformasyonun çeşitli bileşenlerine karşı güncellenebilme gibi özellikleri onun dezenformasyonla mücadelede çözüm üreten bir teknoloji olarak görülmesini sağlamaktadır. Ancak dezenformasyonla mücadele sürecinde yapay zekâ teknolojilerine dayanan imkânların sınırlarının hâlâ kesin olarak çizilemediği ve hatta dezenformatif içeriğin azaltılmasında yapısal kısıtlar ve istenmeyen sonuçlarla karşılaşıldığını da eklemek gerekmektedir (Bontridder & Poullet, 2021; Kertysova, 2018). Dolayısıyla, dezenformasyon ve yapay zekâ ilişkisinin bu çok dinamikli yapısı, yapay zekâ teknolojilerinin sorunlu içeriğin üretimi ve dağıtımı sürecindeki belirleyiciliğinin ve çevrimiçi dezenformasyonun tespit edilip azaltılabilmesi için yapay zekâ sistemlerinden en efektif biçimde nasıl yararlanılabileceğinin anlaşılmasını da gerekli kılmaktadır. Bu odak noktasından hareketle gerçekleştirilen çalışma kapsamında, yapay zekâ sistemlerinin dezenformasyonla mücadele sürecindeki potansiyelinin yapay zekâ uzmanlarının gözünden değerlendirilmesi hedeflenmektedir. Bu yolla, yapay zekâ uzmanları tarafından çevrimiçi dezenformasyonun niteliklerinin ve risklerinin nasıl tanımlandığının ortaya koyulması ve söz konusu dezenformatif içeriklerle mevcut yapay zekâ tekniklerinden yararlanılarak nasıl mücadele edilebileceğine ilişkin bir yol haritasının çizilebilmesi amaçlanmaktadır. Ayrıca ilgili alanyazında yapay zekâ ve dezenformasyon ilişkisini yapay zekâ uzmanlarının görüşleri üzerinden inceleyen bir çalışmayla karşılaşılmadığı düşünüldüğünde, söz konusu araştırma bulgularının hem uluslararası literatüre katkı sağlayacağı hem de dezenformasyonla mücadele sürecinde Türkiye perspektifinden yol gösterici bir nitelik taşıyacağı öngörülmektedir. Bu doğrultuda, aşağıda öncelikle yapay zekâ, algoritma ve dezenformasyon ilişkisi ele alınmakta; sonrasında ise Yapay Zekâ Politikaları Derneği (AIPA) üyesi ve paydaşı olan yapay zekâ uzmanlarıyla gerçekleştirilmiş yarı yapılandırılmış görüşmeler üzerinden ulaşılan çalışma bulguları aktarılmaktadır.

Yapay Zekâ, Algoritma ve Dezenformasyon İlişkisi: Engeller, Olanaklar ve Kaygılar

Kendi kendine öğrenen ve tahmine dayalı biçimde çalışan makine öğrenimi yaklaşımlarıyla yakından bağlantılı olan yapay zekâ sistemleri (Shrestha et al., 2019), karmaşık senaryolarla karşılaşıldığında nasıl ilerlenebileceği, seçenekler arasından en uygun seçimin hangisi olabileceği gibi sorulara cevap vermekte ve kullanıcı eylemlerini izleyerek geçerli öngörülerde bulunmaktadır (Belhadi et al., 2021). Makine öğrenmesiyle desteklenen yapay zekâ sistemi çıktılarının yüksek niteliği, söz konusu teknolojinin dezenformatif içeriğin üretilmesi ve yaygınlaştırılması için aktif olarak kullanılmasına da kaynaklık etmektedir. Günümüzde yapay zekâ aracılığıyla üretilen gerçekçi ancak sentetik içerik yine çevrimiçi platformlar üzerinden dolaşıma sokulmakta ve hedeflenen kullanıcı kitlesiyle farklı kanallar üzerinden buluşturulmaktadır.

Yapay zekâ tekniklerinin çevrimiçi dezenformasyonu iki ana biçimde arttırdığından söz etmek mümkündür (Bontridder & Poullet, 2021, s.3): Bunlardan ilki, yapay zekânın metin, resim, ses veya video içeriğini oluşturmak ya da değiştirmek için sunduğu olanaklardır. İkincisi ise dijital platformlar tarafından kullanıcı katılımını arttırmak için geliştirilen ve dağıtıma sokulan yapay zekâ sistemlerinin, dezenformasyonun çevrimiçi ortamda etkili ve hızlı bir biçimde yayılmasını mümkün kılmasıdır. Yapay zekâdan yararlanılarak, sahte/yanıltıcı içeriğin üretimi ve yaygınlaştırılmasında deepfake içerikleri kullanılmakta; böylelikle kullanıcıların çeşitli amaçlar doğrultusunda yönlendirilmesi hedeflenmektedir. Söz konusu içerikler, Generative Adversarial Network (GAN) olarak adlandırılmakta ve en genel anlamıyla, iki yapay zekâ algoritmasının çıktısı gibi çalışarak mevcut veri kümelerinden algoritmik yeni veri türleri oluşturulmasını sağlamaktadır (Goodfellowet et al., 2014). Örneğin, bir GAN, Donald Trump'ın binlerce fotoğrafını analiz edebilmekte ve ardından analiz edilen görüntülere benzeyen ancak hiçbirinin tam kopyası olmayan yeni bir resim oluşturabilmektedir. Ayrıca bu teknoloji resim, hareketli görüntü, ses ve metin gibi çeşitli içerik türlerine de uygulanabilmektedir (Goodfellowet et al., 2014; Küçükşabanoğlu & Soysal, 2023; Walorska, 2020). Bu durum, dezenformatif içeriğin oldukça inandırıcı bir nitelik kazanmasını sağlamakta ve kullanıcı algısındaki gerçek ve gerçek dışı arasındaki sınırların yapay zekâ teknikleri aracılığıyla muğlaklaştırılmasına neden olmaktadır. Yani deepfake içerikleri, sadece dezenformatik içerik üretilmesinin ötesinde, dijital ortamda karşılaşılan gerçek bilgilerin doğruluğuna şüpheyle yaklaşılması ve geleneksel basın, kamu kurumları, hükûmet idareleri tarafından halka aktarılan meşru bilgilerin güvenilirliğini zayıflatma noktasında da sorun yaratmaktadır (Bontridder & Poullet, 2021; Chesney & Citron, 2019; Greengard, 2019; Karakoç & Zeybek, 2022; Masood et al., 2022). Bu yönüyle, deepfake içerikleri hem sahte içeriği gerçek hem de gerçek bilgiyi sahteymiş gibi göstererek kullanıcının bilgiye erişimini zorlaştırmakta ve edindiği bilgiye güvenini zedelemektedir.

Üretilen dezenformasyonun geniş kullanıcı topluluğuna oldukça etkili bir biçimde yayılmasında da yapay zekâ tekniklerinden yararlanılmaktadır. Bilindiği üzere, web ekosistemi ekonomik bir model üzerine kuruludur ve bu doğrultuda dijital platformlar, bireyin tarama geçmişinin izlenerek reklamların çeşitli hedef kitle segmentasyonlarına uygun olarak hedeflenmesini (third-party tracking, browser fingerprinting gibi) ve kullanıcının platformda daha uzun süre tutulmasına olanak tanıyan yapay zekâ teknolojilerini kullanmaktadır. Söz konusu yöntemler her ne kadar reklamın potansiyel tüketicilerle buluşturulması amacını öncelese de kullanıcıyı önceki aramalarına benzer içeriklerle karşılaştırarak platformda daha fazla tutmayı sağlayan öneri algoritmaları, dezenformasyonun etkin ve geniş düzeyde yayılmasını mümkün kılmaktadır. Çünkü algoritmalar, kullanıcının dijital platformda neyle karşılaştığını, içeriğe ne düzeyde maruz kalacağını ya da içeriğin ne derece dışında tutulacağını belirlemektedir. Bu mekanizma içerisinde; kullanıcı profili (demografik, psikometrik) oluşturma, segmentasyon ve aşırı kişiselleştirilmiş hedefleme (Rosenbach & Mansted, 2018), bireyin doğrudan kendi ilgisiyle ilişkilenen dezenformasyon içeriklerine maruz kalmasına neden olmaktadır.

Dijital platformlarda yer alan kullanıcılar ise çoğu zaman kendilerine çizilen bu yolun farkında olmamakta ve öneri akışı içerisinde kendileri için en ilgi çekici biçimde sunulan dezenformasyonla baş başa kalmaktadır. Yapay zekâ teknolojilerinin gelişimine paralel olarak, kullanıcı verisi analizi ve algoritmaların bireyleri yönlendirici etkisi her geçen gün artmakta; günümüzde neredeyse her bir kullanıcı dijital ortamda gerçeğin farklı bir versiyonunu görmektedir (Bontridder & Poullet, 2021). Facebook'un haber akışı, X'in zaman çizelgesi, YouTube'un öneri sistemi, Netflix'in içerik sıralamaları, Instagram'ın reels akışı, Spotify'in önerileri bireysel kullanıcıların çevrimiçi ortamda ne göreceğini belirleyen içerik şekillendirme algoritmalarının yalnızca birkaç örneğidir (Gül-Ünlü & Kesgin, 2021; Karakoç-Keskin et al., 2023; Marechal & Biddle, 2020). Bu ekosistem, kullanıcı mahremiyeti ya da bilgi edinme hakkı gibi endişelerin yanı sıra belirli kullanıcıları hedeflemek, onları filtre balonuna hapsederek dezenformasyon akışlarına dâhil etmek gibi amaçlarla da kullanılabilmesi bakımından da sorunludur (Akers et al., 2018). Çünkü mikro hedeflemeye dayalı algoritmalar, dezenformasyon ve yanlış bilginin niyetli yayılımını kısa sürede ve geniş ölçekte arttırabilmektedir.

Çevrimiçi dezenformasyon yayılımına kaynaklık eden unsurlardan bir diğeri de sosyal botlardır. Sosyal medya platformlarında, kullanıcıların çevrimiçi davranışlarını taklit eden tam veya yarı otomatik kullanıcı hesapları olan sosyal botlar (Bontridder & Poullet, 2021), sahte kullanıcılar olsalar da etkileri itibarıyla gerçektir. Sosyal botlar tarafından üretilerek dolaşıma sokulan içerik, siyasal çekişmeyi artıran, çevrimiçi söylemi çarpıtan ve manipüle eden bir nitelik taşımaktadır (Lamo & Calo, 2018; Shao et al., 2017; Wang et al., 2018). Özellikle kullanıcıların, sosyal botları gerçek kullanıcılar olarak değerlendirdikleri ve güvendikleri durumlarda, önemli dezenformasyon kaynaklarına

dönüştükleri görülmektedir (Akers et al., 2018). Dolayısıyla, dijital ekosistemde dezenformatif içeriğin üretimi ve yayılımı için kullanılan yapay zekâ teknikleri, çoğu durumda bireyin haberi olmadan gerçeğin değiştirilmiş bir görünümünü sunarak, fikirlerin etkili bir biçimde manipüle edilebilmesi için çeşitli fırsatlar yaratmaktadır. Hedeflenen kullanıcıların ise sistemin işleyiş mekanizmasının nadiren farkında oluşu, genellikle maruz kalınan içeriğin objektif olduğunun düşünülmesi ve tüm kullanıcıların da aynı bilgiyle aynı biçimde karşılaştığına inanılması gibi tüm bileşenler, dezenformasyona kullanıcı düzeyinde eleştirel bir gözle bakılabilmesini zorlaştırmaktadır (Bontridder & Poullet, 2021). Kullanıcının gerçek bilgiye erişiminde karşılaştığı bu zorluk, onun tercihler arasında ne derece özgür ve bilinçli seçim yapabildiği soruları üzerinde düşünülmesini de gerekli kılmaktadır.

Ayrıca, doğrudan dezenformasyon yayma amacına hizmet eden yapay zekâ teknolojilerinin ötesinde sosyal medya platformlarının yapısal özelliklerinin de çevrimiçi ortamda dezenformasyonun yaygınlaşmasına ve kullanıcıların gerçek bilgiye ulaşmalarının engellenmesine neden olduğundan söz etmek olanaklıdır. Bilindiği üzere, bireyin bir konu hakkında kendi görüşünü oluşturabilmesi, onun ilgili tüm bilgilere erişmesini ve bir çıkarımda bulunmasını gerekli kılmaktadır. Oysaki kullanıcıların içerik şekillendirme ve öneri algoritmalarından kaynaklanan bir filtre balonu içerisine hapsedilmeleri, bireylerin sosyal medya platformlarında farklı görüşlere erişemiyor oluşlarının temel nedenlerindendir. Böylelikle bireyler, "diyaloğa yer olmayan bir entelektüel izolasyon ortamına" (Bergamini, 2020, s.10) sürüklenmektedir. Dahası algoritma, kullanıcıların beğendiği haber ve bilgilere öncelik vererek, fikirlerini, zevklerini, alışkanlıklarını güçlendirmekte ve alternatiflere erişimlerini de sınırlandırmaktadır (Bergamini, 2020; Gül-Ünlü & Kesgin, 2021). Yani aslında kullanıcılar, algoritmalar aracılığıyla bir yandan aynı tür içeriğe fazlaca maruz kalırken bir yandan da sistemin döngüsü dışında tutulabilmektedir (Rosenbach & Mansted, 2018) Söz konusu durum, kullanıcının kendisine sunulan içeriğe hapsedilmesine, gerçek ya da alternatif bilgiye erişiminin zorlaştırılmasına veya alternatif görüşlere hiç ulaşamayacak biçimde tek tip bir fikir balonu içerisinde tutulmasına neden olmaktadır.

Diğer yandan, özellikle yakın tarihli alanyazınında yapay zekânın çevrimiçi dezenformasyonu nasıl azaltabileceği yönündeki tartışmaların (Akers et al., 2018; Akhtar et al., 2022; Bontridder & Poullet, 2021; Gupta et al., 2022; Kertysova, 2018; Stiff & Johansson, 2022) da önemli bir yer tuttuğu görülmektedir. Bu kapsamda çevrimiçi dezenformasyonun öncelikle web'in reklam gelirlerine dayalı iş modelinden kaynaklandığı, bu modelin uyarlanmasının dijital ekosistemin manipülasyon sorununu önemli ölçüde azaltacağı belirtilmekte ve kullanıcının bireysel tercihi ve dijital platformların etik sorumluluklarının öncelenerek, yapay zekâ sistemlerinin çevrimiçi içerik ve hesapların düzenlenmesinde yardımcı bir rol üstlenmesine yönelik önlemlerin görünürlük kazanması önerilmektedir (Bontridder & Poullet, 2021). Bahsi geçen önlemlerden ilki, sorunlu/yanıltıcı içeriğin filtrelenmesidir. Filtreleme, içeriğin yüklenmesini veya

yayımlanmasını önlemek amacıyla teknik sağlayıcılar tarafından alınan bir önlemdir ve içeriğin kaldırılması, talep edilmesi üzerine gerçekleştirilir. Yine belirli konular hakkında daha az içerikle karşılaşmak isteyen kullanıcılar doğrudan ağ tabanlı filtreleme yöntemlerinden de yararlanabilir. Son olarak, dijital platformlar tarafından hesapların askıya alınması ya da devre dışı bırakılması, çevrimiçi dezenformasyonla mücadelede kullanılabilecek yöntemler arasındadır. Fakat, teknoloji sağlayıcılar, kullanıcılarının hizmet şartlarını kötüye kullanması, topluluk kurallarına veya mevzuata uymaması durumunda bu önemleri almaktadırlar (Bontridder & Poullet, 2021). Ayrıca bu noktada dijital platformların ön filtreleme ve engellemelerinin ne sıklıkta ve hangi koşullar altında gerçekleştiğinin tam olarak bilinemeyeceğinin de altı çizilmektedir (Marsden & Meyer, 2019). Yine tüm bu önlemler, çevrimiçi platformlarda yer alan dezenformasyonun dolaşıma girmemesini engelleyememekte; sadece yayılımının kısıtlanması anlamına gelmektedir.

Yapay zekâ teknolojilerinin yardımcı rolünün ötesinde, başlı başına çözümün bir parçası olarak nasıl kullanılabileceği üzerinde de önemle durulmaktadır. Konu hakkında Akers ve arkadaşları (2018), çevrimiçi ortamda yanlış ve yanıltıcı içerikleri tespit etmek ve bu tür içerikleri düzenlemek için yararlanılabilecek yapay zekâ tekniklerini şöyle sıralamaktadır: (1) yanlış ve doğru bilgi içeren etiketlenmiş veriler üzerinden makine öğrenmesi aracılığıyla içerik üretim modellerinin baştan sona eğitimi, (2) doğru içeriğin dışarıdan bireyler yoluyla tespiti, (3) konu hakkında bilgi sahibi bireyler ve makine öğrenmesi yoluyla üslubun analiz edilerek yanıltıcı tarzın tespiti, (4) içerik yerine paylaşanın profili, nitelikleri gibi meta veriler yoluyla içeriğin tespiti. Söz konusu öneriler, şüphesiz, çevrimiçi dezenformasyonla mücadele için sorunlu içeriğin tespiti noktasında önemli avantajlar sunmaktadır. Ancak, her birinin kendi içerisinde barındırdığı sınırlılıklar nedeniyle sorgulandığı da görülmektedir (Akers et al., 2018; Bontridder & Poullet, 2021; Kertysova, 2018; Lee vd., 2019; Montoro-Montarroso et al., 2023): Öncelikle, eğitim verisinin içeriğinin nasıl etiketlendiği oldukça belirleyicidir. Çünkü algoritmalar, belirli bir grup içerisindeki bireyler, görüş ya da düşünceler için daha az avantajlı sonuçlar üreterek, insan önyargılarını kopyalama ve hatta otomatikleştirme özelliği taşımaktadır. Ayrıca, her ne kadar sistem doğru ve yanlış olmak üzere etiketlenmiş içeriği birbirinden kolaylıkla ayırabilecek olsa da ihtiyaç duyulan büyük miktardaki veriyi etiketleyebilmek oldukça zorlu bir süreci kapsamakta, hatta çoğu dilde yeterli miktarda eğitim verisi olmadığı görülmekte ve makine öğrenmesi yoluyla yapılacak her atama söz konusu veri kümelerindeki önyargılardan etkilenmektedir. Bu durum yapay zekânın aşırı kapsayıcılık özelliğinden kaynaklanmakta ve aşırı negatif ya da aşırı pozitif değer yükleme nedeniyle yasal/doğru içeriğin engellenmesi, ifade özgürlüğünün kısıtlanması gibi sonuçlar doğurabilmektedir (Marsden & Meyer, 2019; West, 2017). Dahası, mevcut yapay zekâ sistemleri henüz temel bildirimsel ifadeleri tanımlayabilmekte yani ima edilenler, ironiler, karmaşık cümle yapıları içerisindeki vurgular, bağlamsal ve kültürel ipuçları, alaylar, dilsel engeller,

politik ortamlar gibi incelikli dezenformasyon biçimlerini gözden kaçırabilmektedir (Graves, 2018; Vincent, 2019).

Yukarıda yer verilen tüm bu dezavantajları ortadan kaldırabilmek için yanlış ve yanıltıcı içeriğin tespitinde insan kaynaklardan yararlanılması önerilmektedir. Fakat dezenformasyon hacmi büyüdükçe, manuel doğrulamanın çevrimiçi içeriği değerlendirmede giderek daha etkisiz ve verimsiz bir hâl aldığı ile karşılaşılmaktadır (Kertysova, 2018). İçeriğin dışarıdan birey görüşlerine başvurularak tespiti, doğal dil nüanslarından kaynaklanan anlam farklılıklarını ortadan kaldırabilecek potansiyele sahip olmakla beraber, içeriği doğrulayan bireylerin kişisel yargıları ve bilgi kaynaklarının geçerliliği noktasındaki kuşkuyu da gidememektedir. Ayrıca manuel doğrulama, insan kaynağın ruh hâli, ahlaki değerleri, kişisel geçmişi gibi değişkenlerle doğrudan ilişkilenmektedir (Lekach, 2018; Newton, 2019). Bu durum manuel doğrulamanın maliyetinin yanı sıra hataya ve tahmin edilemez sonuçlara açık olmasını da beraberinde getirmektedir.

Bu sebeplerle her iki kaynağın da üslubu analiz etmek ve incelikli dezenformasyonun nitelikli tespiti için kullanıldığı, yani her iki doğrulama tekniği aracılığıyla bir teyit mekanizması oluşturulduğu görülmektedir. Bu teyit mekanizması içerisinde en sık görülen içerikler için yapay zekâ sistemlerine, daha hibrit modeller için ise insan incelemesinin birleşimine başvurulmaktadır. Fakat bu durumda da içeriğin yanıltıcı tarzının varlığı veya yokluğunun her zaman bilginin yanlışlığı veya doğruluğuyla doğrudan ilişkili olup olmadığı gündeme gelmektedir. Son olarak, sadece meta verilerin analizi ise içerikten bağımsız bir incelemenin gerçekleştirilmesine, içeriğin doğasından kaynaklanan yanıltıcı bilgilerin göz ardı edilmesine neden olması bakımından sınırlıdır. Dahası, yapay zekâ sistemleriyle geliştirilerek yaygınlaştırılan yanıltıcı içeriklerle nasıl bir mücadele stratejisinin geliştirileceği, dezenformasyon kaynağının türüyle (deepfake videolarının belirlenmesinde anormal göz kapağı hareketlerini tespit eden teknolojilerin geliştirilmesi, sosyal botların belirlenmesinde hesap davranışlarının izlenmesi gibi) de yakından ilişkilidir (Akers et al., 2018; Kertysova, 2018).

Yapay zekâ sistemlerinin kompleks çalışma biçimi ele alındığında tek bir çözüm önerisinden ziyade çok yönlü ve bütüncül bir yapılanmaya ihtiyaç duyulduğunu söylemek yanlış olmayacaktır. Ayrıca, dezenformasyonla mücadelede yapay zekânın kullanılmasının çeşitli etik sorunlara kaynaklık edebileceği de hatırlatılmalıdır. İlk olarak, dezenformasyon içeriğini tespit etmek için yapay zekâ sistemlerinin kullanılması, herhangi bir konu hakkındaki dezenformasyonun neleri kapsadığının net biçimde tanımlanmasını gerektirir. Fakat dezenformasyon kapsamına nelerin girip nelerin girmediğini tanımlamak, karar vericinin de kim olacağı sorusunu gündeme getirmektedir (Bontridder & Poullet, 2021). Yine yanlış bilginin dezenformasyondan nasıl ayrılacağı, içeriğin dolaşıma sokulma niyetinin ne olduğu ve içeriği tehdit olarak görmeyenlerin nasıl ayrıştırılacağı, ifade ve fikir edinme özgürlüğünün nasıl korunacağı

gibi birçok kaygı nedeni de ortaya çıkmaktadır. Bununla birlikte yukarıda da değinildiği üzere, dezenformasyonun algoritmalar aracılığıyla tespit edilmesi, içeriklerin aşırı biçimde pozitif ya da negatif değerle etiketlenmesine neden olarak meşru içeriğin sansürlenmesine ya da göz ardı edilen dil nüansları sebebiyle dezenformasyonun asimetrik bir müdahalesine kaynaklık edebilecektir. Dolayısıyla, her ne kadar mevcut dijital ekosistem, bireylerin manipüle edilmelerine sınırlamalar getirme zorunluluğunu gerekli kılsa da yanlış ve yanıltıcı bilgilerle mücadele için kullanılan programlama algoritmaları da bireyin görüşlerini özgürce ifade etmesinin önünde yeni engellere neden olabilmesi bakımından problemlidir.

AMAÇ VE YÖNTEM

Araştırma, yapay zekâ sistemlerinin dezenformasyonla mücadele sürecindeki potansiyelinin yapay zekâ uzmanlarının gözünden değerlendirilmesi amacını taşımaktadır. Bu çerçevede, yapay zekâ uzmanları tarafından çevrimiçi dezenformasyonun nitelik ve risklerinin değerlendirilme biçimlerinin ortaya koyulması ve dezenformasyonla mücadelede mevcut yapay zekâ tekniklerinden nasıl yararlanılabileceğine ilişkin bir yol haritasının çizilebilmesi hedeflenmektedir. Bu hedef doğrultusunda yanıt aranan araştırma soruları aşağıdaki gibidir:

AS1: Yapay zekâ uzmanları, çevrimiçi dezenformasyonla mücadele sürecinde yapay zekânın rolünü nasıl tanımlamakta ve değerlendirmektedir?

AS2: Yapay zekâ uzmanları, dezenformatif içeriğin etkisini artıran algoritma kaynaklı süreçlere dair nasıl bir strateji önermektedir?

AS3: Yapay zekâ uzmanları, çevrimiçi dezenformasyonun tespiti ve önlenebilmesi (doğrulama platformu girişimleri, kamu politikalarına adaptasyon, dijital platformların sorumluluğu gibi) için nasıl çözüm yolları öncelemektedir?

AS4: Yapay zekâ uzmanları, çevrimiçi dezenformasyonla mücadele sürecinde yapay zekâ sistemlerinin doğasından kaynaklanan sınırlılıklarını (aşırı kapsayıcılık, bağlamsal/kültürel/politik ipuçlarını kaçırma, ironi, dilsel engeller gibi) nasıl değerlendirmektedir?

Araştırma sorularını yanıtlayabilmek için yarı yapılandırılmış görüşme tekniğinin kullanıldığı betimsel yönteme dayalı bir alan araştırması gerçekleştirilmiştir. Nitel araştırma perspektifine uygun olarak, amaca yönelik örneklem benimsenmiş; Yapay Zekâ Politikaları Derneği (AIPA) üyesi ve paydaşı olan 28 uzmanla yarı yapılandırılmış derinlemesine görüşmeler gerçekleştirilmiştir. Araştırmada yer alan çalışma grubunun profili şöyledir:

Tablo 1. Çalışma Grubu Profili			
Katılımcı Kodu	Cinsiyet	Yaş	Yapay Zekâ Alanındaki Uzmanlığı
U1	Erkek	50	Mühendis
U2	Kadın	43	Akademisyen
U3	Erkek	45	Mühendis
U4	Kadın	46	Akademisyen
U5	Kadın	40	Yönetici
U6	Erkek	46	Gazeteci
U7	Erkek	21	Öğrenci
U8	Erkek	48	Akademisyen
U9	Erkek	35	Veri bilimci
U10	Kadın	41	Akademisyen
U11	Erkek	41	Akademisyen
U12	Erkek	28	Yapay zekâ mühendisi
U13	Erkek	38	Akademisyen
U14	Erkek	38	Girişimci
U15	Erkek	34	Direktör
U16	Kadın	30	Avukat /Hukuk teknolojileri
U17	Erkek	29	Grafik tasarımcı
U18	Kadın	43	Stratejik iletişim yöneticisi
U19	Kadın	30	Akademisyen
U20	Kadın	43	Akademisyen
U21	Erkek	52	Akademisyen
U22	Kadın	22	İstatistikçi
U23	Erkek	23	Veri bilimci
U24	Erkek	22	Yapay zekâ mühendisi
U25	Erkek	25	Yapay zekâ mühendisi
U26	Kadın	43	Akademisyen
U27	Erkek	44	Hukukçu
U28	Erkek	38	Akademisyen

Katılımcı ifadeleri görüşmeler sırasında kayıt altına alınmış, tüm görüşmeler 24.09.2023-07.10.2023 tarihleri arasında tamamlanmıştır. Toplanan veri yeterli derinlik ve doygunluğa ulaştığında ise görüşmeler sonlandırılmıştır. Ayrıca araştırmacının etik sorumluluğu gereği, katılımcılara araştırma hakkında önceden bilgi verilmiş, kişisel

bilgilerinin gizli tutulacağı aktarılmış, katılımcı isimleri anonimliği koruyabilmek için U1, U2 şeklinde kodlanmış, veri toplama süreci öncesinde etik kurul izni alınmıştır.

Verinin analizi sürecinde geçerlilik ve güvenilirliğinin sağlanabilmesi amacıyla belirli adımların izlenmesine dikkat edilmiştir. Öncelikle iç geçerliliğinin sağlanabilmesi için verinin toplanması ve analizi sürecinde çalışmanın literatürü, alandaki benzer araştırma analizleriyle (Akers et al., 2018; Bontridder & Poullet, 2021; Gupta et al., 2022; Kertysova, 2018; Marsden & Meyer, 2019) karşılaştırılmıştır. Katılımcı ifadelerinden doğrudan alıntılara yer verilmiş, çeşitli ifadelerine ilişkin teyit alınmış, katılımcılar yaş, cinsiyet ve meslek grubu bakımından çeşitlilik gösterecek biçimde seçilmiştir. Dış geçerlilik için yine katılımcı ifadelerinden doğrudan alıntılara başvurulmuş, örneklem seçim kriteri (AIPA üyesi veya paydaşı olma) ve katılımcı profiline ilişkin detaylar sunularak bulguların aktarılabilirliği sağlanmaya çalışılmıştır. Verileri inceleme, kodlama, kodlardan temalara ulaşma, temalar arası bağlantılar kurma ve yorumlama süreci her aşamada sıklıkla kontrol edilerek yürütülmüş, veri analiz süreci ayrıntılı biçimde açıklanmış ve sonuçlara nasıl ulaşıldığına ilişkin araştırma şeffaflığı sağlanmaya çalışılmıştır. İç güvenilirliğinin sağlanabilmesi için ulaşılan veriler kaydedilmiş, ortak görüşme protokolüne sadık kalınmış, veriler, belirlenen kod ve kategorilere uygun olarak organize edilmiştir. Dış güvenirliliğin sağlanabilmesi için ise araştırma sürecinin detaylı aktarılmasına dikkat edilerek çalışma bulguları katılımcıların ifadelerinden yapılan doğrudan alıntılarla desteklenmiş ve farklı katılımcı ifadeleri göz ardı edilmeden yansıtılmaya çalışılmıştır.

BULGULAR

Çevrimiçi Dezenformasyonla Mücadelede Yapay Zekâ

Çalışmada yer alan uzmanların neredeyse tümünün (27 katılımcı) yapay zekânın dezenformatif içeriği arttırdığı, yaygınlaştırdığı, fark edilmesini zorlaştırdığı ve daha ikna edici bir nitelik kazandırdığı hususunda hemfikir olduklarını söylemek mümkündür. Örneğin sırasıyla U10 ve U3, yapay zekâ sistemlerinin kullanılmasıyla çevrimiçi dezenformasyonun nasıl bir nitelik kazandığını şöyle aktarmaktadır:

> “Günümüzde yapay zekâ ile dezenformasyon daha inandırıcı ve görünür hâle gelmiştir. Deepfake teknolojisi bu anlamda dikkatle takip edilmelidir. Bu durum, sürekli manipüle edilen bireylerin bilgiye şüphe ile yaklaşmasının yanı sıra güven kaybına da neden olabilmektedir. Dezenformasyondan korunmaya çalışan bireyler doğru ve yeterli bilgiyle arasına mesafede koyabilmektedir.”

> “Generative AI’nin bu konuda oldukça etkisi oldu. Deepfake’ten de öte bir noktaya geçerek günümüzde hiç olmayan bir insanı dijital ortamda

> yaratabilir, var olan bir insana da dilediğimiz gibi istediğimizi yaptırabiliriz. Ayrıca haber üretme ve bu haberleri gerçek dışı kanıtlar üreterek etkin biçimde yayma gibi birçok şeyi de yapabiliriz."

Ayrıca uzmanların önümüzdeki yıllarda yapay zekâ teknolojisinin ilerlemesine bağlı olarak, çevrimiçi dezenformasyonun daha kompleks bir nitelik kazanacağını söyledikleri de görülmektedir. Bu kapsamda, U9 şöyle aktarmaktadır:

> "Yapay zekâ uygulamaları şimdiden ikna edici bir şekilde fotoğraf, ses kaydı ve video oluşturacak teknik kabiliyete erişti. Bu kabiliyetlerin kod yazmadan kullanılabilmesini sağlayan programlar arttıkça dezenformasyon konusunda da artış göreceğiz. Özellikle kendilerine ilişkin yüksek miktarda görüntü ve ses kaydı olan ünlü şahıslar için eğitim verisi fazla olacağından onlarla ilgili sahte görüntü ve ses üretmek zor olmayacaktır."

Diğer yandan, uzmanların her geçen gün gelişen yapay zekâ sistemlerinin çevrimiçi dezenformasyonla mücadelede yeni fırsatlar sunabileceği konusunda da hemfikir oldukları görülmektedir. Bu bağlamda, yapay zekânın dil ve anlam öğrenmesi, doğru veri girişiyle güncel ve güvenilir tutulması aracılığıyla dezenformatif içeriğin kaynağının tespiti ve engellenmesinde çözümün bir parçası olabileceği de belirtilmektedir. Örneğin U10, yapay zekânın "veri etiketleme, deepfake algılama, dezenformasyon algılama, filtreleme, bot tespit etme, içerik etiketleme ve izleme", U4 "bot hesapların tespiti ve içerik filtreleme", U14 "doğrulama platformları, içerik filtreleme, deepfake tespit mekanizmaları" gibi noktalarda çevrimiçi dezenformasyonla mücadelede önemli bir rol üstlenebileceğini ancak bunun için sürekli iyileştirme yaklaşımını esas alan girişimlere ağırlık verilmesi gerektiğini söylemektedir. U24 ise yapay zekânın dezenformatif içeriğin tespitinde nasıl kullanılabileceğini şöyle açıklamaktadır:

> "Yapay zekâ, metinler üzerinde semantik ve dil bilgisi analizleri yaparak sahte haberleri ve yanıltıcı bilgileri tespit edebilir. Bu, özellikle büyük veri kümeleri üzerinde hızlı bir filtreleme sağlamaktadır. Aynı zamanda, deepfake gibi sahte videoların ve manipüle edilmiş görsellerin tespiti için yapay zekâ algoritmaları oldukça etkilidir. Bir bilginin veya haberin internetteki ilk ortaya çıkışını ve yayılışını takip eden yapay zekâ, bilginin güvenilir bir kaynaktan gelip gelmediğini belirleyebilir. Ayrıca, bir haberin farklı kaynaklarda nasıl sunulduğunu karşılaştırarak yanıltıcı bilgi içerip içermediğini değerlendirme kapasitesine sahiptir. Yapay zekâ destekli eğitsel uygulamalar, kullanıcılara dezenformasyon taktiklerini tanımlama ve sahte haberleri ayırt etme becerileri kazandırabilir. Kullanıcılar bir web sitesini ziyaret ettiklerinde veya bir haber okuduklarında, yapay zekâ algoritması yanıltıcı bilgi içerikleri konusunda gerçek zamanlı uyarılar sağlayarak onları bilgilendirebilir."

Dolayısıyla, uzmanların, çevrimiçi dezenformasyonun mevcut yapay zekâ teknolojileriyle daha da ikna edici ve tespiti zor bir nitelik kazandığını düşündüklerini; ancak yine yapay zekâ teknolojinin sunduğu olanaklarla dezenformatif içerikle mücadelenin mümkün olduğunu eklediklerini söylemek mümkündür.

Dezenformatif İçeriğin Etkisini Azaltma: Strateji Önerileri

Uzmanlara, yapay zekâ sistemlerinin çevrimiçi dezenformasyonla mücadelede üstlendiği rolün yanı sıra bireysel kullanıcının dijital platformlarda karşılaştığı algoritma kaynaklı (hedef kitle segmentasyonu, aşırı kişiselleştirilmiş hedefleme, içerik döngüsü dışında tutulma gibi) dezenformatif içeriklere karşı nasıl bir strateji geliştirilebileceği de sorulmuştur. Bu kapsamda, uzmanların bir bölümü tarafından (5 katılımcı) yine yapay zekâ sistemlerinden yararlanılması suretiyle bir koruma mekanizmasının oluşturulması gerektiğinin önerildiği görülmektedir. Örneğin U8 "konuda bireyselleştirilmiş savunma sistemleri (bu tip kişiselleştirilmiş hedefleme sürecine maruz kalınıp kalınmadığı konusunda uyaran sistemler) en mantıklıca çözüm olacaktır." demekte; U14 "Kullanıcının kendi rızasıyla hesabına entegre edeceği, hesaplarıyla doğrudan etkileşime geçen, adresleme/segmentasyon çalışmalarını tespit eden, kullanıcıya olasılık hesabı sunan bir savunma motorunun geliştirilmesi" önerisinde bulunmaktadır.

Bunun ötesinde, katılımcıların büyük bir çoğunluğunun kullanıcıların farkındalık kazanmasının gerekliliği üzerinde durduklarını söylemek mümkündür. Söz konusu farkındalık vurgusu hem yapay zekânın dezenformasyonun yayılması hususunda oynadığı rolün tanınması (17 katılımcı) hem de çevrimiçi dezenformasyonun engellenmesi için gerçekleştirilen girişimler (9 katılımcı) hakkında kullanıcıların bilgilendirilmesini içermektedir. Bu kapsamda U24 şunları aktarmaktadır:

> "Dezenformasyon içeriğinin hedef kitle segmentasyonu ve aşırı kişiselleştirilmiş hedefleme yoluyla etkisinin arttığı bir dijital çağda yaşamaktayız. Bu süreçlere karşı savunma yaratmak için, öncelikle kullanıcıların dijital okuryazarlık seviyelerini artırmak kritik bir adımdır. Kullanıcıların çevrimiçi içeriklerin nasıl ve neden sunulduğunu anlamalarını sağlayacak eğitimler ve farkındalık kampanyaları düzenlenmelidir. (...) En önemlisi, bireylerin kritik düşünme yeteneklerini geliştirmeleri ve karşılaştıkları bilgilere sorgulayıcı yaklaşmaları teşvik edilmelidir."

Dahası, kullanıcı farkındalığının artırılması konusunda önemli bir kavram olarak yapay zekâ okuryazarlığının öne çıktığı görülmektedir. Örneğin U11 "dezenformasyona maruz kalanların yapay zekâ okuryazarı olarak bu yanlış bilgilere karşı koyacak yetkinliğe" ulaşmasının elzem olduğunu; U13 "kullanıcı farkındalığı için veri okuryazarlığı ve yapay zekâ okuryazarlığı eğitim ve denetimlerinin" şart olduğunu; U15 "kullanıcıla-

rın dezenformasyona maruz kalmaması için iyi birer teknoloji okuryazarı olup yapay zekâ özelinde daha fazla yoğunlaşmaları" gerektiğini söylemektedir. U19 ise bireysel kullanıcının algoritma farkındalığı geliştirebilmesinde yapay zekâ okuryazarlığı yetkinliği kazanmasının nasıl bir belirleyiciliğinin olduğunu şu şekilde açıklamaktadır:

> "Kullanıcıların, dijital ortamla ilişkilenme sürecinde algoritmik deneyimlerinde daha fazla kontrol sahibi olmaları ve aynı zamanda algoritmik sistemlerin mevcut zararlarından ve potansiyel risklerinden korunmaları, onlarla mücadele etmeleri için medya okuryazarlığı ile birlikte yapay zekâ/algoritma okuryazarı olmaları gerektiği düşüncesi de motive etmiştir. (...) Edilgen kullanıcıdan eyleyen kullanıcıya geçişin sağlanmasına yönelik olarak algoritma okuryazarlığı eğitimi alan bireylerde; algoritmaları tanıma, algoritmaları şekillendiren etkileşimleri ve davranışları açıklayabilme, kendi eğilimlerini görebilme ve sorgulayabilme, güncel gelişmelerde/haber içeriklerinde ve olaylarında kaynak doğrulama pratiğini ve sorgulama refleksini geliştirmiş olma, kişisel verilerinin kontrolü noktasında çaba sarf etme gibi niteliklerin bulunması gerekmektedir."

Bu çerçevede, yapay zekâ uzmanlarının algoritma kaynaklı dezenformatif içeriğin etkisinin azaltılabilmesi için yapay zekâ sistemlerine başvurulabileceğini hatırlatmalarıyla birlikte, kullanıcı farkındalığının arttırılması gerekliliğinin altını da önemle çizdiklerini söylemek mümkündür. Bahsi geçen kullanıcı farkındalığı ve yapay zekâ okuryazarlığı becerisinin kazanılması, hem sorunlu içeriğin görülebilmesi hem ilgili teknolojik araçların kullanılabilmesi hem de kullanıcının kendi kullanım eğilimlerini görebilme yetkinliğine sahip olabilmesi vurgusunu içermektedir.

Dezenformasyonun Tespiti ve Önlenmesine Yönelik Çözüm Yolları

Çalışma kapsamında, uzmanlara, dezenformatif içeriğin tespiti ve önlenebilmesi kapsamında neler yapılabileceği, devlet ve özel girişimlerin bu konuda nasıl sorumluluk alması gerektiği hakkındaki düşünceleri de sorulmuştur. Çalışmada yer alan uzmanların neredeyse tümü kamusal politikalar çerçevesinde kamu, kullanıcı ve dijital platformlar olmak üzere dezenformasyonla mücadelenin üç eksenli yürütülmesi gerekliliğine dikkat çekmektedir. Örneğin U7, "Kullanıcıların doğru bilgilere ulaşmasında yapay zekâ sistemlerinin neden olduğu zorluklar teknoloji şirketleri, düzenleyiciler ve genel toplum tarafından ele alınmalıdır. Bu sorunların çözümü için teknoloji, şeffaflık ve kullanıcı eğitimi gibi çoklu yaklaşımların birleştirilmesi gereklidir." demekte; U19, "Etik ve sorumlu teknolojileri içerecek biçimde kullanıcı otonomisine dayalı dijital platform yapılarının oluşmasına odaklı regülasyonlara kaynaklık edecek projelerin medya profesyonelleri ve algoritma geliştiricilerin iş birliğinde tasarlanması ve yürütülmesi" gerektiğini eklemektedir.

Dahası, dijital platformların kullanıcıya karşı sorumluluğu ve gerekli kamu politikalarının desteklenmesi noktasında daha aktif rol almaları gerektiği (19 katılımcı) de hatırlatılmaktadır. Örneğin U6 "Kamusal politikalar, bu konuda genel bir çerçeve oluşturmalı ve temel kuralları belirlemelidir. Ancak sosyal medya şirketleri, kendi platformlarında doğrudan önleyici stratejileri geliştirmelidir. Çünkü onlar dezenformasyonun en hızlı yayıldığı ve etkilediği yerlerdir." U19 "Sosyal medya platformları ve web tarayıcıları, kişiselleştirilmiş hedeflemenin nasıl çalıştığına dair şeffaflığı artırmalı ve kullanıcılara bu süreçleri kontrol edebilmeleri için araçlar sunmalıdır" demekte; U3 "sosyal medya şirketlerinin savunma hattının ilk blokunu oluşturması ve sonra kamusal çalışmalar ile desteklenmesi gibi bir hibrit yaklaşım" önermektedir. U24 ise dijital platformların dezenformatif içeriği filtrelemede kullanıcı verilerinden de yararlanılabileceğini eklemektedir:

> "Sosyal medya platformları ve diğer dijital hizmet sağlayıcılar, kullanıcı geri bildirimlerini dikkate alarak algoritma hatalarını düzeltebilir ve yapay zekânın daha doğru ve etkili hâle gelmesine katkıda bulunabilir. Bu yaklaşım hem dezenformasyonun yayılmasını sınırlar hem de kullanıcılara dijital ortamda daha bilinçli hareket etme yeteneği kazandırır."

Hatta bazı uzmanlar tarafından dezenformasyona karşı küresel çapta alınacak önlemlerin dijital platformlardan talep edilmesi gerektiği de belirtilmektedir. Örneğin, U13 "Bu noktada uygulanabilecek çözüm, küresel çapta yapay zekâ geliştirme kuralları uygulamak ve bunları şirketlere zorunlu olarak lanse etmektir." demekte; U2, kamu politikası yoluyla yapay zekâ ürünü dezenformatif içeriklere karşı yaptırım uygulanması gerektiğinden bahsetmektedir.

Doğrulama platformlarının yaygınlaştırılması da dezenformasyonun önlenebilmesi sürecinde üzerinde önemle durulan (15 katılımcı) konulardan biri olmuştur. Örneğin U7, doğrulama platformlarının daha etkin bir rol üstlenmesi gerektiğini ve farklı doğrulama platformlarının birbirinden beslenerek bilgi teyidi gerçekleştirebileceklerini söylemektedir: "Farklı doğrulama platformları arasında iş birliği teşvik edilmelidir. Birçok doğrulama platformu benzer bilgileri farklı kaynaklardan alabilir. Bu bilgilerin birleştirilmesi, doğruluğunu artırabilir." Yine U10 "Kullanıcılara, doğrulama kaynaklarına erişim sağlayan araçlar sunulabilir. Bu, kullanıcıların bilgiyi kendi kaynaklarından doğrulamalarına yardımcı olabilir." diyerek kullanıcının doğrulama platformlarına doğrudan erişim yollarının genişletilmesi gerektiğine işaret etmektedir.

Bunlara ek olarak bazı uzmanların, yapay zekâ sistemlerinin dezenformatif içeriği tespit ve engellemeye yönelik olarak geliştirilebilmesi için konunun uzmanı araştırmacıların desteklenmesi gerekliliği üzerinde durdukları da görülmektedir. Bu kapsamda, farklı dezenformasyon türlerine kaynaklık eden her yapay zekâ sistemi için o konuda uzman kişilerin desteklenmesine ve bunun toplumun tüm tabanına yayılmasını

mümkün kılacak biçimde farkındalığı artıran ve etik değerlere saygı gösterilmesini önceleyen politikalara ihtiyaç duyulduğu belirtilmektedir. Örneğin U14 yapay zekânın "dil ve anlam öğrenebilmesi için NLP (natural language processing/doğal dil işleme) konusunda çalışan ve Türkçe kaynaklara erişebilen girişimlerin, deepfake konusunda ise yüz tanıma ve doğrulama çalışması yapan ekiplerin desteklenmesi" gerektiğini söylemekte; U5 "insan hayatını kolaylaştıran ve etik değerlere saygı gösteren projelerin" öne çıkarılması gerektiğini belirtmektedir.

Uzman görüşleri üzerinden değerlendirildiğinde, kamu ve dijital platformların iş birliği içerisinde olması ve kullanıcıya karşı sorumlulukların bilincinde, etik değerlere saygılı stratejilerin izlenmesi gerekliliği ön plana çıkmaktadır. Dolayısıyla uzmanlara göre sadece dezenformasyona karşı kamu eliyle alınabilecek önlemler değil, dijital platformların kendi içlerindeki kullanıcının doğru bilgiye erişimini önceleyen politikalar da gerekmektedir. Ayrıca bu süreçte doğrulama platformlarının yaygınlaştırılması ve yapay zekânın otonom biçimde dezenformasyonla tespit ve kısıtlama hususunda mücadele edebilmesini olanaklı kılacak teknolojik girişimlerin desteklenmesi gerektiği de eklenmektedir.

Yapay Zekâ Sistemlerinin Sınırlılıklarını Aşmak

Dezenformasyonla mücadele sürecinde yapay zekâ sistemlerinin kendi doğasından kaynaklanan sınırlılıkların nasıl üstesinden gelinebileceği çerçevesinde de uzman görüşlerine başvurulmuştur. Bu kapsamda, uzmanların bir bölümü (8 katılımcı) öncelikle yapay zekâ çalışmalarının ilerlemesine paralel olarak yapay zekâ sistemlerinin işleyişinden kaynaklanan dil ve anlamı anlama, bağlama yönelik çıkarımda bulunabilme gibi sorunların da azalacağına dikkat çekmektedir. Ayrıca uzmanlar, günümüz teknolojisiyle dahi yapay zekâ sistemlerinin çevrimiçi dezenformasyonun azaltılmasında büyük bir etkiye sahip olabileceğini de eklemektedir. Örneğin U6 şöyle aktarmaktadır: "Sonuçta yapay zekâ da teknoloji olarak şu anda "baby step" dediğimiz dönemini yaşıyor. Dolayısıyla normal bir süreçteyiz. Var olan sistemdeki sınırlılıklar bence dezenformasyonu yüksek ölçüde engelleyebilir ve kesinlikle de olmamasına karşın çok daha yararlı olacaktır." Yapay zekâ sistemlerindeki sınırlılıklarının kaldırılabilmesi için ise eğitici veri setinin güçlendirilmesi ve çeşitlendirilmesi üzerinde durulmaktadır. Örneğin U9 şunları söylemektedir: "Yapay zekâ konusunda ülkemizde çalışmaların artmasıyla muhtemelen Türkiye ve Türkçe özelinde eğitildiği için bu limitasyonların çoğu ortadan kalkacaktır. Kalan limitasyonlar da bu yapay zekâ uygulamalarının yayılmasıyla beraber muhtemelen kendiliğinden ortadan kalkacaktır." Yine U15 söz konusu kısıtlılıkların zaman içerisinde giderilmesiyle dezenformasyon hızının önemli ölçüde azalacağını söylemekte; fakat yapay zekâ sistemlerinin sürekli değişme potansiyeli olduğunu da eklemektedir. Yapay zekâ sistemlerinin doğasından kaynaklanan sınırlılıkların azaltılması konusunda ise veri setlerinin önemine şöyle vurgu yapmaktadır: "Ben bu

durumu, insanların kendisinden sonrakilere güzel bir miras bırakma ihtiyacına benzetiyorum. Nasıl ki medeniyetler kendinden öncekilerin üzerine inşa ediliyor ve oradan yükseliyor aynı durum verilerimizin oluşması, doğruluğu ve bizden sonrakilerin de bu veriler üzerinden insanlığın gelişmesi için kullanılması ile birlikte olacaktır."

Bununla birlikte, söz konusu sınırlılıkların aşılmasında kullanıcı ve yapay zekâ sistemlerinin etkileşimini güçlendirecek ve dezenformasyona karşı kullanıcı nezdinde farkındalık oluşturmayı olanaklı kılacak dijital okuryazarlık becerilerinin geliştirilmesi gerekliliği (15 katılımcı) de tekrar hatırlatılmaktadır. Buradaki okuryazarlık becerisi çerçevesinde kastedilen, kullanıcıların yapay zekanın mevcut sınırlılıklarının farkında olabilmesi hâlidir. Örneğin U24 şöyle aktarmaktadır: "Farkındalığın artırılması için, öncelikle yapay zekânın bu sınırlılıklarını anlamak ve bunları kullanıcılara açıklamak önemlidir. Kullanıcılara, algoritmalara tamamen güvenmek yerine kendi kritik düşünme yeteneklerini kullanmaları için teşvik edici eğitimler ve bilgilendirme kampanyaları düzenlenebilir." Yine U7, dezenformasyonla mücadelede yapay zekâ teknolojisinin sınırlılıklarının farkında olunmasının ve bu sınırlılıkların kullanıcı tarafından anlaşılmasının önemli olduğunu; U10 ise farkındalığın artırılmasında kullanıcılara, dezenformasyonu tanıma ve eleme konusunda eğitim verilmesi gerekliliğinin altını çizmektedir.

Bu kapsamda, araştırmada yer alan uzmanların öncelikle yapay zekâ teknolojisinin ilerlemesi ve nitelikli, güncel eğitim verisine ulaşılmasıyla birlikte, yapay zekânın doğasından kaynaklanan sınırlılıkların azalacağını düşündükleri; ancak söz konusu aşamaya gelene dek kullanıcıların dijital okuryazarlık bileşenlerinden biri olan yapay zekâ okuryazarlığı becerilerinin de geliştirilmesi gerektiğini belirttikleri görülmektedir.

SONUÇ

Çevrimiçi dezenformasyonla mücadele sürecinde mevcut yapay zekâ sistemlerinin potansiyelinin yapay zekâ uzmanlarının gözünden değerlendirilmesi ve geliştirilebilecek çeşitli stratejilere yönelik bir yol haritasının çizilebilmesinin hedeflendiği bu çalışmada, öncelikle yapay zekâ sistemlerinin dezenformasyonla mücadele sürecinde aktif biçimde rol alabileceğinin düşünüldüğünü söylemek mümkündür. Bu kapsamda dezenformasyona yönelik tespit ve filtreleme mekanizmalarının ön plana çıktığı görülmektedir. Bahsi geçen öneriler her ne kadar dezenformatif içeriğin dolaşıma girmesini engelleyemese ve dezenformasyon yaratan içeriğin sonradan analizini gerçekleştirse (Montoro-Montarosso et al., 2023) de dezenformatif içeriğin tespiti ve tespit yoluyla manipülasyona karşı durulması noktasında (Bontridder & Poullet, 2021) oldukça önem taşımaktadır.

Bununla birlikte hem algoritmanın çalışma biçiminden kaynaklanan hem de yapay zekâ sistemlerinden yararlanılarak dolaşıma sokulan dezenformatif içerikle mücadele edilmesinde yapay zekâ okuryazarlığı üzerinde durulmuştur. Söz konusu vurgu,

toplumsal farkındalığı ve eleştirel medya tüketimini artırmak açısından önem taşımakta ayrıca yeni bir okuryazarlık türü olarak da yapay zekâ sistemlerine ilişkin farkındalık ve beceri ihtiyacını öne çıkarmaktadır. Yine çevrimiçi dezenformasyonun tespiti ve önlenmesinde kamu ve dijital platformların iş birliği içerisinde olması gerektiği ve kullanıcıya karşı sorumlu, etik değerlere saygılı stratejilerin geliştirilmesi gerekliliği de ulaşılan çalışma sonuçları arasındadır. Bu kapsamda yapay zekâ uzmanları; dezenformasyona karşı alınabilecek önlemin sadece kamu eliyle olmayacağını, dijital platformların da kullanıcıyı önceleyen bir mekanizmayı ekosistemlerine uyarlamaları gerektiğini vurgulamaktadır. Bu doğrultuda dijital platformların kendilerini sadece içerik barındırıcılar olarak konumlandırmalarının ötesinde, içerik doğrulama mekanizmalarının entegrasyonu, içeriklerin topluluk kurallarına uygunluğunun hızlı denetimi, algoritma şeffaflığının öncelenmesi ve hatta bir konu hakkındaki içeriği farklı kanallardan teyit ederek verinin kaynağına ulaşıp doğru bilgiyi teyit edebilecek analiz becerisine sahip yapay zekâ sistemlerinin teşvik edilmesine ihtiyaç duyulduğundan söz etmek mümkündür. Ayrıca doğrudan kamu ya da bağımsız kuruluşlar tarafından sosyal medyada akan mesajlar üzerinden büyük dil modelleriyle desteklenen yapay zekâ programlarının geliştirilmesi ve ihtiyaç hâlinde insan moderasyonuyla kontrol edilerek etiketlenmesi; böylelikle toplumsal/siyasal infial uyandırma potansiyeli yüksek iletilerin önceden tespiti de atılacak önemli adımlar arasındadır. Dahası dezenformasyon tespitinin de ötesinde, yapay zekâ sistemlerinin sorumlu yapay zekâ rolünde kullanımı gerekmektedir. Kullanıcılar sorunlu bir web sitesini ziyaret ettiklerinde ya da güvenilir kaynaklardan gelmeyen içeriklerle karşılaştıklarında yapay zekâ sistemleri algoritmalar aracılığıyla gerçek zamanlı uyarılar sağlayabilme potansiyeline sahiptir. Yine gerekmesi durumunda doğrulamanın ötesinde düzeltici içerikler üretmesini sağlayacak sistemlerin geliştirilebilmesi üzerinde de durulmalı; söz konusu girişimler kamu, üniversite ve sanayi iş birliği içerisinde gerçekleştirilmelidir.

Tüm bu süreç içerisinde, hem web'in iş modelinden (Bontridder & Poullet, 2021) hem de mevcut yapay zekâ sistemlerinin doğasından kaynaklanan sınırlılıkların teknolojik ilerlemelerle ortadan kaldırılmasına dek kullanıcının kendini dezenformasyona karşı koruyabileceği teknik beceriyi kazanması gerekliliği ve algoritma seçimleri hakkında bilgilendirilmelerine duyulan ihtiyaç da çalışma özelinde eklenen bulgular arasında yer almaktadır. Çalışma, kullanıcı failliğinin önemini bir kez daha ortaya koymaktadır. Bu bakımdan yapay zekâ destekli eğitsel programlar, kullanıcıların dezenformasyon taktiklerini tanıma ve sahte haberleri ayırt etme becerilerini kazanabilmeleri için de kullanılabilir.

Şüphesiz, çevrimiçi dezenformasyonla mücadele sürecine yapay zekâ sistemlerinin adaptasyonu yasa yapıcılar, dijital platformlar, doğrulama platformları, bireysel kullanıcılar ve konuyla ilişkilenen diğer tüm paydaşlar ekseninde tek ve mutlak bir çözümün mümkün olamayacağı, değişime açık ve kompleks bir yapıya işaret etmektedir. Ancak, yapay zekâ sistemlerinin gelişimiyle birlikte, çevrimiçi dezenfor-

masyonla mücadelede önemli bir kazanım elde edilebileceği de açıktır. Bu bakımdan tanımları, riskleri, zorlukları ve bağlamları farklılık gösteren bir teknoloji için düzenleme ve politika geliştirme girişimleri dikkatli olmayı ve tüm paydaşlarla sürekli diyalog kurmayı gerektirmektedir. Dahası, dezenformasyonla mücadelede sadece teknolojik çözümlerin yeterli olmadığını da hatırlatmak gereklidir. Yapay zekâ sistemleri kadar kullanıcı deneyimini anlamaya ve her geçen gün tespit edilmesi daha da zorlaşan yeni nesil dezenformasyonu tanımlamayı odağına alan araştırmalara da ihtiyaç duyulmaktadır.

KAYNAKÇA

Akers, J., Bansal, G., Cadamuro, G., Chen, C., Chen, Q., Lin, L., Mulcaire, P., Nandakumar, R., Rockett, M., Simko, L., Toman, J., Wu, T., Zeng, E., Zorn, B. & Roesner, F. (2018). Technology-Enabled Disinformation: Summary, Lessons, and Recommendations. Technical Report UW-CSE, 21

Akhtar, P., Ghouri, A.M., Khan, H.R., ul Haq, M.A., Auan, U., Zahoor, N., Khan, Z., Ashrar, A. (2022). Detecting fake news and disinformation using artifical intelligence and machine learning to avoid supply chain distruptions. Annals of Operations Research, 327, 633-657.

Belhadi, A., Mani, V., Kamble, S.S., Khan, S.A.R., & Verma, S. (2021). Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: an empirical investigation. Annals of Operations Research, 021-03956-x.

Bergamini, D. (2020). Need for Democratic Governance of Artifical Intelligence. Comittee on Political Affairs and Democracy-Council of Europe. Retrieved https://pace.coe.int/en/files/28742 Erişim T. 13 Eylül 2023.

Bontridder, N. & Poullet, Y. (2021). The role of artificial intelligence in disinformation. Data & Policy, 3, e32.

Bouziane, M., Perrin, H., Cluzeau, A., Mardas, J. & Sadeq, A. (2020). Team Buster. ai at CheckThat! 2020 Insights and Recommendations to Improve Fact-Checkin, CLEF 2020.

Chesney, B. & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753.

Funke, D. (2019). These fact-checkers won $2 million to implement ai in their newsrooms. Poynter. Retrieved https://www.poynter.org/fact-checking/2019/these-fact-checkers-won-2-million-to-implement-ai-in-their-newsrooms/

Goodfellowet, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 1-9.

Graves, L. (2018). Understanding the Promise and Limits of Automated Fact-checking. The Reuters Institute for the Study of Journalism at the University of Oxford, February 2018.

Greengard, S. (2019). Will deepfakes do deep damage? Communications of the ACM, 63(1), 17-19.

Gupta, A., Li, H., Farnoush, A. & Jiang, K. (2022). Understanding patterns of COVID infodemic: A systematic and pragmatic approach to curb fake news. Journal of Business Research, 140, 670-683.

Gül-Ünlü, D. & Kesgin, Y. (2021). Tavşan deliği ve siyasal radikalleşme: YouTube kullanıcı önerileri üzerinden bir değerlendirme. In. A. Aydemir (Ed.), Gelenekselden Dijitale Siyasal İletişim Çalışmaları (ss.67-78). Konya: Eğitim Yayınevi.

Jackson, J. (2016). Fake news clampdown: Google gives €150,000 to fact-checking projects. The Guardian. Retrieved https://www.theguardian.com/media/2016/nov/17/fake-news-google-funding-fact-checking-us-election

Karakoç, E. & Kuş, O. & Gül Ünlü, D. (2023). Algoritma farkındalığı ve hayali olanaklar: İnsan Hakları ihlallerinin dijital mekanizması üzerine düşünmek. In. M.A. Göngen & Y. Kesgin (Ed.), Medya ve İnsan Hakları (ss.153-170). İstanbul: Kriter Yayınları.

Karakoç, E. & Zeybek, B. (2022). Görmek inanmaya yeter mi? Görsel dezenformasyonun ayırt edici biçimi olarak siyasi deepfake içerikler. Öneri Dergisi, 17(57), 50-72.

Kertysova, K. (2018). Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. Security and Human Rights, 29(1-4), 55-81.

Küçükşabanoğlu, Z. & Soysal, B. (2023). Yapay zekânın siyaseti. In. U. Demirezen (Ed.), Geleceği Şekillendiren Teknoloji Yapay Zekâ (ss.1-33). İstanbul: Nobel Yayıncılık.

Lekach, S. (2018). The cleaners shows the terrors human content moderators face at work. Marshable, 13 November 2018. https://mashable.com/article/the-cleaners-content-moderators-facebook-twitter-google Erişim T. 13 Eylül 2023.

Lamo, M. & Calo, R. (2018). Regulating Bot Speech. UCLA Law Review, 66, 988-1028.

Marechal, N. & Biddle, E.R. (2020). It's not just the content, it's the business model: Democracy's online speech challenge. A Report from Ranking Digital Rights, New America, 17 March 2020.

Masood, M., Nawaz, M., Malik, K.M., Javed, A., Irtaza, A. & Malik, H. (2022). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. Applied Intelligence, 54, 3974-4026.

Marsden, C. & Meyer, T. (2019). Regulating Disinformation with Artifical Intelligence: Effect of Disinformation Initiatives on Freedom of Experssion and Media Pluralism. European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA).

Montoro-Montarroso, A., Caton-Correa, J., Rosso, P., Chulvi, B., Panizo-Lledot, A., Huertas-Tato, J., Calvo-Figueras, B., Rementeria, M.J. & Gomez-Romero, J. (2023). Fighting disinformation with artifical intelligence: Fundamentals, advances and challenges. Profesional de la Informacion, 32(3), e320322.

Newton, C. (2019) The trauma floor: The secret lives of Facebook moderators in America. The Verge, 25 February 2019. Retrieved https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona Erişim T. 13 Eylül 2023.

Rosenbach, E. & Mansted, K. (2018). Can democracy survive in the information age?. Belfer Center for Science and International Affairs, 30. Retrieved https://www.belfercenter.org/publication/can-democracy-survive-information-age Erişim T. 15 Eylül 2023.

Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A. & Menczer, F. (2017). The spread of misinformation by social bots. ArXiv Preprint ArXiv:1707.07592.

Shrestha, Y.R., Ben-Menahem, S.M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. California Management Review, 61(4), 66–83

Stiff, H. & Johansson, F. (2022). Detecting computer-generated disinformation. International Journal of Data Science and Analytics, 13, 363-383.

Vincent, J. (2019). AI won't relieve the mistery of facebook's human moderators. The Verge, 27 February 2019. Retrieved https://www.theverge.com/2019/2/27/18242724/facebook-moderation-ai-artificial-intelligence-platforms. Erişim T. 15 Eylül 2023.

Walorska, A.M. (2020). Deepfakes and Disinformation. Friedrich Naumann Foundation for Freedom. https://www.freiheit.org/de/consent?dest=https%3A%2Fshop.freiheit.org%2F%23!%2F-Publikation%2F897. Erişim T. 15 Eylül 2023.

Wang, P., Angarita, R. & Renna, I. (2018). Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. In Companion Proceedings of the The Web Conference 2018 (pp. 1557-1561).

West, D.M. (2017). How to combat fake news and disinformation. The Brookings Institution, 18 December 2017. https://www.brookings.edu/articles/how-to-combat-fake-news-and-disinformation. Erişim T. 14 Eylül 2023.

Yazar katkı düzeyi/Author contributions:

Makale Tasarımı: D. Gül Ünlü & Z. Küçükşabanoğlu. Literatür Taraması: D. Gül Ünlü. Veri Toplama ve Analiz: D. Gül Ünlü & Z. Küçükşabanoğlu. Sonuç: D. Gül Ünlü & Z. Küçükşabanoğlu. Son Okuma, Kontrol ve Sorumluluk: D. Gül Ünlü & Z. Küçükşabanoğlu.

Design of article: D. Gül Ünlü & Z. Küçükşabanoğlu. Literature review: D. Gül Ünlü. Data acquisition and analysis: D. Gül Ünlü & Z. Küçükşabanoğlu. Conclusion: D. Gül Ünlü & Z. Küçükşabanoğlu. Final reading, checking and approval: D. Gül Ünlü & Z. Küçükşabanoğlu.

Hakem değerlendirmesi/Peer review:

Dış bağımsız/Externally peer reviewed

Çıkar çatışması/Conflict of interest:

Yazarlar çıkar çatışması bildirmemiştir/The authors have no conflict of interest to declare

Finansal destek/Grant support:

Yazarlar bu makalede finansal destek almadığını beyan etmiştir/The authors declared that this article has received no financial support.