

# Bilgi Çarpıtma Sürecinde Yapay Zekanın İşlevi

## The Function of Artificial Intelligence in the Information Distortion Process



İkram ALPSOY (Asst. Prof. Dr.)  
Ağrı İbrahim Çeçen University, Faculty of Communication  
Ağrı/Türkiye  
ialpsoy@agri.edu.tr

Başvuru Tarihi | Date Received: 10.02.2025  
Yayına Kabul Tarihi | Date Accepted: 01.07.2025  
Yayınlanma Tarihi | Date Published: 30.07.2025

Alpsoy, İ. (2025). Bilgi Çarpıtma Sürecinde Yapay Zekanın İşlevi. *Erciyes İletişim Dergisi*, 12(2), 551-568 <https://doi.org/10.17680/erciyesiletisim.1636521>

### Öz

Yapay zekâ sistemleri, günümüzde insanlığın karşı karşıya olduğu bilgi çarpıtma (dezenformasyon) olgusunda çok önemli bir rol oynamaktadır. Bu sistemlerin yapay zekâ aracılığı ile yapılan gerçekçi düzmece içeriklerin oluşturulmasına imkân sağlamasının yanı sıra art niyetli kişilerce bağlamından koparılan bilgilerin belirlenen kişilere ve geniş bir ölçekte ulaştırılmasını kolaylaştırma imkânı sunması problemi daha da artırmaktadır. Bu duruma tepki olarak çevrim içi bilgi çarpıtmayı belirlemek ve kontrol etmek amacıyla başka yapay zekâ sistemleri oluşturulmuştur. Bu sistemler de özellikle bilgi ve ifade özgürlüğüyle alakalı insan hakları ve etik bağlamındaki hassasiyetlerden kaçamamaktadır. Bu bağlamda çalışma, yapay zekâ ve bilgi çarpıtma ilişkisini betimleyici ve yazın taraması olarak sunmayı amaçlamaktadır. Bunun yanı sıra insanlar, gruplar ve toplumlar arasındaki iletişimde geçmişten farklı bir alanın oluşmasını sağlayan yapay zekânın faydalarına, olumsuzluklarına ve iletişim faaliyetlerindeki olası değişikliklere yapacağı katkılara odaklanılmaktadır. Araştırmada, web'in reklam gelirlerine dayalı iş modelinin bilginin çarpıtılmasında esas sorunu oluşturduğu belirlenmekte, böylelikle bu alanda yapılacak hukuki düzenleme ve denetimin sorunu büyük oranda azaltacağı öngörülmektedir.

**Anahtar Kelimeler:** Dijital Ekosistem, Bilgi Çarpıtma, Yapay Zekâ, İçerik Düzenlemesi, Etik Çıkarım.

### Abstract

Artificial intelligence systems are important in the information distortion phenomenon facing humanity today. These systems make it easy to disseminate realistic fake content on a large scale, which exacerbates the problem. To prevent this, other systems have been created to identify and control information distortion. These systems cannot escape the human rights and ethical sensitivities associated with freedom of information and expression. In this context, the study aims to present the relationship between artificial intelligence and disinformation as a descriptive, theoretical and literature review. It also focuses on the benefits and drawbacks of artificial intelligence, which opens a different field in interpersonal communication than in the past, and its contributions to possible changes in communication activities. The study finds that the business model of the web based on advertising revenues is important in disinformation, and predicts that legal regulations and controls can reduce the problem.

**Keywords:** Digital Ecosystem, Information Distortion, Artificial Intelligence, Content Curation, Ethical Inference.



## Giriş

Bilgi ve iletişim teknolojilerinde son yıllarda meydana gelen hızlı gelişmeler ve bunların her geçen gün yaygınlaşmasıyla birlikte doğru olmayan bilgiler yapay zekâ teknikleri de dahil olmak üzere gerçekçi bir şekilde ve kolaylıkla üretilebilmekte, ulaşılmak istenen kitleye yayılması çok hızlı ve yoğun bir şekilde gerçekleşmektedir. Bu durumun da geniş kapsamlı ve önemli neticeleri olmaktadır. Örneğin Birleşik Krallık'ın 2016 yılında Avrupa Birliği'nden ayrılma referandumu ve Amerika Birleşik Devletleri'nde 2016'da yapılan başkanlık seçimleriyle alakalı olarak medya ekosisteminden vatandaşların fikirlerini ve rey kararlarını etkilemede yararlanılmıştır (Bennett & Livingston, 2020; Howard & Kollanyi, 2016). Sosyal medya aygıtları arasında yer alan Facebook, Rohingya Müslümanlara karşı Myanmar'da nefret oluşturmak isteyenler için elverişli bir platforma dönüşmüş (Volger, 2010), WhatsApp'taki bazı söylentiler Hindistan'da sonu cinayetle biten olaylara neden olmuş (Dixit & Mac, 2018), Fransa'da ise bir profesör sosyal medyada kendisine karşı yürütülen şiddet kampanyası sonucu hayatını kaybetmiştir (Bindner & Gluck, 2020).

Yanlış bilgi, aldatma gayesi olmadan paylaşılan yanıltıcı ve hatalı bilgiyi ifade etmekte buna karşın bilgi çarpıtma (dezenformasyon) ise hedef kitleyi yanıltma amacıyla kamuoyuna aktarılan doğru olmayan, yanıltıcı ve hatalı bilgi olarak tanımlanmaktadır (European Commission, 2018). Dijital yapıdaki gelişmeyle ilintili olmakla birlikte daha geniş bir politik, toplumsal ve teknolojik unsurların kaynaklık ettiği dönüşümün bir ifadesi olan bilgi çarpıtma (Keller & Público, 2021, s. 488), tesadüfi olmayan yanıltıcı bilgi olma yönü ile öteki aldatıcı veya yalan bilgilerden ayrılmaktadır. Bundan dolayı bilgi çarpıtma, belli bir niyetle ve güdüleme ile belli bir fayda sağlamak için doğru olmadığı bilinen ve düzenli bir şekilde aktarılan bilgi olarak betimlenmektedir.

Dijital teknolojiye yaşanan gelişmeler ve sosyal medya aygıtları, yanlış bilginin ve çarpıtılmış bilginin oluşturulmasını ve yaygınlaşmasını daha da çoğaltmış olmakla birlikte Avrupa Birliği kurumları yalnızca bilgi çarpıtmayı hukuki ve teknik usullerle mücadele edilmesi gereken bir tehdit olarak değerlendirmiştir (European Commission, 2018). Bunda, temel hak ve özgürlüklerin, özellikle de ifade ve bilgi edinme hakkının korunması çabası etkili olmaktadır. Aksi takdirde herhangi bir kurum veya kişiyle alakalı gerçekliğin doğru kabul edilmemesi ve yazılı, görsel, işitsel vb. içeriklerin buna göre düzenlenmesi halinde ifade ve bilgi edinme özgürlüğünün olumsuz etkilenmesi durumu ortaya çıkabilmektedir (Chesney & Citron, 2019; Greengard, 2019; Shrestha vd., 2019).

Bilgi ve iletişim teknolojilerinde meydana gelen gelişmeler bilginin oluşturulma ve yayılma biçiminde değişikliğe neden olmuş, yapay zekâ teknikleri de bunda belirleyici olmuştur. Bu teknikler, art niyetli kişiler tarafından bilginin çarpıtılmasını ve yayılmasını kolaylaştırmakta, öte taraftan çevrim içi çarpıtılmış bilgiyle mücadele etmek için de kullanılmaktadır.

Bu bağlamda çalışmada betimsel içerik analizi ve yazın taraması yöntemleri ile yapay zekâ ve bilgi çarpıtma arasındaki ilişkinin açıklanması amaçlanmaktadır. İlk olarak yapay zekâ kavramı ve yapay zekâ aracılığıyla sahte içeriklerin üretilmesi detaylı bir şekilde ele alınmaktadır. Daha sonra yapay zekâ kullanılarak bilginin nasıl çarpıtıldığı ve yayıldığı incelenmektedir. Akabinde yapay zekâ tekniklerinin bilgi çarpıtma sürecindeki rolü üzerine yapılan etik değerlendirmelere ve bilgi çarpıtmaya karşı verilen mücadeleye odaklanılmaktadır. Son olarak ise temel hak ve özgürlükler göz önünde bulundurularak bilgi çarpıtma sorununun etik açıdan irdelenmesi için önerilerde bulunmaktadır.

## 1. Bilgi Çarpıtmanın Oluşturulma ve Yayılma Sürecinde Yapay Zekâ Tekniklerinin Rolü

Amerika Birleşik Devletleri'nde 1955'te John McCarthy, Nathan Rochester, Marvin Minsky ve Claude Shannon (1956) tarafından gerçekleştirilen bir proje neticesinde ortaya çıkan yapay zekâ ile insanlarla aynı şekilde davranış sergileyen akıllı makinelerin oluşturulması amaçlanmıştır (McCarthy vd., 2006). İnsanlar gibi, dil öğrenebilen bir bilgisayar programı olan yapay zekâ, birçok parçadan meydana gelen dizgelerin bütünü olan makinelerin öğrenme, düşünme, problemlere çözüm üretme vb. bilişsel fonksiyonu kapsayan görevleri yapabilecek konuma getirilmesi şeklinde betimlenmektedir (Kolarević, 2022, s. 113). İnsanlar gibi akılcı düşünmesi ve davranması amacıyla oluşturulan bir sistem olan yapay zekâ (Surden, 2019, s. 1310), Kaye'ye (2018) göre ise "karmaşık veya büyük ortamlarda, bir makinenin karar vermesi, model tanınması ve tahmin etmesinin doğruluğunu, hızını ve/veya ölçeğini otomatikleştirmek ve geliştirmek ya da belirli işlerde insan performansının yerine geçmek için kullanılan genel amaçlı bir teknolojidir" (s. 3).

İnsanoğlu, ilk dönemlerden itibaren geçmiş toplumlardaki araçlardan yararlanarak daha iyi ve işlevsel araçlar meydana getirmiştir. Yapay zekâ sistemleri de içerisinde bulunduğumuz yüzyılda öne çıkan araçlar içerisinde yer alarak günlük yaşamda ciddi değişimlere zemin hazırlamıştır. Bilginin gerçekliğinden çıkarılarak çarpıtılması olgusu, yapay zekâ tekniklerinin kullanılması suretiyle çevrim içi ortamda iki şekilde gerçekleştirilmektedir. Birincisi, yapay zekâ teknikleri ses, metin, video, fotoğraf, görüntü vb. içerikleri meydana getirmek ya da mevcut verilerde değişiklikler yapmak için yeni imkanlar sunmaktadır. İkincisi ise kullanıcı kitlenin katılım oranını artırmak amacıyla çevrim içi uygulamaların geliştirdiği ve kullandığı yapay zekâ sistemleri, çevrim içi ortamda bilgi çarpıtmanın etkin ve seri bir biçimde yaygınlaşmasına büyük ölçüde katkı sağlamaktadır. Her iki unsur da belirleyici faktörler olarak probleme katkıda bulunmaktadır (Goodfellow vd. 2014).

### 1.1. Yapay Zekâ Teknikleriyle Sahte İçeriklerin Oluşturulması

Yapay zekâ teknolojisi aracılığıyla üretilen sahte içerikler, deepfake (derin öğrenme ve sahte kelimelerinin birleşiminden oluşur) olarak tanımlanmaktadır. Teknoloji yardımıyla yapılan bilgi çarpıtma ile alakalı olarak (Akers vd., 2018), tahrif edilecek medya varlığını korudukça sahte medyanın da var olacağı, sanat eserleri, ehliyet, diploma, sertifika vb. hususlarla alakalı dijital belgeler üzerinde düzenleme imkânı sunan programların amatör kişilerce kullanımının yaygınlaşmasının da sorunu büyüttüğü değerlendirilmektedir (s.4). Yapay zekâ teknolojilerinde yakın dönemde meydana gelen gelişmeler ses, fotoğraf, video ve metinler üzerindeki manipülasyon imkanını kolaylaştırmış, özellikle ses ve videoların gerçekçilik düzeyleri daha da artmıştır.

Derin Sahteliklerin (Deepfakes), en iyi mevcut veri kümelerinden algoritmik olarak yeni veri türleri oluşturma bir yolu şeklinde betimlenen ve Generative Adversarial Network (GAN) adı verilen bir ağda birlikte çalışan iki yapay zekâ algoritmasının ürünüdür. Örneğin, bu teknoloji binlerce resmi veya videoyu tahlil ederek elde edilen verilere benzeyen ama herhangi birinin tam benzeri olmayan yeni bir video ya da resim oluşturabilmektedir. Başlangıçta ses ve görüntü içeriğine dair düzenlemeler için kullanılan derin sahtekârlıklar ile metin, hareketli görüntüler vb. çeşitli içerik türleri üretilmektedir (Walorska, 2020).

Geçmişte çeşitli uygulamalarla basitçe yapılan görüntü ve metin üzerindeki yönlendirme işlemi, derin sahtekârlıkların ortaya çıkışıyla birlikte sahte ses ve videoların meydana getirilmesini daha kolay ulaşılabilir hale getirmektedir. Bu sistemle istenilen neticeyi elde

etmek için sadece birkaç yüz ses kaydı, resim veya video yeterli olabilmektedir. Bundan dolayı derin sahteliklerin medya organlarında yer bulması toplum yapısını olumsuz yönde etkileyebilmektedir. Çünkü bu teknoloji ile sahte bilginin yayılması kolaylaşabilmekte geleneksel basın kuruluşları aracılığıyla aktarılan doğru bilginin güvenilirliği tartışmalı hale gelebilmekte bu da aktarılan bilgiler üzerinde şüphe oluşturabilmektedir. Bilgi çarpıtma bağlamında insanları yanıltmak ya da doğrulardan uzaklaştırmak isteyenler, gerçek bilgiyi sahte gibi göstermek ya da sahte içerikleri hakikatmiş gibi yansıtmak suretiyle gerçeği bu iki etkili yolla manipüle edebilmektedir.

## 1.2. Bilgi Çarpıtmada Yapay Zekâ Tekniklerinin Etkisi

Bilginin çarpıtılmasının önceden belirlenmiş çok sayıda kişiye ulaşmasını sağlamada, mikro hedefleme özellikle etkili olmaktadır. Web’de yer alan reklamcılığa dayalı ekonomi modelinden dolayı bilgi ve iletişim uygulamaları tarafından verimliliği sağlamak amacıyla reklamların belirli tüketicilere yönelik olmasını sağlayan yapay zekâ teknolojilerinden yararlanan veri takip yöntemleri geliştirilmiştir. Konvansiyonel tarayıcı çerezi tabanlı takip ya da üçüncü taraf izleme, işletmeler tarafından tüketicilerin direkt olarak ziyarette bulunduğu sitelere web sitesi analiz komut klasörleri, sosyal medya widget’ları, reklam ağları vb. içeriklerin konumlandırıldığı bir uygulamadır. Bu yerleştirme ilişkisi, hedef kitlenin başka web sitelerinde üçüncü taraflarca tekrar tanımlanmasına ve bir tarama arşivi oluşturmaya imkân tanımaktadır. Bir başka takip aracı olan tarayıcı parmak izi ise tüketicinin farklı web siteleri arasındaki gezinmelerine yönelik bağlantıyı sağlamak için hedef kitlenin kullandığı tarayıcının parmak izini edinmek amacıyla web tarayıcısına ve işletim sistemine özgü niteliklere sahip izleme şeklidir (Akers vd., 2018; Habes vd., 2021; Jagtap vd., 2015; Wang & Alshboul, 2015). Daha önce de ifade edildiği gibi her iki yöntem de muhtemel tüketicilere hitap eden reklamları tespit etmek amacıyla geliştirilmiştir. Bu bağlamda reklamverenler, ilgi alanları, demografik bilgiler vb. hedefleme değişkenlerini saptamakta, platformun algoritmik sistemleri ise reklamın muhataplarını seçmekte ve platformlarda reklamın nasıl konumlandırılacağını belirlemektedir (Maréchal & Biddle, 2020, s. 13).

Günümüzde takip yöntemleri, tüketici konumundaki herkesi belirli bir muhtevayla örtüşürmek amacıyla yoğunluklu olarak kullanılmaktadır. Büyük çevrim içi platformlar ve web siteleri tarafından benimsenen ve dikkat ekonomisine dayanan ekonomik modele göre, doğrudan ziyaretçilerden ziyade reklam şirketlerinden alınan ücretler yatırımların finansmanını oluşturmaktadır (Festré & Garrouste, 2015). Böylelikle platformun ekonomik kazancı kullanıcının burada geçirdiği zamanla orantılı bir şekilde artmaktadır. Bu süreçte algoritmalar, platformun kullanılmasıyla geçirilen vakti en uygun hale getirmek ve satın alınma ihtimalinin güçlü olduğu ürünleri önermek için çevrim içi olarak yapılan her davranışı kayıt altına almaktadır. Bu durumla ilgili Avrupa Birliği Konseyi tarafından alınan kararlarda belirtildiği üzere bağımsız ve çoğulcu bir medya yapısının devamı adına, algoritmaların tüketicilerin edilgen durumdayken karşılaştıkları medya içeriklerini ve yine tüketicilerin etkin olarak aradıkları sonuçları etkilediği hususu önem arz etmektedir (European Parliament, 2020).

Sanal alemde çevrim içi olan herkes, muhatap olunan veriler üzerinde etkisi her geçen gün artan veri analizi ve algoritmalarından dolayı hakikatin farklı yorumlarıyla karşılaşmaktadır. Youtube’un ve Sportify’nin tavsiye sistemi, Netflix’in film tavsiye akışı, Twitter’in zaman tüneli, Facebook’un haber akışı, Google’ın sıralama sistemi, Instagram’ın hikâye akışı kişisel kullanıcıların çevrim içi ortamda ne ile karşılaşacağını belirleyen içerik düzenleme algoritmalarına örnek olabilmektedir (Maréchal & Biddl, 2020, s. 13).

Bu ekosistem, takip etme ve hedefleme teknolojileriyle birlikte ortaya çıkan gizliliğe dair endişeler, bilgi edinme ve kişilerin bağımsızlığıyla alakalı kaygıları artırmaktadır. Bunun yanı sıra bilgi çarpıtma uygulamaları tarafından, teknoloji konusunda yeterli donanımına sahip olmayan kullanıcıların hedefe konulması, filtreleme araçları oluşturularak bunlardan faydalanılması mümkün olabilmektedir (Akers vd., 2018, s. 5). Örneğin, Facebook'ta kullanılan ilgi alanları ile alakalı kategoriden ücreti karşılığında yararlanan reklamverenler (Sankin, 2020) yapay zekâ sistemlerinin davranışsal ve psikografik kullanıcı profillerini tahlil etme özelliği sayesinde seçmenlere yönelik mikro hedeflemeyi artmışlardır (Bergamini, 2020; Marsden & Meyer, 2019).

Dar kapsamlı algoritmaların oluşturduğu tasarımlar, yanlış artırırken bilgi çarpıtmanın yaygınlaşmasını da güçlendirebilmektedir. Mesela YouTube'da hemen her gün 700 milyon saat video izlenmektedir ve bunların yaklaşık yüzde 70'inde kullanıcılara sonraki süreçlerde hangi görüntüleri izlemeleri hususunda önerilerde bulunan otomatik sistemlerin etkisi bulunmaktadır. İki milyarın üzerinde kullanıcının yer aldığı, kâr amacını öncelediği ve video tercihlerinin izlenme sürelerine göre tekrar tekrar düzenlenmesinden dolayı içeriklerin kalitesinin ikinci planda olduğu YouTube'un algoritmasını ortaya koymayı amaçlayan AlgoTransparency projesi, bu platformun verimli olmayan geri dönüt rotasyonunu ortaya koymaktadır. Bu rotasyona göre ilk olarak zararlı içerik daha iyi performans göstermekte ikinci aşamada işlemleyici sistemler görüntülenme zamanını yeniden düzenlemek için içeriği tavsiye etmekte, son aşamada da bu tarz içeriklerin daha yoğun bir şekilde izlenmesinden dolayı içerikleri hazırlayanların oluşturduğu yeni içeriklerin oranı daha da artmaktadır (Meyerson, 2012).

Bilgi çarpıtmanın etkin bir biçimde yayılmasında art niyetli kişiler tarafından kullanılan sosyal botların (yazılım robotlarının kısaltması) da katkısı bulunmaktadır. Bu botlar, sosyal medya uygulamalarında faaliyet gösteren, iletişim ve birey davranışlarının taklidi amacıyla oluşturulmuş yarı otomatik veya tam kullanıcı hesaplarıdır. İletişim sürecinde beğenme, yorumlama, paylaşma vb. herhangi bir unsuru otomatik hale getiren karmaşık olmayan botlar, otomatik unsurlara sahip kısmen insan tarafından yönlendirilen hesaplar, Microsoft'un Zo1 veya Replika.ai gibi öğrenme becerileriyle donatılmış otonom olarak hareket eden ajanlar vb. çeşitli sosyal bot türleri bulunmaktadır (Assenmacher vd., 2020). Sosyal medyada varlığını sürdüren bu yapay konuşmacıların, piyasaları yönlendirme, politik gerginlikleri artırma, çevrim içi ifadeleri saptırma vb. gerçekçi etkileri bulunmaktadır (Lamo & Calo, 2019; Varol vd., 2017; Zago & Nespoli, 2020). Bu tarz botlar, içerisine yerleşmeye çalıştıkları grup ve topluluğa göre bağlamla alakalı içerikleri yayınlayacak biçimde tasarlanmakta, bir kez güvenilir bir profil kazanıp itimat edilebilir olarak değerlendirildiklerinde ise çarpıtılmış bilgiyi etkin bir biçimde yayabilmektedir (Akers vd., 2018, s. 5).

### **1.3. Bilgi Çarpıtma Sürecinde Yapay Zekâ Tekniklerinin Rolü Üzerine Etik Değerlendirmeler**

Dijital ekosistemde kullanılan yapay zekâ tekniklerinin insanları geniş ölçekte ve etkin bir şekilde yönlendirmek için yeni fırsatlar oluşturduğu düşünüldüğünde, uzun bir dönemdir etkisi görülen bir problem olan bilgi çarpıtmayla alakalı birden fazla etik kaygı kendini göstermekte bu da durumu daha da kötü hale getirmektedir. Halihazırda var olan dijital sistemin olumsuz yönde etkilediği ilk etik olgu, insan onurudur. Bu temel değere göre, bireyler kendi içerisinde bir amaç olarak kabul edilmeli ve kesinlikle bir araç olarak değerlendirilmemelidir (EDPS Ethics Advisory Group, 2018). Buna karşın algoritmalar, içerikteki niteliği dikkate almaksızın ve etkileşimi yeniden uyarlamak için bireylerin dikkatine sunulanları verilerin birleştirilmesiyle meydana getirilen profillere dayalı



olarak düzenleyecek şekilde programlandığında, bu kişiler iktisadi hedefler için sadece birer araç olarak kabul edilmektedir. Bunun yanı sıra bu algoritmalar art niyetli kişilerce görüşlerin etkili biçimde manipülasyona uğratılması ve böylelikle istenilen hedefe varılması için kullanılabilir. Ayrıca dijital sistemde yer alan yapay zekâ teknikleri, çoğunlukla kullanıcılardan habersiz bir şekilde gerçeği ters yüz etmekte ve bu kişilerin kanaatlerinin başarılı bir biçimde yeniden oluşturulması adına fırsatlar oluşturmaktadır. Bu süreçte yapay zekânın hedefinde olan kişiler, dijital ekosistemin ender olarak farkındadır, genellikle çevrim içi gördükleri bilgilerin nesnel olduğunu ve başka kullanıcıların da evrensel olarak bunlarla karşılaştıkları yanılsamasına düşmektedirler (EDPS Ethics Advisory Group, 2018).

İkinci olarak, çevrim içi bilgi çarpıtmanın yaygın bir şekilde görünmesiyle birlikte bilgiye ulaşmanın güçleşmesi, insanların bağımsız olabilmeleri için öncelikli bir ön şart olan özgür ve akıllıca kararlar alma yetilerini azaltmaktadır. Algoritmaların bireylere gösterilecek içeriği tespit etmek için kişisel bilgilerden faydalanması ve özgür bir şekilde karar alma imkânlarına zarar vermesi, verilerin korunması hakkını ve gizliliğini olumsuz etkileyebilmektedir.

Bireylere toplumsal problem ve hakikatlerle alakalı gelişmelerden haberdar olmalarını ve bunları anlamalarını kolaylaştırarak kararlarını bilinçli bir şekilde almaları imkânı sunan, bundan dolayı da özerklik konusunda hayati öneme sahip bilginin sahte içeriklerle sunulması (Hanot & Michel, 2020), bireylerin farkında olarak ya da olmayarak birçok filtreye uğramalarına, bireylerin yalnızca kendileriyle benzerlik gösteren görüş ya da inançla karşılaştığı ortam olan yankı odalarına sıkıştırılmalarına ve dijital ekosistemden yararlanan bilgi çarpıtmaların hedefi olmalarına yol açmakta, bu da bilgiye erişimi zorlaştırmaktadır. Çevrim içi ortamda bilginin çarpıtılmasının yayılmasının da etkisiyle haberlere duyulan güven seviyesinin azalması bireylerin karşı karşıya kaldığı bilgi eksikliğine katkıda bulunmaktadır (Newman vd., 2023).

Böylelikle bireyler diyaloga yer olmayan bir izolasyona maruz kalabilmekte kullanıcıların onayladığı haber ve bilgilere öncelik veren algoritmalar, görüşleri, alışkanlıkları, zevkleri vb. pekiştirmekte ve diğer kanaatlere ulaşmayı sınırlandırabilmektedir (Bergamini, 2020, s. 10).

Bireyleri filtre baloncukları ve yankı odalarına hapsedmek gerçek bilgi için tehlikeli olabilmektedir. Bireyler, bazen sosyal botlar tarafından şişirilen kendi görüşlerine sahip insan gruplarıyla giderek daha fazla etkileşime girdikçe ve gerçekliğin önyargılı bir versiyonuyla manipüle edildikçe, diğer kültürlerin varlığını kabul etme ve onları anlamada zorlanabilmektedir. Bu durum radikalleşmeyi, dolayısıyla barış ve istikrar için gerekli olan birbirlerine karşı hoşgörü eksikliğini beslemektedir (Perrigo, 2019; Wright, 1997).

## 2. Yapay Zekâ Teknikleri Aracılığıyla Çevrim içi Bilgi Çarpıtma ile Mücadele

Çeşitli teknolojik yöntemlerin geliştirildiği günümüzde çevrim içi bilgi çarpıtmanın kaygı uyandırıcı neticelerine tepki olarak, arama motorları ve sosyal medya aygıtlarından yanlış ve yanıltıcı bilgilerin yayılmasına daha fazla yanıt vermeleri beklenmektedir. Bu bağlamda yanlış, yanıltıcı ya da hatalı içerikleri saptamak ve bu tür içerikleri çevrim içi olarak hazırlamak amacıyla yapay zekâ teknikleri üzerine yapılan araştırmalar yoğunlaştırılmaktadır. Yapay zekâ sistemlerinin, teknolojide her geçen gün artan ilerlemeler ve dijitalleşmeyle beraber gücü her geçen gün artan çarpıtılmış bilgiden

doğru olmayan bilgiyi ayırt etme hususundaki kısıtlılığı ifade ve bilgi özgürlüğü açısından sorun oluşturmaktadır. Öte taraftan vatandaşların doğru ve güvenilir bilgilere erişmesinde önemli bir sorun olarak değerlendirilen bilgi çarpıtmayı önlemek adına hayata geçirilen önlemlerin kişilerin bilgi ve ifade hürriyetini sınırlayıcı bir hale dönüşmesi ise bilgi çarpıtmayla benzer sorunlara neden olabilmektedir.

## 2.1. Yapay Zekâ Tekniklerinin Çevrim içi Bilgi Çarpıtmanın Saptanması İçin Geliştirilmesi

Teknoloji destekli bilgi çarpıtmaların belirlenmesi ve yanlış bilgi içermesine karşın derin sahte olmayan metinlerin tespit edilmesi için dört teknikten yararlanılmaktadır. İlk olarak, makine öğrenimi ile etiketli verileri, yani yanlış ve doğru bilgi içeren metinleri kullanarak uçtan uca modelleri eğitmek mümkündür. Böylelikle sistem daha sonra bu iki metin türü arasında doğrudan ayırım yapabilmektedir. Fakat, böyle bir modelin başarılı bir şekilde çalışması için gerekli olan ve elde edilmesi zor olabilen önemli oranda etiketli verinin yanı sıra, modelin ortaya koyduğu üretim açıklanabilirlikten yoksun olmakta ve veri kümelerindeki önyargılardan etkilenebilmektedir. İkinci aşamada, olgusal yanlışlıkların belirlenmesi daha etkili olabilmektedir. Bunu, metnin muhtevasını harici argümanlarla karşılaştırma yaparak gerçekleştirmek olası olmakla birlikte doğal dilin ayrımlarını biçimlendirmek kolay olmadığı için bu iş insan editörler tarafından daha iyi yapılmaktadır. Üçüncü olarak, yanıltıcı üslubun analizinin yapılması sonucu bir metnin amacını ortaya çıkarmak, eğitilmiş insan uzmanlar tarafından ya da makine öğrenimi yoluyla gerçekleştirilebilmektedir. Ancak aldatıcı üslubun varlığı ya da yokluğu her zaman sağlanan bilginin doğruluğu veya yanlışlığıyla alakalı olmamaktadır. Dördüncü olarak da metnin içeriği yerine kullanıcının nitelikleri, profili vb. meta verilerin analizi de incelenmektedir (Akers vd., 2018; Murphy & Flynn, 2022; Yu vd., 2021; Zhang, 2022).

Marsden & Meyer'e (2019) göre kullanılan ya da kullanılmaya hazır bilgi çarpıtma unsurlarını saptamak amacıyla oluşturulan metin analizi programları, metinden ortaya çıkan çoklu, kompleks ve tutarsız anlamları ayırtmanın güçlüğünden dolayı yanlış yönelimli olabilmektedir (s. 1-2) pluralism and the functioning of a democratic polity. The study examines the trade-offs in using automated technology to limit the spread of disinformation online. It presents options (from self-regulatory to legislative. Bundan dolayı, doğru olmayan, yanıltıcı ve hatalı verilerin yer aldığı metinlerin doğru bir biçimde saptanması için hâlâ insanların müdahale etmesi gerekmektedir.

Derin sahtelikleri belirlemek amacıyla birçok teknik öngörülmüştür. Bunlardan birincisi, gerçek içerik ile sahte içeriği ayırt edebilecek teknolojiler geliştirmektir. Bundan dolayı derin sahtelikleri oluşturan algoritmalarla benzer algoritmaların kullanılması mümkündür. Araştırmacılar, bazı eğitim veri setlerinde gözlerin kapalı olmaması nedeniyle yapay zekâ tarafından meydana getirilen görüntülerdeki aktör ve aktrislerin gözlerini kırpmadığını gözlemlemiştir. Buna çözüm üretmek amacıyla anormal göz kapağı hareketi kalıplarını arayan teknolojiler geliştirilmiştir. Ancak, bu yöntem yaygınlaştığı zaman yeni derin sahte videolar, göz kırpan insanları içerecek şekilde uyarlanmıştır (Kertysova, 2019; Walorska, 2020). Derin sahtelikleri saptamak amacıyla öngörülen ikinci teknik, oluşturulan içeriğin yayılmadan önce doğrulanmasıdır. Ses, video ve görüntülerin oluşturuldukları zaman dijital olarak etiketlenmesi halinde bu etiket doğru olmayan sahte içerikle karşılaştırmak için bir referans olarak kullanılabilir (Kertysova, 2019, s. 71). Ancak, derin sahtelikler, daha önce var olan herhangi bir etiketlenmemiş içerikle meydana getirilebilmekte bu durum da yaklaşımı verimsizleştirebilmektedir.

Derin sahtelikleri belirlemek için etkili araçların eksik olması ve bunların çevrim içi bilginin doğruluğu ve geçerliliği üzerindeki olası etkisinden dolayı 2016'da yapay zekâ üzerine araştırmalar yapan Microsoft, DeepMind ve Google, Apple, Facebook, IBM ve Amazon tarafından kurulan Yapay Zekâ Ortaklığı (PAI), yönlendirilmiş ve yapay içeriğin saptanması da dâhil olmak üzere yanlış/çarpıtılmış bilgi çözümlerini güçlendiren projeler geliştirmek ve öneriler sunmak amacıyla Yapay Zekâ ve Medya Bütünlüğü Yönlendirme Komitesi'ni hayata geçirmiştir. Komite tarafından yapılan Deepfake Tespit Mücadelesi ile yapay zekâ tarafından oluşturulan ve manipüle edilen videoları tespit eden teknik çözümlerin geliştirilmesinin teşvik edilmesi (Partnership on AI, 2020).

Sosyal botları belirlemek amacıyla faaliyetler yürüten Komite tarafından hazırlanan ve teknolojiye beslenen bilgi çarpıtmanın ele alındığı raporda, buluş yaklaşımıyla genellikle insan kullanıcılar ile bot hesaplar arasındaki ayrımları tespit etmek için makine öğreniminden yararlanıldığı ortaya konulmaktadır. Buna göre sosyal ağ yapısı, gönderi metinleri ve hesap verilerinden yararlanmanın yanı sıra doğal dil işleme teknikleriyle profillerdeki metin içerikleri analiz edilmektedir. Sosyal medya platform sağlayıcıları için gönderileri görüntülemek için harcanan zaman, yapılan arkadaşlık isteklerinin sayısı vb. hesap davranışlarını izleyebilmek önemli bir kazanım sağlamaktadır (Akers vd., 2018, s. 5). Yapay zekâ teknikleri ile bot hesaplar da dâhil olmak üzere gerçek olmayan hesapların tespitinde başarılı olunduğu gözlemlenmektedir. Örneğin 2020'de Facebook'un harekete geçirdiği sahte hesapların yüzde 99,60'ı kullanıcıların henüz haberi olmadan şirket tarafından bulunmuş ve işaretlenmiştir (Toplum Standartları Uygulama Raporu, 2024).

## 2.2. Geliştirilen Yapay Zekâ Teknikleri ile Çevrim İçi İçeriğin Düzenlenmesi

Çevrim içi içerik ve hesaplara yönelik düzenlemelerde yapay zekâ tekniklerinden yararlanılmaktadır. Bu teknikler arasında yer alan içeriğin filtrelenmesi, paylaşım yapılması için içeriğin yüklenmesini ya da yayınlanmasını engellemek amacıyla teknik tedarıkçiler tarafından alınan bir önlemdir. İçeriğin kaldırılması ise istek, farkındalık ya da emir üzerine yapılmaktadır. Bilginin çarpıtılmasına karşı mücadele etmede içeriğe filtre uygulanması veya kaldırılması yöntem olarak etkili olmakla beraber içerik paylaşımını engellediği için aynı zamanda en yaygın yöntemler arasında yer almaktadır (Nuñez, 2020, s. 785-786). İçeriğin engellenmesi ile kullanıcıların içeriklere ulaşması güçleştirilebilmekte, buna karşın sanal özel ağ (VPN) ile bu yöntem aşılabilmektedir. İçeriğin önceliklendirilmesi, belirli kişi, grup, topluluk ya da konularla alakalı daha az içerik görmeyi tercih eden kullanıcılar tarafından ya da ağ tabanlı çözümler ve algoritmalar yardımıyla otomatik sıralama kullanılarak yapılabilmektedir. Hesapların devre dışı bırakılması ve askıya alınması, çevrim içi bilgi çarpıtmayla mücadele etmek için kullanılabilecek bir başka yöntemdir (Brkan, 2019, s. 6). Teknoloji hizmet tedarık edicileri, kullanıcıların kanunlara uymadığı ya da hizmet şartlarını kötü amaçlı kullandıklarında bu önlemleri almaktadır. Örneğin, Facebook ve Twitter Amerika Birleşik Devletleri eski Başkanı Donald Trump'ın hesabını platformların hüküm ve şartlarına uymadığından dolayı askıya almıştır (Marsden & Meyer, 2019, s. 38, 41) pluralism and the functioning of a democratic polity. The study examines the trade-offs in using automated technology to limit the spread of disinformation online. It presents options (from self-regulatory to legislative.

Bilgi çarpıtmanın önlenmesinde merkezi olmayan web (Dweb) teknolojileri gibi farklı yenilikçi teknolojiler de geliştirilmektedir. Dweb teknolojileri, kullanıcılar yerine internet



şirketleri tarafından merkezi olarak tutulan muazzam veri tabanlarının çözümlenmesini sağlamaktadır. Bu da kullanıcıları özel ve devlet gözetiminden korumaktadır (Harbinja & Karagiannopoulos, 2019).

### 2.3. Yapay Zekâ Teknikleri Aracılığıyla Çevrim İçi Bilgi Çarpıtma ile İlgili Etik Çıkarımlar

Dijital alandaki çarpıtılmış bilgiye karşıt olarak yapay zekâ tekniklerinin geliştirilmesindeki amaç meşru olmakla birlikte bunların kullanımı da birçok etik kaygıyı barındırmaktadır. Bu tarz tekniklerden dijital ekosistemde yararlanmadan önce bunları değerlendirmeye tabi tutmak önem taşımaktadır.

Bilgi çarpıtma öğelerinin saptanmasında yapay zekâ sistemlerinin kullanılması, bu olgunun kapsadığı unsurlar üzerinde anlaşmaya varılmış resmi bir tanım gerektirmektedir. Buna karşın problemi betimlemek her zaman kolay olmamaktadır (Hanot & Michel, 2020). Bu da toplumlarda hangi hususların yasal ya da yasa dışı, istenilen veya arzu edilmeyen olduğunun belirlenmesinde karar verici merci ile alakalı soruları gündeme getirmektedir. Kandırmaca, propaganda ve karikatürize etmeyi çarpıtılmış bilgiden ayırmak sorunlu olabilmekte ve bazı tanımlar bu tür içerikleri bilgi çarpıtma sorununa dâhil etmektedir. Ayrıca problemin muhtevası da tartışmalıdır. Doğru bilgiye ulaşmada yaşanabilecek olası yetersizlik, bilgi çarpıtma meselesinin ötesine geçmektedir. Herhangi bir içeriği paylaşan birçok kişi veri kaynağının yayıncısının başlangıçtaki maksadından habersiz olabilmekte doğru olmayan bilgilendirme de toplumun tüm kesimlerini etkileyebilmektedir. Ancak yanlış bilgiyle mücadele etmek, daha önce de belirtildiği gibi ifade ve bilgi edinme özgürlüğünü önemli ölçüde sınırlamaktadır.

Bununla beraber, paylaşımcının niyetinin belirlenmesinin gerekliliğinden dolayı bilgi çarpıtmadan doğru olmayan bilgiyi ayırt etmek söz konusu olduğunda güçlükler ortaya çıkmaktadır. Çevrim içi bilgi çarpıtmayla mücadele amacıyla geliştirilen yapay zekâ teknikleri, paylaşımı yapan kişinin maksadına yönelik herhangi bir ayırım yapmaksızın tüm yanıltıcı, hatalı ve yanlış bilgileri tespit etmektedir. Bundan dolayı, yapay zekâ sistemlerinin içeriği otomatik bir şekilde düzenlemesine izin vermek, ifade ve bilgi edinme özgürlüğünü ciddi şekilde etkilemektedir. İnsanların sürece dâhil olması da yapay zekâ sistemlerinin verdiği bilgilere duyulan güvenden dolayı bu durumu değiştirmemektedir. Bunun için yapay zekâ sistemlerinin paylaşım yapan kişinin zararlı amacını değerlendirmesini sağlama ihtimalini sorgulamak önem arz etmektedir. Çünkü içeriği meydana getiren ya da bu verileri kamuya açık hale getirenlerin niyeti belirsiz olabileceğinden dolayı makine tarafından yapılan bir paylaşımın içeriğinin arka planındaki düşünceyi, hassas moderasyon eyleminin gerektirdiği kesinlik derecesiyle değerlendirmesini sağlamak zor olmaktadır.

Yanlış veya aldatıcı bilginin ifade edilmesi olumsuz bir durum olmakla birlikte insanların manipüle edilmesi kastıyla bilinçli olarak yanıltıcı ya da hatalı bilgilerin yayılmasında teknolojinin kötü amaçlı kullanımı daha önemli bir sorun olarak değerlendirilmektedir. Bu nedenle yapay zekâ teknikleri, yapay zekâ tarafından üretilen öneri, içerik ve sosyal botların kullanımı vb. zararlı kullanımların saptanmasında kullanılabilir.

Bununla birlikte Marsden & Meyer'e (2019) göre dijital platformlarda daha önceden hangi şart ve sıklıkta filtreleme ya da mâni olma işleminin yapıldığı net olarak belirlenmemektedir. Bunlardan bazıları yapay zekâ güdümlüdür fakat içeriğin illegalliğinin ya da toplumda yer alan kurallara ihtilafının nasıl tespit edildiği ya da yazı, görüntü, fotoğraf vb. içeriklerin sıkı denetimindeki aşırılığın engellenmesi amacıyla ne tür

tedbirlerin alındığı belirgin değildir (s. 45)pluralism and the functioning of a democratic polity.The study examines the trade-offs in using automated technology to limit the spread of disinformation online. It presents options (from self-regulatory to legislative. Buna karşın içeriğin hazırlanmasındaki saydamlık, ifade hürriyeti bakımından hukukiliğini değerlendirmek için birincil önemdedir. Bu nedenle, herhangi bir içerik ayarlamasının açıklanabilir olması gerekli olmakla birlikte yapay zekâ sistemlerinin çalışma biçimlerindeki kompleks yapının yetersiz veri çıktılarına neden olma durumundan dolayı bu veriler sorunlu olabilmektedir.

Birleşmiş Milletler Eğitim, Bilim ve Kültür Kurumu'nun (United Nations Educational, Scientific and Cultural Organization-UNESCO) Herkes İçin Bilgi Programı'na göre önceki yıllarda görüşlerin aktarılabilmesine ve verilerin paylaşılabilesine imkân sunan teknolojiler aynı zamanda hali hazırdaki verileri engellemek ve farklı düşünceleri ifade edenleri saptamak ve müdahalede bulunmak için de kullanılabilir. Oysaki düşünce ve ifade özgürlüğü hakkı, bireylerin düşüncelerini başka kişilere aktarma olanağı bulunmadığında değerini kaybetmektedir. Bilgi ve iletişim teknolojileri bu etkileşimin yapılabileceği kamusal bir alan oluşturmak amacıyla kullanılmakta veya insanların muhataplarıyla iletişim sağlama yetisini kısıtlayarak ifadeyi sınırlandırabilmektedir (Rundle & Conley, 2007, s. 15-16).

Ayrıca yanlış ve temelsiz verileri saptamak amacıyla programlanan yapay zekâ sistemleri yanlış negatif ve yanlış pozitif eğilimli olabilmektedir. Yanlış pozitifler yani hatalı, aldatıcı ve yanlış içeriğin tespiti yanlış olması ifade özgürlüğünü etkilemektedir. Bu durum makine tarafından yanlış bir şekilde çarpıtılmış bilgi olarak kayıt altına alınan yasal bir içeriğin sansüre uğramasına neden olabilmektedir (Marsden & Meyer, 2019, s. 17) pluralism and the functioning of a democratic polity.The study examines the trade-offs in using automated technology to limit the spread of disinformation online. It presents options (from self-regulatory to legislative. Buna karşın yanlış negatif ve pozitif eğilimler ayrımcılık oluşturabilmekte bu da eşit olma halini etkileyebilmektedir. Bu durum algoritmalarındaki önyargılardan kaynaklanabilmektedir.

İçeriğin otomatik bir şekilde düzenlenmesinde yapay zekâ tekniklerinden yararlanılması halinde bazı sorunlar oluşabilmektedir. İçeriğin sınırlandırılmasını yönelik algoritmaların kullanımı, yetkili kaynaklara öncelik verilmesini öne çıkardığında bu durum medya çoğulculuğu bakımından problemlere neden olabilmektedir. Böylesi bir usul yeni öğelerin var olmasını güçleştirmektedir. Bu kaygılar, muhtevanın güvenilirliğini ve kalitesini betimlemekten kaçınarak bilgi ve ifade hürriyetini gözetmenin ehemmiyetine ilişkin endişelerle bağdaşmaktadır.

Çevrim içi bilgi çarpıtmayla mücadele amacıyla geliştirilen yapay zekâ sistemlerine yönelik etik değerlendirmelerin dikkatle ele alınması önem arz etmekle birlikte moderatörlük yapan insanların çok önemli ve zor bir işi yürüttükleri unutulmamalıdır. Buna rağmen insanların yaptığı kontroller de yanlış ya da belli olmayan neticelere açık olabilmekte ve peşin hükümlü olmak da insan belleğini etkilemektedir.

Bunun yanı sıra bilgi çarpıtma sorunu çevrim içi arama motor ve platformlarının ilerisine geçmektedir. Kullanıcı sayısının çok olduğu kapalı grup ve topluluklar içerisinde bilgi paylaşımına ve kapalı bir gruptan diğerine bilgi transferine imkân sağlayan anlık mesajlaşma ortamları da bilgi çarpıtmanın yayılmasında etkili olabilmektedir (ERGA, 2020, s. 44). Öte taraftan özel iletişimdeki bilgi çarpıtmaya karşı koymak özel hayatın gizliliği hususundaki ciddi endişeleri de beraberinde getirmektedir.

## Sonuç

Bilginin çarpıtılması, verilerin doğrulanması sürecinde uzun bir zamandır varlığını devam ettiren bir problemdir. Web’de bulunan dijital sistemdeki yapay zekâ teknolojileri de bu sıkıntının derinleşmesine iki şekilde etki etmektedir. Birincisi, insanları önemli oranda ve güçlü bir biçimde yönlendirmek amacıyla art niyetli kişiler tarafından kullanılabilir. İkincisi ise bu tarz içeriklerin yaygınlaşmasını direkt olarak arttırmaktadır. Yapay zekâ sistemleri etkileşimi daha da güçlendirmek için programlanmıştır, bundan dolayı hatalı bilgilendirmenin yaygınlaşmasına etki eden esas unsur web’deki iş modelidir. Bunun yanı sıra, yaygın olan sosyal medya botları aracılığıyla da doğru olmayan içerikler her geçen gün daha gerçekçi olabilmektedir.

Bilgi çarpıtmanın toplumlar nezdindeki tüm negatif yansımalarına tepki olarak, çarpıtılmış bilginin çevrim içi yayılmasının önlenmesi adına arama motorları ve sosyal medya uygulamaları daha fazla tedbir almaktadır. Bunun için problemin çözümlenmesi amacıyla bazı yapay zekâ sistemleri geliştirilmiştir. Bununla beraber, yapay zekâ sistemlerinin yaygınlaşmasına katkı sağladığı çarpıtılmış bilgilerdeki içeriklere karşı koyabilmek için diğer yapay zekâ sistemlerinden yararlanılmasının beklenen etkiyi göstermeme ihtimali yeni kaygılara yol açmaktadır (Hildebrandt, 2018). Manipülasyon araçları, global ölçekte kabul gören etik değerlere zarar verdiği için önlenmelidir. Bunu yaparken de etik değerler ile ifade ve haber alma özgürlüğü gibi temel haklara yönelik tehlikelere önem verilmesi gerekmektedir.

İfade özgürlüğü ile çevrim içi içeriğin kontrol edilmesi arasında var olan problemlerin çözümü aşamasında yapay zekâ sistemlerinden denetleme için yararlanılması, sorunu artırabilmektedir. Bunu önlemek için sistemde var olan yapay zekâ uygulamalarının etik değerler ve temel haklarla uyumlu olması gerekmektedir. Ayrıca, arama motorlarının ve platformların reklam verenlerin ödediği ücretlere yönelik beklentileri devam ettiği müddetçe hedef kitlelerin katılımının hedeflenmesinden dolayı web’in iş şablonunun bu duruma uyarlanması sürecin önemli bir parçası durumundadır.

Haber alma ve ifade hürriyetine değer vererek gerçek bilgiye ulaşma imkânını artırmak için çoğulcu yapısı önem arz eden medya organları ve bu kurumlarda çalışan gazetecilere destek verilmesi gerekmektedir. İnsanlara bilgiyi denetimli bir biçimde sunan kamu bilgi edinme sisteminin kurulması da önerilebilmektedir. Buna karşın herhangi bir biçimde bu durum insanların düşüncelerini aktarma yoğunluğunu ya da medyadaki çoğulcu yapının etkinliğini önlememelidir.

Yapay zekâ teknikleri yardımıyla içerik düzenlemesi yapılması doğru bir bakış açısı olmamakla birlikte bu tarz sistemler; sosyal medya botlarının doğru ve başarılı bir şekilde saptanması, yapay zekâ tarafından oluşturulan verilerin tespit edilmesi, kullanıcıları art niyetli mikro hedefleme kullanımından koruyarak bağlamından koparılmış bilgiye maruz kalmalarının önlenmesi vb. dijital sistemin yönlendirmelerini etkisizleştirmek amacıyla kullanılabilir.

Bu tedbirlerle birlikte kullanıcıların sunulan verileri doğru bir şekilde değerlendirmesi için medya okuryazarlığı önemlidir. Enformasyonun eleştirel yorumlanmasıyla alakalı eğitimin önemi artmakla beraber, bilginin gündelik yaşamın merkezine konumlandığı günümüzde bu eğitim daha da gerekli hale gelmiştir. Bunun yanı sıra yapay zekâ sistemleriyle birlikte tüketici tecrübesini kavramaya ve gelişen teknolojiyle birlikte

saptanması giderek güçleşen bilgi çarpıtmanın betimlenmesini önceleyen çalışmaların da önem kazandığı görülmektedir.

### 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*. April 2019.
- Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., & Grimme, C. (2020). Demystifying Social Bots: On the Intelligence of Automated Social Media Actors. *Social Media and Society*, 6(3). <https://doi.org/10.1177/2056305120939264>
- Bennett, w. L., & Livingston, S. (2020). The Disinformation Age. İçinde *The Disinformation Age*. <https://doi.org/10.1017/9781108914628>
- Bergamini, D. (2020). *Need for democratic governance of artificial intelligence*. Parliamentary Assembly.
- Bindner, L., & Gluck, R. (2020). *Social Media and the Murder of Samuel Paty, Global Network on Extremism and Terrorism*. <https://gnet-research.org/2020/11/06/social-media-and-the-murder-of-samuel-paty/>
- Brkan, M. (2019). Freedom of Expression and Artificial Intelligence: On Personalisation, Disinformation and (Lack Of) Horizontal Effect of the Charter. *SSRN Electronic Journal, February*, 1–17. <https://doi.org/10.2139/ssrn.3354180>
- Chesney, B., & Citron, D. (2019). Deep Fakes : A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107, 1753–1820.
- Dixit, P., & Mac, R. (2018). *How WhatsApp Destroyed a Village*. Buzzfeed News. <https://www.buzzfeednews.com/article/pranavdixit/whatsapp-destroyed-village-lynchings-rainpada-india>
- EDPS Ethics Advisory Group. (2018). *Towards a Digital Ethics - Report*. European Data Protection Supervisor.
- ERGA. (2020). *ERGA Report on disinformation: Assessment of the implementation of the Code of Practice*. 1–73. European Regulators Group for Audiovisual Media Services.
- European Commission. (2018). *2018 Code of Practice on Disinformation*. <https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>
- European Parliament. (2020). European Parliament Resolution on Strengthening Media Freedom: The Protection of Journalists in Europe, Hate Speech, Disinformation and the Role of Platforms 2020/2009(INI). [https://www.europarl.europa.eu/doceo/document/TA-9-2020-0320\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-9-2020-0320_EN.html)
- Festré, A., & Garrouste, P. (2015). The 'Economics of Attention': A History of Economic Thought Perspective. *OEconomia*, 5–1, 3–36. <https://doi.org/10.4000/oeconomia.1139>
- Goodfellowet, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Ben gio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 1–9.

- Greengard, S. (2019). Will deepfakes do deep damage? *Communications of the ACM*, 63(1), 17–19. <https://doi.org/10.1145/3371409>
- Habes, M., Ali, S., Elareshi, M., Tahat, K. M., & Ziani, A. (2021). Understanding Users' Social TV Content using Artificial Intelligence Approach: A Survey. *International Conference of Modern Trends in ICT Industry: Towards the Excellence in the ICT Industries, MTICTI 2021*, 2, 1–7. <https://doi.org/10.1109/MTICTI53925.2021.9664762>
- Hanot, M. & Michel, A. (2020). Entre menaces pour la vie en société et risques réglementaires, les fake news: un danger pour la démocratie? İçinde *Vie privée, liberté d'expression et démocratie dans la société du numérique* (Yves Poull, Sayı 47, ss. 155–208).
- Harbinja, E., & Karagiannopoulos, V. (2019). *Web 3.0: The Decentralised Web Promises to Make the Internet Free Again. The Conversation*. <https://theconversation.com/web-3-0-the-decentralised-web-promises-to-make-the-internet-free-again-113139>
- Hildebrandt, M. (2018). Primitives of Legal Protection in the Era of Data-Driven Platforms. *SSRN Electronic Journal*, 2(2), 1–21. <https://doi.org/10.2139/ssrn.3140594>
- Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2798311>
- Jagtap, N. P., Patil, K. A., Shakil, S. S., & Ingle, N. S. (2015). Mobile Activity Monitoring System Using Android Spy. *Ijarcce*, 4(2), 158–162. <https://doi.org/10.17148/ijarcce.2015.4235>
- Kaye, D. (2018). *A /73/348\*. October*.
- Keller, I., & Público, D. (2021). *Don't Shoot the Message : Regulating Disinformation Beyond Content Assunto Especial Dossiê – Democracia , Fake News e Eleições Don ' t Shoot the Message : Regulating Disinformation Beyond Content*.
- Kertysova, K. (2019). Artificial Intelligence and Disinformation. *Security and Human Rights*, 29(1–4), 55–81. <https://doi.org/10.1163/18750230-02901005>
- Kolarević, E. (2022). The influence of Artificial intelligence on the right to freedom of expression. *Pravo - teorija i praksa*, 39(1), 111–126. <https://doi.org/10.5937/ptp2201111k>
- Lamo, M., & Calo, R. (2019). Regulating bot speech. İçinde *UCLA Law Review* (C. 66, Sayı 4). <https://doi.org/10.2139/ssrn.3214572>
- Maréchal, N., & Biddl, E. R. (2020). It's Not Just the Content, It's the Business Model: Democracy's Çevrim içi Speech Challenge. *New America*, 1–53.
- Marsden, C., & Meyer, T. (2019). Regulating disinformation with artificial intelligence: effects of disinformation initiatives on freedom of expression and media pluralism. İçinde *Panel for the Future of Science and Technology European Science-Media Hub* (Sayı March).
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.
- Meyerson, E. (2012). *YouTube Now: Neden İzlenme Süresine Odaklanıyoruz*. <https://blog.youtube/news-and-events/youtube-now-why-we-focus-on-watch-time/>



- Murphy, G., & Flynn, E. (2022). Deepfake false memories. *Memory*, 30(4), 480–492. <https://doi.org/10.1080/09658211.2021.1919715>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2023). Reuters Institute Digital News Report. *Reuters Institute for the study of Journalism*, 1–160.
- Núñez, F. (2020). Disinformation Legislation and Freedom of Expression. *UC Irvine Law Review*, 10(2), 784–797.
- Partnership on AI. (2020). *The Deepfake Detection Challenge: Insights and Recommendations for AI and Media Integrity*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/[https://partnershiponai.org/wp-content/uploads/2021/07/671004\\_Format-Report-for-PDF\\_031120-1.pdf](https://partnershiponai.org/wp-content/uploads/2021/07/671004_Format-Report-for-PDF_031120-1.pdf)
- Perrigo, B. (2019). *The World Wide Web Turns 30 Today. Here's How Its Inventor Thinks We Can Fix It*. *Time*. <https://time.com/5549635/tim-berners-lee-interview-web/>
- Rundle, M. & Conley, C. (2007). *Ethical Implications of Emerging Technologies: A Survey*. <https://unesdoc.unesco.org/ark:/48223/pf0000149992>
- Sankin, A. (2020). *Want to Find a Misinformed Public? Facebook's Already Done It*. The Markup.
- 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. <https://doi.org/10.1177/0008125619862257>
- Surden, H. (2019). Artificial intelligence and law: An overview. *Georgia State University Law Review*, 35(4), 1304–1337.
- Toplum Standartları Uygulama Raporu. (2024). *Community Standards Enforcement Report*. <https://transparency.facebook.com/community-standards-enforcement#fake-accounts>
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Çevrim içi human-bot interactions: Detection, estimation, and characterization. *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, 280–289. <https://doi.org/10.1609/icwsml.v11i1.14871>
- Volger, H. (2010). Human Rights Council. İçinde *A Concise Encyclopedia of the United Nations* (ss. 329–336). <https://doi.org/10.1163/ej.9789004180048.i-962.273>
- Walorska, A. M. (2020). *Deepfakes & Disinformation Analysis*.
- Wang, Y., & Alshboul, Y. (2015). Mobile security testing approaches and challenges. *2015 1st Conference on Mobile and Secure Services, MOBISecSERV 2015, February 2015*. <https://doi.org/10.1109/MOBISecSERV.2015.7072880>
- Wright, R. (1997). The Man Who Invented The Web. *Time Magazine*, 149(20), ?? <http://search.proquest.com.biblioteca.uniandes.edu.co:8080/docview/217285729/abstract/139E9055DA87AA7EF4E/21?accountid=34489>
- Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A Survey on Deepfake Video Detection. *IET Biometrics*, 10(6), 607–624. <https://doi.org/10.1049/bme2.12031>

- Zago, M., & Nespoli, P. (2020). *ITASEC20 – Presentation of Screening Out Social Bots Interference: Are There Any Silver Bullets. February*. <https://doi.org/10.13140/RG.2.2.32173.92649>
- Zhang, T. (2022). Deepfake generation and detection, a survey. *Multimedia Tools and Applications*, 81(5), 6259–6276. <https://doi.org/10.1007/s11042-021-11733-y>

# The Function of Artificial Intelligence in the Information Distortion Process

İkram ALPSOY (Asst. Prof. Dr.)

## Extended Abstract

With the rapid developments in information and communication technologies in recent years and their widespread use day by day, inaccurate information can be produced realistically and easily, including artificial intelligence techniques, and its dissemination to the desired audience is very fast and intensive.

Information distortion, which is an expression of the transformation caused by a wider range of political, social and technological factors, although it is related to the development in the digital structure, is distinguished from other deceptive or false information by being non-accidental misleading information. Therefore, information distortion is described as information that is regularly transmitted with a certain intention and motivation to provide a certain benefit, even though it is known to be untrue.

Developments in information and communication technologies have led to changes in the way information is created and disseminated, and artificial intelligence techniques have also been decisive in this. These techniques facilitate the creation and dissemination of disinformation by malicious people, and on the other hand, they are also used to combat çevrim içi disinformation.

In this context, the study aims to present the relationship between artificial intelligence and disinformation descriptively, theoretically and methodologically. In addition, the study focuses on the benefits and drawbacks of artificial intelligence, which enables the formation of a different field than the past in communication between people, groups and societies, and its contributions to possible changes in communication activities.

In this research, firstly, different artificial intelligence techniques that increase the disinformation problem are analysed, and then artificial intelligence techniques developed in response to this problem are examined. Finally, suggestions are made for the ethical examination of the disinformation problem by taking into account fundamental rights and freedoms.

Artificial intelligence, which emerged as a result of a project carried out by John McCarthy, Nathan Rochester, Marvin Minsky and Claude Shannon in the United States in 1955, aimed to create intelligent machines that behave in the same way as humans.

Artificial intelligence, which is a computer programme that can learn a language like humans, is described as bringing machines, which are the whole of strings consisting of many parts, to a position to perform tasks that include cognitive functions such as learning, thinking, producing solutions to problems, etc. According to Kaye, artificial intelligence, which is a system created to think and behave rationally like humans, is 'a general-purpose technology used to automate and improve the accuracy, speed and/or scale of a machine's decision-making, pattern recognition and prediction in complex or large environments, or to replace human performance in certain jobs.'

Fake content produced through artificial intelligence technology is defined as deepfake. Regarding the disinformation produced with the help of technology, he assesses that fake media will exist as long as the media to be falsified continues to exist, and that the widespread use by amateurs of programmes that enable editing on digital documents related to artworks, driving licences, diplomas, courses, etc. increases the problem. Recent developments in artificial intelligence technologies have improved the possibility of manipulation on audio, photographs, videos and texts, and the realism levels of audio and videos have increased even more.

Considering that artificial intelligence techniques used in the digital ecosystem create new opportunities to manipulate people on a large scale and effectively, multiple ethical concerns related to disinformation, a problem that has been in effect for a long time, manifest themselves, making the situation even worse. The first ethical phenomenon negatively affected by the current digital system is human dignity. According to this fundamental value, individuals should be recognised as ends in themselves and never as means. On the contrary, when algorithms are programmed to edit what is brought to the attention of individuals based on their profile, which is created by combining data, without taking into account the quality of the content and in order to re-tailor the interaction, these individuals are considered as mere means to economic ends.

In response to the worrying consequences of çevrim içi disinformation, search engines and social media devices are expected to respond more to the spread of false and misleading information. In this context, research on artificial intelligence techniques to detect false, misleading or inaccurate content and to prepare such content çevrim içi is intensifying. The limitation of artificial intelligence systems in distinguishing inaccurate information from disinformation, which is becoming more and more powerful with the ever-increasing advances in technology and digitalisation, poses a problem in terms of freedom of expression and freedom of information.

Although disinformation is a long-standing problem in the process of verifying information, artificial intelligence technologies in the digital system on the web deepen this problem in two ways. Firstly, they can be used by people with malicious intentions to manipulate people in a significant and powerful way. Secondly, it directly increases the spread of such content. Artificial intelligence systems are programmed to enhance interactivity, so it is the business model on the web that is the main driver of misinformation proliferation. In addition, through the widespread use of social media bots, inaccurate content can become more and more realistic.

Although content editing with the help of artificial intelligence techniques is not a correct point of view, such systems can be used to neutralise the directions of the digital system, such as accurate and successful detection of social media bots, detection of content created by artificial intelligence, identification of the use of malicious micro-targeting in order to expose users to disinformation, etc. What is important here is to prevent the malicious use of technology, rather than content, in order to strengthen the message conveyed.

Along with these measures, media literacy is important for users to correctly evaluate the data presented. Although the importance of education related to the critical interpretation of information is increasing, this education has become even more necessary today, when information is at the centre of daily life.

In addition, it is seen that studies that prioritise understanding the consumer experience with artificial intelligence systems and the description of information distortion, which is increasingly difficult to detect with developing technology, have gained importance.

**Keywords:** Digital Ecosystem, Information Distortion, Artificial Intelligence, Content Curation, Ethical Inference.

Bu makale **intihal tespit yazılımlarıyla** taranmıştır. İntihal tespit edilmemiştir.

This article has been scanned by **plagiarism detection softwares**. No plagiarism detected.

Bu çalışmada “**Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi**” kapsamında uyulması belirtilen kurallara uyulmuştur.

In this study, the rules stated in the “**Higher Education Institutions Scientific Research and Publication Ethics Directive**” were followed.

Araştırma tek bir yazar tarafından yürütülmüştür.

The research was conducted by a single author.

Çalışma kapsamında herhangi bir kurum veya kişi ile **çıkar çatışması** bulunmamaktadır.

There is no **conflict of interest** with any institution or person within the scope of the study.