ISSN: 2980-2776 **https://e-tamga.com/**

TAMGA

Türkiye Göstergebilim Araştırmaları Dergisi
Turkish Journal of Semiotic Studies
Journal des Recherches Sémiotiques de Turquie

DOI: 10.5281/zenodo.8092216

OPEN ACCESS

Derin sahtekarlıkların göstergebilimsel ideolojisi*

[The semiotic ideology of deepfakes]
[L'idéologie sémiotique des deepfakes]

Massimo LEONE**

Geliş Tarihi (Received): 23.04.2023 - **Kabul Tarihi (Accepted)**: 26.05.2023 - **Yayın Tarihi (Published)**: 30.06.2023
Makale Türü: Araştırma makalesi - **Article Type:** Reserach article - **Type de l'article:** l'article de recherche

Özet

Makale, yeni sembolik değişim teknolojilerinin temelini oluşturan, anlam ideolojilerini tespit edebilen, göstergebilim odaklı bir iletişim felsefesini teşvik etmektedir. Tarih boyunca geçirdikleri evrim, sahte olanın retoriğine ilişkin önemli değişiklikleri imlemektedir. İnsan türünün kurucu bir unsuru olan bu durumun koşulları, dijital ve yapay zekânın yükselişiyle kökten değişmiştir. Bu makale, çekişmeli üretici ağların göstergebilimsel ideolojisi ile bunların sahtelerinin üretimi ve alımlanması açısından sonuçlarına odaklanmaktadır. Bu yeni rahatsız edici metinsel ürünler çoğunlukla eğlenceli görünmektedir. Ancak makalenin vardığı sonuca göre, insanların bunları tespit edemeyecek hale gelmesi, an meselesidir. Bu nedenle göstergebilim, dijital sahtecilikteki gelişmeler nedeniyle ortaya çıkacak 'epistemolojik krize' acilen odaklanmaya davet edilmektedir.

Anahtar Kelimeler: Derin sahtekarlıklar, göstergebilimsel ideoloji, sahtelik, yapay zekâ, Çekişmeli Üretici Ağ (ÇÜA)

Abstract

The article promotes a semiotically-oriented philosophy of communication, able to detect the ideologies of meaning that underpin the new technologies of symbolic exchange. Their evolution throughout history implies important alterations as regards the rhetoric of the fake. This is a constitutive element of the human species, yet its conditions are radically modified in the digital sphere and by the raise of artificial intelligence. The article focuses on the semiotic ideology of generative adversarial networks and their consequences in terms of production and reception of deepfakes. These disquieting new textual products are mostly seen as entertaining; yet, the article concludes, it is just a matter of time before humans will be unable to detect them. Semiotics is therefore called to urgently concentrate on the 'epistemological crisis' that will be brought about by advances in the digital fake.

Keywords: Deepfakes, semiotic ideology, fake, artificial intelligence, Generative Adversarial Network (GAN)

Résumé

L'article promeut une philosophie de la communication orientée sémiotiquement, capable de détecter les idéologies du sens qui sous-tendent les technologies de l'échange symbolique. Leur évolution au cours de l'histoire implique des changements importants en ce qui concerne la rhétorique du faux. Il s'agit d'un élément

* Bu makale, Avrupa Birliği'nin Horizon 2020 araştırma ve yenilik programı (hibe sözleşmesi no. 819649-FACETS) kapsamında Avrupa Araştırma Konseyi'nden (ERC) fon alan bir projenin sonucudur.
** **Sorumlu Yazar:** Massimo LEONE, Turin Üniversitesi, Felsefe ve Eğitim Bilimleri Bölümü, İtalya, massimo.leone@unito.it, https://orcid.org/0000-0002-8144-4337.

constitutif de l'espèce humaine, dont les conditions sont pourtant radicalement modifiées par le numérique et l'essor de l'intelligence artificielle. L'article se concentre sur l'idéologie sémiotique des réseaux adverses génératifs et leurs conséquences en termes de production et de réception des deepfakes. Ces nouveaux produits textuels, perturbants, sont le plus souvent considérés comme divertissants; pourtant, conclut l'article, ce n'est qu'une question de temps avant que les humains ne soient incapables de les détecter. La sémiotique est donc appelée à se concentrer d'urgence sur la « crise épistémologique » que les progrès du faux numérique est susceptible d'engendrer.

Mots-clés: Deepfakes, idéologie sémiotique, faux, intelligence artificielle, Réseaux Adverses Génératifs (RAG)

> Bir nesne hakkında yalan söylendiğinde, sadece o anla ilgili olan nesne değil, hepsi çarpıtılır. (Picard, Max. 1955. *Der Mensch und das Wort*. Erlenbach-Zürich: E. Rentsch: 51).

1. Giriş

Göstergebilim odaklı bir dijital iletişim felsefesi, yeni aygıtların, süreçlerin ve anlamlı eserlerinin yaratılmasının altında yatan örtük ideolojileri ortaya çıkarmak için insanın anlamlama dizgelerinin uzun tarihi içinde anlam teknolojilerini okumayı amaçlar. Yapay zekâ bir istisna değildir, çünkü gelişimi genellikle zekânın ne olduğu, nasıl çalışması gerektiği ve dünyada ne tür sonuçlar üretmesi gerektiği konusunda belirli önyargılarla desteklenmektedir.

Her kültür ve her tarihsel dönem, sahte olanın üretiminde benimsediği belirli göstergebilimsel yöntemlerle ayırt edilir[1]. İnsan türü, kasıtlı olarak ampirik gerçekliğe karşılık gelmeyen temsillere yol açmak için doğuştan gelen bir kapasiteye sahiptir. Bununla birlikte yanlışın teknolojileri ve dilleri, zaman ve mekân içinde değişmektedir. Dijital teknoloji, telematik iletişim, özellikle yapay zekâ ve derin öğrenme ile sahtenin insan kültürünü belirlediği bir eşikten geçmektedir.

Dijital dünyada insan kültürü 'mutlak sahte' alanına girmektedir. Bu durum her şeyden önce, teknolojinin maddi özelliklerinden kaynaklanmaktadır: Gerçekte dijital temsillerin nesnesi olabilen her şey, ontolojik referans olmaksızın da nesne olabilir. Ontolojisi henüz var olmayan bir gelecekte yaşlı bir yüzün üretilecek herhangi bir dijital görüntüsü, dijital simülasyonun şimdisinde yeniden inşa edilebilir. İkincisi, mutlak sahteliğin alanı niceliksel birikimin gücünden kaynaklanır: gençleşmiş bir yüzün görüntüsü sosyal ağlarda o kadar yoğun ve viral bir şekilde dolaşabilir ki sonunda kimliğini internette temsil eder duruma gelebilir. Üçüncüsü, mutlak sahtecilik alanı yeni yaratım yöntemlerinden kaynaklanmaktadır: Önceden sahtecilik oyunu, sahteciler ve uzmanlar arasında oynanırdı (örneğin sanat alanında); şimdi ise bu oyun giderek artan bir şekilde algoritmalar tarafından oynanmakta ve sonuçları büyük ölçüde öngörülememektedir.

Sahte olanın yaratılmasında uygulanan yapay zekâ her zaman belirli bir nesneye, yani ana arayüz ve kişiler arası iletişim için en önemli insan bedeninin bir bölümü olan yüze uygulanmıştır (Leone, 2021, *El rostro*).

2. Yöntem

Göstergebilim; nesnesi sahtecilik, yüz ve dijital temsilin kesişiminde yer alacak olan bir çalışmayı yürütmek için mükemmel bir donanıma sahiptir. Sahte olana gelince, göstergebilimin tüm kurucu babaları bu konuyu ele almıştır (Ousmanova, 2004, s. 1). Amerikan geleneğinde Charles S. Peirce (Cooke, 2014, s. 2) Fransız dergisi *Communications'ın* "vraisemblable" kavramına adanmış özel sayısından başlayarak yapısal göstergebilimin ana sesleri: Tzvetan Todorov, Gérard Genette, Christian

[1] Massimo Leone, yakında çıkacak.

Metz, Julia Kristeva, Gérard Genot, Roland Barthes ve diğerleri (Todorov, 1968) bu konuya değinmiştir. Baudrillard da bu konuya geri döndü (1987; 2000). Yakın zamanda, 11-14 Haziran 2019 tarihlerinde Lyon'da düzenlenen Fransız Göstergebilim Derneği Kongresi'nde Jacques Fontanille tarafından 'Hakikat sonrası ve demokrasi' konulu bir yuvarlak masa toplantısı düzenlendi (Di Caterino, 2020). Umberto Eco, 'sahte' kavramı üzerine kapsamlı bir şekilde yazmış (1995), göstergebilim dergisi *Versus'un* "Sahte, Kimlik ve Gerçek Şey" konulu özel sayısının editörlüğünü yapmış (1987; Eco, Prieto, Calabrese ve diğerlerinin denemeleriyle) ve ayrıca konuyu çok sayıda deneme ve romanında ele almıştır (*Foucault'nun* Sarkacı; *Prag Mezarlığı*; *Sıfır Numara*, s. 3). Yuri M. Lotman, sahte ve sahte olmayan kavramları ile sahtecilik konusunu defalarca ele almıştır (Andrews, 2003, s.101; Makarychev & Yatsyk, 2017).

3. Araştırmanın doğuşu

Yüzün dijital temsilleri üzerine göstergebilimsel araştırmalar, özellikle de yüzün yapay zekâ tarafından temsil edilmesiyle ilgili olarak giderek artmaktadır. Bununla birlikte, sentetik yüzlerin yaratılmasının altında yatan "göstergebilimsel" ideolojilerin bir analizini geliştirmek için son yıllarda bu alandaki anlam pratiklerinde devrim yaratan algoritmaların kökenine bakmak gerekir. Özellikle de kurucu metinlerine, genç Ian J. Goodfellow'un Montreal Üniversitesi'nde doktora öğrencisiyken 10 Haziran 2014'te yayınladığı "Generative Adversarial Nets" başlıklı makalesine dönmeliyiz (Goodfellow vd., 2014).[2] Goodfellow, bir grup bilgisayar bilimleri doktoralı arkadaşıyla birlikte, iki modelin aynı anda eğitildiği bir karşıt süreç yoluyla üretken modelleri tahmin etmek için yeni bir çerçeve önerdi: Verilerin dağılımını yakalayan bir üretken model ve bir örneğin üretken modelden ziyade eğitim verilerinden gelme olasılığını tahmin eden bir ayırt edici model. Üretici karşıt model, yapay zekâ ve derin öğrenmede "yapay yüzler" (Leone, 2021, *Prefazione*) ve derin sahtekarlıkların oluşturulması da dahil olmak üzere çığır açan uygulamalara yol açmıştır.

Göstergebilim, yapay zekâ çalışmalarına zaten uygulanmıştır. Bununla birlikte, yapay zekânın sonuçlarına ve ürünlerine odaklanmıştır. Oysa göstergebilim aracılığıyla ideolojik ön kabullerini ve işleyişinin derin yapısını incelemek gereklidir. Goodfellow'un üretken yapay zekâ şeması iki örnek arasındaki bir karşıtlık olarak düşünülmüştür; bu nedenle yapısal göstergebilim çerçevesi onun anlaşılabilirliğine büyük katkı sağlayabilir. GAN'ların (Generative Adversarial Networks) soyut mimarisinde iki ana eyleyen vardır: birincisi bir veri örüntüsünü inceleyen ve ondan türetilebilecek bir metin üreten bir üretici eyleyen; ikincisi ise bu şekilde üretilen metni inceleyen ve bunun veri örüntüsünden mi yoksa üretici eyleyenden mi geldiğini değerlendiren bir ayırt edici eyleyendir. Dolayısıyla epistemik açıdan bakıldığında, üretici eyleyen doğru olmayanı doğru olarak göstermeyi ve dolayısıyla doğru olduğuna inandırmayı amaçlarken ayırt edici eyleyen doğru olmayanı doğru olarak göstermeyi ve dolayısıyla yanlış olduğuna inandırmayı amaçlar.

Üreteç pg'nin x verisi üzerindeki dağılımını öğrenmek için giriş gürültü değişkenleri pz (z) üzerinde bir önsel tanımlarız ve ardından veri alanlarında bir eşlemeyi G (z; θg) olarak temsil ederiz; Burada G, θg parametreli çok katmanlı bir algılayıcı tarafından temsil edilen türevlenebilir bir fonksiyondur. Ayrıca tek bir skaler üreten ikinci birçok katmanlı algılayıcı D (x; θd) tanımlarız. D (x), x'in pg'den ziyade veriden gelme olasılığını temsil eder. D, hem eğitim örneklerine hem de G'deki örneklere doğru etiketi atama olasılığını en üst düzeye çıkarmak için eğitilir. G aynı zamanda log (1 - D (G (z))'yi en aza indirmek için eğitilir.)

4. Araştırma bulguları

Üretici karşıt ağların (GAN'lar) kurucu makalesi göstergebilim merceğinden okunduğunda özellikle iki unsur göze çarpmaktadır:

i. İfade ettiği yapay zekâ anlayışı karşıtlık fikrine dayanmaktadır (iş birliği ya da basit rekabet değil);

[2] Araştırmacı, zaman içinde dünyanın yapay zekâ ve özellikle de derin öğrenme konusunda uzman haline geldi ve şu anda Apple'ın Özel Projeler Grubu'nda makine öğrenimi departmanının yöneticiliğini yapmaktadır.

ii. Yeni derin öğrenme mimarisini en iyi açıklayan metafor, kalpazan ve uzman (özellikle madeni para konusunda) metaforudur.

Yapay zekânın bu yeni mimarisi artık pek çok profesyonel ve sosyal alanda uygulama alanı bulmaktadır. Özellikle de durağan ya da hareketli yüzlerin bilgisayar tarafından üretilen görüntü ve videolarının yaratılmasında giderek artan bir şekilde yine bilgisayar tarafından üretilen genellikle yüz ifadeleri, jestler, hareketler, sözlü konuşma parçaları, şarkılar ve danslar gibi çoklu işaret sistemleriyle ifade edilen kafalar, bedenler ve bağlamlarla ilişkilendirilmektedir. Bundan dolayı her iki yön de daha fazla felsefi ve göstergebilimsel yansımayı hak etmektedir.

GAN şeması, aynı Goodfellow tarafından 2014 yılında önerilen metaforla okunabilir: D ve G sırasıyla bir bilen ve bir kalpazan gibi davranır. Kalpazan dolaşımdaki parayı inceler ve onu üretmeye çalışır; bilen ise kaynağını bilmeden kalpazan tarafından üretilen parayı inceler ve onun sahte mi yoksa gerçek para mı olduğunu anlamaya çalışır. Ancak bunu yaparken bilen kişi kalpazana, gerçek paradan ayırt edilmesi daha da zor olan sahte para yaratmada faydalı olacak bilgiler verir. Bilen kişi gerçek ve sahte para arasında giderek daha fazla ayrım yapmayı da öğrenir. Sanat piyasası metaforu, bu üretim ve ayrımcılık sarmalı fikrini etkili bir şekilde yakalayabilir: Bir kalpazan sahte Modigliani'yi dolaşıma sokmaya çalışırken, bir uzman da bunları gerçek Modigliani'den ayırt etmeye çalışır. Ancak bunu yaparken ikincisi birinciye eserleri en iyi nasıl taklit edeceği konusunda bilgi verir; tersi durumda, birincisi de ikinciden İtalyan sanatçının eserlerini nasıl taklit edeceğini öğrenir.

Kişi kendine bu sarmalın gözlemci aktörünün doğası hakkında soru sormalıdır. Üretken modelin ürünleri sadece ayırt edici model tarafından değil, aynı zamanda, en azından ilk etapta, GAN'ların muhatabı ile örtüşen bir insan muhatap tarafından da onaylanmaktadır. Modeller bir insan muhatap tarafından programlanır, ancak 'davranışları' tamamen öngörülebilir değildir, özellikle de insan bilişi ile yapay zekâ arasındaki hesaplama boşluğu nedeniyle. Dolayısıyla insan programcı, üreten model ile ayırt edici model arasındaki etkileşimin ürünlerinin hem alıcısı hem de muhatabıdır. Buna ek olarak bu profesyonel gözlemci aktörün ötesinde, üretici modelin ürünlerini kaynağını bilmeden alacak olanlardan oluşan bir başka aktör daha vardır. Yukarıda açıklanan spiral, bu profesyonel olmayan gözlemcinin epistemik belirsizliğini arttırmak üzere tasarlanmıştır.

İlk metafor çerçevesinde daha basit bir şekilde ifade etmek gerekirse sahteci ve uzman arasındaki rekabet, sahte olan ancak özellikle sarmalın dışındaki gözlemci aktör tarafından bu şekilde tanınması giderek zorlaşan para veya sanat eserlerini dolaşıma sokar. Artık bu şekilde tanımlanamayan bir sahtenin kitlesel dolaşımı, otantik sanat eserlerine, otantik paraya epistemik itibarsızlık getirir. Bu belki de 'sahtecilik sarmalının' en büyük tehlikesidir.

Bazı araştırmacılar GAN'lara olumlu bir ışık tutmaya çalışmış ve iç diyalektiklerinin daha çok öğretmen ve öğrenci arasındaki diyalektiğe benzediğini öne sürmüşlerdir. Dolayısıyla, üretici model bir veritabanından güvenilir temsiller üretmeye çalışan bir öğrenci gibi olurken ayırt edici model bu temsilleri inceleyen ve değerlendiren bir öğretmen gibi olacaktır. Bu kısmen doğrudur ancak farkı yaratan şey, GAN'ların dünyasında üretici modelin temsillerinin öğrenme bağlamına atıfta bulunmadan dolaşmaya başlamasıdır.

Dijital ve analog sahtecilik arasındaki fark da burada yatmaktadır. İnsan türü doğası gereği kasıtlı olarak gerçekliğin sahte temsillerini, yani belirtisel bir kökenden yoksun olmakla birlikte, özellikle görüntüsel bir anlam etkisi yaratarak bunu taklit eden temsilleri üretme yeteneğine sahiptir. Bu yetenek muhtemelen türün biyolojik evrimi tarafından uyarlayıcı olarak seçilmiştir çünkü potansiyel olarak tehlikeli durumları deneysel olarak deneyimlemek zorunda kalmadan zihinsel olarak deneyimlemesine izin vermiştir. Ayrıca kendisini yırtıcılardan korumasına veya potansiyel avını tuzağa düşürmesine de olanak sağlamıştır. Bu hem bitki hem de hayvan türlerinde bulunmayan bir yetenektir.

Örneğin, hayvanların en dikkat çekici özelliklerinden biri, diğer kuşların ve çeşitli doğal unsurların seslerinin yanı sıra bir kameranın, motorlu testerenin, yangın alarmının, hidrolik silindirin tetiklenmesi gibi insan çevresinden gelen sesleri de taklit etme yetenekleridir. Ancak insan türünde, dilde ve dil aracılığıyla ifade edilen bu kapasite, kasıtlı ve yanlış temsillere estetik zevk ve değer atfetme

yeteneğinden oluşan, muazzam bir kurgusal metin üretimini tetikleyen bir tür 'uyarlama'ya yol açmıştır. Dijital, insan türünün sahte olanla ilişkisinin tarihinde önemli bir niteliksel ve niceliksel değişiklik getirmektedir.

5. Dijital sahteciliğin göstergebilimsel özellikleri

İlk olarak, dijital olan, göstergebilimsel belirtisi tamamen programlanabilir olan, değişken bir maddiyatla donatılmıştır ki dijital öncesi metinlerin belirtisinde asla böyle bir durum söz konusu değildir. Nesnesiyle belirtisel bir ilişkisi olan herhangi bir dijital temsilin, bu ilişki olmadığında bile aynı şekilde yeniden üretilebileceği anlamına gelir; resim elbette var olmayan yüzleri simüle edebilir ve yine de ontolojik yüz ile boyanmış yüz arasındaki boşluk her zaman belirgin olacaktır, dijitalde durum böyle değildir. Dijital teknoloji, fotoğrafın belirleyici özelliği olan belirtisellik duygusunu özümser ve belirtiselliğin yokluğunda yeniden üretir; aynı zamanda fotoğrafik görüntünün inşasına tam bir programlanabilirlik getirir. Resim var olmayan nesneleri temsil edebilir ama insanları onların varlığına inandıramaz; analog fotoğrafçılık insanları temsil ettiği nesnelerin varlığına inandırabilir ama var olmayan nesneleri temsil edemez, en azından etkili bir şekilde temsil edemez; dijital fotoğrafçılık ise insanları temsil ettiği var olmayan nesnelerin varlığına inandırabilir.

İkincisi, yapay zekânın ve özellikle GAN'lar tarafından derin öğrenmenin dijital temsillerinin maddi belirtisinin üretimine uygulanması, onları insan değerlendirmesinden uzaklaştırmaktadır. Sahtecilik insan türüyle özdeştir ancak türün tarihinde ilk kez insan olmayan ajanlar, değerlendirilmesi giderek insanlardan kaçan ve bunun yerine giderek yapay zekâ aracılığıyla yürütülen bir değerlendirmeye emanet edilen bir sahtecilik üretimine konumuna getirilmiştir.

Üçüncüsü, dijital sahtecilikler geçmişe kıyasla eşi benzeri görülmemiş bir kolaylıkla çoğaltılıp dolaşıma sokulabiliyor ve bu niceliksel boyut aynı zamanda niteliksel bir değişime de yol açıyor: Sanki otantik sanat kendini sürekli ve hızlı bir şekilde kopya üretimi üzerinde çalışan sonsuz sayıda sahteciye karşı savunmak zorundaymış gibi.

Dijital sahte, uzun vadede 'dijital gerçek'ten ayırt edilemez hale gelmeye mahkûmdur; örneğin yüzler söz konusu olduğunda bir yüzün dijital fotoğrafından, biyolojik, ontolojik bir yüzden mi yoksa sentetik bir görüntüden mi üretildiğini anlamak artık mümkün olmayacaktır. Göstergebilim, gerçekliğe uygunluk olarak mantıksal 'hakikat' kavramını sorunsallaştırma eğiliminde olması ile birlikte 'gerçeklik etkisi' üreten göstergebilimsel koşulları da dikkate alır. Ancak, ontolojik bir gerçeklik varsaymaksızın "gerçeklik etkisi" retoriğini açıklamak kaçınılmaz çıkmazlara yol açar. Benzer şekilde, analog bir fotoğrafın 'gerçeklik etkisi' sorunsallaştırılabilir ancak dijital teknolojinin ve özellikle de görüntü oluşturmaya uygulanan dijital derin öğrenmenin ortaya çıkışının, gerçeklik etkisine sahip bir göndergesel görüntü ile tam olarak aynı etkiyi üreten sentetik bir görüntü arasında ayrım yapma olasılığını zayıflattığını da kabul etmek gerekir.

6. Derin sahtecilikten göstergebilimine doğru

Dijital sahteciliğin tespit edilemeyen doğası, giderek daha karmaşık ve sosyal açıdan merkezi metinlerde kendini gösterdikçe endişe verici bir hâl almaktadır. Bu açıdan bakıldığında, derin sahtekarlık fenomeni üzerinde acilen düşünülmesi gerekmektedir. Birincisi, insan toplumlarının işleyişinin merkezinde yer alan bir nesnenin, yüzün dijital simülasyonunu içerdiği için; ikincisi, bu nesneyi yalnızca durağan görüntüde değil, aynı zamanda hareketli görüntüde ve giderek artan bir şekilde, örneğin dudak hareketlerinin sentetik temsili yoluyla bağlamında ve işlevlerinde simüle ettiği için.

2019 yılında, çevrimiçi sentetik medya tespiti ve izlenmesi için derin öğrenme ve bilgisayarla görme teknolojileri sağlayan Amsterdam merkezli bir siber güvenlik şirketi olan Deeptrace –o zamandan beri Sensity olarak yeniden adlandırıldı– "Deepfake" başlıklı bir rapor yayınladı ve o sırada görüngünü çevrimiçi olarak hızla büyüdüğünü, derin sahtekarlık videolarının sayısının çevrimiçi 14.678 video rakamına ulaşana kadar yedi ay boyunca neredeyse iki katına çıktığını belirtti. Sensity'nin

işbirlikçilerinden Francesco Cavalli tarafından 8 Şubat 2021'de yayınlanan bir blog yazısı, çevrimiçi sahte video sayısının 2018'den bu yana katlanarak arttığını ve yaklaşık her altı ayda bir ikiye katlandığını ortaya koyuyor. 2020 Aralık ayında Sensity tarafından 85.047 derin sahtekarlık videosu tespit edildi.

Sensity, şu anda sistematik olarak derin sahtekarlık üreten 516 kaynağı izlemekte olup, bugüne kadar 3.231 tanınmış kişiyi hedef alan 118.232 "görsel tehdit" üretildiğini belirtir. Derin sahtekarlıkların hedeflerinin %42'si ABD'de; %10.3'ü İngiltere'de; %6'sı Hindistan'da; %5.7'si Güney Kore'de; %5.6'sı Japonya'dadır. En çok hedef alınan sosyal ve profesyonel faaliyetler ise eğlence sektörü, %55.9; moda, %23.9; siyaset, %4.6; spor, %4.5; üst düzey endüstri yöneticileri, %3.1. 2018 Deeptrace raporu aynı zamanda pornografide rıza dışı derin sahtekarlığın öne çıktığını ve o dönemde toplam çevrimiçi derin sahtekarlık videolarının %96'sını oluşturduğunu tespit etmiştir. Derin sahtekarlık pornografiye adanmış dört ana web sitesinin, dünyanın dört bir yanından yüzlerce kadın ünlüyü hedef alan videoların 134 milyondan fazla görüntülendiği de tespit edilmiştir. Ayrıca, "deepfake" terimi, "Deepfakes" adlı bir Reddit kullanıcısının 2017'nin sonlarında ünlülerin yüzlerini porno videolara dönüştürmesine olanak tanıyan bir derin öğrenme algoritması geliştirdiğini iddia etmesinin ardından yaygın olarak kullanılmaya başlandı.

Derin sahtekarlıklar siyasi alanda da önemli bir etkiye sahiptir. Batı medyasında çok az yer bulan Gabon ve Malezya'daki en az iki önemli vakada derin sahtekarlıklar, özellikle hükümetin örtbas ettiği ve siyasi karalama kampanyası yürüttüğü iddialarında merkezi bir rol oynamıştır. İlk vaka bir askeri darbe girişimiyle sonuçlanırken, ikincisi yüksek profilli bir politikacının hapisle tehdit edilmesine yol açmıştır.

Geleneksel olarak genel medya adli tıpına adanmış araştırma alanı, artık görüntü ve videolarda yüz manipülasyonunu tespit etmeye yönelik artan çabalara odaklanmaktadır. Bu çabanın bir kısmı, biyometrik anti-gözetim ve modern veritabanı tabanlı derin öğrenme üzerine yapılan önceki araştırmalara dayanmaktadır. Tespit yöntemlerinin değerlendirilmesini standartlaştırmak amacıyla, yüz manipülasyonu tespiti için otomatik bir ölçüt önerilmiştir. Bu ölçüt özellikle, rastgele bir seviyede ve sıkıştırma boyutunda yüz manipülasyonlarının önde gelen temsilcileri olarak DeepFakes, Face2Face, FaceSwap ve NeuralTextures'a dayanmaktadır.

Günümüzde, verilerin kamuya açık olması ve otomatik kodlayıcılar (AE'ler) ve aslında *Üretken Çekişmeli Ağlar* (GAN) gibi birçok manuel düzenleme adımını ortadan kaldıran derin öğrenme tekniklerinin evrimi sayesinde var olmayan yüzleri otomatik olarak sentezlemek veya bir görüntü veya videodaki bir kişinin gerçek yüzünü manipüle etmek giderek daha kolay hale gelmektedir. Sonuç olarak, ZAO ve FaceApp gibi açık yazılımlar ve mobil uygulamalar, bu alanda önceden deneyim gerektirmeden herkesin sahte görüntüler ve videolar oluşturmasına olanak sağlamaktadır.

Adli medyada sahte görüntülerin tespit edilmesine yönelik geleneksel yöntemler genellikle şunlara dayanmaktadır:

i. Kamera içinde üretilen "parmak izleri", yani optik lens, renk filtresi dizisi, enterpolasyon, sıkıştırma vb. gibi hem cihazlar hem de yazılımlar tarafından kameralar tarafından ortaya konan içsel dijital parmak izlerinin analizi;
ii. Kes-yapıştır veya gömme işlemleri gibi düzenleme yazılımları tarafından ortaya konan dış parmak izlerinin analizi gibi kamera dışında üretilen "parmak izleri".
iii. Kopyalama ve yapıştırma işlemleri veya görüntünün farklı unsurlarının entegrasyonu, bir videodaki kare hızının azaltılması vb. gibi düzenleme yazılımları tarafından sunulan harici parmak izlerinin analizi gibi kamera dışında üretilen "parmak izleri".

Bununla birlikte, Ruben Tolosana ve *diğerlerinin* 2020 tarihli "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection" başlıklı makalesinde de belirtildiği üzere, geleneksel sahte CGI tespit yöntemleri tarafından dikkate alınan özelliklerin çoğu, belirli eğitim senaryosuna büyük ölçüde bağlıdır ve bu nedenle beklenmedik koşullara karşı etkili değildir.

Derin sahtekarlığın sosyal etkisi yeni bir çalışma konusu olmakla birlikte, giderek artan sayıda araştırmacının dikkatini çekmektedir. Jeffrey T. Hancock ve Jeremy N. Bailenson, 2021 yılında *Cyberpsychology, Behavior, and Social Networking* dergisinin "The Social Impact of Deepfakes" başlıklı özel bir sayısının editörlüğünü yapmıştır. Bu konudaki teknik durum hâlâ yeterince gelişmemiştir. 2021 yılında, Nanyang Teknoloji Üniversitesi Wee Kim Wee İletişim ve Enformasyon Okulu'nda araştırmacı olan Saifuddin Ahmed, "Who Inadvertently Shares Deepfakes? Analyzing the role of political interest, cognitive ability, and social network size" başlıklı bir makale yayınladı. Amerika Birleşik Devletleri ve Singapur'da toplanan anket verilerini kullanan bu çalışma, siyasi ilgi, bilişsel yetenek ve sosyal ağ büyüklüğünün yanlışlıkla derin sahtekarlık paylaşımındaki rolünü incelemektedir. Sonuçlar, daha dar siyasi ilgi alanlarına ve daha az bilişsel yeteneğe sahip olanların yanlışlıkla derin sahtekarlık paylaşma olasılıklarının daha yüksek olduğunu göstermektedir. Sonuçlar ayrıca siyasi ilgi ve derin sahtekarlık paylaşımı arasındaki ilişkinin ağ büyüklüğü tarafından yönetildiğini de göstermektedir. Dolayısıyla, siyasi olarak ilgili vatandaşların derin sahtekarlık paylaşma olasılığı daha büyük sosyal ağlarda artmaktadır.

Derin sahtekarlıkların psikolojik ve psikososyal etkilerine ilişkin ampirik kanıtlar hâlâ azdır. Bununla birlikte, sanal gerçeklikte "Doppelgänger" yaratılmasından ilginç içgörüler elde edilebilir. Sanal gerçeklikte kişinin kendisinin bir simülakrını izlemesi, katılımcıların kendilerini temsil edildiğini gördükleri eylemleri gerçekten gerçekleştirdiklerine inandıkları sahte anıların kodlanmasına neden olmaktadır; diğer deneyler bu simülakrların marka tercihi veya sağlık davranışı üzerindeki etkisini göstermektedir. Daha 2009 yılında Segovia ve Bailenson, *Media Psychology* dergisinde "Virtually True: Children's Acquisition of False Memories in Virtual Reality" makalesini; aynı sayıda Fox ve Bailenson, "Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors" makalesini; daha sonra 2014 yılında Ahn ve Bailenson, *Journal of Marketing Theory and Practice* dergisinde "Self-Endorsed Advertisements: When the Self Persuades the Self" makalesini yayınlamıştır.

Derin sahtekarlıkların sosyal etkilerini incelemek için benimsenen yaklaşımlar çeşitlilik göstermektedir. Young Ah Lee *ve arkadaşlarının* (2021) "To Believe or not to Believe: Framing Analysis of Content and Audience Response of Top 10 Deepfake Videos on YouTube" başlıklı makalesi, YouTube'daki en popüler 10 derin sahtekarlığa tarihsel bir bakış sunmakta ve izleyici yorumları aracılığıyla dilsel tepkileri analiz etmektedir. Catherine Francis Brooks'un (2021) "Popular Discourse Around Deepfakes and the Interdisciplinary Challenge of Fake Video Distribution" başlıklı makalesi, derin sahtekarlıkların alımlanmasını ölçmek için 2018 yılında Reddit'i inceliyor ve bu verileri olumsuz kullanım durumlarına olası çözümler önermek için kullanıyor. Justin D. Cochran ve Stuart A. Napshin'in (2021) "Deepfakes: Awareness, Concerns, and Platform Accountability" başlıklı çalışması, öğrencilerin derin sahtekarlıklar konusundaki farkındalık ve endişelerinin yanı sıra platformların bu yeni teknolojiyi düzenleme çabalarındaki hesap verebilirlik derecelerini değerlendirmek için anket yapıyor.

Diğer makaleler, derin sahtekarlıkların benlik algısı üzerindeki psikolojik dinamiklerine ilişkin bazı ilk bilgiler sunmaktadır. Wu, Ma ve Zhang (2021), genç kadınların kendi görüntülerini bir ünlününkiyle karıştıran bir derin sahtekarlığa maruz kalmadan önce ve sonra kendi görünümlerini nasıl değerlendirdiklerini incelemiştir. Bu deneyler, benlik algısı üzerinde olumlu etkiler göstermiştir. Bir başka çalışma (Weisman & Peña 2021), bir yapay zekâ programı tarafından oluşturulan yeniden yapılandırılmış bir versiyona maruz kalmanın yapay zekâya olan güveni nasıl etkilediğini araştırmaktadır. Katılımcının yüzüyle konuşan bir kafaya maruz kalması, yapay zekâya olan duygulanım temelli güveni azaltmaktadır.

7. Sonuç

Çoğunlukla, derin sahtekarlıklar hâlâ gülümsetiyor, her ne kadar komiklik bazen rahatsız edicilikle bağlantılı olsa da ancak, birkaç istisna dışında, derin sahtekarlıklar hâlâ *trompe l'oeil* işlevi görüyor: Eğlendiriyorlar çünkü aldatmacalarını fark ediyorsunuz. Bununla birlikte, derin sahtekarlık üretmenin teknik koşulları göz önüne alındığında, insanları dışlayan algoritmik aşırılıkla, diğer makinelerin

aldatmacalarını ortaya çıkarmak için makinelerin kullanılması anlamına gelse bile, aldatmacanın kesinlikle tespit edilemez olması sadece bir zaman meselesidir. Birçok insan toplumunun tekilliğe karşı bir siper olarak diktiği yüz, yakında tüm dijital temsillerinde istenildiği zaman tahrif edilebilir hâle gelecektir. Üç boyutlu derin sahtekarlıkların ya da yapay zekâya bağlanabilen yapay biyolojik yüzlerin üretimindeki ilerleme, sahte yüzlerin ayırt edilmesini daha da karmaşık hale getirecektir. Sahte olanın söylemine diğerlerinden daha fazla odaklanmış bir disiplin olan göstergebilim, "dijital sahtelerin" çoğalmasının yol açabileceği epistemolojik savrulma üzerine, özellikle de birlikte yaşamaya yönelik bir arayüz olarak yüzün temsiline ilişkin olarak acilen düşünmeye çağrılmaktadır. Göstergebilim, dijital sahteciliğin yeni zorluklarını hesaba katmak ve son derece yanlış olanın anlamını kavramak için en azından kısmen kendini yenilemek zorunda kalacaktır.

Kaynakça

Ahn, S. J.-G., & Bailenson, J. (2014). Self-Endorsed Advertisements: When the Self Persuades the Self. *The Journal of Marketing Theory and Practice*, 22(2), 135-136.

Andrews, E. (2003). *Conversations with Lotman: Cultural Semiotics in Language, Literature, and Cognition [Toronto Studies in Semiotics and Communication]*. University of Toronto Press.

Baudrillard, J. (1987). Au-delà du vrai et du faux, ou le malin génie de l'image. *Cahiers internationaux de sociologie, nouvelle série*, 82, 139-145.

Baudrillard, J. (2000). *The Vital Illusion: The Wellek Library Lectures*. New York.

Brooks, C. F. (2021). Popular Discourse Around Deepfakes and the Interdisciplinary Challenge of Fake Video Distribution. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 159-163.

Cochran, J. D., & Napshin, S. A. (2021). Deepfakes: Awareness, Concerns, and Platform Accountability. Cyberpsychology, *Behavior, and Social Networking*, 24(3), 164-172.

Cooke, E. F. (2014). Peirce and the 'Flood of False Notions'. In T. Thellefsen, B. Sørensen, & C. De Waal (Eds.), *Charles Sanders Peirce in His Own Words: 100 Years of Semiotics, Communication and Cognition [Semiotics, Communication and Cognition, 14]* (pp. 325-331). De Gruyter Mouton.

Di Caterino, A. (2020). Fake News: Une mise au point sémiotique [*Actes Sémiotiques*, 123]. Retrieved from https://www.unilim.fr/actes-semiotiques/6445

Eco, U. (1987). Fakes, Identity and the Real Thing [Special issue]. *Versus*, 46.

Eco, U. (1995). *Faith in Fakes: Travels in Hyperreality*. Minerva.

Fox, J., & Bailenson, J. N. (2009). Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. *Media Psychology*, 12(1), 1-25.

Goodfellow, I. J., et al. (2014). *Generative Adversarial Networks*. Retrieved from https://arxiv.org/abs/1406.2661

Hancock, J. T., & Bailenson, J. N. (2021). The Social Impact of Deepfakes [Special issue]. Cyberpsychology, Behavior, and Social Networking, 149-152. Retrieved from https://stanfordvr.com/pubs/2021/the-social-impact-of-deepfakes/

Leone, M. (2021). Prefazione / Preface. In M. Leone (Ed.), Volti artificiali / Artificial Faces [Lexia: International Journal of Semiotics, 37-8] (pp. 9-25). Aracne.

Leone, M. (2021). El rostro aumentado: Trayectorias tecnologicas de lo falso. In H. Valdivieso & L. Rojas Parma (Eds.), Next: Imaginar el Post-Presente: Filosofía, arte y tecnología en la cultura digital (pp. 55-76). Universidad Católica Andrés Bello.

Leone, M. (Forthcoming). Semioethics of the Visual Fake. In T. Andina & T. Dreier (Eds.), The Ethics of Digital Images [Bild und Recht, 5]. NOMOS.

Makarychev, A. S., & Yatsyk, A. (2017). Lotman's Cultural Semiotics and the Political: Reframing the Boundaries. Rowman & Littlefield International.

Ousmanova, A. (2004). Fake at Stake: Semiotics and the Problem of Authenticity. Problemos, 66(1), 80-101.

Segovia, K. Y., & Bailenson, J. N. (2009). Virtually True: Children's Acquisition of False Memories in Virtual Reality. Media Psychology, 12(4), 371-393.

Todorov, T. (Ed.). (1968). Recherches sémiologiques le vraisemblable [Special issue]. Communications, 11.

Tolosana, R., et al. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. Retrieved from https://arxiv.org/abs/2001.00179

Weisman, W. D., & Peña, J. F. (2021). Face the Uncanny: The Effects of Doppelganger Talking Head Avatars on Affect-Based Trust Toward Artificial Intelligence Technology are Mediated by Uncanny Valley Perceptions. Cyberpsychology, Behavior, and Social Networking, 24(3), 182-187.

Wu, F., Ma, Y., & Zhang, Z. (2021). 'I Found a More Attractive Deepfaked Self': The Self-Enhancement Effect in Deepfake Video Exposure. Cyberpsychology, Behavior, and Social Networking, 24(3), 173-181.

YoungAh Lee, et al. (2021). To Believe or Not to Believe: Framing Analysis of Content and Audience Response of Top 10 Deepfake Videos on YouTube. Cyberpsychology, Behavior, and Social Networking, 24(3), 153-158.