Bölüm 8

Dijital Enformasyon Çağında Sentetik Kitle Manipülasyonu: Deepfake (Derin Sahtelik) Ürünleri

Serdar Kuzey Yıldız[1]

Özet

Yeni medya ile dönüşen iletişim bilimi yeni form ve araçlar kazanmış, fakat bu dönüşüm her zaman olumlu sonuçlar doğurmamıştır. İleri teknoloji tüm avantajlarına rağmen etik tartışmalarla da sık sık gündeme gelmektedir. Enformasyonun üretiminden dağıtımına kadar tüm aşamada çığır açan gelişmelere tanıklık edilen post-truth (hakikat ötesi) çağ ile beraber iletişim aksiyonları farklı bir düzleme oturmuş; alanyazında yeni kavramlar ve tartışmalar ortaya çıkmıştır. Türkçe'ye "derin sahtelik" adıyla taşınan "deepfake" kavramı da üzerinde güncel tartışmaların yaşandığı bir alana dönüşmüştür. Derin öğrenme teknikleri kullanılarak eğitilmiş yapay zekâ ile ortaya konulan deepfake ürünlerinin gerçekliğin çarpıtılması, dezenformasyonun yayılması ve çoğu zaman topluma mal olmuş figürlerin, siyasi liderlerin ya da popüler kültür ikonlarının itibarlarının zedelenmesi gibi amaçlarla kullanıldığı görülmektedir. Deepfakelerin, dijital enformasyon çağında haber medyasının güvenilirliğine de gölge düşürebileceği ortaya atılan iddialar arasındadır. Betimsel analiz yöntemi ile kapsamlı bir doküman taraması yapılan bu çalışmada, deepfake teknolojisinin kavramsal arka planı, dijital uzamlardaki kullanım biçimleri, potansiyel tehlikeleri ve olanakları güncel istatistikler ve olgularla açıklanmaya çalışılmıştır. Çalışmanın bir diğer motivasyonu ise konuyla ilgili Türkçe literatürdeki sınırlı sayıdaki tartışmaya kapsamlı bir bakış açısı ile katkıda bulunmaya çalışmaktır. Araştırma sonucunda, deepfake ürünlerinin insan teknolojiyle olan ikileminde yeni bir tartışma başlattığı sonucuna ulaşılmıştır. Ayrıca kavramın olası tehlikelerinden korunabilmek için mahremiyet, demokrasi, siyaset ve ulusal güvenlik gibi pek çok hassas başlıkta denetim ve kontrol mekanizmalarına, iyi bir medya okuryazarlığı eğitimine ve karşı teknolojilere ihtiyaç duyulduğu görülmüştür.

1 İstanbul Aydın Üniversitesi, ORCID: 0000-0001-5891-5682, skuzeyyildiz@aydin.edu.tr

https://doi.org/10.58830/ozgur.pub63.c226

Giriş

Yapay zekâ ile manipüle edilmiş, derin öğrenme (DÖ) kullanılan deepfake videolarının yaygın olarak bilinen ilk örnekleri Kasım 2017'de ortaya çıkmış, o tarihten beri haber medyası ve dolayısıyla kamuoyu, bu yeni kitle manipülasyon aracı olan video türüne atıfta bulunmak için "deepfake" (derin sahtelik) terimini kullanmaya başlamıştır. Haber medyası sıklıkla deepfakelerin gerçek ve yalan arasındaki çizgiyi kalıcı olarak bulanıklaştırarak videonun hakikat iddiasını yok etmeye hazırlandığını iddia etmektedir. Gazeteciler ve araştırmacılar ise deepfakelerin demokrasiyi yok etme becerilerine dikkat çekmektedir. Deepfakelerin şöhretinin bu kadar olumsuz olmasının arkasında bir dizi teşebbüs yatmaktadır. Seçimleri tahrif etmeye çalışma, ulusal güvenliği tehlikeye atma veya toplumda yerleşmiş olan yaygın şiddeti kışkırtma gibi pek çok kötüye kullanım bu teknolojinin olumsuz bir imaj edinmesine neden olmuştur.

İletişim ve teknoloji tarihinde hemen hemen her yeni gelişme etrafında oluşan panik havası, uzmanların hukuki, ekonomik veya söylemsel güç elde etmesi için bir açılım da yaratmıştır. Bugün deepfake konusunda da bu tavır değişmemiştir. Kişinin sesi ve benzerliği de dâhil olmak üzere bedenlerin sayısallaştırılması için giderek daha rutin hâle gelen derin sahtekârlıklar uygulanmaktadır. Bunlar yalnızca birkaç yüz görüntüden oluşan eğitim verisi gerektiren makine öğrenmesi (MÖ) tekniğinden oluşmaktadır. Deneysel makine öğrenimine dayanan deepfakeler, görsel-işitsel manipülasyon spektrumunun bir ucunu temsil etmektedir. Deepfake sürecinin, yani sentetik ve yanıltıcı medya yaratmanın diğer biçimi "cheapfake/shallowfake" (ucuz sahtelik) ürünlerinden oluşmaktadır. Bu yöntemle yapılan manipülasyonların maliyeti ucuzdur, erişilebilir yazılımlara dayanmakta veya hiçbir yazılıma dayanmamaktadır.

Post-truth (hakikat ötesi) çağın getirdiği gerçekliğin yeniden üretimi ve dezenformasyon sorunu günümüz medyasının ve dijital gazeteciliğin en büyük açmazlarından biridir. Bir yandan habere ve bilgiye erişim saniyeler içerisinde birden fazla mecra üzerinden eşzamanlı gerçekleşirken diğer yandan teyit sorunu yaşanmaktadır. Olguların ve güvenilirliğin sorgulandığı durumlarla sıklıkla yüzleşilmektedir. Bilgiye erişimin boyutlarını yeniden şekillendiren yeni medya araçları, yalan ya da sahte içeriklerin özellikle sosyal medyada yayılımında da doğrudan sorumludur. Tam da bu nedenle bazı odaklar tarafından "yalanlar çağı" olarak adlandırılan post-truth dönemi, yalanın âdeta kurumsallaştırıldığı bir yöne doğru evrilmekte ve bu durum medyaya duyulan güveni zedelemektedir.

Sahte içeriklerin üretilmesini kolaylaştıran -başka bir deyişle gerçeği yeniden üreten ya da çarpıtan- teknolojik araçlar ve yazılımlar medya tüketicileri açısından gözle ayırt edilmesi oldukça zor bir yalan distopyasını inşa etmektedir. Böylece neyin gerçek neyin yalan olduğu hakkındaki kanı giderek muğlak bir hâl almaktadır. Dezenformasyon sürecinin yeni parçalarından biri olarak da görebileceğimiz deepfake ürünlerinin akıllı telefonlarda yer alan basit uygulamalar yoluyla bile rahatça üretilebilir hâle gelmesi tehlikenin yaygınlaştığına işaret etmektedir. Giderek daha da gelişen deepfake ürünleri gerçeğinden ayırt edilmesi zor "replikalar" üretilmesini mümkün kılmaktadır. Bu durum elbette sadece eğlence amaçlı kullanılmamakta; siyasetten sinemaya kadar uzanan geniş bir alanda kamuoyuna mal olmuş bazı isimlerin itibarı zedelenmektedir. Bu durum etik değerlerin bir kez daha masaya yatırılması ve hatta sosyal mühendislik tartışmalarının başka bir kulvarda alevlenmesi anlamına gelmektedir.

Bugün teknoloji ve siberetik uzmanları, kanun yapıcılar, gazeteciler ve akademisyenler, bu teknolojinin gelecekteki rolünü belirleyebilmek adına birtakım düzenlemelerle, tasarımlarla ve kültürel normlarla deepfakelerin durdurulamaz ilerleyişine bağlı olarak ortaya çıkan tehditlere yanıt aramaktadır. Betimsel analiz yöntemi ile kapsamlı bir literatür taraması yapılan bu çalışmada da söz konusu teknolojisi tüm yönleriyle ele alınmaya çalışılmıştır. Deepfake kavramının arka planı, olumlu-olumsuz kullanım biçimleri irdelenerek güncel verilerle ve olgularla desteklenmiştir. Araştırmada son yıllarda farklı bir alt başlıkta tartışmaya açılan "cheapfake/shallowfake" (ucuz sahtelik), "deepfake geography" (derin sahte coğrafya) kavramlarına da yer verilmiş, bu anlamda da alanyazına katkı sunulması amaçlanmıştır. Ayrıca, kişisel mahremiyetten demokrasi kültürüne, siyasetten ulusal güvenliğe kadar pek çok alanda etkisi görülen deepfake teknolojisinin potansiyeli de objektif bir biçimde değerlendirilmeye çalışılmıştır.

1. Deepfake (Derin Sahtelik) Kavramının Temelleri

Yapay zekâ tabanlı video ve ses işleme teknolojisine dayanan deepfake videoları, yakın dönemin en çok tartışılan medya manipülasyon araçlarından biri olarak öne çıkmaktadır. Türkçe'de "derin sahtelik" olarak kullanılan, post-truth döneminin getirilerinden biri olarak da düşünülebilecek olan kavram yeni medya gazeteciliği açısından da potansiyel bir risk alanıdır. Deepfake ifadesine temel oluşturan, yapay zekâ ile manipüle edilmiş bilinen ilk amatör örnek bir Reddit kullanıcısının 2017 yılının Kasım ayında yaptığı sosyal mecra paylaşımına dayanmaktadır. "Deepfakes" takma adlı bu Reddit kullanıcısı, aralarında Gal Gadot ve Scarlett Johansson gibi tanınmış Hollywood aktrislerinin de yer aldığı bir dizi ünlü yüzün dijital manipülasyon

yoluyla pornografik içeriklerdeki bedenlere dâhil edildiği videolar yüklemiştir. Bu olaydan sonra haber medyası insan yüzlerini ve bedenlerini manipüle etmek için DÖ ya da MÖ denilen metodun kullanıldığı bu türden içerikleri tanımlarken deepfake terimini kullanmaya başlamıştır.

Ancak deepfakeler, daha büyük bir görsel-işitsel manipülasyon alanının sadece bir bileşenidir. Söz konusu görsel-işitsel manipülasyonlar, hem yapay zekâya dayalı son teknolojileri kullanarak gerçeğinden ayırt edilmesi çok zor olan ürünleri hem de üretiminde görüntüleri hızlandırma, yavaşlatma, kesme, yeniden sahneleme ya da yeniden bağlamsallaştırma gibi daha geleneksel ve basit metotların kullanıldığı "cheapfake" adı verilen "ucuz sahtelikleri" kapsamaktadır. Bu bağlamda, deepfake teknolojisi, kullanılan yöntemler ve seviye açısından bakıldığında iki gruba ayrılmaktadır. Günümüzde pek çok medya uzmanı post-truth çağın getirisi olan deepfakeleri yaklaşmakta olan bir tür bilgi kıyametinin habercisi olarak kabul etmektedir (Warzel, 2018). Dahası deepfake ürünlerinin sıklıkla seçimleri olumsuz etkileyerek demokrasi kültürüne zarar verme, ulusal güvenliği tehlikeye atma veya toplumda yer edinen yaygın şiddeti körükleme gibi olası sonuçları da masaya yatırılmaktadır (Vincent, 2018).

MÖ, büyük veri analitiği bilgisayar görüşü duyarlılığından otonom kontrol sistemlerine kadar pek çok alanda çeşitlilik gösteren hayati uygulamaları tasvir etmektedir. Fakat ne yazık ki, DÖ tekniklerinin gelişmesiyle birlikte, MÖ odaklı sistemlerin gizliliğine, gücüne ve güvenliğine yönelik riskler de paralel biçimde gelişmiştir. Deepfake, sentetik/sahte bir ortam oluşturmak için görüntüleri, akustiği ve videoyu değiştirmek için DÖ algoritmalarını kullanan bir "sahtecilik" biçimidir. Yüz görüntüleri üst üste bindirilebilir, yüz hareketleri oluşturulabilir, yüzler değiştirilebilir, yüz ifadeleri hareket ettirilebilir, yeni yüzler üretilebilir ve hedef bireyin konuşması üzerinde hedef kişinin benzer şekilde hareket ettiği bir video oluşturmak için bir konuşmacının videosunda kaynak kişi ile sentezlenebilir.

Bilgisayarların dijital görüntülerden veya videolardan nasıl bir anlam kazanabileceğiyle ilgilenen disiplinler arası bilimsel bir alanı olan "Bilgisayarlı Görü" (Computer Vision), görüntülerin işlenmesiyle ilgili karmaşık bir alandır ve PC'lere görsellerden/imajlardan bilgi elde etme becerisi sağlamaktadır. Bu teknolojinin günümüzdeki en yaygın kullanım biçimleri arasında otonom araçlar, sağlık sektöründe hastalıkların teşhis edilmesi ve fotoğraf etiketleme önerileri için Facebook tarafından yüz tespiti gibi uygulamalar gösterilebilir. Deepfake teknolojisi de Bilgisayarlı Görü şemsiye tanımı altında değerlendirilmektedir.

Son yılların popüler kelimelerinden olan deepfake kavramının temeli, 1997 yılına dayanmaktadır. Bregler vd. (1997) "Video Rewrite: Driving Visual Speech with Audio" adını taşıyan bildirileri ile bir ses çıkışından yeni bulunan yüz simülasyonları üretebilen "Video Rewrite Program" adlı çığır açıcı bir projeyi tarihe kaydetmeyi başarmıştır. Bu makaledeki fikirler, konseptin gerçeğe dönüşmesi anlamında orijinalliğini korumaktadır ve deepfake teknolojisinin alt yapısının oluşturulmasında temel çalışmalardan biri olarak kabul görmektedir. 2001'de Cootes vd. (2001) tarafından yayınlanan bir başka ünlü makalede de bir şekli bir görüntüye sığdırmak için kapsamlı bir istatistiksel prototip kullanan aktif görünüm modellerinde (AAM) algoritması yayınlanmış, yüz eşleştirme ve izleme alanında literatüre önemli bir katkı sağlanmıştır.

Theis vd., 2016 yılında "Face2Face" adını verdikleri yöntemi ortaya koymuşlardır. Bu sistemde hedef videonun ağız bölgesi bir aktörle değiştirilerek anlık bir animasyon yaratılmaya çalışılıyor ve bu video ses içermiyordu (Thies vd., 2019). Benzer şekilde 2017 yılında Suwajanakorn vd. eski Amerika Birleşik Devletleri (ABD) başkanı Barack Obama'yı sentezledikleri çalışmalarında grafik iyileştirmelerini daha fazla animasyon, doku ve ifadeyle geliştirmişlerdir. Her iki çalışmanın amacı farklı olsa da bir yandan fotogerçekçi görünüm için grafik uyumluluğu yenilenirken öte yandan literatüre eklenen bu görüşler ve denemeler işleme ve dönüştürme sürelerini iyileştirmiştir. Sözü edilen bilimsel yayınlar deepfakelerin geliştirilmesinde gerçek birer kilometre taşı olmuşlardır (Westerlund, 2019).

Bir Reddit kullanıcısı tarafından kelimenin popüler terminolojiye taşındığı 2017 yılından bu yana dayandığı teknoloji giderek gelişen deepfake terimi, "sentetik medya uygulamalarını" ve StyleGAN (gerçek görünen ama var olmayan insan görüntüleri) gibi yenilikçi kreasyonları içerecek şekilde anlam genişlemesine maruz kalmıştır. Terimin kapsamı son zamanlarda manipüle edilmiş kayıtlı konuşmalara kadar uzanmıştır (Thies vd., 2020; Suwajanakorn vd., 2017).

Son zamanlarda ise yapay zekâ ve bilgisayar görüşündeki ilerlemeler ve tüm bu gelişmelerin basit bir akıllı telefon uygulamasına kadar indirgenmiş olması deepfake manipülasyonları ile istismarların üretiminin ve dağıtımının kolay hâle gelmesi sonucunu doğurmuştur. Materyallerin çekişmeli üretici ağ (Generative Adversarial Network – GAN) tekniği kullanılarak kaynak imajlar ve videoların üzerine diğer imaj ve videoların bindirilmesi yoluyla üretilmesi, istismarların ünlü isimlerden sıradan bireylere doğru yayılmasına neden olmuştur (Vaccari ve Chadwick, 2020). Bu yöndeki genel eğilimin

de çoğunlukla pornografik ve politik deepfake üretme yönünde olduğu görülmektedir.

2. Rakamlara Yansıyan Tehlike ve Somut Örnekler

Geçtiğimiz yıllarda deepfake teknolojisinin MÖ ile manipüle edilmiş CEO seslerinin ABD'li şirketlerden para çalmak amacıyla kullanıldığı vakaların olduğu medyaya yansımıştır. 2019 yılında bir hacker grubu Almanya merkezli bir şirketin CEO'sunun sesini yapay zekâ tabanlı bir yazılımla taklit ederek kendi hesaplarına yüklü bir ödeme yapılmasını sağlamışlardır. 2018'de hasta olduğu açıklanan Afrika ülkesi Gabon'un Devlet Başkanı Ali Bongo'nun öldüğü yönünde kuşkular doğuran uzun süren sessizliği bir video ile bozulmuş; Ocak 2019'da geleneksel yeni yıl konuşmasıyla halkın karşısına çıkan Bongo'nun videosu üzerinde deepfake tartışmalarının yapılması öldüğü iddialarını daha da güçlendirmiştir. Hükûmete duyulan güveni zedeleyen bu tartışmalardan tam bir hafta sonra ordu videoyu gerekçe göstererek başarısız bir darbe girişimi başlatmıştır. İncelemeler sonucunda videoda manipülasyon olduğu sonucuna ulaşılamamıştır. Fakat kuşku ve deepfake yaftası bu örnekte demokrasiye galip gelmiştir. Bunlara benzer vakaların artış göstermesi sonucu, hükûmetler ve çeşitli kuruluşlar deepfake tehdidini fark ederek bu yalan fırtınasını engelleyebilmek için bir dizi önlemi hayata geçirmişlerdir.

2017'de deepfake ürünlerinin görünür hâle gelmesinden bu yana, kötü amaçlı yazılımlarda 1990'lardaki ilk günlere benzer şekilde patlayıcı bir büyümeye tanık olunmuştur. 2019'dan bu yana, çevrim içi deepfake sayısı 14.678'den 145.227'ye yükselmiştir. Bu da yıllık yaklaşık %900'lük şaşırtıcı bir büyüme anlamına gelmektedir. Bunların çoğu popüler sosyal medya platformlarında dolaşıma girmiştir ve aralarında 6 milyara yakın izlenme toplayan içerikler mevcuttur. Bu üstel büyüme, yapay zekâ algoritmalarındaki gelişmeler ve daha düşük bilgi işlem maliyeti ile web'deki verilerin katlanarak büyümesine yol açmıştır. Mobil uygulamalara kadar genişleyen deepfake kültürü, ucuza, düşük bariyerde ve ölçekte oluşturulabilen derin sahtekârlıkların da önünü açmıştır. Basit ürünleştirilmiş çözümler artık sezgisel kullanıcı arabirimleriyle birlikte sunulmakta ve yalnızca tek bir görüntü kullanılarak derin sahtelik ürünleri oluşturulabilmektedir. Bu tür bir teknolojinin kötü niyetli kullanımı, Forrester'ın deepfake dolandırıcılığının 2020'de şimdiden 250 milyon dolara ulaşacağını tahmin etmesine neden olmuştur (Pollard, 2019).

Derin sahtekârlıkları izlemek ve tespit etmek için DÖ teknolojileri geliştiren Amsterdam merkezli bir şirket olan Sensity (eski adıyla Deeptrace)

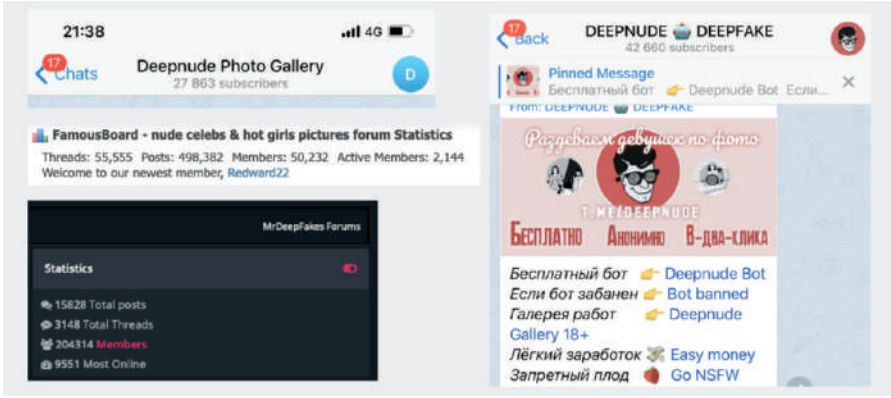
tarafından 2019 yılında yayınlanan rapora göre internette toplam 14.678 adet deepfake videosu bulunmaktaydı. Ayrıca elde edilen bulgular, çevrim içi derin sahtelik videolarının toplam sayısının hızla arttığını ortaya koymaktadır. Bu ölçüm, Aralık 2018'de yapılan önceki ölçüme (7.964 adet) göre neredeyse %100'lük bir artışı temsil etmekteydi. Ayrıca bu ürünlerin %96'sının pornografik içerikler taşıdığı görülmektedir. Taramalarda ilk dört özel deepfake pornografi sitelerindeki toplam video görüntüleme sayısı ise 134,364,438 olarak çarpıcı boyuttadır. Şirket, web'deki deepfake pornografinin çevrim içi varlığını, yüklenen videoların sayısını ve bu videoları barındıran web sitelerinin sayısını izleyerek ölçmüştür (Ajder vd., 2019: 1). Araştırma, deepfake pornografinin birkaç farklı web sitesinde önemli bir çevrim içi varlığa sahip olduğunu ortaya çıkarmıştır. Bu web siteleri özel deepfake pornografi web siteleri ve ana akım pornografi web siteleri olmak üzere iki kategoriye ayrılmıştır. Yine aynı raporda deepfake pornografi ekosisteminin neredeyse tamamen, keşfedilmiş toplam videoların 13.254'ünü barındıran özel deepfake pornografi web siteleri tarafından desteklendiğini ortaya koymuştur. Buna karşılık, raporun yayınlandığı tarihte ana akım pornografi web siteleri yalnızca 802 video barındırmaktadır. Bu özel deepfake pornografi web sitelerinin giderek artan sayısı, deepfake pornografinin büyüyen bir iş fırsatını temsil ettiği de raporda ifade edilmiştir (Ajder vd., 2019: 6).

Bu noktaya nasıl gelindiğine dair başka bir veri daha bulunmaktadır. Deepfake oluşturma toplulukları ve forumları, deepfakelere ve deepfake oluşturma yazılımlarına artan erişilebilirliğin arkasındaki temel itici güç olarak karşımızdadır. Bu kreasyon topluluklarının ve forumlarının çoğu, deepfake oluşturmakla ilgilenen kişiler için bir başlangıç noktası gibidir ve daha deneyimli içerik oluşturucular arasında iş birliğini kolaylaştırmaktadır. Sensity'nin raporuna göre yaratıcı toplulukların ve forumların çoğu derin sahte pornografi web siteleri ile Reddit, 4chan, 8chan ve Voat gibi forum tabanlı web sitelerinde bulunmaktadır. 4chan ve 8chan gibi web sitelerinden bazılarının yasa dışı ve etik olmayan faaliyetlere ev sahipliği yapmakla ünlü olduklarını hatırlatmakta fayda vardır. Tüm bu yapıyı düşündüğümüzde yaklaşık 100 bin üye sayısından söz edilmektedir (Ajder vd., 2019: 4). Konu deepfake olduğunda masada duran en bariz tehdidin rıza dışı/sahte pornografi ile kadınlara yönelik olduğu görülmektedir. Genellikle ünlüleri hedef alan bu sahte ürünler zaman içerisinde intikam pornografisine kadar eğilim gösterir duruma gelmiştir. Hollywood'da hayatını kaybeden oyuncuların suretlerini yeni filmlerde kullanabilmek amacıyla kullanılması da deepfake teknolojisi sayesinde olmuştur. Fakat yapay zekânın gelişimiyle birlikte amatör ve düşük bütçeli versiyonları mobil cihazların uygulama

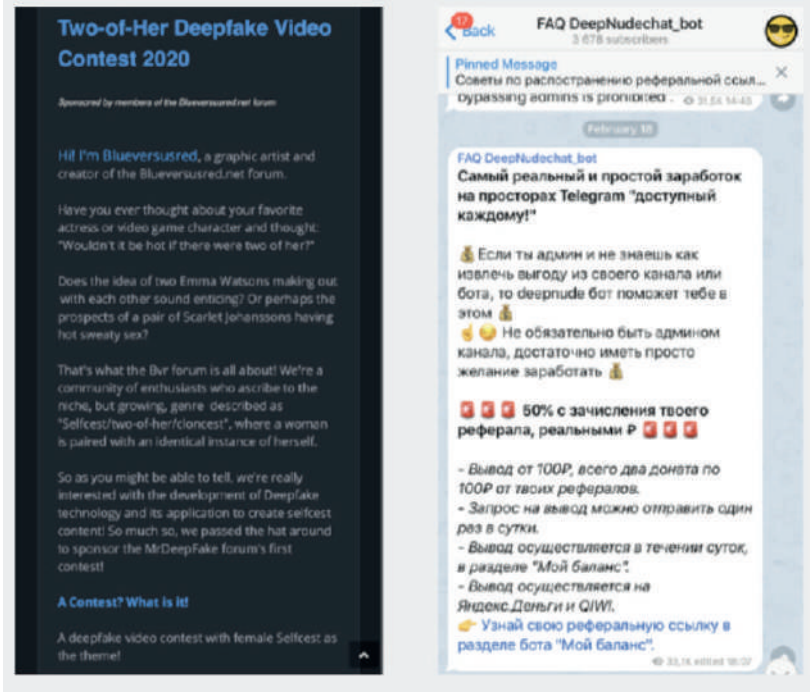
mağazalarına dâhil olması işleri karmaşık bir hâle getirmiştir. 2017 yılında aralarında Michelle Obama, Ivanka Trump ve Emma Watson'ın da olduğu pek çok ünlü kadın yüzü, yetişkin filmlerinde kullanılmış ve bu teknolojinin karanlık yüzü ortaya çıkmıştır. Ayrıca deepfake kültürünün yükselişine dair endişeler kimlik sahteciliği, diplomatik itibar suikastleri ve demokrasiye yönelik olarak çeşitlilik göstermektedir.

Deepfake tehdidinin ulaştığı son boyutu raporlayan bir diğer şirket ise Sentinel'dir. Son teknoloji bir yapay zekâ tespit platformu geliştirerek demokrasileri dezenformasyon kampanyalarından, sentetik medyadan ve bilgi operasyonlarından korumaya yardımcı olmak için hükûmetler, uluslararası medya kuruluşları ve savunma kurumlarıyla birlikte çalışan şirketin merkezi e-devlet, dijital kimlik ve siber güvenlikte dünya lideri olan Estonya'da yer almaktadır. Eski NATO yapay zekâ ve siber güvenlik uzmanları tarafından kurulan Sentinel; Jaan Tallinn (Skype'ın kurucu ortağı ve DeepMind'ın ilk yatırımcısı) ve Taavet Hinrikus (Skype'ın kurucu ortağı) dâhil olmak üzere birinci sınıf yatırımcılar tarafından desteklenmektedir. Kritik bilgi tedarik zincirini doğrulayarak 1 milyar insanı bilgi savaşından koruyarak internet için güven katmanı olmayı amaçlayan şirket, 2020 yılında yayınladığı kapsamlı bir raporla durumun ciddiyetini gözler önüne sermiştir (Tammekänd vd., 2020: 2).

Sentinel tarafından yayınlanan 2020 yılına ait kapsamlı rapor, çalışmanın bir diğer veri kaynağını teşkil etmektedir. Dünyada benzer çalışmalar yürüten kısıtlı kaynaklar olduğundan Sentinel'in raporu bu anlamda en güncel ve güçlü verileri barındırmaktadır. Rapor hazırlanırken 30'dan fazla porno içerikli web sitesi analize tabi tutulmuş ve oldukça büyük bir pornografik deepfake arşivi ile karşılaşılmıştır. Raporda web'de çevrim içi pornografik deepfake sayısı 27.271 olarak belirtilirken bunların çoğunda ünlülerin ya da nüfuzlu kişilerin kullanıldığı açıklanmıştır. Pornografik deepfakelerin %69'unun özel deepfake porno sitelerinde barındırıldığına da vurgu yapılmıştır. Sentinel'in raporunda bahsedilen çarpıcı bir ayrıntı daha vardır. Özel pornografik deepfake sitelerine ek olarak, deepfake porno yaratıcılarından oluşan yer altı toplulukların da varlığından bahsedilmektedir (Tammekänd vd., 2020: 12).



Fotoğraf 1. Pornografik deepfakelerin satın alınabileceği veya paylaşılabileceği ve taleplerin iletilebileceği forumlardan ve Telegram sohbet gruplarından örnek ekran görüntüleri.



Fotoğraf 2. Deepfake pornografisi konusunda yarışma yürüten bir forum (LHS görüntüsü) ve topluluk tabanını büyütmek için parasal bir ödülle bir tavsiye programının reklamını yapan Telegram sohbet grubunun (RHS görüntüsü) ekran görüntüsü.

Sentinel'in raporuna göre yukarıda örneklendirilen bu toplulukların toplam 100 binden fazla üyesi vardır ve pornografik derin sahtekârlıkların geliştirilmesiyle uğraşmaktadırlar. Şifreli Telegram sohbet grupları da ticaretle uğraşmaktadır. Bu yapıda herkes özel pornografik deepfakeler geliştirmek için para ödeyebilmektedir. Hatta bir grubun topluluklarını büyütmek için para kazandıran bir tavsiye programı bile mevcuttur. Ayrıca forumlar, en iyi pornografik derin sahtekârlığı kimin oluşturabileceğini belirlemek için ödüllü yarışmalara aktif olarak katılmaktadır.

Tehditler bununla sınırlı değildir. Kadınlar bu karanlık dünyanın en büyük mağdurları konumundadır ve bu kadınların her zaman ünlü bir isim taşımasına gerek yoktur. Pornografik deepfakelerde kullanan kadınlara yönelik şantaj ve fidye sahtekârlığı saldırılarında da artış yaşanmaktadır. Pornografik deepfakeler oluşturmaya adanmış sitelerine ev sahipliği yapan karanlık web ile kötü amaçlı yazılım pazarlarının erken dönemleri arasında paralellik kurulabilmektedir. Uzun vadede bu tehdit web'de en az bir yüz fotoğrafı olan hemen hemen herkesin pornografik deepfakelerin kurbanı olabileceği ihtimalini akıllara getirmektedir. Araştırma bulguları, pornografik derin sahtekârlıkların kurbanları üzerinde kalıcı olumsuz etkilere sahip olduğunu göstermektedir. En savunmasız durumda olanlar kadınladır. Mağdurlar, videoların sahte olduğunu bilmelerine rağmen travma yaşayabilmektedirler. Rana Eyyub ile Noelle Martin örneklerinde olduğu gibi iş olanakları ya da kişisel güvenlik tehdit edilebilmektedir (Delfino, 2019; India Today, 2018; ABC Australia, 2018).

"Araştırmacı bir gazeteci olarak beni itibarsızlaştırmaya yönelik çabalar bir yanlış bilgilendirme kampanyasıyla başladı. Sonra yüzüm bir porno videoya dönüştürüldü". Bu sözler gazeteci Rana Ayyub'a ait. The Huffington Post'ta başından geçenleri kaleme aldığı "I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me" başlıklı makalesinde sosyal medyada çok fazla nefret söylemine maruz kaldığından bahsetmektedir. Kendisine bunun yalnızca çevrim içi nefret olduğunu ve asla çevrim dışı tacize dönüşmeyeceğini söyleyerek bunu her zaman görmezden gelmeye çalıştığını dile getiren Ayyub 2018'in Nisan ayında durumun değiştiğini anlatmaktadır. 8 yaşındaki Keşmirli bir kıza tecavüz vakasının ardından BBC ve Al Jazeera'de Hindistan'ın çocuklara cinsel tacizde bulunanları koruyarak nasıl kendi kendini utandırdığına dair bir konuşma yapmıştır. Ertesi gün sosyal medyada kendisine ait olduğu iddia edilen bir dizi sahte tweet'in dolaşmaya başladığını görmüştür. "Hindistan'dan nefret ediyorum", "Hintlilerden nefret ediyorum", "Pakistan'ı seviyorum" diye başlayan sahte tweet'ler "Çocuk tecavüzcülerini seviyorum ve bunu İslam adına yapıyorlarsa destekliyorum" gibi oldukça tehlikeli bir yöne kaymıştır.

Üstelik bu ekran görüntüleri, doğrulanmış hesap anlamına gelen mavi tik taşımaktadır. Ayyub, tweet'lerin sahte olduğunu netleştirmek ve insanlara kanmamalarını söylemek için gerçek hesabına bir açıklama yazmak zorunda kalmıştır. Fakat ertesi gün taciz ve istismar farklı bir boyut kazanacaktır. İktidardaki siyasi partiden bir kaynak "WhatsApp'ta bir şeyler dönüyor, sana göndereceğim ama üzülmeyeceğine söz ver" diyen bir mesaj göndermiştir. Gönderilen içerik bir deepfake pornodur ve yüzü kullanılan kadın kendisidir. Ayyub, videoyu ilk açtığında yüzünü görünce şoke olmuştur ama kısa bir an sonra aslında kadının kendisi olmadığını fark etmiştir. Hindistan gibi bir ülkede bunun çok önemli olduğunu söyleyen Ayyub, derin sahteliğin siyasi çevrelerde neden dolaştığını sorduğunda videonun parti içindeki insanlar tarafından yayıldığı cevabını almıştır. Sosyal medyadan yüzlerce taciz mesajları almaya başlayan Ayyub, videonun 40 binden fazla paylaşıldığını söylemiştir. Daha sonra Twitter'da videonun ekran görüntüsünün yanında Ayyub'un numarası da ifşa edilmiştir. Bu noktadan sonra Ayyub, WhatsApp üzerinden yüzlerce taciz mesajı ve parayla seks teklifi almaya başlamıştır. Günün sonunda üzüntüden ve şoktan hastaneye kaldırılan Ayyub, sonunda davasını üstlenmeyi kabul eden yüksek profilli bir feminist avukatla temasa geçmeyi başarmıştır. Fakat karakola gittiklerinde polis tutanak tutmamıştır, çünkü videoyu paylaşan kişiler siyasi isimlerden oluşmaktadır. Dahası polisler talihsiz gazetecinin önünde videoyu izleyerek sırıtmışlardır. Polisler ancak olayların basına yansıtılması tehdidinden sonra işlem yapmayı seçmişlerdir. Sonrasında bir sulh hâkimine ifade veren ve kanıtları gösteren Ayyub, uzun bir bekleyişten sonra Birleşmiş Milletler'in (BM) olaya müdahale etmesiyle biraz olsun rahatlamıştır. Baskı altındaki hükûmet kendi itibarını korumak adına tacizin boyutunu yavaşlatmıştır. Ayyub, makalesini şu sözlerle noktalamaktadır: "İronik olan şu ki, videonun yayınlanmasından yaklaşık bir hafta önce bir editörün Hindistan'da deepfake tehlikesinden bahsettiğini duydum. Ne olduğunu bile bilmiyordum, bu yüzden Google'da arattım. Bir hafta sonra başıma geldi. Uzun bir süre bunun hakkında konuşmadım çünkü daha geniş kitlelerin benimle empati kurmayacağından veya bana sempati duymayacağından, ancak onu daha fazla keşfetmek isteyeceklerinden endişelendim. Deepfake ürünlerinin bu tür bir popülerlik kazanmasını istemedim. Ama ne yazık ki, son birkaç hafta içinde çok yüksek profilli kadın film yıldızlarının sayısız derin sahte videosunu gördüm, bu yüzden bunu önlemek için çok geç gibi geliyor. Bu çok, çok tehlikeli bir araç ve onunla nereye gittiğimizi bilmiyorum." (Ayyub, 2018).

Fotoğraf 3. Deepfake pornografisinin kurbanı olan gazeteci Rana Eyyub, sahte içeriğin viralliğine karşı savaşmak için Twitter'dan seslenme yolunu seçmiştir.

Noelle Martin de kadınlara yönelik deepfake şiddetinin sembol isimlerinden biri olmuştur. Fotoğraflarının sosyal medyasından çalındığını ve müstehcen pornoya dönüştürüldüğünü öğrendiğinde âdeta şoke olmuştur. Daha önce hiç erkek arkadaşı olmayan, herhangi bir mahrem görüntü kaydı olmayan Martin olaydan sonra dizüstü bilgisayarını polise götürmüştür. Fakat o zamanlar gerçek veya sahte mahrem görüntülerin yayılmasını engelleyen hiçbir yasa yoktur. Dolayısıyla polisin veya herhangi bir devlet kurumunun yapabileceği hiçbir şey yoktur. Martin, bu nedenle görüntülerin yer aldığı her sitenin web yöneticisine e-posta gönderip içeriklerin kaldırılmasını talep etmek gibi zor bir göreve başlamıştır. Ancak giderek daha açık hâle gelen görüntüler giderek yayılmıştır ve tacizin boyutu daha da kötüleşmiştir. Fail, Martin'in pornografik deepfake videolarından birini oluşturmak için onun 17 yaşındayken çekilmiş bir fotoğrafını kullanmıştır. Şu anda 27 yaşında bir avukat olan Martin, web sitelerine yayından kaldırma emri vermenin yeterli olmadığını söylemektedir: "Bir görüntü bir kerede, tek bir lokasyonda, tek bir web sitesinden kaldırılabilir; ancak bu, o içeriğin tamamen kaldırıldığı anlamına gelmez. Çünkü yine de yeniden ortaya çıkabilir, çoğalabilir ve araması kurbana bırakılır." (Mills, 2021).

Fotoğraf 4. Deepfake kurbanı olan Noelle Martin, deepfake pornografisi ve daha geniş görüntü tabanlı tacizle mücadelenin savunucusuna dönüşmüş durumdadır.

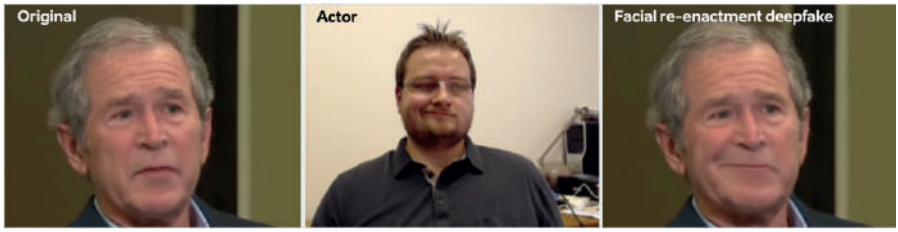
Sosyal medya hesaplarından kadınların görüntülerini bulmak için internette gezinen, derin sahte pornolar oluşturan ve ardından onlara karşı fidye saldırıları gerçekleştiren otomatik botların varlığı bilinmektedir. Bu tür şantaj ve fidye saldırıları, faillerin kurbanlarının pornografik deepfakelerini halka açıklama tehdidiyle kripto para birimlerinde ödeme talep etmesiyle birlikte artış göstermiştir. ABD'de yapılan bazı önemli araştırmalar ilginç sonuçlar ortaya koymuştur. Amerikalıların 10'da 9'u derin sahtekârlıkların yarardan çok zarar verebileceğine inanmaktadır (Allen, 2019). Her 10 kişiden 7'si sahte haberlerin devlet kurumlarına olan güvenlerini etkilediğine inanmaktadır (Mitchell vd., 2019). Ayrıca, ABD hükûmeti tarafından 2019'da derin sahtekârlıkların üstesinden gelmek için 10'dan fazla girişim başlatılmış ve yasalar çıkarılmıştır (Tammekänd vd., 2020). Facebook, Amazon ve Microsoft tarafından başlatılan deepfake algılama mücadelesi çalışmaları için 10 milyon dolardan fazla bütçe ayrılmıştır (Meta AI, 2020).

3. Deepfake Türleri

DÖ algoritmalarının kullanılmasıyla gelişmiş deepfake ürünleri çıplak gözle ve kulakla ayırt edilememektedir. Ancak algılama teknolojisi ile manipüle edilmiş materyalin gelişmiş biçimleri algılanıp gerçek dışı olarak etiketlenebilmektedir (Rössler vd., 2019). Deepfake teknolojisi o kadar gelişmiştir ki artık farklı türlerinden söz etmek mümkündür.

3.1. Yeniden Yüz Canlandırma (Facial Re-Enactment)

Yeniden yüz canlandırma, bir videoda hedef öznenin yüz ifadelerinin bir kaynak aktörden sağlanan girdiye dayalı olarak manipüle edilmesini içermektedir. İlk adımda hedef öznenin yüz özellikleri yapay zekâ tarafından öğrenilmektedir. İkinci adımda oyuncu, yüz ifadelerinin girdisini sağlamaktadır. Son adımda ise aktörün girdisi sentetik hareketler oluşturmak için kullanılmaktadır. Bu teknik, beraberinde başka bir teknolojiyi daha gündeme taşımıştır. Nöral Ses Kuklası (Neural Voice Puppetry) tekniği, bir kaynak kişinin veya dijital asistanın sesi verildiğinde hedef kişinin fotogerçekçi bir çıktı videosunun üretilebildiği son teknoloji bir yöntemdir (Thies vd., 2020). Hedef aktörün videolarını, kaynağı bilinmeyen herhangi bir aktörün sesiyle veya metinden konuşmaya üretilebilen sentetik seslerle sentezlemeye olanak tanımaktadır.



Fotoğraf 5. Eski ABD Başkanı George W. Bush'un bir videosunda, daha önce kaydedilmiş bir röportajdan alınan yüz ifadeleri, bir aktörün ifadesi kullanılarak değiştirilmiştir.

3.2. Yüz Değiştirme (Face Swapping)

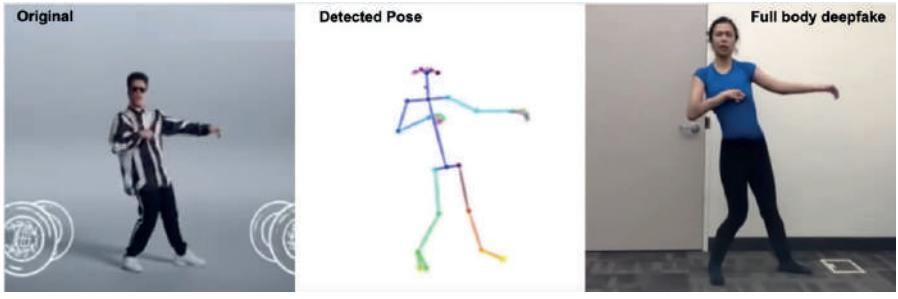
Yüz değiştirme, bir kaynak öznenin yüzünün hedef özne yüzüyle değiştirilmesini içeren tekniktir. Deepfake teknolojisi sayesinde, bir hedef videodaki özne tamamen kaynak özne olarak hareket edebilmektedir. Bu yöntemin ilk adımında önce kaynak yüze ait özellikler daha sonra hedefin yüz özellikleri yapay zekâ modeli tarafından öğrenilmektedir. Son aşamada ise kaynak yüz özellikleri hedefin üzerine bindirilerek manipüle edilmektedir.

Fotoğraf 6. Yüzü aktris Kate McKinnon ile değiştirilen politikacı Elizabeth Warren'ın kullanıldığı yüz değiştirme videosundan bir kesit.

Bu yöntemin en popüler kullanımı, yukarıda görüldüğü gibi bir kişinin (orijinal özne) yüzünün hedef bir kişiye gerçekçi bir şekilde bindirilmesi ile ortaya çıkmaktadır. Bu tür bir derin sahtekârlığın geliştirilebilmesi sınırlı miktarda yapay zekâ eğitim verisi ve son kullanıcı düzeyinde teknoloji ile mümkün hâle gelmiştir. Bu eylem aynı zamanda bir videodaki kaynak oyuncuya, tamamen sentetik ve hiper gerçekçi eylemlerle hedef özne olarak bir şeyler söyletilip yaptırılabileceği anlamına gelmektedir (Agarwal vd., 2019).

3.3. Tam Vücut Değiştirme (Full Body Swapping)

Tam vücut derin sahtelik yöntemi tıpkı diğer örneklerde olduğu gibi yüz değiştirme ve yüz canlandırma tekniklerini içermektedir. Fakat bu tekniğin ayrıca yapay zekâ teknolojisinin bedensel eylemleri de gerçekleştirebilmesi adına tüm vücudun sentetik olarak oluşturulması yönünde genişlediği görülmektedir. Bu tekniğin ana çalışma prensibi de üç aşamadan oluşmaktadır. Öncelikle yapay zekâ modeline kaynağın ve hedefin vücut özellikleri öğretilmektedir. Daha sonra hedef ve kaynak gövde öğeleri eşleştirilerek uyum sağlanması amaçlanmaktadır. Son olarak hedef, manipüle edilmiş vücut unsurları ile yapay zekâ tarafından yeniden oluşturulmaktadır. Dolayısıyla bu teknikte konuşan yüzlerin yanında hareket eden vücutlar söz konusudur.



Fotoğraf 7. Hareketin saptanması (ortadaki kare) ve hedefin gövdesinin yeniden oluşturulması (sağdaki kare) aşamasında sanatçı Bruno Mars'ın erişime açık olan orijinal müzik videosunu (soldaki kare) kullanan bir tam vücut deepfake örneği.

Tam vücut deepfakelerinde de bir vücut tarafından gerçekleştirilen çeşitli pozları oluşturan unsurların öğrenilmesi için yapay zekâ kullanılmaktadır. Bu pozların bir dizisi, daha sonra hedefin sentetik olarak oluşturulmuş gövdesine uygulanan hareketi oluşturmaktadır. Bu tekniğin çıktıları insanların yüz manipülasyonlarının ötesinde çeşitli sahte eylemler gerçekleştirmeleri için tamamen sentetik olarak üretilmesinin mümkün olduğu anlamına gelmektedir. Bu metodun dayandığı teknoloji henüz ileri seviyede değildir; dolayısıyla tam vücut deepfakelerini oluşturmak diğer ürünlerden daha zordur (Chan vd., 2019).

3.4. Ses Deepfakeleri (Audio Deepfakes)

Yapay zekâ kullanılarak seslerin sentetikleştirildiği ve gerçeğinden ayırt edilemeyecek biçimde sentezlenerek yeniden üretildiği bu yöntemde hedefe ait orijinal ses parçaları sahte ürünlere temel oluşturmaktadır. Bu orijinal kayıtlar genellikle basın röportajlarından ve çeşitli kamusal açıklamalardan oluşmaktadır. Bu alanda çalışan Descript, Resemble.ai ve Baidu gibi şirketler, sesleri klonlayabilen ve sentezleyebilen ürünleri pazara sunmuştur.

Fotoğraf 8. Ses deepfakelerinin üretilmesinde kullanılan ses klonlama teknolojileri sunan bazı özel şirketler.

Pazardaki teknolojilere bakıldığında öne çıkan şirketler arasında yer alan Descript, basit bir transkripti düzenleyerek herkesin mevcut herhangi bir sesi manipüle etmesine izin vererek yeni bir versiyon üretmede kullanılabilen ücretsiz, ürün hâline getirilmiş bir hizmet geliştirmiştir. Benzer şekilde Baidu adlı şirket, Deep Voice 3 adını verdiği hizmeti ile bir ses örneğine dayalı olarak hedef bir sesin yalnızca 3.7 saniye gibi kısa bir süre içerisinde klonlanabilmesini mümkün kılmıştır (Cole, 2018; Wiggers, 2019). Bu durum dramatik derecede olumsuz ve tehlikeli sonuçlar yaratabilecek güçtedir. Sentetik olarak yeniden üretilen ses deepfakeleri, duyulanın yapay zekâya mı yoksa gerçek bir insana mı ait olduğu konusunda algıları körelten bir inandırıcılığa sahiptir. Bu bağlamda sahte telefon görüşmelerinin ya

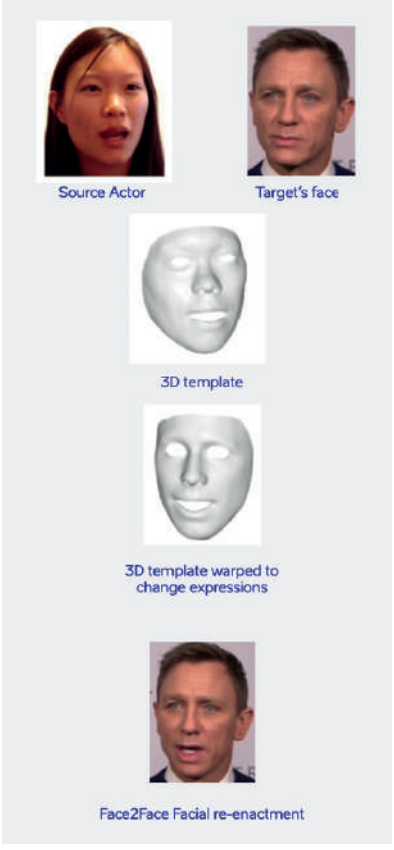
da sesli mesajların yaratabileceği hasarı tahmin etmek güç değildir (Statt, 2019). Yaşanan gerçek olaylar bu tehlikeyi doğrular niteliktedir. 2019 yılında siber güvenlik firması Symantec, ses tabanlı derin sahtelik dolandırıcılığı içeren ve milyon doları aşan kayıpla sonuçlanan yaşanmış olaylara dikkat çekmektedir (Stupp, 2019). 2019 yılı Mart ayında yaşanan olayda suçlular bir şirketin CEO'sunun sesini taklit etmek için yapay zekâ tabanlı yazılım kullanarak sahtekârlıkla hesaplarına 220 bin Euro transfer talep etmişlerdir. İngiltere merkezli bir enerji firmasının CEO'su, kendisinden fonları bir Macar tedarikçiye göndermesini isteyen firmanın Alman ana şirketinin genel müdürü olan patronuyla telefonda konuştuğunu sanmıştır. Şirketin sigorta şirketi Euler Hermes Group SA'ya göre arayan kişi, talebin acil olduğunu söyleyerek yöneticiyi bir saat içinde ödemeye yönlendirmiş; fakat Euler Hermes, kurban şirketlerin isimlerini vermeyi reddetmiştir. Avrupa'daki bu ses sahtekârlığı saldırısı, suçluların açıkça yapay zekâdan yararlandığı ilk siber suç olmuştur.

3.5. Sentetik İnsanlara Dayalı Deepfakeler

Bu yöntemde çekişmeli üretici ağlar (GAN) kullanılarak tamamen sentetik insan yüzleri oluşturulabilmektedir. Bu şekilde üretilen derin sahtelik ürünleri, diğer metodların ürünleriyle karşılaştırıldığında sahteciler için daha görünmez bir yol sunmaktadır. Böylece sahtelik için bir koruma katmanı eklenmiş olmaktadır. Çünkü her görüntü benzersizdir ve ters görüntü aramasıyla dahi bir kaynağa kadar izlenememektedir. GAN tarafından üretilmiş olan sentetik insanlar ve derin sahtekârlık teknikleri, sentetik insanlara dayalı derin sahtekârlıklar oluşturmak için birleştirilebilmektedir. Bu teknik, var olmayan insanların görüntülerini ve videolarını yapay olarak oluşturmayı mümkün kılmaktadır. Yapay zekâ bilgi işlem sistemleri konusunda önde gelen şirketlerden olan Kaliforniya merkezli Nvidia, rakip ağlara dayanan stil tabanlı bir araca ilişkin en son sürümü yayınlamış ve sınırlı eğitim verileriyle sentetik olarak yeni insan yüzleri oluşturmak kolaylaşmıştır (Karras vd., 2020). Fakat yakın tarihte yapılmış olan bir araştırma çarpıcı bir gerçeği gündeme taşımıştır. Imperial College of London'da yapılan araştırmaya göre artık sadece tek bir imaja/görsele dayanarak sentetik video üretmek mümkün hâle gelmiştir. Deepfake ürünleri sesle eşleştirildiğinde, tarihi figürler ya da kaynak görüntüler açısından arşiv video içeriğe erişimin sınırlı olduğu durumlarda bu yöntemin faydalı olacağı gerçektir. Elbette bu teknolojinin beraberinde getirdiği tehlikenin boyutları da aynı ölçüde genişleyecektir.

3.6. Face2Face

Face2Face, video akışlarında yüz hareketlerini değiştirebilen gelişmiş bir gerçek zamanlı yüz canlandırma sistemidir. Bu sistemde çıktılar oluşturulurken üç boyutlu (3B) model rekonstrüksiyon ve görüntü tabanlı işleme teknikleri birleştirilmektedir. Face2Face tekniğinin çalışma prensibi yüz işaretlerini izleyerek hem kaynağın hem de hedefin yüzünün 3B modelini oluşturmaya dayanmaktadır. Videonun süresi boyunca hem kaynak hem de hedef için ifadeyi değiştiren dudak hareketleri, kaşlar, yüzdeki en uzak açılar gibi belirli noktalar izlenmektedir. İfadelerin model üzerinde oluşturulması için takip edilen yer işaretleri daha sonra 3B modele uygulanmaktadır. Daha sonra hedefin yüzünü 3B model üzerinde yeniden oluşturmak için kaynak oyuncunun uygulanmış ifadeleri kullanılmaktadır (Thies vd., 2019).



Fotoğraf 9. Face2Face tekniği ile kaynak aktörlerin 3B şablonunun değiştirildiği ve ifade manipülasyonunu gerçekleştirmek için hedefin yüzüne yansıtıldığı yüz canlandırması örneği.

4. Cheepfake / Shallowfakes (Ucuz Sahtelik)

Deepfake teknolojisi günümüzde tedirgin edici bir boyuta ulaşmıştır. Fakat her derin sahtelik ürünü yüksek fotogerçekçiliğe sahip değildir. Teknoloji ve uygulamalar açısından değerlendirildiğinde deepfakelerin daha basit ve ucuz denilebilecek versiyonları olan "cheapfake" ya da "shallowfake"ler (ucuz sahtelik) her ne kadar kusursuz ürünler olmasa da inandırıcılık konusunda tehlikelidir. Üstelik deepfakelere göre üretimleri daha hızlı ve ucuzdur. Cheapfake ürünlerinde yapay zekâ teknolojisinin nadiren kullanıldığı görülmektedir. Çünkü kullanımında ustalık ya da ileri teknik beceri gerektirmeyen basit video düzenleme uygulamaları kullanılarak bu tür sahte ürünler oluşturmak mümkündür. Cheapfake, görece basit maskeleme tekniği ile hazırlanmış görüntülerden oluşmaktadır. Bu tip videoların sahte olduğunu anlamak için herhangi bir uzmanlık gerekmeksizin ilk bakışta kolaylıkla anlaşılabilir olmaları en büyük özellikleridir.

Ucuz sahteler veya sığ sahteler, kolayca bulunabilen araçlar kullanılarak üretilen sahte medyalardır. Sahtecilik karşıtı araştırmacılara göre cheapfake'ler, ucuz ve oluşturması basit olan bir fabrikasyon medya sınıfıdır. Genellikle fotoğraf kolajlarını içermektedir. Birleştirilmiş, yavaşlatılmış, hızlandırılmış veya başka herhangi bir şekilde düzenlenmiş kötü amaçlı videolar da bu grupta değerlendirilmektedir. Yeniden bağlamsallaştırma cheapfake'ler için yaygın olarak kullanılan başka bir tekniktir. Adobe Photoshop, Movie Maker, Audacity ve evde kullanıma yönelik diğer yazılımlar bunları üretmek için kullanılabilmektedir. Bununla birlikte, bazı durumlarda FaceApp veya diğer basit ve ucuz uygulamalar kullanılarak bir cheapfake oluşturulabilmekte ve bir yüz tanıma sistemini kandırmak için kullanılabilmektedir.

Cheapfake ürünlerini geliştirmenin bazı yolları vardır. Anlatımı veya bağlamı değiştirmek için içerik kaldırılabilir. Örneğin, videolardaki çerçeveler ya da klipler yeniden düzenlenmektedir. Bir diğer yöntemde içerik, anlatıyı ya da bağlamı değiştirmek için manipüle edilmiştir. Örneğin videodaki kareler veya klipler hızlandırılmış ya da yavaşlatılmıştır. Bir diğer yöntemde anlatıyı veya bağlamı değiştirmek için içerik eklenmiştir. Bu durumda bir videoya çerçeveler veya klipler eklenebilmektedir.

Sosyal medyada viral olan ve dünya kamuoyunun dikkatini çeken bazı cheapfake'ler bu alandaki sembol gelişmelerdir. 2018 yılında Facebook'ta viral olan bir videoda acil iniş yapan Beijing Capital Airlines'a ait olduğu iddia edilen, acil iniş yapan bir uçak yer almaktadır. Esasında video, paylaşımdan bir yıl önce bir film yapımcısı ve animatör tarafından yapılmıştır. Bilgisayarda oluşturulmuş bir klip cheapfake yöntemiyle gerçek bir haber görüntüsü hâline getirilerek sanal ortamda dolaşıma sokulmuştur. Kısa sürede 14 milyon

izlenmeyi aşan video yayıncı kuruluşa 225 bin yeni takipçi de kazandırmıştır (Fowler, 2018).

Fotoğraf 10. 2018 yılında Facebook'ta viral olan bir videoda acil iniş yapan Beijing Capital Airlines uçağına ait olduğu iddia edilen cheepfake videosundan bir kare.

ABD Temsilciler Meclisi Başkanı Nancy Pelosi'nin yer aldığı bir videoda da cheapfake sahteciliği yapıldığı ortaya çıkmıştır. 2019 yılında çekilen orijinal video bir cheapfake videoya dönüştürülmüştür ve Pelosi'nin konuşurken gevelediği gösterilmiştir. Videoda Pelosi'nin konuşma hızı %75 oranında yavaşlatılarak ses perdesi değiştirilmiştir. Sonuç olarak ortaya Pelosi'nin sarhoş olduğu izlenimini veren bir cheapfake video çıkmıştır. Bu sahte ürün de büyük ilgi görmüş ve sanal ortamda 3 milyonu aşan bir izlenme oranına ulaşmıştır (Harwell, 2019).

Fotoğraf 11. 2019 yılına Nancy Pelosi'nin Center for American Progress konuşmasına yönelik teknik analizler, videonun orijinal hızının yaklaşık yüzde 75'ine kadar yavaşlatıldığını göstermiştir.

Benzer bir örnek 2020 yılında ABD Başkanı Joe Biden'ın yer aldığı bir videoda karşımıza çıkmaktadır. Konuşmanın özünde Biden, şiddet kültürü ve tarihsel kökenleri üzerine bir konuşma yapmaktadır. Ancak yeniden düzenlenerek gerçek bağlamından koparılan bir video ile Biden'ın beyaz milliyetçi fikirleri savunduğu öne sürülmüştür. Oldukça dikkat çeken cheapfake video sadece 48 saat içerisinde 1.6 milyon kez izlenmiştir (Dale, 2020). Politikacıların konuşmalarını onları küçük düşürmek veya itibarlarını zedelemek için tasarlanmış videolara dönüştürmek yeni bir durum değildir. Fakat ses ve görsellerin tamamen değiştirilmesi, sahte haberler için endişe verici yeni bir durum olarak karşımızdadır.

5. Deepfake Geography (Derin Sahte Coğrafya)

Yapay zekânın sayısız teknik avantajının yanı sıra son yıllarda fabrikasyon Küresel Konumlama Sistemi (GPS) sinyalleri (Tippenhauer vd., 2011), sosyal medyadaki sahte konum bilgileri (Zhao ve Sui, 2017), çevrim içi oyun botlarının simüle edilmiş yörüngeleri ve coğrafi ortamların sahte fotoğrafları (Isola vd., 2017) gibi yapay zekâ ve Coğrafi Bilgi Sistemleri (GIS) biliminin yakınsamasının kaygı verici ve beklenmedik uygulamaları da olmuştur. Sonuç ne olursa olsun, deepfake ürünleri hem kişisel mahremiyet hem de devletlerin ulusal güvenlikleri için ciddi bir tehdittir (Chesney ve Citron, 2019).

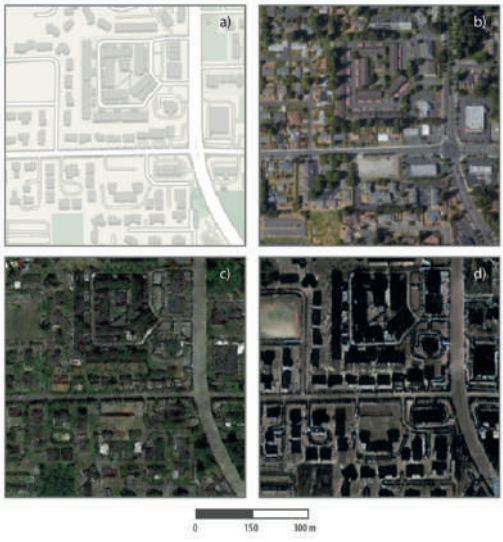
Deepfake kavramı kapsamında hayal edilmeyen başka bir tür daha ortaya çıkmıştır: Deepfake Geography. Türkçe'de "Derin Sahte Coğrafya" olarak isimlendirilen bu yeni deepfake türü, yine yapay zekâ tarafından oluşturulan, aslında dünya üzerinde hiç var olmamış sahte şehir manzaraları ve silüetleri, tesisler, çeşitli yapılar ile kırsal görüntülerden oluşmaktadır. ABD Savunma Bakanlığı'na bağlı bir muharebe destek teşkilatı olan Ulusal Jeouzamsal İstihbarat Teşkilatı'ndaki (NGA) üst düzey yöneticilerden Todd Myers, yapay zekânın, kötü amaçlarla uydu görüntülerinde eserler oluşturmak için sahneleri ve pikselleri manipüle etmek için kullanıldığına dikkat çekmektedir (Tucker, 2019).

Terim ilk olarak yapay zekâ tarafından üretilen sahte bir dijital coğrafi ortamı tanımlamak ve yaratabileceği tehditler konusunda uyarmak için ortaya çıkmış olsa da (Maclenan, 2018), günümüzün sahte coğrafya tartışmalarına ışık tutabilecek ölçüde çalışmalar ortaya koyan ilk coğrafyacılardan biri olan Monmonier önemli bir isimdir. Bilim insanı, "How To Lie With Maps" adlı ünlü kitabında haritaların (ya da jeo uzamsal verilerin) gerçek dünyayı çarpıtarak temsil ettiği çeşitli yolları sistematik olarak açıklamıştır (Monmonier, 1991). Bunun yanında düşmanın moralini sarsmak için gerçek savaş durumlarını

çarpıtarak gösteren savaş zamanındaki propaganda haritaları da yine ilk sahte coğrafya örnekleri arasında gösterilebilir (Herb, 1997).

Tıpkı bir insanın yüzünün sahteciliğe kurban edilmesinde olduğu gibi yapay zekâ tarafından üretilmiş sentetik uydu görüntüleri de son derece tehlikeli ve yanıltıcı olabilir. Orman yangını, sel gibi doğa afetler konusunda aldatmacalar yaratmak veya gerçek uydu görüntülerine dayanan haber içeriklerinin itibarsızlaştırılmasında kullanılabilirler. Jeopolitik açıdan kritik öneme sahip sınırlar ya da siyasi çekişmeli bölgeler üzerinden tasarlanacak olası bir derin coğrafya ürünü diplomatik ve askeri krizlere sebep olabilir. Sahte uydu görüntüleri, askeri planlamacıları yanlış yönlendirmek için ya da sıcak savaşın seyrini değiştirmek için kullanılabilir. Örnekleri çoğaltmak mümkün olmakla beraber sahte uydu medyası yeni bir dezenformasyon biçimi olarak kabul edilmektedir.

Washington Üniversitesi'nde coğrafya profesörü olan Bo Zhao ve meslektaşları yakın geçmişte bu görüntüleri üreten ve tespit eden kendi deneylerini içeren "derin sahte coğrafya" konusunda alanında bir ilk olan bir makale yayınlamışlardır. Zhao, The Verge'e verdiği demeçte, amacın uydu görüntülerinin mutlak güvenilirliğinin işlevini aydınlatmak ve derin sahte coğrafyanın potansiyel etkisi konusunda kamuoyu farkındalığını arttırmak olduğunu söylemektedir (Vincent, 2021).



Fotoğraf 12. Diğer şehirlerin manzara özelliklerine sahip Tacoma'daki bir mahallenin sahte uydu görüntüleri. (a) Orijinal CartoDB (web haritalama ve mekansal veri bilimi araçları sağlayan bir yazılım) deseni, (b) uyumlu/eşdeğer uydu görüntüsü deseni, (c) Seattle ve (d) Pekin'in görsel kalıplarından oluşan sahte uydu görüntüsü.

Fotoğraf 12'de sunulduğu gibi Zhao ve meslektaşları, yapay zekâ tarafından üretilmiş kendi sahte uydu görüntülerini oluşturabilmişlerdir (Zhao vd., 2021). Bu sahte ama inandırıcı fotoğrafları daha iyi anlamak için Zhao ve meslektaşları, üretken bir rakip ağ veya GAN, genellikle derin sahtekârlıklar oluşturmak için kullanılan bir tür makine öğrenen bilgisayar modeli tasarlamışlardır. Esasen bu bir zekâ oyununda rekabet etmek için tasarlanmış bir çift sinir ağıdır. Bu yöntemde jeneratör olarak bilinen sinir ağlarından biri, binlerce gerçek görüntüyle edindiği deneyime dayanarak sahte uydu görüntüleri üretmektedir. Diğer sinir ağı ise ayrımcı, renk, doku ve keskinlik gibi uzun bir kriter listesini analiz ederek sahtekârlıkları tespit etmeye çalışmaktadır. Bu tür birkaç denemeden sonra, nihai sonuç neredeyse gerçeklikten ayırt edilemez görünmektedir. Zhao ve meslektaşları, Washington, Tacoma'nın bir haritasıyla başlamışlar, ardından Seattle ve Pekin'in görsel modellerini haritaya aktarmışlardır. Melez görüntüler elbette dünyanın hiçbir yerinde var olmayan imajlardır, ancak türetildikleri gerçek uydu görüntüleri kadar meşru görünmektedirler.

Makalelerinde deepfakeleri yeni bir meydan okuma olarak sunmaktan ziyade söz konusu teknolojiyi bin yıl öncesine dayanan uzun bir sahte coğrafya tarihi ile çerçeveleyen Zhao ve meslektaşları, Babilliler gibi eski uygarlıklar tarafından tasarlanan mitolojik coğrafyalardan savaş sırasında dağıtılan modern propaganda haritalarına kadar çeşitli ilkel örneklerden bahsetmektedir. Zhao ve diğerlerine göre başka bir ilginç örnek ise "kâğıttan kasabalar" ve "tuzak sokaklar" konusudur. Bunlar, çalışmalarını çalan rakipleri yakalamak için haritacılar tarafından haritalara eklenen sahte yerleşimler ve yollardır. Sahte coğrafya ürünlerini asırlık bir fenomen olarak tarif eden Zhao, yeni teknoloji ile yeni zorlukların doğduğunu, eğitimsiz gözlerin derin sahte uydu görüntülerini gerçekçi bulabildiğini itiraf etmektedir. Sahte uydu görüntüleri üretmenin insanların sahte videolarını üretmekten daha kolay olduğunu belirten Zhao, uydu görüntülerinin çoğunun profesyoneller veya hükûmetler tarafından üretildiğinden dolayı halkın genellikle bunları gerçek kabul ettiğinden söz etmektedir. Bunun yanı sıra, deepfakelerin arkasındaki teknoloji de sadece kötü olarak görülmemelidir. Zhao, aynı makine öğrenimi taktiklerinin görüntü çözünürlüğünü iyileştirebileceğine, iklim değişikliğini modellemek için gereken bir dizi fotoğraftaki boşlukları doldurabileceğine veya hâlâ büyük ölçüde insan denetimi gerektiren harita oluşturma sürecini kolaylaştırabileceğine vurgu yapmaktadır.

6. Tehditleriyle ve Olanaklarıyla Deepfake Teknolojisi

Deepfake teknolojisi ilerledikçe gerçeğin üzerinde sallanan yalan kılıcı daha da keskinleşmektedir. Deepfake kavramı üzerinden büyüyen tartışmalar

sonucunda Amazon, Meta ve Microsoft gibi büyük teknoloji şirketleri ortaklık içerisinde deepfakeleri tespit etme mücadelesine girişmişlerdir. Daha sonra Microsoft, yapay olarak manipüle edilmiş medyayı tespit etmek için bir araç yayınlamıştır (Burt, 2020). Bugün olduğu gibi yakın gelecekte de çözülmeyi bekleyen bazı riskler bulunmaktadır. Bunların başında dezenformasyon gelmektedir. Medya içeriklerine ve materyallere etki gücü açısından yaklaşıldığında ilk sırada izleyicilerin gösterilene güçlü bir tepki geliştirmesine yol açan video içerikler gelmektedir. Post-truth çağın en büyük yanılgılarından biri olan "görmek inanmaktır" söyleminin sahte içeriğin ve bilginin yayılmasındaki payı oldukça büyüktür. İleri ya da ucuz teknolojilerle manipüle edilmiş ve değiştirilmiş videolar hızlı biçimde yayılırken içeriğin orijinal olmadığı konusundaki itirazlar doğrulanana kadar dezenformasyonun kamuoyunda açtığı yara büyümektedir.

Medya okuryazarlığı bu anlamda başka bir önem kazanmaktadır. Medyadaki içeriklerin dijital manipülasyona uğrayıp uğramadığı sorusu zihinleri sürekli meşgul ederken bireylerde içeriğe karşı duyulan zihinsel bir yorgunluğa, güven duygusunun yok olmasına ve sonunda da kayıtsızlığa neden olacaktır. Post-truth çağında bireyler şüphe süzgeçlerini her dönemkinden daha da sık kullanmak zorundadır. Sürekli zihinsel bir eleme yapan, içerikleri teyit etmek zorunda kalan hedef kitle, zaman içerisinde bu tutumunu sürdürmekte güçlük çekecektir. Deepfake ürünleri de bu sürece negatif olarak katkı sağlamaktadır. Günümüzde en güvenilir kaynakların, ünlü isimlerin, devlet kurumlarının ve kamuya mal olmuş şahsiyetlerin bile zaman zaman yanılarak sahte olan içerikleri paylaştıkları da göz önüne alındığında gerçek üzerinde devam eden bu belirsizliğin kişilerin doğrulama refkleksleri üzerinde aşınmalar yaratması kaçınılmazdır. Böylesine devasa bir bilgi setinin çok farklı medyumlar üzerinden kitlelere aktığı bir çağda, sahte içerikler bireylere ve şahıslara haklarında medyada sunulan kimi iddiaları yalanlama, asılsızlaştırma ve gerçeği saptırma yeteneği sunmaktadır. Zararlar bunlarla da sınırlı değildir. Deepfakeler uygun koşullar altında bireylere veya kuruluşlara zarar vermek için kullanılabilmektedir. Herhangi bir suça ilişkin dolaşıma sokulan videolar, bazı kesimleri fiziksel şiddete yönlendirebilir. Ya da bir devlet temsilcisinin deepfake videosuyla diplomatik kriz çıkarılabilir ya da terör saldırısı gibi sahte tehditlerle halk paniğe sürüklenebilir. Bunun yanı sıra pornografik içeriklerden oluşan derin sahtekârlıklar telafisi oldukça zor travmalara ya da psikolojik zararlara neden olarak itibarlara zarar verebilir. Deepfakeler adli vakalarda ya da davalar sırasında gerçek bir delil olarak sunulabilir. Bu durum, davanın sonucunu etkileyebilir, yargı makamlarına ve kamuoyuna şüphe ekebilir veya süreci tamamen tıkayabilir.

Dolandırıcılık, ahlaki sömürü ya da askeri/siyasi casusluk faaliyetleri gibi çok çeşitli amaçlar için de deepfake kullanılabilmektedir. Yüklü finansal işlemlere izin verilmesini sağlayan büyük dolandırıcılıklar, belirli haber ve içeriklerin gündemden düşmesi için dağıtılan sentetik gündem ürünleri ya da ünlü isimlerden sade vatandaşlara kadar hemen herkese yönelik hazırlanan intikam pornoları bunlardan bazılarıdır. Ayrıca içeriğin doğruluğunun teyit edilmesi için yeterli zamanın olmadığı genel seçim gibi kritik durumlarda bireyleri ya da grupları sabote edebilecek şekilde kararları etkileyebilir. Haberciler açısından da sorumluluk büyüktür. Enformasyon yarışında geride kalmamak adına henüz doğrulanmamış bilgiyi ve içeriği dolaşıma sokan haberciler, yayıncılar ve kuruluşlar deepfake paylaşımları nedeniyle varlıklarının en önemli unsuru olan güvenilirliklerini kaybedebilirler. Özellikle sosyal medyada hızla yayılan deepfake içerikler yankı odaları yoluyla karşıt gruplar arasındaki ayrışmayı artırabilir ve toplumdaki fikirsel çözülmeyi ya da kutuplaşmayı hızlandırabilir. Topluluklar arasında azınlıkta kalmış gruplar, belirli bir inanç sistemine ya da dine mensup olan veya belirli bir yaşam tarzının sembolü olarak görülen bireyler tehlike altında kalabilirler.

Deepfake teknolojisinin yukarıda sözü edilen birçok olumsuz sonucu ve potansiyel tehdidi bulunsa da her teknoloji gibi iyi amaçlara hizmet edildiğinde bazı fırsatlar da teknolojinin doğasında taşımaktadır. Bir HBO belgeseli olan Welcome to Chechnya adlı yapımda deepfake teknolojisi kullanılmıştır. Belgesel için kameralara röportaj veren gönüllülerin gerçek yüzleri olası suçlamaların, şiddet olaylarının ya da yargılamaların önüne geçebilmek için hayatta kalan anonim LGBTQ+ bireylerin yüzleriyle değiştirilmiştir ve bu pelerin kimliklerin oluşturulabilmesi için deepfake teknolojisinden yararlanılmıştır. Deepfake teknolojisi, hayatını kaybetmiş sanatçıların, aktörlerin, kanaat önderlerinin ya da tarihi şahsiyetlerin canlandırılmadında ve hatta müzisyenler ile aktörlerin seslerinin ve görüntülerinin yeniden oluşturulması için kullanılabilmektedir. Ayrıca yaratıcı alanda çalışan profesyoneller için görüntü/video sentezinin maliyetini önemli ölçüde düşürmektedir. Bu, büyük bütçelere ve pahalı VFX (görsel efekt) teknolojisine ihtiyaç duyulmadan sektör için bir demokratikleştirme anlamı da taşımaktadır. Son olarak kişinin duygusal ifadelerinin ve seslerinin sentetik olarak daha yüksek bir doğruluk düzeyinde üretilmesini mümkün kılmaktadır. Samsung'un Neon'u gibi teknolojiler, yapay zekâ destekli sanal varlıklar oluşturmak için bunu kullanırken oyun şirketleri, oyun içi deneyimi geliştirmek için kendi 3D oyun avatarlarını oluşturmak için bir oyuncunun özçekiminin nasıl kullanılabileceğini araştırmaktadır. Deepfakeler gelişmekte olan bir endüstrinin parçasıdır ve teknoloji olgunlaştıkça daha meşru kullanım durumlarının ortaya çıkmasını beklemek mümkündür.

Sonuç

Betimsel analiz yöntemi ile kapsamlı bir literatür taraması yapılan bu çalışmada deepfake teknolojisi tüm yönleriyle ele alınmaya çalışılmıştır. Deepfake olgusunun kavramsal arka planı, avantajları ile dezavantajları irdelenmiştir. Çalışmada "cheapfake/shallowfake" (ucuz sahtelik), "deepfake geography" (derin sahte coğrafya) kavramlarına da yer verilmiş, bu anlamda da literatüre katkı sunulması amaçlanmıştır. Çalışma sonucunda deepfake ürünlerinin kişisel mahremiyetten demokrasi kültürüne, siyasetten ulusal güvenliğe kadar pek çok alanda etkisinin olduğu, kötüye kullanımı nedeniyle kavrama bakışın son derece kaygı verici ve negatif olduğu saptanmıştır.

Çevrim içi ve çevrim dışı sınırlar ortadan kalktıkça ve temassız yaşam tarzları yeni normal hâline geldikçe, çeşitli iletişim biçimleri ve ticari işlemler çevrim içi olarak gerçekleşmektedir. Özellikle kişisel bilgiler ve sözleşmeler içeren önemli belgelerden arkadaşlarla paylaşılan günlük fotoğraflara kadar çok çeşitli dosya türleri internette sürekli olarak iletilmektedir. Yakın zamana kadar, bilgi operasyonları, bir dizi bilgi kanalında yaymak için manipülatif içerik işleyen insan uzmanlar tarafından yürütülen bir savaş disipliniydi. Bununla birlikte, yapay zekânın ilerlemesiyle birlikte, artık neredeyse sıfır maliyetle, hiper kişiselleştirilmiş ve geniş ölçekte gerçekleştirilebilen, yapay zekâ tarafından oluşturulan bilgi savaşı kampanyalarının devrilme noktasındayız. Deepfakeler, yapay zekâ tarafından üretilen sentetik video, ses, görüntü ve metin kullanımı yoluyla bu yeni nesil bilgi operasyonlarının arkasındaki birincil güçtür.

En basit hâliyle bir sosyal medya profiline sahip olan herkes, deepfake teknolojisini kullananlar için açık potansiyel bir hedeftir. Sosyal medya devleri de kullanıcılarını hizmet şartlarına gömülmüş, mekanik biçimde okunmadan onay verilen yumuşak rıza metinlerinden güç almaktadır. Onay verilen izinler, verilerin toplanmasını web'de gezinme veya reklamlara tıklama alışkanlıklarının çok ötesinde bir noktaya doğru taşımıştır. Çünkü sosyal medya, deepfake teknolojisi için yüzlerin ve vücutların yer aldığı devasa bir veri seti işlevi görmektedir. Amatör programcılar, bu görüntüleri eğitim verileri olarak kazımak için şimdiden listenin başında yer almaktadır.

Yapay zekâ tarafından oluşturulan deepfake videolar, görseller, sesler ve haber makaleleri, verilere dayalı olarak kişinin tercihlerine göre son derece kişiselleştirilmek üzere oluşturulduğu bir dünyanın yeni formlarıdır. Sıfıra yakın bir maliyetle ve yüz milyarlarca ölçekte üretilen bu biçemler, sosyal medya gibi ucuz dağıtım kanallarıyla birleştiğinde katlanarak artan bir hızla bireyleri manipüle etme, ekonomiye zarar verme ve demokratik süreçleri sekteye uğratma gücüne sahiptir. Bu üstel büyüme, dünyadaki dijital bilgilerin

çoğunun yakın gelecekte yapay zekâ tarafından üretilmesine yol açacaktır. Bu durum sade vatandaşlar için internette bir güven katmanı ihtiyacını daha da belirgin hâle getirecektir.

Bu çalışmada örnek olarak sunulan şantaj, itibar suikasti, finansal dolandırıcılık gibi amaçlara hizmet eden deepfake vakaları, var olan örneklerin yalnızca medyada görünen ve popüler olan kısmıdır. Şimdiye kadar, haberler ve kamuoyu tartışmaları en çok ünlülerin, politikacıların ve tanınmış kişilerin taklit edilmesini içeren vakalara odaklanmıştır. Ancak bu teknikler daha erişilebilir hâle geldikçe, özel kişilere, özellikle de politik, sosyal veya ekonomik açıdan savunmasız olanlara zarar verme riski de aynı oranda artış gösterecektir. Bu görüntülerin etkili bir şekilde nasıl anlaşılacağı sorusu çok önemlidir. Derin sahtekârlıklar ve dezenformasyonla etkili bir şekilde mücadele etmek için halkın iyi eğitimli ve bilgilendirilmiş olması gerekmektedir. Hükûmetler, sivil toplum ve kuruluşlar, toplumun derin sahtekârlıklara yatkınlığını önlemek, teknolojik karşı önlemleri finanse etmek ve gerekli mevzuat taslağını hazırlamak için derhal eğitime yatırım yapmalıdır. Deepfake teknolojilerinin kötü etkilerini önlemenin en iyi yöntemi, gazetecilere, hükûmet yetkililerine ve genel olarak halka deepfake ve dezenformasyon hakkında bilgi vermektir. Çıplak gözle tespit edilemeyecek bir deepfake ürününe bakarken bile, kaynağı, onu yayan tarafın güvenilirliğini ve arkasındaki bağlamı anlamak, bireylerin ve gazetecilerin görünenin ve duyulanın ne kadar doğru olduğu konusunda karar vermesine yardımcı olabilir.

Kaynakça

ABC Australia. (2018, August 31). *My fake naked body: one woman's story of image-based abuse*. 18.12.2022 tarihinde https://www.abc.net.au/radionational/programs/earshot/noelle- martin-personal-story-of-revenge-porn-and-deepfakes/11417940 adresinden alınmıştır.

Agarwal, S., Farid, H., Gu, Y., He, M. and Li, H. (2019, June 15-20). *Protecting world leaders against deep fakes*, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 10.12.2022 tarihinde https://farid.berkeley.edu/downloads/publications/cvpr19/cvpr19a.pdf adresinden alınmıştır.

Ajder, H., Patrini, G., Cavalli, F and Cullen, L. (2019, September). *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace. 20.12.2022 tarihinde https://regmedia.co.uk/2019/10/08/deepfake_report.pdf adresinden alınmıştır.

Ayyub, R. (2018, November 21). I was the victim of a deepfake porn plot intended to silence me. *Huffington Post*. 18.12.2022 tarihinde https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316 adresinden alınmıştır.

Boylan, J.F. (2018, October 17). Will deep-fake technology destroy democracy? *The New York Times*. 18.12.2022 tarihinde https://www.nytimes.com/2018/10/17/opinion/deep-fake-technology-democracy.html adresinden alınmıştır.

Bregler, C., Covell, M. and Slaney, M. (1997, August 3-8). *Video rewrite: Driving visual speech with audio*. 24th Annual Conference on Computer Graphics and Interactive Techniques, LA, California, USA, 10.12.2022 tarihinde https://doi.org/10.1145/258734.258880 adresinden alınmıştır.

Burt, T. (2020, September 1). *New steps to combat disinformation*. 18.12.2022 tarihinde https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/ adresinden alınmıştır.

Chan, C., Ginosar, S., Zhou, T. and Efros, A.A. (2019, October 27-November 2). *Everybody dance now: Motion retargeting video subjects*. 2019 International Conference on Computer Vision, Seoul, Korea (South), 10.12.2022 tarihinde https://arxiv.org/pdf/1808.07371.pdf adresinden alınmıştır.

Chesney, B. and Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954.

Cole, S. (2018, March 7). 'Deep voice' software can clone anyone's voice with just 3.7 seconds of audio. *Vice*. 18.12.2022 tarihinde https://www.vice.com/en_us/article/3k7mgn/baidu-deep-voice-software- can-clone-anyones-voice-with-just-37-seconds-of-audio adresinden alınmıştır.

Cootes, T.F., Edwards, G.J. and Taylor, C.J. (2001). Active Appearance Models. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 23(6), 681–682. https://doi.org/10.1109/34.927467.

Dale, D. (2020, January 3). Fact check: Widely viewed Twitter video misleads on Biden's comments. *CNN*. 18.12.2022 tarihinde https://www.cnn.com/2020/01/03/politics/biden-clip-inaccurate-white-nationalism-fact-check/index.html adresinden alınmıştır.

Delfino, R.A. (2019). Pornographic deepfakes: The case for federal criminalization of revenge porn's next tragic act. *Fordham Law Review*, Article 2, 88(3), 886-938. https://flickread.com/edition/html/ 5a4b7cdbe3415#18.

Fowler, G.A. (2018, October 18). I fell for Facebook fake news. Here's why millions of you did, too. *The Washington Post*. 18.12.2022 tarihinde https://www.washingtonpost.com/technology/2018/10/18/i-fell-facebook-fake-news-heres-why-millions-you-did-too/?utm_source=reddit.com adresinden alınmıştır.

Funke, D. (2018, June 6), A potential new marketing strategy for political campaigns: Deepfake videos. *Poynter*. 18.12.2022 tarihinde https://www.poynter.org/news/potential-new-marketing-strategy-political-campaigns-deepfake-videos adresinden alınmıştır.

Harwell, D. (2019, May 24). Faked Pelosi videos, slowed to make her appear drunk, spread across social media. *The Washington Post*. 20.12.2022 tarihinde https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/ adresinden alınmıştır.

Herb, G. H. (1997). *Under the map of Germany: Nationalism and propaganda 1918–1945*. London: Routledge.

India Today. (2018, November 21). I was vomiting: Journalist Rana Ayyub reveals horrifying account of deepfake porn plot. 20.12.2022 tarihinde https://www.indiatoday.in/trending-news/story/journalist-rana-ayyub-deepfake-porn-1393423-2018-11-21 adresinden alınmıştır.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A. (2017, July 21-26). *Image-to- image translation with conditional adversarial networks*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 10.12.2022 tarihinde https://doi.org/10.1109/ CVPR.2017.632 adresinden alınmıştır.

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, 1(3), 138–147. https://doi.org/10.1109/TTS.2020.3001312.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T. (2020, June 13-19). *Analyzing and improving image quality of StyleGAN*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recogni-

tion (CVPR), Seattle, WA, USA, 10.12.2022 tarihinde https://doi.org/10.1109/CVPR42600.2020.00813 adresinden alınmıştır.

Maclenan, A. (2018, January). Fake geography. *GeoConnexion International Magazine*. Retrieved from https://flickread.com/edition/html/5a4b7cdbe3415%2318#18 in 10.12.2022.

Meta AI. (2020, June 25). *Deepfake Detection Challenge Dataset*. 23.12.2022 tarihinde https://deepfakedetectionchallenge.ai/ adresinden alınmıştır.

Mills, T. (2021, July 21). Devices hacked, women and girls blackmailed as cyber abuse grows. *The Sydney Morning Herald*. 20.12.2022 tarihinde https://www.smh.com.au/national/devices-hacked-women-and-girls-blackmailed-as-cyber-abuse-grows-20210720-p58b8l.html adresinden alınmıştır.

Mitchell, A., Gottfried, J., Stocking, G., Walker, M. and Fedeli, S. (2019, June 5). *Many Americans say made-up news is a critical problem that needs to be fixed*. Pew Research Center. 20.12.2022 tarihinde https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/ adresinden alınmıştır.

Monmonier, M. (1991). *How to lie with maps*. Chicago: University of Chicago Press.

Pollard, J. (2019, October 30). *Predictions 2020: This time, cyberattacks get personal*. 06.12.2022 tarihinde https://www.forrester.com/blogs/predictions-2020-cybersecurity adresinden alınmıştır.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019, October 27-November 2). *FaceForensics++: Learning to detect manipulated facial images*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 13.12.2022 tarihinde https://doi.org/10.48550/arXiv.1901.08971 adresinden alınmıştır.

Statt, N. (2019, September 5). Thieves are now using AI deepfakes to trick companies into sending them money. *The Verge*. 20.12.2022 tarihinde https://www.theverge.com/2019/9/5/20851248/deepfakes-ai-fake-audio-phone-calls-thieves-trick-companies-stealing-money adresinden alınmıştır.

Stupp, C. (2019, August 30). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal*. 20.12.2022 tarihinde https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 adresinden alınmıştır.

Suwajanakorn, S., Seitz, S.M. and Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning Lip Sync from Audio. *ACM Transactions on Graphics*, Article 95, 36(4), 1-13. https://doi.org/10.1145/3072959.3073640.

Tammekänd, J., Thomas, J. and Peterson, K. (2020, October). *Deepfakes 2020: The Tipping Point*. Sentinel. 20.12.2022 tarihinde https://thesentinel.ai/media/Deepfakes%202020:%20The%20Tipping%20Point,%20Sentinel.pdf adresinden alınmıştır.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. and Nießner, M. (2019). Face2Face: Real-time face capture and reenactment of RGB videos. *Communications of the ACM*, 62(1), 96–104. https:/doi.org/10.1145/3292039.

Thies, J., Elgharib, M., Tewari, A., Theobalt, C. and Nießner, M. (2020, August 23-28). *Neural voice puppetry: Audio-driven facial reenactment*. 2020 ECCV: European Conference on Computer Vision, Glasgow, United Kingdom, 13.12.2022 tarihinde https://doi.org/10.48550/arXiv.1912.05566 adresinden alınmıştır.

Tippenhauer, N. O., Pöpper, C., Rasmussen, K. B. and Capkun, S. (2011, October 17-21). *On the requirements for successful GPS spoofing attacks*. CCS '11: 18th ACM Conference on Computer and Communications Security, Chicago, Illinois, USA, 12.12.2022 tarihinde https://doi.org/10.1145/2046707.2046719 adresinden alınmıştır.

Toni Allen, T. (2019, September 19). *Dodging deception & seeking truth online,* Who Is Hosting This. 08.12.2022 tarihinde https://perma.cc/2LJN-D3UP adresinden alınmıştır.

Vaccari, C. and Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media and Society*, 6(1), 1–13. https://doi.org/10.1177/2056305120903408.

Vincent, J. (2018, September 14). US lawmakers say AI deepfakes 'Have the potential to disrupt every facet of our society'. *The Verge*. 22.12.2022 tarihinde https://www.theverge.com/2018/9/14/17859188/ai-deepfakes-national-security-threat-lawmakers-letter-intelligence-community adresinden alınmıştır.

Vincent, J. (2021, April 27). Deepfake satellite imagery poses a not-so-distant threat, warn geographers. *The Verge*. 22.12.2022 tarihinde https://www.theverge.com/2021/4/27/22403741/deepfake-geography-satellite-imagery-ai-generated-fakes-threat adresinden alınmıştır.

Warzel, C. (2018, February 11). Believable: The terrifying future of fake news. *BuzzFeed*. 22.12.2022 tarihinde https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news adresinden alınmıştır.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 39–54. https://doi.org/10.22215/timreview/1282.

Wiggers, K. (2019, December 17). Resemble AI launches voice synthesis platform and deepfake detection tool. *The Verge*. 22.12.2022 tarihinde https://venturebeat.com/business/resemble-ai-launches-voice-synthesis-platform-and-deepfake-detection-tool/ adresinden alınmıştır.

Zhao, B., and Sui, D. Z. (2017). True lies in geospatial big data: Detecting location spoofing in social media. *Annals of GIS*, 23(1), 1–14. https://doi.org/10.1080/19475683.2017.1280536.

Zhao, B., Zhang, S., Xu, C., Sun, Y. and Deng, C. (2021). Deep fake geography? When geospatial data encounter artificial intelligence. *Cartography and Geographic Information Science*, 48(4), 338-352. https://doi.org/10.1080/15230406.2021.1910075.