Derin Sahte (Deepfake) Teknolojisinin Temel Çalışma Mekanizması ve Etik Tartışmalar

Sinem Sarak
Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi
Esenler, İstanbul - Türkiye
sinem.sarak@std.yildiz.edu.tr

Özet

Derin sahte (deepfake) teknolojisi, yapay zekanın hızlı gelişimiyle birlikte dijital içerik üretiminde önemli bir araç haline gelmiştir. Çalışmada, derin sahte teknolojisinin temel prensipleri ve çalışma mekanizmaları açıklanmış, etik ve etik dışı kullanım alanları ele alınmıştır. Derin sahtenin medya, sanat, eğitim ve sağlık gibi alanlarda yaratıcı ve faydalı uygulamalar sunduğu, ancak aynı zamanda mahremiyet ihlali, manipülatif içerikler, kimlik hırsızlığı ve bilgi kirliliği gibi ciddi tehditlere yol açabileceği vurgulanmıştır.

Bu bağlamda, bireylerin hem kendilerini hem de çevrelerini korumak adına medya okuryazarlıklarını artırmaları; teknolojiyi geliştirenlerin, geliştirdikleri teknolojinin kullanım alanlarına ilişkin sorumluluk alması ve düzenleyici kurumların etkili politikalar oluşturması gerektiği belirtilmiştir. Derin sahte teknolojisinin kötüye kullanımının önlenmesi, bu teknolojinin sunduğu fırsatlardan toplumların güvenle yararlanabilmesi için kritik bir öneme sahiptir. Çalışma, derin sahte içeriklerin zararlarını minimize etmek ve olumlu etkilerini artırmak adına iş birliği ve farkındalığın artırılması gerektiğini savunmaktadır.

Anahtar Kelimeler: Derin sahte (deepfake), dijital manipülasyon, dijital etik, medya okuryazarlığı

Giriş

Bilgisayar teknolojilerinin gelişimi, insanların hayatını kolaylaştıran birçok sanal aracın da doğmasını sağlamıştır. Son yıllarda yapay zekanın gelişmesi ve sanal araçlarla birlikte kullanılması sayesinde bu sanal araçlar, günlük yaşamdaki birçok konuya çözümler sunabilir hale gelmiştir. Yapay zeka araçlarının hızla gelişmesiyle beraber insanlar yapay zekayı benimsemiş ve önemli bir kesim, gündelik işlerinde yapay zeka araçlarını kullanır hale gelmiştir. Günümüzde üretken yapay zeka; öğrencilerin ödevlerinden akademik çalışmalara, sanal asistanlardan film ve diziler için müzik ve görsel efekt üretimine kadar birçok farklı alanda kullanılmaktadır. İnternete erişimin kolaylaşmasına ek olarak üretken yapay zekanın yaygın kullanımı içerik üretim sürecinin hızlanmasına neden olmuştur.

Bu hızlanmanın, bilgiye olan erişimi kolaylaştırmak ve sistematik bilgi birikimini artırmak gibi olumlu sonuçları olsa da aynı zamanda yarattığı olumsuz sonuçlar da bulunmaktadır. İnsanların ürettikleri içeriklerden para kazanabiliyor ve toplumun dikkatini çekebiliyor olmaları, sıklıkla etik ve hukuki sınırlamaları ihlal edecek içerikler üretmelerine sebebiyet vermektedir. Yoğun bilgi ve içerik üretimi, niteliksiz ve yanıltıcı bilgilerin yanı sıra belirli etnik gruplara veya siyasi görüşlere yönelik aldatıcı içeriklerin ortaya çıkmasına yol açmakta ve bu durum, bilgi ekosistemini kirleterek bir "bilgi kıyameti" yaratmaktadır. Üretilen içeriklerin doğruluğunun her zaman denetlenememesinin yanı sıra ayrıştırıcı ve insan haklarını ihlal eden yanlış bilgi ve haberlerin hızla yayılması, bu kıyametin tehlikesini daha da artırmaktadır.[1]

Bu içeriklerin üretiminde kullanılan teknolojilerden birisi de derin sahte (deepfake) teknolojisidir. Derin sahte her ne kadar eğlence ve medya endüstrilerinde yaratıcı, insanları eğlendirici ve sektörün gelişimine katkı sağlayıcı içerikler üretme amacıyla kullanılsa da özellikle kötü niyetli internet aktörleri tarafından bu teknolojinin dijital uzamlarda kullanılması sonucunda ortaya çıkan gerçek olmayan ama sahte olmayacak kadarda inandırıcı olan içerikler görerek ve işiterek öğrenen ya da haber alan kullanıcıların gerçeklik algılarını yönlendirmektedir.[2]

Bu çalışmada derin sahte teknolojisinin etik sorunları üzerine yazılmış olan yerli ve yabancı makaleler taranmış, konuyla ilgili kitap ve haberler değerlendirilmiştir. Çalışmada, okuyucunun derin sahte teknolojisinin temel prensibini anlayabilmesi amacıyla, derin sahte teknolojisinin tanımı ve tarihçesi sunulmuş, geliştirilmesinde kullanılan en yaygın algoritmalardan olan "Çekişmeli Üretici Ağ (GAN)" algoritması kısaca açıklanmıştır.

Ayrıca derin sahtenin etik ve etik dışı kullanımlarına örnekler verilerek okuyucunun derin sahte teknolojisinin olumlu ve olumsuz özelliklerini görerek tarafsız bakmasına imkan sağlanmıştır. Bunlara ek olarak bireylerin kendilerini bu teknolojinin kötüye kullanımına karşı koruyabilecekleri anlatılmıştır.

Derin Sahte Nedir ve Nasıl Çalışır?

Derin sahte terimi ilk defa Reddit isimli sosyal medya platformunda anonim bir kullanıcı tarafından açılan bir forum başlığı olarak kullanılmıştır.

Terim, hedef videodaki kişi ile kaynak videodaki değişim tekniğini ifade eden derin öğrenme (deep learning) ve sahte (fake) terimlerinin beraber kullanılmasıyla türetilmiş bir kelimedir.[3] Kavram, sosyal medyada ortaya çıkıp geliştiğinden ötürü tanımı konusunda fikir birliği sağlanamamıştır. Kimi kaynaklarda derin sahte; (yapay zekanın bir alt dalı olan) derin öğrenme alanındaki metotlar kullanılarak bir videonun, fotoğrafın veya ses dosyasının yapay şekilde değiştirilmesi sonucu ortaya çıkan görsel, işitsel içerik ya da aynı yöntemlerle sıfırdan yaratılan yapay bir fotoğraf, ses, video veya görüntüdür. [4] Bir başka deyişle derin-kurgu, bir bireyin resim veya videolarının gerçek hâlleri kullanılarak yeniden üretilmesi ile ortaya çıkan, dijital ortamlarda beğeni ve yeniden paylaşım yoluyla hızlı yayılan, yapay/sentetik bir medyadır. [5] İlk dönemlerinde etik dışı içeriklerin üretimine yönelik kullanıldığından ötürü bazı kaynaklar ise derin sahteyi zararlı, kasten veya kasta dayalı olmayan yanlış bilgi üretiminde veya yayılmasında kullanılan sentetik medya olarak tanımlamaktadır.[1] Ancak günümüzde derin sahte teknolojisi, etik kullanım alanlarını da kapsayacak şekilde geliştiğinden bu tanım kapsayıcılığını yitirmiştir.

Derin sahte ile üretilen içerikleri 4 grup altında toplamak mümkündür:

1- Yüz değiştirme (Face replacement): Derin sahte teknolojisinin en yaygın kullanım biçimidir. Bir kişinin yüzünün dijital olarak başka bir kişiye kopyalanmasıdır.

2- Yüz şekillendirme (Face re-enacment): Bir kişinin yüzünün; başka birisiyle değiştirilmeksizin, mimiklerinin veya yüz özelliklerinin manipüle edilmesidir.

3- Yüz oluşturma (Face generation): Gerçek olmayan bir insana ait bir yüz üretilmesidir.

4- Konuşma üretimi (Speech synthesis): Bir kişinin ses modelinin oluşturulması ve kişinin hiç söylemediği şeyleri söylemiş gibi gösterilmesidir. [6]

Derin Sahte Teknolojisinin Ortaya Çıkışı

Derin sahte terimi ilk kez 2017 yılında aynı isimli anonim bir kullanıcı tarafından açılan bir forumda kullanıldı. Açılan forum, internette mevcut cinsel içerikli videolara Scarlett Johansson, Maisie Williams, Taylor Swift, Aubrey Plaza, and Gal Gadot gibi ünlü kadın oyuncuların yüzleri yerleştirilerek oluşturulan videoların paylaşılması için kullanıldı.[7] Forumun sahibi; kullandığı algoritmayı Nvidia ve Google gibi açık kaynaklı geliştirilmiş benzer araçlardan faydalanarak oluşturmuş bir "yapay zeka ilgilisiydi".[1]

İçeriklerin üretiminde kullanılan kodların açık kaynaklı olarak paylaşılması forum içeriklerinin hızla artmasına ve çeşitlenmesine neden oldu. Forum içerisinde Game Of Thrones ve Harry Potter serilerinde yer alan çocuk oyuncuların (Arya Stark ve Emma Stone) derin sahte ile yapılmış videolarının paylaşılması sonucu forum Reddit tarafından kalıcı olarak kapatıldı. Forum sahibi ise anonim olduğundan ötürü tespit edilememiştir.[1]

Forum kapatılmış ve içerikleri silinmiş olsa da paylaşılmış kodlar açık kaynaklı olduğundan ötürü derin sahte teknolojisi gelişimini sürdürmüştür. Bu kodlar kullanılarak DeepFaceLab and Face Swap gibi birçok derin sahte uygulaması geliştirilmiştir. Bu uygulamaların herkese açık ve herhangi bir bilgi birikimi gerektirmeksizin kullanılabilecek şekilde geliştirilmiş olması yalnızca bireylerin mahremiyetini ihlal eden içerikler değil, aynı zamanda sahte haberler üretmek ve dezenformasyon yaymak için de kullanılabilecek bir araç haline gelmiştir. Özellikle politik figürlerin ses ve görüntülerini manipüle ederek yanıltıcı içerikler oluşturulması bazı etnik grupları ve politik görüşleri hedef gösterme hatta seçim süreçlerini etkileme amacıyla da kullanılmıştır.[10] [11]

Derin sahte teknolojisi yaygınlaştıkça etik sayılabilecek kullanımları da olmuş, reklamcılık, eğlence, eğitim gibi farklı sektörlerde de kullanılmaya başlanmıştır. 2020 yılında State Farm şirketi tarafından ilk defa derin sahte kullanılarak bir reklam filmi çekilmiştir.[1] Ülkemizde de bir banka tarafından oyuncu Kemal Sunal'ın derin sahte ile üretilen görüntüleri reklam filminde kullanılmıştır.[8] Ayrıca Welcome to Chechnya isimli

belgeselde yer alan LGBTİ+ bireylerin kimliklerini korumak amacıyla da derin sahte kullanılmıştır.[9] Anında çeviri yapabilme özelliği ile iletişimde dil engelini kaldırma, özellikle tıp eğitiminde simülasyonlar oluşturarak öğrenimi pekiştirme gibi çok farklı kullanımları da bulunmaktadır.

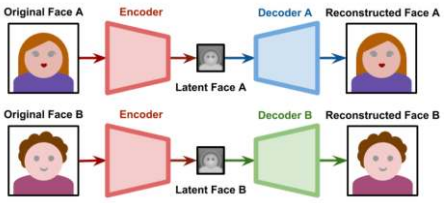
Derin Sahte Teknolojisinin Temel Çalışma Prensibi:

Derin sahtede **autoencoder** denilen bir tür yapay sinir ağı kullanılır. Autoencoder'ın temel amacı bir kişinin yüzünün bulunduğu veri setlerinin içerisinden gerekli yüz özelliklerini (kaş yapısı, göz rengi ve şekli, çene ve ağız yapıları gibi) daha düşük boyutlu bir gizli alan (lantent space) temsiline indirgemek ve daha sonrasında bu temsilden orijinal veriyi yeniden inşa etmektir. Autoencoder'ın iki ana bileşeni bulunur:

Kodlayıcı (Encoder): Veriyi gizli alan temsiline indirme işlemi bu bileşende gerçekleştirilir. Görüntüdeki önemli özellikler sıkıştırılır.

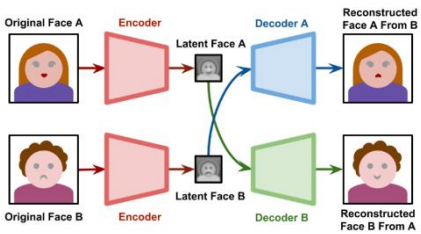
Çözücü (Decoder): Oluşturulan gizli alan kullanılarak orijinal görüntüyü oluşturan kısımdır.

Bir derin sahte videosunda, bir kişinin yüzü, öncelikle kodlayıcı aracılığıyla sıkıştırılır ve bu sıkıştırılmış temsilde sadece yüzün önemli özellikleri (örneğin, göz rengi, kaş yapısı, çene şekli) saklanır. Çözücü bu bilgiyi kullanarak hedef kişinin yüzünü bu özelliklere uygun şekilde yeniden oluşturur.[12] Görsel 1'de iki ayrı autoencoder yapısı gösterilmiştir.



Görsel 1: Birbirinden bağımsız iki autoencoder [13]

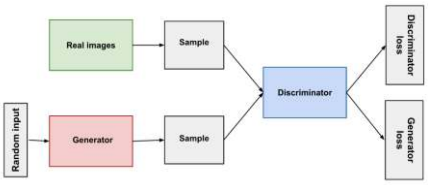
Bir görüntüde yer alan bir yüzün başka bir yüz ile değiştirilmesi, iki farklı sinir ağının üretmiş olduğu gizli alanların birbirlerine yönlendirilmesiyle gerçekleştirilir. Bu işlem iki ağın, aynı kodlayıcıyı kullanması ve her bir ağın kedisine ait bir çözücüsünün olması ile mümkün olur. Örneğin görsel 2'de yer alan çözücü A sadece



Görsel 2: Derin sahte görüntüsünün autoencoder ile oluşturulması [13]

A kişisinin ve çözücü B sadece B kişisinin yüzleriyle eğitilir. Bu süreçte, çözücü A yalnızca A kişisinin yüz özelliklerini yeniden oluşturmak için optimize edilirken, çözücü B aynı şekilde B kişisinin yüz özelliklerini yeniden inşa etmek üzere eğitilir. Eğitim süreci tamamlandıktan sonra ağlardan birinin ürettiği gizli alan diğer ağın çözücüsüne yönlendirilir. Bu sayede A kişisinin yüz özellikleri B görseline yerleştirilir ve yüz değişim işlemi gerçekleştirilir.

Oluşturulan görüntülerin daha gerçekçi olmasını sağlamak amacıyla iki farklı sinir ağı birbirleriyle etkileşimli şekilde kullanılır. Bu yapıya **çekişmeli üretici ağ (generative adversarial network ; GAN)** denir. Bu yapıda iki farklı ağ bulunur. Bu ağlar sırasıyla üretici (generator) ağ ve ayrıştırıcı (discriminator) ağ olarak adlandırılır. Üretici ağ görüntü üretip ayrıştırıcı ağa yönlendirir ve ayrıştırıcı ağ bu görüntünün gerçek olup olmadığını tanımlamaya çalışır. Görüntünün yapay olduğu anlaşılırsa üretici ağa dönülür ve tekrardan üretim işlemi gerçekleşir. Bu döngü ayrıştırıcı ağ, üretilen görüntüyü gerçek olarak algılayana kadar devam eder. Bu işlem görsel 3'te şematik olarak gösterilmiştir.



Görsel 3: GAN modelinin şematik gösterimi [14]

Derin Sahte Teknolojisinin Kullanımları

Derin sahte teknolojisi ilk çıktığında sadece etik dışı kullanılmış olsa da günümüzde yaygınlaşmış ve birçok farklı alanda kullanılmaya başlanmıştır. Bu başlık altında derin sahtenin etik ve etik dışı kullanımları incelenecektir.

Derin Sahte Teknolojisinin Etik Kullanımları:

Derin sahte teknolojisinin kullanımı günümüzde en çok medya ve eğlence sektöründe görülmektedir. Derin sahte teknolojisi ile zararsız eğlendirici sosyal medya içeriklerinin oluşturulmasının yanı sıra film ve dizilerde ölmüş oyuncuların canlandırılabilmesi uygulamalara imkan sağlanmaktadır. Ayrıca oyuncuların yaşlandırılması veya gençleştirilmesinde de bu

teknoloji oldukça faydalıdır. Geleneksel olarak, bu amaçlar için bilgisayar tabanlı görüntü (CGI) uygulamaları, profesyonel makyaj teknikleri, özel ayarlanmış lensler ve karmaşık çekim düzenekleri kullanılmaktadır. Ancak derin sahte teknolojisi, bu süreçleri hem hızlandırmakta hem de maliyetleri azaltmaktadır. Örneğin bir filmde Netflix tarafından bir aktöre birçok CGI uzmanı ve çeşitli teknikler kullanılarak yapılan gençleştirme işlemi gerçekleştirilmiştir. Buna karşılık, bir YouTube kullanıcısı, derin sahte teknolojisini kullanarak aynı işlemi yalnızca yedi gün içinde uygulamayı başarmıştır.[15] Görsel 4'te, Netflix'in gençleştirme çalışmasının (sol) ve bir YouTube kullanıcısının derin sahte yöntemiyle yaptığı yaşlandırma çalışmasının (sağ) karşılaştırması görülmektedir. Bu, derin sahte teknolojisinin geleneksel yöntemlere kıyasla sunduğu potansiyeli açıkça ortaya koymaktadır.

Derin sahte teknolojisi, sanat alanında da kullanılmaktadır. Özellikle müzelerde ve sergilerde ölmüş sanatçıların canlandırılması, bu teknolojinin dikkat çeken uygulamalarından biri haline

Görsel 4:The Irish Man filmindeki bir oyuncunun Netflix tarafından CGI kullanılarak (sol) ve YouTube kullanıcısı tarafından derin sahte ile(sağ) gençleştirilmiş hali [15]

gelmiştir. Derin sahte sayesinde, sanat tarihinin önemli figürleri izleyicilerle adeta birebir iletişim kuruyormuş gibi gösterilebilmektedir. Bu sayede ziyaretçilere eşsiz bir deneyim sunulabilmekte ve sanata olan ilgi artırılabilmektedir. Örneğin, Salvador Dali'nin bir müzenin ziyaretçileriyle iletişim kurduğu bir projede derin sahte teknolojisiyle bir canlandırma yapılmıştır.[16] Bu uygulama, sanatseverlere hem eğitici hem de interaktif bir deneyim sağlarken, sanatçının eserlerine olan ilgiyi de artırmıştır. Benzer projeler, müzelerin ve galerilerin ziyaretçi deneyimini zenginleştirmek için derin sahte teknolojisini nasıl kullanabileceğini göstermektedir.

Derin sahte teknolojisi sağlık alanında da kullanılmaya başlanmıştır. Taipei Medical Üniversitesi'nde yapılan bir çalışmada hastaların şikayetlerini yüz ifadelerinden anlayabilmek için derin sahte yöntemleri kullanılmış ve bu sayede doktor hasta arası iletişimin geliştirilmesi amaçlanmıştır.[17] Başka bir araştırmada ise tıbbi patoloji alanında kullanılmak üzere histolojik görüntüler sentezlenmesi amacıyla derin sahte kullanılmıştır. Derin sahte kullanılarak üretilen görüntülerin gerçek dokularla neredeyse ayırt edilemez olduğu ve patologlar tarafından tanısal doğrulukla kullanılabileceği gözlemlenmiştir. Bu sayede hasta gizliliği ihlal edilmeden veya preparat doku örneklerine ihtiyaç kalmaksızın doktor adaylarının eğitimleri için gerekli veri kümeleri oluşturabilme imkanının sağlanmasının önü açılabilmiştir. [19]

Özetle derin sahte teknolojisinin etik kullanımı birçok farklı alanda insanların hayatlarına katkı sağlamaktadır. Bu teknolojinin doğru ve etik bir şekilde yönlendirilmesi, mevcut birçok sorunun çözümüne katkı sağlayabilir. Bu nedenle, etik kurallar çerçevesinde geliştirilen ve kullanılan derin sahte teknolojisinin, farklı sektörlerde yarattığı olumlu etkiler giderek artacaktır.

Derin Sahte Teknolojisinin Etik Dışı Kullanımları:

Derin sahte teknolojisinin yanlış şekillerde kullanılması insanlara bireysel ve toplumsal olarak çok farklı boyutlarda zarar verebilmektedir. Bu zararlar genel olarak dört ana başlıkta incelenebilir:

1- Mahremiyet İhlali:

Derin sahte teknolojisinin en yaygın etik dışı kullanımlarından biri, izinsiz olarak bireylerin yüzlerinin cinsel içeriklerde kullanılmasıdır. Özellikle kadınlar, bu tür içeriklerin hedefi haline gelmekte ve ciddi itibar kaybı, psikolojik zarar ve mahremiyet ihlali yaşamaktadır. Sentisy.ai isimli şirketin raporuna göre internette yer alan derin sahte videolarının %96'sını cinsel içerikli videolar oluşturmaktadır. Cinsel içerikli videolarda yüzleri izinsizce kullanılan bireylerin %100'ünü kadınlar oluşturmakta ve bu kadınların %99'unu eğlence sektöründe yer alan oyuncular veya şarkıcılar oluşturmaktadır.[20] Kadınların yüzlerinin izinsiz şekilde cinsel içeriklerde kullanılması, sadece bireysel düzeyde büyük bir mahremiyet ihlali ve psikolojik travma yaratmakla kalmayıp, aynı zamanda dijital dünyada kadınların itibarının zedelenmesine de neden olmaktadır. Bu veriler, özellikle kadınların, derin sahte teknolojisi tarafından hedef alınmasının, toplumsal cinsiyet eşitsizliği ve dijital güvenlik sorunlarına yol açtığını göstermektedir.

2- Yanıltıcı ve Manipülatif İçerikler:

Siyasi figürlerin veya hassas grupların derin sahte ile yapılan görüntüleri sosyal medyada sıklıkla paylaşılarak dezenformasyona neden olmaktadır. Bu tür içeriklerle, yaşanmamış olayların yaşanmış gibi gösterilebilir ve toplum algısı yanlış yönlendirebilmektedir. Örneğin, bir siyasi liderin asla söylemediği bir konuşmanın ona atfedilmesi, destekçileri ve karşıtları arasında yanlış anlamalara ve çatışmalara neden olmakla birlikte toplum içerisindeki kutuplaşmanın da artmasına neden olmaktadır. Özellikle seçim dönemlerinde, sahte videolar ve ses kayıtları kullanılarak seçmenlerin algıları manipüle edilmekte ve demokratik sistemin temelini oluşturan özgür irade üzerinde ciddi etkiler bırakılmaktadır. Barack Obama'nın Donald Trump hakkında hakaret içerikli bir açıklama yapıyormuş gibi gösterildiği sahte video, bu tür manipülatif içeriklere örnek olarak gösterilebilir. Bu video, gerçekte hiç yaşanmamış bir olayın geniş çapta tartışılmasına neden olmuş ve siyasi figürlerin itibarını zedelemek için derin sahte kullanımının etkisini ortaya koymuştur.[21]

Derin sahte ile üretilen bu tür yanıltıcı haberlerin seçmenlerin tercihleri üzerinde doğrudan etkili olduğuna dair çalışmalar bulunmaktadır. Örneğin bir çalışmada, 2017 Almanya parlamento seçimlerinde, yanlış bilgilere inanan bireylerin, ana akım siyasi partilere olan desteklerini kaybederek aşırı sağ partilere yöneldiği tespit edilmiştir. [22] Benzer şekilde Rusya- Ukrayna arasındaki savaş sırasında Rusya devlet başkanı Vladimir Putin ve Ukrayna devlet başkanı Volodymr Zelenskyy'nin gerçek olmayan videoları internete gerçekmiş gibi servis edilmiş ve basın bu yolla manipüle edilmiştir.[3] Bu tür manipülatif içerikler toplumdaki kutuplaşmanın artmasına ve toplumsal güvenin sarsılmasına neden olmaktadır.

3- Kimlik Hırsızlığı ve Dolandırıcılık:

Derin sahte kullanılarak insanların ses ve görüntülerinin oluşturulması, kimlik hırsızlıklarında ve bireylerin kandırılmasında da kullanılmaktadır. Özellikle insanların sesleri taklit edilerek yapılan 'phone phishing' saldırılarında derin sahte seslerin kullanılmasıyla birlikte insanları paniğe sevk etmek ve kandırmak daha kolay hale gelmiştir. Buna örnek olarak Kaliforniya'da yaşayan Gary Schildhorn isimli adamın, oğlunun sesi kullanılarak kandırılması ve dolandırılması örnek verilebilir.[23] . Bu tür vakalar ülkemizde de görülmüştür. Bir yayıncı olan Revanch'ın kimlik bilgilerinin derin sahte teknolojisiyle taklit edilerek dolandırıcılık amacıyla kullanılmıştır. Olayda, yayıncının annesinin cep telefon numarası ve sesi taklit edilerek sahte bir acil çağrı bırakılmış ve bu durum, hem mağduru hem de ilgili kurumları ciddi şekilde etkilemiştir. [5]

Dolandırıcılık vakalarının başka bir örneği ise derin sahte videolarla ünlü kişilerin veya iş dünyasından liderlerin sahte açıklamalar yaptığının gösterilmesidir. Bu içerikler, hem kimliği kullanılan ünlü kişilerin itibarına zarar vermekte hem de yanlış bilgilendirme yoluyla maddi ve manevi kayıplara yol açmaktadır.[24] Benzer bir şekilde, dolandırıcılar tarafından derin sahte teknolojisiyle, şirketlerin üst düzey yöneticilerine ait sahte video ve ses kayıtları kullanılarak şirketlerden para sızdırma girişimlerinde bulunulmuştur. [25]

4- Toplumsal Güven Kaybı ve Bilgi Kirliliği:

İnternette üretilen sahte içeriklerin git gide daha gerçekçi hale gelmesi insanları tedirgin etmektedir. Öte yandan sahte haberlerin kutuplaştırıcı ve kışkırtıcı özelliklerinden kaynaklı hızlı yayılıyor olması da etkilerini artırmaktadır. Sahte medya içerikleri ile bilgi kirliliği yaratılması içinde yaşanılan dünyanın algılanmasını zorlaştırır, insanlar gerçek ile sahteyi ayırt edemez hale gelir. Tartışmalar ortaya çıkar ve insanlar taraf seçmeye zorlanırlar. Sadece tartışmaları kazanmaya odaklı hale gelirler ve kendi görüşlerini destekleyen her türlü içeriği ,gerçekliğini sorgulamaksızın, destekler ve paylaşırlar.[1] Bu durum sosyal medya ve basın yayın kuruluşlarına güveni azaltmakla birlikte insanları paniğe sürükleyen bir ortamın oluşmasına sebep olmaktadır.

Derin Sahtenin Olumsuz Etkilerinden Korunma Yolları

Derin sahte içeriklerin yarattığı tehditlere karşı bireylerin farkındalık kazanması ve bu içeriklerden korunması önemlidir. Bireylerin medya okuryazarlıklarını geliştirmeleri sahte haberlerin tespiti konusunda kritik bir rol oynamaktadır. İnsanların internette karşılaştıkları içeriklere şüpheyle yaklaşmaları ve bir içeriğin doğruluğuna güvenmeden önce farklı kaynaklardan araştırmaları

gerekir. Derin sahte içeriklerin tespitinde yapay zeka destekli araçlardan yararlanmak da önemlidir. Derin sahte teknolojisinin yaygınlaşmasıyla birlikte, bu içeriklerin gerçek olup olmadığını analiz eden Sensity.ai gibi araçlar geliştirilmiştir. Bu tür doğrulama araçları, bir videonun veya görüntünün manipüle edilip edilmediğini teknik yöntemlerle analiz edebilir. Bunun yanında, görsel ya da ses içeriğinin oluşturulma tarihi ve kaynağı da şüpheli durumların aydınlatılmasında kullanılabilir.

Medya okuryazarlığının yanı sıra bireylerin kendi mahremiyetlerini korumak için de önlemler almaları gerekir. Kullanıcıların kendi görsellerini çevrim içi ortamda paylaşmadan önce potansiyel riskleri değerlendirmesi önemlidir. Özellikle, yüksek çözünürlüklü yüz fotoğraflarının kamuya açık platformlarda paylaşılması, bu görsellerin kötüye kullanılma ihtimalini artırabilir. Paylaşılan fotoğraflarda yüz filtreleri veya filigranlar kullanmak, kötü niyetli kullanıcıların bu görselleri manipüle etmesini zorlaştırabilir.

Bireysel farkındalığın artırılmasının yanı sıra, eğitim programları aracılığıyla toplum genelinde dijital güvenlik ve medya okuryazarlığı bilinci geliştirilmelidir. Bu tür programlar, yalnızca bireyleri değil, aynı zamanda işletmeleri ve kurumları da hedefleyerek geniş çapta koruma sağlamaya yönelik çözümler sunabilir. Bu sayede derin sahte kullanılarak yapılan dolandırıcılıkların da azaltılabilmesi sağlanabilir.

Sonuç

Derin sahte teknolojisi, sunduğu imkanlar ve yol açtığı etik sorunlarla tartışmalı bir teknoloji olarak dikkat çekmektedir. Çalışma içerisinde, derin sahte teknolojisinin hem olumlu hem de olumsuz yönleri incelenmiş, bu teknolojiye dair etik ve etik dışı kullanım örnekleri sunulmuştur. Derin sahtenin medya, sanat, eğitim ve sağlık gibi birçok alanda yaratıcı ve faydalı uygulamalar sunduğu ve bu örneklerin gün geçtikçe çeşitlendirilebileceği açık bir şekilde görülmüştür. Ancak öte yandan, mahremiyet ihlali, manipülatif içerikler, kimlik hırsızlığı ve toplumsal güven kaybı gibi olumsuz etkileri de ciddi bir tehdit oluşturmaktadır.

Bu noktada, bireylerin ve toplumların medya okuryazarlığını geliştirmesi, dijital etik ilkelerini benimsemesi ve derin sahte teknolojisinin kötüye kullanımına karşı farkındalık kazanması büyük önem taşımaktadır. Ayrıca, yasal düzenlemelerin yapılması ve derin sahte içeriklerin tespitinde yapay zeka destekli doğrulama araçlarının geliştirilmesi gereklidir.

Sonuç olarak, derin sahte teknolojisinin sorumlu bir şekilde kullanımı, toplumlar için yeni fırsatlar yaratırken, etik dışı kullanımının sınırlanması, bu teknolojinin olası zararlarının önüne geçmek için kritik bir adımdır. Derin sahte teknolojisinin bireyleri mağdur etmeyecek şekilde kullanılması ve toplumsal faydayı ön planda tutacak şekilde yönlendirilmesi, bu teknolojinin sunduğu potansiyelin en iyi şekilde değerlendirilmesini sağlayacaktır. Bu nedenle, bireylerin, kurumların ve devletlerin iş birliği içinde çalışarak etik standartlar geliştirmesi, farkındalık eğitimleri düzenlemesi ve alınan önlemleri yasalarla denetlemesi gerekir. Derin sahte teknolojisi, bilinçli ve etik bir şekilde kullanıldığında, birçok alanda kayda değer gelişmelere yol açabilir. Ancak, kötüye kullanımı önlenmediği takdirde, toplumsal düzen ve bireysel haklar üzerinde ciddi olumsuz etkiler yaratma potansiyeline sahiptir. Bu nedenle, kötüye kullanımının önlenmesi ve potansiyel zararların minimize edilmesi için teknolojiyi geliştirenler, kullananlar ve düzenleyici kurumların iş birliği içinde hareket ederek gerekli adımları atmaları hayati önem taşımaktadır.

Referanslar

[1] N. Schick, Deep Fakes and the Infocalypse: What You Urgently Need To Know, 1st ed., New York, Twelve Book Group, 2020.

[2] T. Elitaş, *Dijital Manipülasyon 'Deepfake' Teknolojisi Ve Olmayanın İnandırıcılığı*, Hatay Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Vol. 19 , Iss. 49, pp. 113-128, 2022.

[3] B. Akmeşe, *The Artificial Intelligence Dimension of Digital Manipulation Deepfake Videos: The Case of the Ukrainian-Russian People*, Contemporary Issues of Communication, Vol. 2, Iss. 2 , pp. 76-85, 2023.

[4] C. Yavuz, *Deepfake (Derin Sahte) Yapay Zekâ Tarafından Üretilen Yeni Nesil Görsel–İşitsel İçerik,* 1. Baskı, Ankara, Seçkin Yayınevi, 2022.

[5] Ş. Özdemir, *Yeni Nesil Tehdit: Derin Kurgu (DeepFake),* TRT Akademi Dergisi, Vol. 19, Iss. 13, pp. 905-917, 2021.

[6] Centre for Data Ethics and Innovation (CDEI), *Deepfakes and audio-visual disinformation*, CDEI Snapshot Series, 2019. [Online] Available: https://apo.org.au/node/267076.

[7] N. Gardiner, *Facial re-enactment, speech synthesis and the rise of the Deepfake*, Honours Thesis, Edith Cowan University, 2019. [Online] Available: https://ro.ecu.edu.au/theses_hons/1530

[8] Ziraat Bankası, "157. Yıl Reklam Filmi," [Online. Available: https://www.ziraatbank.com.tr/tr/bankamiz/basin-odasi/bankamiz-reklam-filmleri. [Accessed: Nov. 27, 2024].

[9] J. Doe, *Deepfake tools find a redeeming role*, New York Times, Sec. C, p. 2, New York edition, Jul. 4, 2020. [Online]. Available: https://www.nytimes.com/2020/07/04/deepfake-tools-role.html.

[10] B. McKernan, *Muharrem İnce: Turkish presidential candidate withdraws over alleged sex tape*, The Guardian, May 11, 2023. [Online]. Available: https://www.theguardian.com/world/2023/may/11/muharrem-ince-turkish-presidential-candidate-withdraws-alleged-sex-tape. [Accessed: Nov. 27, 2024].

[11] E. Villasenor, *Is seeing still believing? The deepfake challenge to truth in politics*, Brookings Institution, [Online]. Available: https://www.brookings.edu/articles/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/. [Accessed: Nov. 27, 2024].

[12] İ. Anıkaydın, *Deepfake Uygulamalarının Hukuki Boyutu*, Düzce Üniversitesi Sosyal Bilimler Dergisi, Vol. 12 Iss. 2, pp. 736-747, (2022).

[13] M. Rougaïbi, "*Essai de maîtrise*," M.S.V.D. thesis, Faculté de Droit et Faculté des Sciences, Université de Sherbrooke, Québec, 2024.

[14] Google Developers, "*GAN yapısı*," [Online]. Available: https://developers.google.com/machine-learning/gan/gan_structure?hl=tr. [Accessed: Nov. 27, 2024].

[15] Gideo Vames, *The Irishman De-Aging: Netflix Millions VS. Free Software!* , YouTube, 2020. [Online]. Available: https://www.youtube.com/watch?v=dyRvbFhknRc&t=238s. [Accessed: Nov. 27, 2024].

[16] D. Lee, *Salvador Dalí lives: Museum brings artist back to life with deepfake technology*, The Verge, May 10, 2019. [Online]. Available: https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum [Accessed: Nov. 27, 2024].

[17] H.C.Yang, A.R. Ragmanti, C.W.Huang, Y.C.J.li , *How Can Research on Artificial Empathy Be Enhanced by Applying Deepfakes?*, Journal of Medical Internet Research, Vol 24, Iss. 3, 2022. [Online]. Available: https://www.jmir.org/2022/3/e29506

[18] C. Botella, R. M. Baños, H. Guillén, and A. García-Palacios, *Treatment of complicated grief using virtual reality: A case report*, ResearchGate, vol. 7, no. 3, pp. 475-482, Jan. 2008.

[19] L. Herrera-Hernandez et al., *Deepfake histologic images for enhancing digital pathology*, Lab Invest, vol. 103, p. 100006, 2023.

[20] H. Ajder, G. Patrini, F. Cavalli, and L. Cullen, *The State of Deepfakes: Landscape, Threats, and Impact*, Deeptrace (Currently known as Sensity.ai), Sep. 2019.

[21] T. Dobber, N. Metoui, D. Trilling, N. Helberger, and C. de Vreese, *Do (microtargeted) deepfakes have real effects on political attitudes?* The International Journal of Press/Politics, Vol. 26, Iss. 1, pp. 69–91, 2021.

[22] F. Zimmermann, M. Kohring, *Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election*, Political Communication, Vol. 37, Iss. 2, pp. 215–237, (2020).

[23] A. Blanco, *A father is warning others about a new AI 'family emergency scam*, The Independent, 2024. [Online]. Available: https://www.independent.co.uk/news/world/americas/ai-phone-scam-voice-call-b2459449.html. [Accessed: Nov. 27, 2024].

[24] T. Hsu, Y. Lu, *No, That's Not Taylor Swift Peddling Le Creuset Cookware*. The New York Times, Jan. 9, 2024. [Online]. Available: https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html. [Accessed: Nov. 27, 2024].

[25] M. Somers, *Deepfakes, explained,* MIT Sloan School of Management, 2020. [Online]. Available: https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained. [Accessed: Nov. 27, 2024].

[26] A. Chadwick, C. Vaccari, and B. O'Loughlin, *Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing*, New Media & Society, Vol. 20, no. 11, pp. 4255–4274, 2018.