

GÖRMEK İNANMAYA YETER Mİ? GÖRSEL DEZENFORMASYONUN AYIRT EDİCİ BİÇİMİ OLARAK SİYASİ DEEPFAKE İÇERİKLER

IS SEEING ENOUGH TO BELIEVE? POLITICAL DEEPFAKE CONTENTS AS A
DISTINCTIVE FORM OF VISUAL DISINFORMATION

Elif KARAKOÇ* 
Burcu ZEYBEK** 

Öz

Yapay zekâ temelli iletişim teknolojilerinin gelişimi, insan ve makine arasındaki ilişkide, bilginin manipülasyonu noktasında önemli riskler taşımaktadır. Deepfake, söz konusu ilişkide oluşturulduğu teknolojinin ismi gibi “yapay” olmaktan uzaklaştığı ve gün geçtikte gerçek gibi algılandığı bir dünya yaratmaktadır. Temel olarak yapay zekâ destekli görsel dezenformasyon olarak tanımlanan deepfake, bilgiye ilişkin güvenilirlik sorunsalını gündeme getirmiştir. Bu nedenle dezenformasyon içerikli paylaşımların bilgi manipülasyonu konusundaki risklerini tartışan literatüre, görsel dezenformasyon biçimi olarak sayılabilen deepfake penceresinden bakmanın önem arz ettiği düşünülmektedir. Deepfake, gerçek videolara çok benzeyen sentetik videoların seri üretimini mümkün hale getirmiştir. Türkiye dışında diğer ülkelerde siyasi söylem içeren deepfake videolara ulaşmak mümkündür. Bu çalışmanın çıkış noktası ise Türkiye’deki dezenformasyona ilişkin literatüre, deepfake bağlamında kapsamlı bir katkı sunmaktır. Buradan hareketle görsel dezenformasyonun bir biçimi olarak deepfake yaratma sürecinin temelini oluşturan teknoloji ele alınarak, deepfake teknolojisinin işitsel – görsel medyayı manipüle etmesini mümkün kılan yöntemler aktarılmış ve deepfake videoların nasıl işlediğini anlamayı sağlayacak bilgiler derlendikten sonra deepfake temelli manipülatif içeriklerin, siyasal iletişim bağlamındaki çıkmazları tartışılmıştır.

Anahtar Kelimeler: Deepfake, görsel dezenformasyon, bilgi manipülasyonu, siyasal iletişim

* Yeditepe Üniversitesi, Radyo, Televizyon ve Sinema, elif.karakoc@yeditepe.edu.tr

** Doğuş Üniversitesi, Halkla İlişkiler ve Tanıtım, bzeybek@dogus.edu.tr

Abstract

The development of artificial intelligence-based communication technologies carries significant risks in the relationship between human and machine, in means of information manipulation. Deepfake creates a world where the technology in question is far from being “artificial”, as the name of the technology it has created and perceived as real day by day. Deepfake, which is defined as artificial intelligence-based visual disinformation, has brought the problem of reliability inherent to information. For this reason, it is deemed to evaluate the literature that discusses the risks of disinformation content sharing on information manipulation with a deepfake perspective, which can be considered as a form of visual disinformation. Deepfake has made it possible to mass-produce synthetic videos that are very similar to real videos. It has become possible to reach deepfake videos containing political discourse in countries other than Turkey. The starting point of this study is to present a comprehensive contribution in the context of the literature concerning the disinformation deepfake in Turkey. From this point of view, the technology that forms the basis of the deepfake creation process as a form of visual disinformation has been discussed, and the methods that enable deepfake technology to manipulate the audio-visual media are presented. After compiling the information that will enable us to understand how deepfake videos work, the dilemmas of deepfake-based manipulative contents in the special context of political communication are discussed.

Keywords: Deepfake, visual disinformation, information manipulation, political communication

1. Giriş

Dijital kimliğe bürünmek, giderek daha gerçekçi ve ikna edici bir hal alırken, deepfake teknolojisi, söz konusu kimlik bürünme sürecinin en ileri teknolojisi (Chesney & Citron, 2019, s. 1758) olarak karşımıza çıkmaktadır. Deepfake, derin öğrenme ve sahte terimlerinin birleşiminden oluşmakta temel olarak görsel manipülasyon algoritmalarını eğitmek için derin öğrenme tekniklerinin kullanılması anlamına gelmektedir. (Temir, 2020, s. 1012) Yapay zekâ teknolojisindeki derin öğrenme modellemesini kullanarak sesleri suretlerle bir araya getiren deepfakeler, abartısız bir şekilde, bir kimsenin sözlerini başka herhangi birinin ağzında dile getirebilmektedir. Bireyler, kendi gözleriyle gözlemlediği ve kendi kulaklarıyla duydukları şeylerin, genellikle daha yüksek bir doğruluk seviyesine sahip olduğuna inanmaktadır. Ancak bu durum, deepfake videolarla ciddi bir zorluk sürecine geçiş yapmış ve görmenin artık inanmak anlamına gelmiyor oluşunu vurgular hale gelmiştir (Beridze & Butcher, 2019). Nitekim deepfakeler, bir kişinin hiç söylemediği veya yapmadığı bir şeyi söylüyormuş veya yapıyormuş gibi görünmesini sağlamak için oluşturulmuş videolardır (Villasenor, 2019).

Son yıllarda deepfake videoların sosyal medyada paylaşımı, artış göstermekte; tahmini mümkün olmayan bir hızda ve büyüklükte kullanıcılara ulaşabilmektedir. Bu durum, özellikle sahte haberlerde de sıkça karşılaşıldığı gibi kitlesel dezenformasyona yol açma potansiyelini taşımaktadır (Perrot & Mostert, 2020, s. 33). Kamuoyunu manipüle edici bir etkiye sahip olan siyasi deepfake içerikler, bilginin hem ses hem de hareketli görüntü aracılığıyla aktarımı ile gerçekçinden ayırt edilemeyecek bir derin sahteliğe sahiptir. Yapay zekâ teknolojilerinin her alanda kolaylıkla kullanılıyor oluşu ve internet ağının interaktif yapısı, deepfake videoların oluşum sürecini ve erişilebilirliğini kolaylaştırmış

bu da beraberinde bir dizi siyasal, teknolojik ve yasal sorunu gündeme getirmiştir (Giasano, 2019; Villasenor, 2019; Kietzman vd., 2020). Gerçeğin yeniden üretildiği deepfake içerikli videolarla, birçok farklı alanda karşılaşmak mümkündür. Mona Lisa'yı, Salvador Dali'yi yıllar sonra hayattaymış gibi konuşturabilen söz konusu videolar, sanat alanı için olumlu gibi görünürken (Yadav & Salmani, 2019; Giasano, 2019) bu durum, siyasal iletişim açısından bambaşka anlamlara ve sonuçlara sebebiyet verme olanağına sahiptir.

Deepfake algoritmaları, genellikle foto-gerçekçi görüntüler ve videolar oluşturmak için derin öğrenme modelleri eğitmek adına büyük miktarda görüntü ve video verisi gerektirmektedir. Ünlüler ve politikacılar gibi halka mal olmuş kişiler ise çok sayıda çevrimiçi video ve görüntüye sahip olduğundan, bu kişiler deepfakelerin ilk hedefleri konumundadır (Nguyen vd., 2019, s. 1). Dünyanın herhangi bir noktasında, herhangi bir ülkenin liderinin sahte bir konuşmasını içeren bir deepfake videonun sosyal medyadan paylaşımının yapılması, bu alandaki çarpıcı durumu özetler niteliktedir. Bu nitelikteki tek bir videonun paylaşımı bile çevrimiçi ortamda manipülatif söylemlerin hızlıca yayılımını sağlayabilme potansiyelini taşımaktadır. Ayrıca bu durum, kamuoyunu yanlış yönlendirerek demokrasinin kalitesi üzerinde de ciddi riskler ve tehditler doğurabilmektedir. Buradan hareketle çalışmada, deepfake içeriklerin ne olduğu, söz konusu teknolojinin nasıl işlediği ve siyasal iletişim bağlamında yarattığı çıkmazlar tartışılmış; siyasi deepfake içeriklerin, siyasal kampanyalar için taşıdığı riskler ve tehditler irdelenmiştir.

2. Gerçeğe Karşı Deepfake Teknolojisinin Yükselişi

Gün geçtikçe etki alanını genişleten yapay zekâ teknolojileri, medya tarihi boyunca değişmeyen bir konu olarak medya manipülasyonunun, değişen itici güçleri olarak karşımıza çıkmaktadır. Yapay zekâ teknolojileri ve makine öğrenimindeki son gelişmeler, sahte içeriğin gerçekçiliğinde ve oluşturulabileceği erişilebilirlikte dramatik bir artışa yol açmıştır (Agarwal, 2019, s. 38). Derin öğrenme modelleri, kendi gerçekliğini üretebilen bir sahtelikten bahsetmemize neden olmaktadır. Gerçek ve sahte kavramının ve/veya algısının iç içe geçtiği günümüzde tartışılan kavramlardan biri olarak deepfake yani derin sahte, bu bağlamda dikkat çekici bir uygulama olarak akademik dünyanın da odağına yerleşmiş durumdadır. Peki, yapay zekâ destekli multimedya dezenformasyonu (Whyte, 2020) olarak tanımlanan deepfake videolardaki derin sahteliği, gerçekliğin yitirilişi ve makine öğreniminin imkânları dâhilinde yeniden üretimi bağlamında nasıl anlamlandırabiliriz?

Medyanın gerçek üzerinde yarattığı etkilerden söz eden Baudrillard'ın (2014) teorize ettiği hiper gerçek evren, gerçeklikten yoksun gerçeklik modelleri inşa etmekteydi. Yani Baudrillard'a göre gerçek, minyatürleşmiş hücreler, matrisler, bellekler ve komut modelleri tarafından üretilmekteydi (2014, s. 13). Makine öğrenimi komutlarının verilere kazandırdığı hiper güçlerle, artık Baudrillard'ın tahayyülünün de ötesinde bir gerçek sunumuyla karşı karşıya olduğumuzu söylemek yanlış olmayacaktır. Zira, makine öğrenimi modelleriyle oluşturulan içerikler, gerçeğin aynısı olarak kendini sunmakta, gerçek ve gerçek olmayan arasındaki ayrımı hiç olmadığı kadar silikleştirmektedir. Hiper

gerçekçi güce sahip deepfake videolar (Chawla, 2019), sahte ve gerçeği sentezleyerek yeni bir gerçek tanımı yaratmıştır.

Bir başka gerçeklik tanımı olan *post truth* da deepfakeelerin barındırdığı gerçeklik içeriği ile ilişkilendirilebilir. Nesnel gerçeklerin kamuoyunu şekillendirmede duygulardan ve kişisel inançlardan daha az etkili olması durumunu ifade eden bir kavram olarak *post truth*, gerçeklerin siyasi görüşümüze tabi olması sebebiyle artık gerçeğin var olmadığı (Mcintyre, 2018) bir dünya resmi çizmektedir. Deepfake videolar, fazla gerçekçi göründükleri için gerçeklik anlayışımızda karışıklıklar yaratmaktadır. Bu videolar, kendi gözlerimizle gördüklerimizin güvenilirliğine inanma eğilimimizi kullanarak, kurguyu görünürdeki gerçeğe dönüştürebilmektedirler. Böylece gerçeğin kendisi anlaşılmaz hale gelmekte; gerçek olana dair güven zayıflamakta ve neyin gerçek neyin sahte olduğundan emin olunamamaktadır (Villasenor, 2019). Bu durum, *post truth* söylemiyle ilişkili olarak karşımıza çıkmaktadır (Giosanu, 2019). Zira, günümüzde kamuoyunu manipüle etmek için yanlış bilgilendirme kampanyaları yürüten kötü niyetli aktörlerin önderlik ettiği ve dijital dezenformasyon ile karakterize edilen bir dönemde *post truth* çağını deneyimlemekteyiz (Qayyum vd, 2019; Zannettou vd., 2019). Yaşadığımız dünyayı, *post truth* çağının riskleri altında yorumlayan Murata vd.'ne göre (2020), yapay zekâ teknolojilerinin *post truth* işleyişine uygun olarak kontrol edilmesi ya da kullanılması, sosyal bölünme ve demokrasi yıkımı gibi risklerle bizi karşı karşıya bırakmaktadır. Yani *post truth*ta içkin olan gerçeklerin saf halleriyle değil, inanılması beklendiği gibi ya da inanıldığı bilindiği gibi aktarılması durumu, deepfake ile daha “sahte” ve paradoksal olarak daha “gerçek” bir boyuta taşınmıştır. Bilginin dezenformatif aktarım pratikleri, söylemi, gerçekten öte bir boyuta taşınmanın yanı sıra artık görüntüleri de bu boyuta taşıyarak inanılabilirliği arttırmanın peşindedir. Bu noktada, söz konusu aktarım pratikleri içinde görsel dezenformasyonun etkili bir örneği olan deepfake içeriklerin oluşum sürecini anlamının yerinde olacağı düşünülmektedir.

Deepfake terimi, 2017 yılının sonlarında “deepfakes” isimli bir Reddit kullanıcısının paylaşım-ları aracılığıyla gündeme taşınmıştır. Şubat 2018’de bu kullanıcı, deepfake biçiminde porno içerikleri paylaştığı için Reddit tarafından engellenmiş (Kirchengast, 2020, s. 4) ancak o tarihten sonra deepfake videolara yönelik ilginin artmasının önüne geçilememiştir. Deepfake içerikler, fotoğraf, ses, video ve ses – video tiplerinde üretilebilmektedir:

Tablo 1: Deepfake İçerik Türleri ve İşlevleri

Deepfake türü	İşlevi
Fotoğraf deepfakeeleri	Yüz ve vücut değiştirmek, yüzde değişiklik yapmak, yüzü başka birinin yüzü ile değiştirmek ve/veya karıştırmak.
Ses deepfakeeleri	Ses değiştirmek veya taklit etmek, yeni bir metin üzerinden hedef alınan sesi konuşturmak.
Video deepfakeeleri	Bir videodaki kişinin yüzünü başka bir kişinin yüzüyle değiştirmek, bir kişinin görsel eylemlerinin başka bir kişinin yüzü ve vücuduna aktararak yansıtmak.
Ses – video deepfakeeleri	Dudak senkronu sağlamak yani konuşan bir yüz içeren videoda konuşan kişinin ağız hareketlerini ve kelimelerini değiştirmek.

Kaynak: (Kietzmann vd., 2020).

Yapay zekâ desteğiyle tamamen sentetik (Witness, 2018) hale gelen deepfake “gerçek”liği, yapay sinir ağlarına dayalı bir makine öğrenme biçimi olan derin öğrenme süreci ile gerçekleşmektedir. Deepfake video oluşturma süreci, 2014 yılında makine öğrenimi direktörü Ian Goodfellow tarafından icat edilen ve bir makine öğrenimi tekniği olan GAN’a (Generative Adversarial Network) dayanmaktadır (Beridze & Butcher, 2019; Giosanu, 2019; Westerlund, 2019). İlk olarak mevcut veri kümelerinden algoritmik yeni veri kategorileri oluşturmak için kullanılan GAN, sonrasında milyonlarca insan fotoğrafını işleyerek mevcut olmayan insan görselleri yaratan bir model halini almıştır.¹ Özetle GAN’ın amacı, gerçeğe ayırt edilemeyecek kadar benzeyen yapay görüntüler sentezlemektir (Giosanu, 2019, s. 68).

Derin öğrenme modellerinde kullanılan algoritma türlerinden biri olarak deepfake algoritmaları, gerçek olandan ayıramayacak sahte videolar ve görüntüler oluşturmayı (Nguyen vd., 2019) sağlayan talimatlar dizisidir. Deepfake işlemini, derin öğrenme modelleri ile mevcut görüntüleri ve videoları birleştirip kaynak görüntüleri veya videolara yerleştirerek sentezleme işlemi (Giasano, 2019, s. 67) olarak özetlemek mümkündür.

Derin öğrenme modelleri, iki boyutlu görüntülerden üç boyutlu derinliği tahmin edebilmekte; banka çekleri, plakalar, vergi formları, resmi belgeler vb. üzerindeki rakamları ve harfleri tanıyabilmektedir. İşlenen verilerin eğitim süreci, benzeri görülmemiş bir başarıya yol açabilmektedir. Ancak derin öğrenme sürecinde deepfakeleri ortaya çıkaran sentezlenen veriye ulaşabilmek için çok fazla bağlantılı veriye ihtiyaç duyulmaktadır. Ünlülerin ve siyasi liderlerin deepfakelerin başlıca hedefinin olmasının sebebi de budur. Çünkü çevrimiçi ortamlarda bu kişilere ait verileri e işleyecek kadar geniş bir resim ve video kütüphanesine ulaşmak mümkündür (Kietzmann vd., 2020, s. 138 – 139).

Westerlund (2019), paylaşılma amaçları doğrultusunda deepfake üreticilerini dört temel kategoride sıralamaktadır:

1. Deepfake videoları eğlence/ hobi olarak üreten topluluklar,
2. Siyasi aktörler ve aktivistler,
3. Manipülatif kötü niyetli aktörler,
4. Televizyon şirketleri gibi meşru aktörler.

İlk olarak eğlence amacıyla yaratılan deepfake teknolojisi, popülerlik kazanmaya başladıkça, ünlülerin ve dünya liderlerinin yüzlerini algoritmik bir tasarımla sentezleyerek manipülatif konuşmalar sergilemelerine yol açacak biçimde kullanılmaya başlanmıştır (Yadav & Salmani, 2019, s. 855). Birçok çalışma çevrimiçi dezenformasyon kampanyalarının sınırlarına dair iç görüler sunmaktadır. Dobber vd. ise (2021) deepfake dezenformasyonunun Twitter’daki troller ve yanlış haber sunan

1 Bknz: <https://thispersondoesnotexist.com/> Site her yenilediğinde GAN tarafından yaratılan bir insan fotoğrafı ile karşılaşılmaktadır. Bunun yanı sıra GAN, mevcut olan bir kişinin fotoğrafını işleyerek, o kişinin hiç var olmayan bir görüntüsünü de fotoğraf olarak yansıtılabilmektedir.

makalelerden çok daha büyük etkilere sahip olabileceğini savunmaktadır. Vaccari ve Chadwick (2020) Birleşik Krallık'ta yapılan bir çalışmada deepfakeelerin neyin gerçek neyin sahte olduğu noktasında bireylerin aklını karıştırdığını, bunun sonucunda da sosyal medyadaki haberlere güvenin azaldığını bulmuşlardır. Zimmermann ve Kohring (2020) ise bireylerin haber medyasına olan güvenleri azaldıkça dezenformasyona inanma olasılıklarının arttığını, zaman içinde bu durumun oy verme davranışlarını da etkileyebileceğini gözlemlemişlerdir.

Çalışmamızda konu edindiğimiz üzere deepfakeeler, bilginin aktarımı noktasında riskler taşımaktadır. Ancak bu işleyecek teknolojinin kendisinin tarafsız olduğu gerçeğini de vurgulamak yerinde olacaktır. Teknolojinin iyiye ya da kötüye yönelik kullanımına içkin paradoksal durum, şüphesiz deepfakeeler için de geçerlidir. Deepfake teknoloji, eğitim ve sanat alanında efektif olarak kullanılabilir (Yadav & Salmani, 2019; Giasanu, 2019). Ancak deepfakeelerin kitlesel dezenformasyon yaratımı ve sosyal medya etkisiyle yayılımı noktasında büyük riskler taşıdığından söz etmek gerekmektedir. Kietzman vd. (2020, s. 136), deepfakeelerin temel risklerini, “sahte içeriğin daha inandırıcı geliyor oluşu ve sahte içerik oluşturmanın daha kolay oluşu” şeklinde özetlemiştir.

Deepfake dezenformasyonunun toplumsal açıdan zarar verici etkileri olduğuna inanmak için çeşitli nedenler bulunmaktadır. Tek başına bu durum dahi deepfake dezenformasyonunun bilimsel incelemelere tabi tutulmaya değer olduğunu göstermektedir. Birincil olarak, deepfakeeler gerçekçi dezenformasyon olabilir. Otomatik olarak hazırlanmış görüntüler ve sesler, en az gerçek ses ve görüntüler kadar ikna edici olabilmektedir. Sıradan bir yurttaş ise bu kurguyu gerçekten ayırt edemeyebilir. İkincisi, deepfakeeler yanlış bilgiyi, çarpıtılmış bilgiyi ya da kötü niyetli olarak bilinçli yayılan bilgiyi genişletmek kastıyla da kullanılabilir. Herhangi bir yapımcı, Papa'nın Donald Trump'ı desteklediğinin görülebileceği ve duyulabileceği ya da kamu sağlığından sorumlu bir yetkilinin açık ve seçik olarak aşılardan kesinlikle otizme neden olduğunu iddia ettiği bir deepfake hazırlayabilir. Üçüncü olarak ise, deepfakeeler aynı zamanda etkili birer dezenformasyon biçimi olabilir. Eğer siyasi bir aktör yeterli eğitim verisine sahip olursa çok kısa bir sürede aynı kişinin pek çok farklı ancak gerçekçi deepfakeini hazırlayabilir. Mikrohedefleme (PMT) teknikleri ile birlikte deepfakeeler özellikle etkili hale gelebilir (Dobber, vd., 2021, s.70). Henüz durum bu noktada olmasa da, (mikrohedefli) deepfakeelerin yakın gelecekte çok kuvvetli dezenformasyon biçimleri teşkil edecek özelliklere sahip olduğunu da belirtmek gerekmektedir.

Deepfake videolar, kamuoyunu manipüle etmek, bir ülkenin kurumlarına olan güveni azaltmak amacıyla birtakım dezenformasyon kampanyalarına dahil olabilirler. Çalışmamızın ilerleyen bölümlerinde de sıkça vurgulayacağımız üzere deepfakeeler, seçimlere müdahale etmeyi ve sivil kargaşa eklemeyi amaçlayan silahlı dezenformasyonlardır (Westerlund, 2019). Siyasal kullanımlarının yanı sıra deepfakeeler, siber kriminal niteliğe de sahiptirler. Para piyasaları ve hisse senedi manipülasyonları, sahte sesler kullanarak telefonda dolandırıcılık, bu niteliklere örnek olarak verilebilmektedir (Westerlund, 2019). Deepfake fotoğraf ve videolar ile ilgili en büyük endişelerden biri, bu içeriklerin pornografik içerik (derin sahte pornografi) oluşturmak için kullanılmasıyla ilgilidir (Giosanu, 2019). Deepfake videolar, kişilere yönelik hakaret yayma, sahte haberlerin dolaşımına neden olma, manipülatif

yapısı sebebiyle demokrasiye ve yargı sistemlerine zarar verme potansiyeli taşıyarak (Buo, 2020) davaların bile seyrini değiştirebilecek niteliktedir (Yadav & Salmani, 2019, s. 855). Deepfake videolar, bireysel itibarı ve halkın güvenini ciddi şekilde zedeleme potansiyeline de sahiptir. Ek olarak, gerçeklere yönelik fikir birliğini aşındıran deepfake içerikler, sosyal bölünmeyi şiddetlendirebilir. Her düzeyde bilgiye ve/ veya habere olan güveni zayıflatarak yaygın bir kinizme ve ilgisizliğe yol açabilir (Beridze & Butcher, 2019).

Deepfake içeriklerin taşıdığı riskler, deepfakelerin üretiminin erişilebilirliğine paralel olarak oldukça kolay biçimde internet kullanıcılarına yayılmaktadır. Nitekim, deepfake teknolojisine benzer teknolojiler kullanan Snapchat, FaceApp gibi popüler uygulamalar, sahte ve gerçek sentezlemesini kullanıcıların kendi yüzleriyle –derin öğrenmeye dair herhangi bir teknik bilgi kullanmadan – deneyimlemesini sağlayan uygulamalardır. Ancak, bu içeriklerin toplumun bir kısmını ya da tamamını ilgilendiren konuları ya da kişileri hedef alarak kötü niyetle oluşturuluyor oluşu, politik çerçevede bir dizi risk ve tehdidi de beraberinde getirmektedir. Dolayısıyla bu risk ve tehditlerin ortaya çıkarılarak çözümlenmesine yönelik adımların atılması, bilgi güvenliği açısından oldukça önemlidir.

3. Siyasal İletişimde Dezenformasyonun Görsellik ile Yenilenen Gücü

Siyasal aktörlerin, iktidara gelmek ya da iktidardaki varlıklarının meşrutiyetini sağlamak için kamuoyunu yönlendirmeye yönelik girişimleri olarak tanımlanan siyasal iletişim alanında, geleneksel tek yönlü iletişime hâkim olan süreçten farklı özelliklere sahip yeni iletişim teknolojileri ve özellikle bu teknolojiler kapsamında etkinleşen sosyal ağlar, söz konusu alanın öne çıkan araçlarıdır. Sosyal ağlar, seçmenleri etkilemek üzere üretilen görüntülü ve/ veya yazılı içeriğin hızla yayılmasını, alternatif bir gündem oluşmasını, video ve fotoğrafların paylaşılabilmesini ve güncel içeriklere erişilmesini mümkün kılmaktadır (Westerman, Spence & Van Der Heide vd., 2014). Sosyal ağların siyasal iletişimde gittikçe daha fazla merkezi bir rol oynuyor olması gerçeği göz önünde bulundurulduğunda, dezenformatif içeriklerin özellikle demokratik süreçler kapsamında oluşturduğu risklere de dikkat çekilmektedir (Thorson & Wells, 2016; Vaccari, 2017). Sosyal ağlar ile küresel çapta bir artış yaşayan dezenformatif içerikler, pek çok ülkede ciddi etkilere sebebiyet vermeye başlamış ve siyasal iletişimde de kitlesel bir tehdit unsuru haline gelmiştir. Özellikle son yıllarda sosyal ağlar aracılığıyla kişisel verilerin kullanımı ile oluşturulan manipülatif siyasal iletişim ortamı, “Cambridge Analytica” örneğinde de görüldüğü gibi, ABD Başkanlık seçimlerinden Brexit referandumuna kadar pek çok önemli olayda yanıltıcı, kışkırtıcı haberin ve dezenformasyonun kullanıldığını göstermiştir (Aktan, 2018). Sosyal ağlar aracılığıyla dezenformasyonu arttıran bir diğer unsur, özellikle Twitter’da iç siyasette kamuoyu oluşturmak için kullanılan bot hesaplardır. Philip Howard bot hesapları, daha çok internetteki tartışmaları yönlendirmek isteyen politikacıların kullandığını belirtmektedir. Bot hesaplar aracılığıyla bir siyasetçiye desteğin mevcut durumdan daha yüksek olduğu algısı yaratılabilmektedir (Baerthlein, 2016).

Toplumların derin siyasi çekişmeleri ve kutuplaşmaları dezenformasyonun paylaşımı ve tüketimine katılım ile doğrudan bağlantılı olduğu için, dezenformasyonun farklı biçimleri doğru bir şekilde anlaşılmalıdır. Artık pek çok siyasi parti taraftarının, rakip partinin mensuplarına yönelik tutumunu yoğun olumsuz duygular karakterize etmekte ve bu durum, çevrimiçi mecrada bilgi paylaşımında da karşılık bulmaktadır (Messing & Westwood, 2012). Sonuç olarak, günümüzün çevrimiçi tüketicileri şüpheli, manipülatif içeriklere yoğun bir şekilde maruz kalmaktadır. Özellikle bu içeriklerin görsel ve işitsel olması halinde daha etkili olacağı bilinmektedir. Garbe (1990), bu konuda referans niteliği taşıyan bir deneyde, televizyon izleyicilerinin, görsel mesajları sözlü mesajlardan daha doğru bir şekilde hatırladığını bulgulamıştır. Grabe ve Bucy (2009), görüntü kesitlerinin (adayların görüldüğü fakat duyulmadığı klipler), seçmenlerin fikirlerini şekillendirmek açısından “ses kesitlerinden” (adayların konuşmalarının, konuşma görüntüleri ile veya bu görüntüler olmadan duyulduğu içerikler) daha güçlü olduğunu belirtmiştir. Prior (2013) ise gerçekleri hatırlamaya yönelik soruların hem görsel hem de sözlü bilgileri içermesi durumunda, anket katılımcılarının daha yüksek bilgi seviyeleri gösterdiğini bulgulamıştır.

Görseller, yurttaşların anılarını zihinlerinde tekrar canlandırıp hatırlamalarına yardımcı olarak bilgi aktarımını geliştirmektedir. Stenberg (2006), bireylerin görsel bilgileri sözlü bilgilerden daha dolaysız bir şekilde ve daha az çabayla işlediklerini aktarmıştır. Witten ve Knudsen (2005, s. 490), algılanan “kesinlik” noktasında görsel bilginin diğer duyuşsal verilerden daha etkili bir şekilde entegre edildiğini savunmaktadır. “Gerçeklik hüristiği”ne göre (Frenda vd., 2013; Sundar, 2008) bireyler, ses ve görüntü içeriklerinin, günlük deneyimin “gerçek dünyasına” metin içeriklerinden daha fazla benzediğini düşündüğü için, yanıltıcı görsellerin yanlış algılar yaratma olasılığı sözlü içeriklerden daha fazladır.

Görüntülerin ve sesli-görsel içeriklerin anlaşılmasının ve işlenmesinin, yazılı metinlerden daha kolay olması, “üst bilişsel deneyim” konusunu, yani bilgileri işlemek gibi görevlere gösterdiğimiz tepkileri şekillendiren ve deneyim yoluyla türetilmiş, düşüncelerimiz hakkındaki duyguları gündeme getirmektedir (Schwarz vd., 2007). Böyle bir deneyimde, “akıcılık” konusu, insanların yanlış bilgilere neden inandığını anlamak açısından özellikle önem taşımaktadır. İnsanların aşına olarak algıladıkları mesajları, gerçek gibi kabul etme olasılıkları daha yüksektir (Berinsky, 2017). Aşinalık yani “doğru gibi gelme etkisi”, içeriklerin daha kolay özümsemesini ve dolayısıyla daha inanılır olarak değerlendirilmesini sağlayan bir akıcılık duygusunu açığa çıkarmaktadır (Newman vd., 2015). Barari, Lucas & Munger’in (2021) siyasi deepfake içerikleri konu edindiği araştırmada, bireylerin siyasi içerikleri işlemede birtakım bilişsel özelliklerin ve becerilerin öne çıktığı bulgulanmıştır. Araştırmaya göre (Tablo 2), siyasi parti taraftarları, motive edilmiş akıl yürütme mekanizması yani önceki inançlarla tutarlılığa dayanan bilginin seçici olarak kabul edilmesi ile deepfake içerikleri alımlama potansiyeline sahiptir. Ek olarak, genel ölçekte, dijital okuryazar olmayan kişilerin deepfake içeriklerin yarattığı manipülatif etki ortamında risk altında olduğu anlaşılmaktadır.

Tablo 2: Deepfake manipülasyonuna karşı varsayımsal olarak hassasiyete sahip alt gruplar

Alt Gruplar	Manipülasyonu Sağlayan Mekanizmalar
Yetişkinler (65+)	Dijital bilginin doğruluğunu değerlendirememesi
Siyasi parti taraftarları	Motive edilmiş akıl yürütme
Cinsiyetçiler	Motive edilmiş akıl yürütme
Düşük bilişsel yansımaya sahip kişiler	Yargılamada sezgiye aşırı güvenme
Düşük siyasal bilgiye sahip kişiler	1. Siyasi olayların akla yatkınlığını değerlendirememesi 2. Siyasi aktörlerin gerçek yüz özelliklerini tanıyamama
Düşük dijital okuryazarlık oranına sahip kişiler	1. Dijital bilgilerin doğruluğunu değerlendirememesi 2. Deepfake teknolojisi hakkında sızır ya da sınırlı farkındalık
Düşük doğruyu saptama becerisine sahip kişiler	Medyanın gerçek üretimine karşı sızır ya da sınırlı dikkat etme
Deepfake hakkında bilgisizlik olan kişiler	Deepfake hakkında sızır farkındalık

Kaynak: (Barari, Lucas & Munger 2021).

Siyasi deepfake içeriklerin bireyler tarafından işlenmesinin ya da algılanmasının yanı sıra bu içeriklerin sosyal ağlarda dolaşıma girmesinde, paylaşım davranışları da önem arz etmektedir. Örneğin, videolar ve durağan görüntüler, Twitter’da haberler ve çevrimiçi imza kampanyalarından daha fazla yayılmaktadır (Goel vd., 2015, s. 186). Birleşik Devletler 2016 Başkanlık Seçimleri sırasında Donald Trump ile Hillary Clinton’ın görüntü ve videolar içeren tweetleri, dikkate değer ölçüde daha fazla beğeni ve retweet almıştır (Pancer ve Poole 2016).

Medyanın dezenformasyonla yollarının kesiştiği noktadaki deepfake içerikler, hâlihazırda ünlü ve/ veya kamuya mal olmuş kişileri betimliyorsa teknik gerçekçilikleri dolayısıyla ve aşinalık aracılığıyla, akıcılığın yaratılabileceği ciddi sorunların endişe verici hale gelmesine yol açabilmektedir. Siyasi aktörlerin kendi bireysel açıklamasıymış gibi gösterilen sahte içeriklerin, doğrulanmasına yerince dikkat edilmemesi durumunda, video yayını yapan ve yayan kullanıcıların kendileri de dezenformasyon kampanyaları tarafından kullanılarak “pazarlama zincirinin bir parçası” haline gelebilirler. Bu bağlamda dezenformasyon kavramını ortaya koyarken, dezenformasyonla ilgili kaygıların oldukça ani bir şekilde ortaya çıktığı siyasi, iletişimsel ve ekonomik ortama dair doğru ve net bir bakış açısı kazanmak hedeflenmelidir.

BuzzFeed, 2018 yılında eski ABD Başkanı Barack Obama’nın Oval Ofis’e benzeyen bir yerde doğrudan kameraya konuşurken gösterildiği bir video yayınlamıştır. Videonun ilk 35 saniyesinde sadece Obama’nın yüzü gösterilmektedir. Obama, kendi kişiliğinden beklenmeyecek nitelikte birkaç açıklama yapmakta ve: “Başkan Trump tam bir pislik.” dedikten sonra kısa bir duraklamanın ardından şöyle devam etmektedir: “Şimdi. . . görüyorsunuz ya, ben böyle şeyleri, en azından halka hitap ettiğim bir konuşmada asla söylemezdim, ama başka biri söylerdi . . . başka biri. . . örneğin Jordan Peele.” Bu noktada BuzzFeed videosunun gerçek niyeti açığa çıkmaktadır. Konuşan aslında Obama değildir. Ekran bölündükten sonra solda Obama, sağda ise Amerikalı ünlü aktör ve yönetmen Jordan Peele görünmektedir. Obama’nın ve Peele’in yüz ifadeleri ve dudak hareketleri “gerçekçi” bir şekilde örtüşmektedir. Peele’in prodüksiyon ekibi, yapay zekâdan yararlanarak Obama’nın yüzünü kendi yüzünü yansıttak şekilde yeniden yapılandırmıştır. Bu video, “Obama’nın Bu Videoda Neler Söylediğine

İnanamayacaksınız!” başlığı ile internette hızla yayılarak, Facebook’ta 5 milyon kez görüntülenip, 83.000’in üzerinde paylaşım almıştır. Ayrıca video, YouTube’da 5 milyondan fazla ve Twitter’da ise yaklaşık 5 milyon kez görüntülenmiş ve neredeyse 52.000 kişi tarafından retweet edilmiştir (Twitter, 2018).



Resim 1: BuzzFeed tarafından yayınlanan Obama için yapılan deepfake video görüntüsü

Kaynak: <https://www.youtube.com/watch?v=cQ54GDm1eL0>

Rusya Devlet Başkanı Vladimir Putin ve Kuzey Kore lideri Kim Jong-Un’un Amerikalıları uyaran söylemlerini içeren iki siyasi içerikli deepfake video, 29 Eylül 2020’de sosyal medyada yayınlanmıştır. Kasım ayında ABD başkanlık seçimleri için üretilen bu deepfake içerikler, seçim manipülasyonu oluşturmak noktasında sentetik gerçeklikte söylemler ve görseller sunmaktadır. Videolar, Amerikalıları demokrasilerinin ne kadar kırılgan olduğu konusunda uyarmakta ve olası saldırılara karşı oy kullanma haklarının korunmasına yönelik teşvik edici söylemleri içermektedir.



Resim 2: Kuzey Kore lideri Kim Jong Un ve Rusya Devlet Başkanı Vladimir Putin’in Amerikalıları demokrasilerinin tehlikede olduğu konusunda uyaran deepfake video görüntüsü

Kaynak: (Kahn, 2020).

Birleşik Krallık Kraliçesi II. Elizabeth de deepfake içeriklere konu olan ünlü isimlerden biridir. 25 Aralık 2020’de geleneksel Noel konuşmasının hemen ardından yayınlanan videoda dans ettiği görülen kraliçe, aynı zamanda ailesi ile ilgili kişisel söylemlerde de bulunmaktadır. Prens Harry ve eşi Meghan Markle’in Kuzey Amerika’ya taşınması ile ilgili olarak “Birisinin sizin yerinize Kanadalılarla birlikte olmayı tercih ettiğini söylemesi kadar acı veren çok az şey vardır” yorumunu yapan Kraliçe’nin deepfake içerikli görüntüleri, ülkede muhafazakâr kesimin tepkilerine neden olmuştur.



Resim 3: Tik Tok danslarından birini yapan Kraliçe Elizabeth’in deepfake video görüntüsü

Kaynak: BBC, 2020.

Siyasilere yönelik bir diğer deepfake içerik ise, Belçika Başbakanı’nın videosudur. Belçika Başbakanı Sophie Wilmès’in ekonomik ve iklim krizleriyle acil mücadele ihtiyacından bahsettiğini gösteren bir deepfake videosu, Extinction Rebellion (XR)² Belçika tarafından dolaşıma sokulmuştur. Wilmès videoda, “Koronavirüs görmezden gelemeyeceğimiz bir alarm zili... Pandemiler, daha derin bir ekolojik krizin sonuçlarından biridir.” şeklinde konuşmaktadır.



Resim 5: Belçika başbakanı Sophie Wilmès’ Covid-19’u iklim kriziyle ilişkilendiren deepfake video görüntüsü

Kaynak: The Brussels Times, 2020.

2 XR, 2018 yılında küresel bir iklim hareketi olarak ortaya çıkmıştır. Hareketin amacı, hükümetlerin aldığı kararlara sivil itaatsizlik göstererek çevreyi korumaktır. Herhangi bir örgüt yapısı bulunmayan XR, küresel alanda kendilerine destek vermek için protestolara katılan herkesi aktivistlerden biri olarak kabul etmektedir.

Sonuç olarak görsel dezenformasyonun bir biçimi olarak deepfake videolar, siyasal iletişimde de güvenilirlik sorununu gündeme getirmektedir. İnsanların gerçeklik algısını etkileyerek, onları manipüle edebilen deepfake içerikler, işitsel ve görsel medyanın desteğiyle özellikle kamuoyunu etkili ve kolay biçimde yönlendirir hale gelmiştir. Çünkü insanlar tarafından izlenen ilk içerik, doğru ya da yanlış değerlendirmesine takılmaksızın algıyı yönlendirebilmektedir.

4. Deepfakeler ile Siyasal Kampanyaların Önündeki Riskler ve Tehditler

İnternetin görsel dönüşümünün ve video destekli görsel dezenformasyonun yeni bir biçimi olarak siyasi deepfake içerikler, gazetecilik, yurttaşların yetkinliği ve demokrasinin kalitesi üzerinde ciddi riskler ve tehditler doğurabilmektedir (Bennett & Livingston, 2018; Chadwick, Vaccari & O'Loughlin 2018; Flynn, Nyhan & Reifler 2017; Rojecki & Meraz, 2016; Waisbord, 2018). Bilhassa, deepfake teknolojisinin kolay erişilebilir olması ve bu teknolojiyle üretilen içeriklerin çevrimiçi ortamda yayılma hızının yüksek olması, demokratik politik katılım noktasında tehditler taşımaktadır (Hall, 2018, s. 59). Kaynağı ve doğruluk değeri şüpheli içeriklerin demokrasinin işleyişini bozabileceği korkusu; hükümetler, vatandaşlar ve bilim insanlarında bu olguyu anlama ve bu olguya mücadele etme motivasyonu doğurmuştur.

Westerlund (2019), deepfakelerin topluma ve siyasi sisteme yönelik tehditlerini aşağıdaki gibi özetlemiştir:

- 1) Gerçeği sahte haberlerden filtrelemeye çalışan gazetecileri engeller.
- 2) Propaganda yaparak ve seçimlere müdahale ederek ulusal güvenliği tehdit eder.
- 3) Vatandaşların bilgiye olan güvenine zarar verir.
- 4) İnsanlar ve kuruluşlar için siber güvenlik sorunları gündeme getirir.

Vaccari ve Chadwick (2020), çalışmalarında, siyasi deepfake videoların belirsizliğe neden olabileceğini ve bunun da sosyal medyadaki haberlere olan güveni azaltabileceği sonucuna ulaşmışlardır. Uzun vadede bu etkiler, sorunlu normlar ve davranışlar ortaya çıkararak çevrimiçi kültüre yayılma potansiyeli taşımaktadır. Güvenin azaldığı bağlamlarda bireylerin işbirliği yapma olasılığını da düşürecek olan bu durum, kutuplaşmış siyasete ve çatışmalara işaret etmektedir (Vaccari & Chadwick, 2020, s. 3). Bu noktada deepfakelerin sahte haberlerle ilişkisini kurmak da güç değildir. Zira deepfakeler de tıpkı sahte haberler gibi bir güvensizlik iklimi yaratmaktadır.

Deepfake videolar konusundaki yasal düzenlemeler, teknolojik yeniliklerin gerisinde kalmaktadır. Örneğin, ABD'de birkaç eyalet, konu ile ilgili bazı mevzuatlar çıkarmış ancak bu mevzuatlar, "kapsam, cezalar ve odak" açısından bir eyaletten başka bir eyalete değişiklik gösterdiğinden dolayı kapsayıcı olamamıştır (Ruiz, 2020). Gelecekte çıkarılacak mevzuatların, deepfake içeriğin genel meşruiyetini, içeriği yaratmak için kullanılan algoritmaların uygunluğunu ve veri sahipliğini – verilerin kime ait olduğu, verilerin nerede, ne kadar süre saklandığı, nasıl korundukları ve sonunda nasıl

ortadan kaldırıldıkları – değerlendirmesi beklenmektedir. Yürürlükteki bazı yasalar, deepfakeelerde tasvir edilen bireylere odaklanmaktadır; örneğin New York'ta bir yasa tasarısı, dijital miras ve ölüm sonrası gizlilik ve ölmüş bir kimsenin dijital benzerliği ile ilgili hakların nasıl yönetileceği ile ilgilenmektedir (Howes, 2018).

Sosyal medya platformları, fikirler, süreçler ve düşüncelerle ilgili rekabet haklarını koruyup sürdürme çabasıyla, düzenleyici sorumluluğunu genellikle politika yapıcılardan önce yönetmektedir. Örneğin Facebook, 2019 yılında “Deepfake Tespit Mücadelesi Yarışması” düzenlemiştir. (Coldewey, 2020) *Deep Trust Alliance* ise 2020 yılında ve gelecekte Birleşik Devletlerde gerçekleşecek başkanlık seçimlerinin deepfake içerikler dolayısıyla sekteye uğramaması için ciddi çalışmalar yürütmüştür (Solsman, 2020). Teknoloji şirketleri, gerçek içeriklerin deepfake içeriklerden ayrılarak tanımlanmasına yardımcı olmak adına içeriklerine filigran veya dijital imzalar koyabilmektedir. Kapsayıcı bir yasal çerçevenin olmaması ve buna bağlı olarak ortaya çıkan belirsizlik ortamı, deepfake video üretimini potansiyel olarak sorunlu hale getirmektedir. Diğer yandan kampanya yürütücüler tarafından ilgili tüm tarafların onayı ile yaratılan hilesiz deepfake içeriklerin, henüz yasa ve yönetmeliklere tabi olmaması, siyasal kampanyaların, deepfake teknolojilerinden yararlanan içeriklerin oluşturulması ve dağıtımı üzerinde yaratıcı bir kontrole sahip oldukları anlamına gelmektedir.

Seçmenlerin deepfake içerikleri algılaması ve yorumlaması ile ilgili olarak, yapay zekâ tarafından oluşturulan içeriklerle bağlantılı risklerin anlaşılması için çok daha fazla çalışmaya ihtiyaç vardır. Yapay olgu vasıtasıyla araçsal değer sunmaya odaklanan teknologlar, dijital teknolojinin paradoksal etkilerini gözden kaçırabilmektedir. Örneğin, Google'ın Deplex ürünü ile insansı sanal telefon yaratma başarısı (Chen & Metz 2019), insanların, telefonda konuştukları kişinin gerçek bir insan olmadığını anlamalarının oldukça zaman almasına neden olmuştur.

Deepfakeeler ile sentetik, sentetik olmayandan ayırt edilemez hâle gelmiştir ve bu nedenle, gelecekteki araştırmalar, siyasi deepfake içeriklerin seçmenler için gerçekçi olup olmayacağını değerlendirme noktasında önem taşımaktadır. Benzer olarak, insanların aşırı kişiselleştirilmiş içeriklere nasıl tepki verecekleri henüz açıkça bilinmemektedir. Kampanya yapıcılar, her bir seçmen için daha önce hiç olmadığı kadar kişiselleştirilmiş video yarattıklarında seçmenlerin, gizliliklerinin algılanan ihlalinin nasıl yorumladığı ya da yorumlayacağı da bir araştırma sorunsalı olarak karşımıza çıkmaktadır.

Dijitalleşmeyle birlikte siyasi içeriklerle ilişkilendirilebilecek en büyük sorunlardan biri mesenformasyon, dezenformasyon ve sahte haberlere yönelik şüphecilik eğilimlerinin ortaya çıkmasıdır (Nyilasy, 2019). Zira sahte içeriklerle sıkça karşılaşmak, sahte içeriği tespit etme ve tepki verme noktasında birtakım boşluklar yaratabilmektedir. Örneğin “Yalancının Kazancı (Liar's Dividend)” olarak adlandırılan kavram (Chesney & Citron, 2019), bazı politikacıların yanlış bilgiden ve sahte içerikten yararlandığını öne sürmektedir. “Yalancının kazancı”, bu yararlanma durumunun, sahteciliğe olan inancı körükleyen tehlikeli bir etkiye sahip olabileceğini ortaya koymaktadır. Başka bir örnek ise “Uyuyan Etkisi (Sleeper Effect)” (Hovland ve Weiss, 1951) kavramıdır. Bu etkiye göre bireyler, sahte içeriğe maruz kaldıklarını bilseler hatta güvenilirliğin yanıltıcı olduğunu görseler dahi, ikna edici

etkiler, bireylerin algılama süreci üzerinde uzun süre geçerli olmaya devam etmektedir. Başka bir deyişle seçmenler, yapay zekâ tarafından oluşturulan siyasi içeriklere karşı şüpheli olsa bile, bu içerikteki iletiler, uzun vadede daha ikna edici olabilmektedir.

Deepfake içeriklerin oluşturduğu tehdit ve risklere karşı deepfake tespit teknolojisi, gittikçe büyüyen bir araştırma alanı haline gelmiş durumdadır. Agarwal ve Farid (2019, s. 44), deepfake videoların tespit edilmesine yönelik gerçekleştirdikleri çalışmalarında, deepfake videolarda konuşan kişilerin yüzlerindeki farklı ve tutarsız yüz ifadelerine dikkat çekmişlerdir. Perot ve Mostert (2020), deepfake görüntülerdeki gölgelendirme gibi birtakım görsel uyumsuzluklarının; Wagner ve Blewer (2019), yüz ifadeleri ile baş hareketleri arasındaki uyumsuzluğun ve kameraya yöneltilen bakışlardaki göz temasına yönelik tutarsızlığın, sezgisel olarak fark edilebilir olduğunu altını çizmiştir. Yadav ve Salmani'ye göre deepfake algoritmaları, insan gözünün kırpma hareketlerini taklit etmede çok başarılı değildir. Bu sebeple kişinin göz hareketlerine odaklanılarak deepfake videonun tespiti sağlanabilir (2019, s. 855). Bu bağlamda, görsel bilgi okuryazarı olmanın, deepfake videoların yaratacağı manipülasyonun etkisinin azaltılmasında büyük rol oynayacağını altı çizilmektedir (Wagner & Blewer, 2019, s. 44). Ayrıca deepfake yaratımını sağlayan derin öğrenme teknolojisi, aynı zamanda deepfake videoların tespit edilmesi noktasında da kullanılabilir durumdadır. Deepfake tespitine yönelik en kapsamlı çalışmalardan biri ABD İleri Savunma Araştırma Projeleri Kurumu (DARPA)'da *Media Forensics* programı tarafından sürdürülmektedir (Villasenor, 2019). Geliştirilen yazılımlar, derin öğrenmeyi kullanarak videolardaki sahte içerikleri ve gerçek olmayan yüz hareketlerini tespit etmekte böylece deepfake videoları ayırt edebilmektedir (Yadav ve Salmani, 2019). Deepfake videoların tespit edilmesinde “tekrarlayan sinir ağları” (recurrent neural networks) da kullanılmaktadır (Guera ve Delp, 2018). Yapay sinir ağları, algoritmaları oluşturan bileşenlerden biridir ve deepfake videoların yapısındaki sahte içeriklerin tespit edilmesinde de kullanılabilir durumdadır (Yadav & Salmani, 2019).

Türkiye’de de deepfake tehditlerine yönelik siber güvenlik alanı oluşturmak adına çalışmalar gerçekleştirilmektedir. Türk Asya Stratejik Araştırmalar Merkezi (TASAM)’ın 2019 tarihli “Sentetik Gerçeklik Teknolojisi (Deepfake): Derin – Sahte Ürün ve Savunma Ekosistemi İnşası” başlıklı raporda, ülkelerin deepfake içeriklere karşı savunmasız olduğu ve Türkiye’nin de bu büyük tehditten korunması gerektiğine dair aktarımlar yapılmıştır. Aynı zamanda raporda, Türkiye’nin deepfake tespit teknolojisindeki araştırma çalışmalarından söz edilmiştir. Raporda da adı geçen Zemana adlı şirketin geliştirdiği *deepware*, deepfake içeriklerin tespitinde kullanılan bir yazılım olarak piyasaya sunulmuştur (<https://deepware.ai/>).

Deepfake tehdidinin oluşturduğu riskler bağlamında bu risklerin azaltılması ve/veya yok edilmesi üzerine çalışmalar sürmektedir. Elbette bu noktada etkili sonuçların elde edilmesi, deepfake videoların tespitine yönelik teknolojik çözümlerin kullanım yaygınlığının sağlanması doğrultusunda gerçekleşecektir.

5. Sonuç

“Görmek, sadece görmektir; inanmak değil.”

George MacDonald, The Princess and the Goblin

Yukarıda yaptığımız alıntı, yerleşik yargılara sahip zihnimiz ile deepfake videolar arasındaki ilişkiye vurgu yapmaktadır. Zira, gerçeklik sonrası çağda zihnimiz, gördüğü şeye inanma noktasında güvenilirliğe dayalı engellemelerle karşılaşmaktadır. Sahtenin gerçekle sentezlendiği deepfake videolar, gerçeğe yönelik güvenimizi sarsmaktadır. Bugün dijital çağda karşılaştığımız her bilgi kırıntısına karşı şüpheli davranıyor oluşumuz ya da tam tersi, gördüğümüz ve inanmaya açık olduğumuz her görseli benimliyor oluşumuz, deepfake videoların gerçeklik algımıza yönelik meydan okumasını açığa çıkartmaktadır.

Deepfake içeriklerin oluşturduğu en başat sorun, manipülatif yapısı gereği gerçeklik algısında yaratacağı olası tahribatlarıdır. Bu sorun doğrultusunda bireyler, izlediklerinin gerçek mi yoksa sahte mi olduğu konusunda kararsız hale geleceklerdir. Bireylerin, birbiriyle çelişen görüşlere sahip oldukları bir seçimle karşı karşıya kaldıklarında ortaya çıkan kararsızlık gibi belirsizlik de yine deepfake içeriklerin etkilerini sağlayacak itici güçler arasında sayılabilir. Kararsızlıktan farklı olarak belirsizlik, bir seçim yapmak için yeterli bilgi olmadığında deneyimlenir ve dolayısıyla yeni bilgilerin sunulması ile belirsizliğin üstesinden gelinir. Deepfake içerikler, doğru bilgiye ulaşmanın maliyetlerini ve bunun sonucunda belirsizliği arttırabilir. Bu nedenle aldatıcı deepfakelerin, içerdikleri bilgiler hakkında belirsizlik yaratıp yaratmadıklarına odaklanmanın önem taşıdığı düşünülmektedir.

Deepfakelerin, dezenformasyon içerikli diğer kaynaklar ve/ veya uygulamalar gibi temel bir “gerçeklik” zeminine kurulamayacağı varsayımının vatandaşlar arasında gelişmesine yol açabileceği düşünülmektedir. Araştırmalar, “kaosa duyulan ihtiyacın”, sonuçları umursanmadan internetteki yanlış siyasi söylentilerin itici güçlerinden biri olduğunu ortaya koymaktadır (Petersen, Osmundsen & Arceneaux vd., 2018). Neyin gerçek ve neyin sahte olduğuna dair belirsizliğin ortaya çıkması, siyasal iletişimin temel stratejik hedefi haline gelmiştir. Kötü niyetli aktörlerin dijital söylem içerisine dahil ettiği birçok çelişkili, anlamsız ve kafa karıştırıcı mesajın birikimli etkisi (Chadwick, Vaccari & O’Loughlin, 2018) sistematik bir belirsizlik hali yaratabilir. Bu bağlamda, deepfake içeriklerin belirsizlik yaratıp güven duygusunu azaltıp azaltmadığına odaklanmak da deepfake konulu araştırmalar açısından göz önünde bulundurulmalıdır.

Derin öğrenme algoritmaları ile ortaya çıkan deepfake içerikler, görsel işitsel sunularla, sosyal medyada her an karşımıza çıkabilecek manipülatif içeriklerdir. Bilginin yalnızca söylemsel olarak değil görsel olarak da dezenformasyona katılıyor oluşu, bu bilgileri deneyimleyen toplumlar için yanlış bilgilendirme ikliminin yanı sıra bireylerde oluşan gerçeklik ilgisizliğine (Westerlund, 2019) de katkı sağlamaktadır. Gerçeğe karşı oluşan şüphecilik, hem gerçek hem de sahte içeriklere olan güvenin erozyona uğramasına yol açmaktadır.

Deepfake videoların kötü niyetli üretiminin ve yayılımının neden olduğu risklerin/ tehditlerin önlenmesi noktasında ise etkili çözümlere ihtiyaç vardır. Deepfakelerin tespit yöntemleri gibi teknik yanlarına ilişkin hatırı sayılır ölçüde literatür bulunmakla birlikte şu ana kadar deepfake siyasi iletişim alanı içinde çok az çalışılmış durumdadır. Var olan bir deepfake'i inceleyerek, deepfakelerin bireylerin kafasını neyin gerçek neyin sahte olduğu noktasında karıştırarak kamusal tartışmaları zehirlediğini göstermiştir. Bildiğimiz kadarıyla, kişilerin kendilerinin yaptığı deepfakelerin bireylerin siyasi tutumları üzerindeki etkileri henüz hiç çalışılmamıştır. Deepfake dezenformasyonunu iyice anlamak ve demokratik toplumlarımız için teşkil ettiği tehditlere karşı koymak, örneğin, deepfake ile savaşmak için daha bilgili stratejiler oluşturmak için bu etkiler üzerine daha çok bilgi sahibi olmak gerekmektedir. Söz konusu dezenformatif içeriklere yönelik kamu bilincinin iyileştirilmesi, dijital okuryazarlık eğitiminin yaygınlaştırılması ve ilgili içeriklerin üretimine dair yasal yaptırımların getirilmesi, başlıca çözüm önerileri olarak sayılabilir. Bunun yanı sıra çalışmamızda de-faacto belirttiğimiz üzere deepfake içeriklerin kitlesel dezenformasyona yol açmasında sosyal medya çok büyük role ve sorumluluğa sahiptir. Bu sebeple deepfake içeriğin tespitini sağlayacak teknolojilerin sosyal medya platformları ile uyumlu çalışıyor olması da içeriklerin önlenmesi noktasında önem taşımaktadır.

Yapay zekâ teknolojileri, deepfake kaynaklı dezenformasyon sorununun yaratıcısı olduğu gibi aynı zamanda bu sorunun doğrudan çözümü olacak teknolojilerin de kaynağıdır. Bu sebeple söz konusu teknolojinin imkânlarının, deepfake içeriklerin yarattığı manipülasyonu önlemesi için kullanımının yaygınlaşması gerekmektedir.

Bu makalenin yazıldığı tarih itibarıyla, Türkiye'de deepfake içeriklerin seçime etkileri üzerine gerçekleştirilmiş akademik bir araştırma bulunmamaktadır. Deepfake aracılığıyla yayılan dezenformasyon özellikle seçim zamanlarında büyük zorluklar teşkil edebilir, zira konuya aşına olmayan gözler için deepfake ile gerçek bir video arasındaki farkı görmek çok da mümkün değildir. Herhangi bir siyasi aktör, kendi amaçlarına ulaşabilmek için siyasi bir skandala neden olabilir ya da rakibinin itibarını zedeleyebilir. Yurttaşlar bir deepfake ile karşılaştıktan sonra, örneğin, deepfake içinde gösterilen siyasetçiye ya da söz konusu siyasetçinin partisine karşı tutumlarını değiştirebilir. Bu durumun sonucunda, yurttaşlar doğru olmayan bilgilere dayanarak ve muhtemelen de deepfake'in arkasındaki siyasi aktörün amaçlarına hizmet edecek biçimde oy vermiş olurlar. Bu durum özellikle demokratik kurumların meşruiyetine, kamusal tartışmaların niteliğine ve kötü niyetli siyasi aktörlerin iktidarına yönelik birtakım soruları gündeme getirebilir. Dolayısıyla deepfake içerikler, siyasi iletişim bağlamında daha yaygın olarak kullanıldığında, seçmenlerin algısının nasıl değişeceğini zamanla göreceğiz. Yine de kesin olan bir şey varsa, o da teknoloji gelişmeye devam edip siyasi iletişimde yeni yollar açtıkça, uygulayıcıların ve akademisyenlerin araştırmaları gereken çok fazla alanla karşı karşıya olduklarıdır.

KAYNAKÇA

- Agarwal, S., Farid, H., Gu, Y., Mingming, H. Nagano, K., Li, H. (2019). Protecting world leaders against deep fakes. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 38-45.
- Aktan, S. (2018). Türkiye Sahte Haber ve Dezenformasyonda Zirveye Oturdu <https://tr.euronews.com/2018/06/15/turkiye-sahte-haber-ve-dezenformasyonda-zirveye-oturdu> Erişim tarihi: 10.04.2021.
- Baerthlein, T. (2016). The Rise of Political Bots On Social Media. <https://www.dw.com/en/the-rise-of-political-bots-on-social-media/a-19450562> Erişim Tarihi: 10.04.2021.
- Barari, S., Lucas, C., Munger, K. (2021). Political deepfake videos misinform the public, but no more than other fake media." OSF Preprints. doi:10.31219/osf.io/cdfh3
- Baudrillard, J. (2014). *Simülakrlar ve Simülasyon*. O. Adanır (Çev.). Ankara: Doğu Batı.
- BBC, (2020): "Deepfake queen to deliver Channel 4 Christmas message", <https://www.bbc.com/news/technology-55424730>, Erişim Tarihi: 15.06.2021.
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139
- Beridze, I., & Butcher, J. (2019). When seeing is no longer believing. *Nature Machine Intelligence*. doi:10.1038/s42256.019.0085-5.
- Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47(2), 241–262.
- Buo, A. S. (2020). The emerging threats of deepfake attacks and countermeasures. DOI: 10.13140/RG.2.2.23089.81762
- Chadwick, A., Vaccari, C. & O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, 20(11), 4255–4274.
- Chadwick, A. (2019). The new crisis of public communication: Challenges and opportunities for future research on digital media and politics. *Online Civic Culture Center*. <https://www.lboro.ac.uk/research/online-civic-culture-centre/news-events/articles/o3c-2-crisis/> Erişim Tarihi: 02.04.2021.
- Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6), 4–8.
- Chen, C. Metz. (2019). Google's Duplex uses A.I. to mimic humans (Sometimes). *The New York Times*. <https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html> Erişim tarihi: 03.01.2021
- Chesney, B. & Citron, D. (2019). Deep fakes: looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820.
- Coldewey, D. (2020). Facebook's 'deepfake detection challenge' yields promising early results. TechCrunch. <https://social.techcrunch.com/2020/06/12/facebooks-deepfakedetection-challenge-yields-promising-early-results/>. Erişim tarihi: 10.12.2020.
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?. *The International Journal of Press/Politics*. Vol. 26(1) 69–9
- Guera, D. & Delp, E. (2018). Deepfake video detection using recurrent neural networks in IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zeland.

- Flynn, D. J., Nyhan, B. & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38, 127–150.
- Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, 49(2), 280–286.
- Giasano, A. (2019). Deep fakes: a challenge of the post-truth era. *Romanian Cyber Security Journal*. 2 (1), 67 – 74.
- Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2015). The structural virality of online diffusion. *Management Science*, 62(1), 180–196.
- Grabe, M. E. & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. Oxford University Press.
- Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication*, 40(3), 134–156.
- Hall, H. (2018). Deepfake videos: When seeing isn't believing. *Catholic University Journal of Law and Technology*, 27(1), 51-76.
- Howes, S. A. (2018). Digital replicas, performers' livelihoods, and sex scenes: Likeness rights for the 21st century. *Columbia Journal of Law & Arts* 42, 345-349.
- Hovland, C. I. & W. Weiss. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly* 15, no.4: 635–650.
- Kahn, J. (2020). These deepfake videos of Putin and Kim have gone viral. <https://fortune.com/2020/10/02/deepfakes-putin-kim-jong-un-democracy-disinformation/> Erişim Tarihi: 16.06.2021.
- Kietzmann, J., Lee, W. L., McCarthy, P. I. ve Kietzmann, C. T. (2020). Deepfakes: Trick or treat?. *Business Horizons*. 63, 135 – 146.
- Kirchengast, Tyrone (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law*, 29(3), 308–323.
- Mcintyre, L. (2018). *Post – truth*. Cambridge: MIT Press
- Messing, S. & Westwood, S. J. (2012). Selective exposure in the age of social media: Endorsements Trump Partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042-1063.
- Murata, K., Orito, Y., Yamazaki, T. Ve Shimizu, K. (2020). Post – Truth society: The ai – driven society where no one is responsible. J. Borondo – Pelegrin, M. Arias – Oliva, K. Murata, A. M. Lara Palma (Eds.). *Paradigm Shifts in ICT Ethics Proceedings of The ETHICOMP 2020*, 18. International Conference on the Ethical and Social Impacts of ICT, Lograno, Spain, June 2020
- Newman, E. J., Garry, M., Unkelbach, C. ve Bernstein, D. M., Lindsay, D., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1337–1348.
- Nguyen, T. T., Nguyen, M. C., Nguyen, T. D., Nguyen, T. D., Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. arXiv preprint arXiv:1909.11573
- Nyilasy, G. (2019). Fake news: When the dark side of persuasion takes over. *International Journal of Advertising* 38, no.2: 336–42.
- Pancer, E., & Poole, M. (2016). The popularity and virality of political social media: Hashtags, mentions, and links predict likes and retweets of 2016 US presidential nominees' tweets. *Social Influence*, 11(4), 259–270.

- Perot, E., Mostert, F. (2020). Fake it till you make it: an examination of the US and English approaches to persona protection as applied to deepfakes on social media. *Journal of Intellectual Property Law & Practice*, 15 (1), 32 – 39.
- Petersen, M. B., Osmundsen, M., ve Arceneaux, K. (2018). A “need for chaos” and the sharing of hostile political rumours in advanced democracies. PsyArXiv Preprints. <https://psyarxiv.com/6m4ts/>
- Prior, M. (2013). Visual political knowledge: A different road to competence? *Journal of Politics*, 76(1), 41–57.
- Rojecki, A., & Meraz, S. (2016). Rumors and factitious informational blends: The role of the web in speculative politics. *New Media & Society*, 18(1), 25–43.
- Ruiz, D. (2020). Deepfakes laws and proposals flood US. Malwarebytes Labs. <https://blog.malwarebytes.com/artificial-intelligence/2020/01/deepfakes-laws-and-proposalsflood-us/>. Erişim tarihi: 12.01.2021
- Qayyum, A., Qadir, J., Janjua, M. U., & Sher, F. (2019). Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News. *IT Professional*, 21(4), 16–24.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, 39, 127–161.
- Solsman, J. (2020). Deepfakes’ threat to 2020 US election isn’t what you’d think. *CNET*. <https://www.cnet.com/features/deepfakes-threat-to-the-2020-us-election-isnt-what-youthink/>. Erişim tarihi: 15.01.2021
- Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, 18(6), 813–847.
- Sundar, S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. Metzger & A. Flanagin (Eds.), *Digital media, youth, and credibility*, MIT Press. 73–100.
- TASAM. (2019). “Sentetik Gerçeklik Teknolojisi (Deepfake): Derin – Sahte Ürün ve Savunma Ekosistemi İnşası Raporu”. Brains Türkiye Uygulama Programı. Erişim Adresi: https://tasam.org/tr-TR/Icerik/61765/sentetik_gerceklik_teknolojisi_deep_fake_derin-sahte_urun_ve_savunma_ekosistemi_insasi_ Erişim tarihi: 23.02.2021
- Temir, E. (2020). Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. *Selçuk İletişim Dergisi*. 13 (2), 1009 – 1024.
- The Brussels Times, (2020). “XR Belgium posts deepfake of Belgian premier linking Covid-19 with climate crisis”. <https://www.brusselstimes.com/news/belgium-all-news/politics/106320/xr-belgium-posts-deepfake-of-belgian-premier-linking-covid-19-with-climate-crisis/>. Erişim Tarihi: 18.06.2021.
- Thorson, K., ve Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3), 309–328
- Twitter. (2018, April 17). You won’t believe what Obama says in this video! https://twitter.com/BuzzFeed/status/986.257.9917_99222272 Erişim tarihi: 02.02.2021.
- Vaccari, C. (2017). Online mobilization in comparative perspective: Digital appeals and political engagement in Germany, Italy, and the United Kingdom. *Political Communication*, 34(1), 69–88.
- Vaccari, C. & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 1 – 13.
- Villasenor, J. (2019). Artificial intelligence, deepfakes, and the uncertain future of truth Erişim adresi (10.02.2021): <https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/> Erişim tarihi: 15.01.2021

- Yadav, D., & Salmani, S. (2019). Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network. 2019 International Conference on Intelligent Computing and Control Systems (ICCS). Jun 27, 2019 – Jun 28, 2019, Secunderabad, India.
- Wagner, L. T., Blewer, A. (2019). “The word real is no longer real”: Deepfakes, gender, and the challenges of AI- Altered video. *Open Information Science*. 3, 32–46.
- Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism Studies*, 19(13), 1866–1878.
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11): 40-53.
- Westerman, D., Spence, P. R. ve Van Der Heide, B. (2014). Social media as information source: recency of updates and credibility of information. *Journal Of Computer-Mediated Communication*, 19(2), 171 – 183. <https://doi.org/10.1111/jcc4.12041> Erişim tarihi: 04.04.2020.
- Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*, 5 (2), 199 – 217.
- Witness. (2018, June 11). Mal-uses of AI-generated synthetic media and deepfakes: Pragmatic solutions discovery convening. [https:// http://www.mediafire.com/file/q5juw7dc3a2w8p7/Deepfakes_Final.pdf/file](https://http://www.mediafire.com/file/q5juw7dc3a2w8p7/Deepfakes_Final.pdf/file) Erişim tarihi: 18.01.2021
- Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, 48(3), 489–496.
- You Won't Believe What Obama Says In This Video! <https://www.youtube.com/watch?v=cQ54GDm1eL0> Erişim tarihi: 16.06.2021
- Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The Web of False Information. *Journal of Data and Information Quality*, 11(3), 1–37.
- Zimmermann, F. ve Kohring, M. (2020). “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* 37:215–37.

IS SEEING ENOUGH TO BELIEVE? POLITICAL DEEPFAKE CONTENTS AS A DISTINCTIVE FORM OF VISUAL DISINFORMATION

Elif KARAKOÇ* 
Burcu ZEYBEK** 

Artificial intelligence technologies, with their ever-increasing sphere of influence, are considered as the changing drivers of media manipulation, which is an ever-present subject over the course of the history of media. Latest developments in artificial intelligence technologies and machine learning have led to a dramatic increase in the realism of fake content and the accessibility that it might create (Agarwal, 2019, p. 38). As assuming digital identity becomes more realistic and convincing, deepfake technology is observed to be the most advanced technology of this impersonation process (Chesney & Citron, 2019 p. 1758). Deepfakes are “real” videos with fake content that contain multimedia disinformation (Whyte, 2000). Utilizing the deep learning modelling in artificial intelligence technology to combine sounds with images, deepfakes are literally capable of articulating the words of someone in the mouths of any other person. Individuals believe that what they observe with their own eyes and hear with their own ears generally have a better level of accuracy. However, this brought along a challenging process with regard to deepfake videos, underlining the fact that seeing no longer means believing (Beridze & Butcher, 2019). Because deepfakes are videos, which are created to make things someone never said or did appear like she/he said or did (Villasenor, 2019). Having a hyperrealistic power, deepfake videos have gone beyond the foresight of Baudrillard (2014) and created a new definition of reality by synthesizing what is fake and what is real. Deepfake can also be interpreted with the concept of post truth, which refers to the fact that objective facts are less effective than emotions and personal beliefs in shaping the public opinion. As deepfake videos pave the way for a thought pattern dominated by the concept of post truth, a vision of world where there is no more a truth is drawn, since the ‘truths’ are subject only to our political views (Micintyre, 2018).

* Yeditepe University, Radio, Television and Cinema, elif.karakoc@yeditepe.edu.tr

** Doğuş University, Public Relations and Publicity, bzeybek@dogus.edu.tr

The term deepfake became a current issue with the posts shared by a Reddit user called “deepfakes” in the late 2017. In February 2018, this user was banned from Reddit because of sharing porn content in the form of deepfake (Kirchengast, 2020, p. 4), but the increasing interest in deepfake videos could not be prevented afterwards. In recent years, the sharing of deepfake videos in social media has increased, and such content has reached to users at an unpredictable speed and magnitude. This situation has the potential to lead to mass disinformation, particularly as it is frequently observed in fake news (Perot & Mostert, 2020, p. 33). The fact that artificial intelligence technologies are easily used in all fields and the interactive structure of internet network facilitated the formation process and accessibility of deepfake videos, which brought along a series of political, technological and legal issues (Giasano, 2019; Villasenor, 2019; Kietzman et al., 2020). Deepfake videos, which reproduce the reality, can be encountered in several different fields. While these videos, which make Mona Lisa and Salvador Dali speak, as if he is alive again after years, may look like a positive aspect for the field of art (Yadav & Salmani, 2019; Giasano, 2019), this has completely different implications for political communications in terms of the potential results it may cause. Deepfake algorithms generally require a vast amount of image and video data for training deep learning models in order to create photo-realistic images and videos. Since public characters like celebrities and politicians have several online videos and images available, these persons are the first targets of deepfakes (Nguyen et al., 2019, p. 1). The fact that a deepfake video, which features the speech of the leader of any country in any part of the world, is a striking example of the alarming state of events in this field. The sharing of a single video with such qualities has the potential of accelerating the propagation of manipulative narratives in online environment. Deepfake contents may pose serious risks and threats on journalism, competence of citizens and quality of democracy (Bennett & Livingston, 2018; Chadwick et al., 2018; Flynn et al., 2017; Rojecki & Meraz, 2016; Waisbord, 2018). As armed disinformation tools aiming to intervene in elections and elicit civil unrest (Westerlund, 2019), deepfakes have a cyber-criminal quality in addition to their political use. Manipulation of money markets and stock certificates, phone frauds by using fake sounds, are some of the examples of this (Westerlund, 2019). One of the biggest concerns about deepfake photographs and videos is the fact that such contents are being used for creating pornographic content (deepfake pornography) (Giosanu, 2019). Deepfake videos, with their potential to spread insults against people, circulate fake news, and harm democracy and judicial systems due to their manipulative structure (Buo, 2020) can even change the course of legal cases (Yadav & Salmani, 2019, p. 855). Deepfake videos also have the potential to seriously damage individual reputation and public trust. In addition to this, deepfake contents may aggravate social division by eroding the consensus on what is real. This may cause a widespread cynicism and apathy, undermining the trust in information and/or news at every level (Beridze & Butcher, 2019). Political communication researches have findings, which suggest that images have a strong effect on the perception of individuals (Graber, 1990; Grabe & Bucy, 2009; Prior, 2013). The fact that understanding and processing images and audio-visual contents is easier than written texts, brings into question the concept of “metacognitive experience”, which refers to the emotions about our ideas, derived through experience with the capacity of our reactions to tasks such as processing information (Schwarz et al.,

2007). In such an experience, the concept of “fluency” is particularly important for understanding why people believe in false information. People are more likely to accept messages, which they perceive as familiar, to be true (Berinsky, 2017). Familiarity, that is “sounding right effect”, elicits a feeling of fluency that allows internalizing contents more easily, and therefore, regarding them as more credible (Newman et al., 2015). If political videos with deepfake content already celebrities and/or public figures, they may make the serious problems posed by fluency far more alarming due to their technical sense of reality and familiarity. Introducing the problem of security in political communication, deepfake videos have the potential of influencing public opinion effectively and easily with the support of audio and visual media, especially in election campaigns and/or discourses with power motive by affecting people’s sense of reality in a manipulative way. Based on this, the study discusses the nature of deepfake contents, how such technologies operate and the problems caused by this technology, and addresses the risks and threats posed by political deepfake contents on political campaigns, and current and possible solutions to these problems.