YENİ MEDYA
DERGİSİ
Hakemli, uluslararası, e-Dergi

Sayı Vol , 17, Güz Autumn, (2024)
https://dergipark.org.tr/tr/pub/yenimedya

Deepfake ve Dezenformasyon: Sentetik Siyasi Videoların Yanıltma, Belirsizlik ve Haberlere Duyulan Güven Üzerindeki Etkisini Araştırma*

Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News**

Cristian Vaccari • **Andrew Chadwick**

Çeviri Translation — Çev. : Şeyda KOÇAK KURT

ÖZ

Yapay Zeka (YZ) günümüzde "deepfake" olarak bilinen ve gerçek videolara çok benzeyen sentetik videoların kitlesel olarak oluşturulmasını mümkün hale getirmektedir. Görsel iletişimin gücü ve belirsizliğin kamusal söyleme duyulan güveni sarsmada oynadığı rol ile ilgili teorileri birleştirerek, deepfake'lerin çevrimiçi dezenformasyona olası katkılarını açıklıyoruz. Birleşik Krallık nüfusunu temsil eden geniş bir örnekleme yeni deneysel uygulamalar yaparak, insanların deepfake'lere ilişkin değerlendirmelerini karşılaştırdık. Katılımcıların deepfake'ler tarafından yanlış yönlendirilmekten çok, belirsizlik hissetme olasılıklarının daha yüksek olduğunu ve ortaya çıkan bu belirsizliğin sosyal medyada yer alan haberlere duyulan güveni azalttığını bulduk. Deepfake'lerin genelleştirilmiş kararsızlık ve sinizme katkıda bulunabileceği ve demokratik toplumlarda çevrimiçi yurttaşlık kültürüne yönelik mevcut zorlukları daha fazla arttırabileceği sonuçlarına ulaştık.

Anahtar Kelimeler: Mezenformasyon, Dezenformasyon, Belirsizlik, Siyasi Deepfake'ler, Çevrimiçi Yurttaşlık Kültürü.

ABSTRACT

Artificial Intelligence (AI) now enables the mass creation of what have become known as "deepfakes": synthetic videos that closely resemble real videos. Integrating theories about the power of visual communication and the role played by uncertainty in undermining trust in public discourse, we explain the likely contribution of deepfakes to online disinformation. Administering novel experimental treatments to a large representative sample of the United Kingdom population allowed us to compare people's evaluations of deepfakes. We find that people are more likely to feel uncertain than to be misled by deepfakes, but this resulting uncertainty, in turn, reduces trust in news on social media. We conclude that deepfakes may contribute toward generalized indeterminacy and cynicism, further intensifying recent challenges to online civic culture in democratic societies.

Keywords: Misinformation, Disinformation, Uncertainty, Political Deepfakes, Online Civic Culture.

* 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, 6(1), 1-13. https://doi.org/10.1177/2056305120903408. (Çev. Şeyda Koçak Kurt).

Hindistan, Nisan 2018: Dünyanın en popüler mobil anlık mesajlaşma platformu WhatsApp'ta bir video viral oldu. Bir CCTV kamerasından çekildiği anlaşılan görüntülerde, sokakta kriket oynayan bir grup çocuk görülmektedir. Birden, motosikletli iki adam çocukların yanına yaklaşır ve en küçük çocuklardan birini kapıp hızla uzaklaşır[1]. Bu "kaçırma" videosu, geniş çaplı bir kafa karışıklığı ve panik yarattı ve en az dokuz masum insanın ölümüne neden olan 8 hafta süren bir çete şiddetine yol açtı (BBC News, 2018).

Kan davalarının fitilini ateşleyen bu görüntüler zekice hazırlanmıştı ancak sahteydi – çünkü Pakistan'da çocuk kaçırma olaylarına karşı farkındalık yaratmak üzere tasarlanmış bir halk eğitim kampanyası videosunun düzenlenmiş haliydi. Eğitim videosu kaçırma olayıyla başlamakta ancak kısa süre sonra aktörlerden biri motosikletten inmekte ve izleyicileri çocuklarına sahip çıkmaları konusunda uyarmaktadır. Hindistan'da viral olan sahte videoda ise, bu "önemli açıklama" kısmı kesilmiş ve geriye sadece bir çocuğun kaçırılışının şok edici derecede gerçekçi bir videosu kalmıştı.

Aynı ay içinde BuzzFeed'de, eski ABD Başkanı Barack Obama'nın Oval Ofis gibi görünen bir yerde doğrudan kameraya konuştuğunu gösteren bir video yayınlandı. Videonun ilk 35 saniyesinde sadece Obama'nın yüzü görülür. Karakterine aykırı birkaç açıklamanın ardından, Obama bombayı patlatır: "Başkan Trump tam bir ahmaktır". Kısa bir duraklamanın ardından, "Şimdi... gördüğünüz gibi, ben asla böyle şeyler söylemem, en azından halka açık bir konuşmada, fakat bir başkası söyleyebilir... Jordan Peele gibi" der. Bu noktada BuzzFeed'in videosunun gerçek niyeti ortaya çıkar. Konuşan kişi aslında Obama değildir. Solda Obama'yı, sağda ise ABD'li ünlü aktör, komedyen ve yönetmen Jordan Peele'i gösteren bölünmüş bir ekran belirir. Obama'nın ve Peele'in yüz ifadeleri ve dudak hareketleri mükemmel bir şekilde eşleşmektedir. Peele'in prodüksiyon ekibi yapay zeka (YZ) kullanarak Obama'nın yüzünü dijital olarak yeniden yapılandırmıştır. Peele Obama'yı sesiyle taklit ederken, YZ Peele'in yüzünü yapay olarak eşleştirmekte ve bunun çevrimiçi videonun nasıl manipüle edilebileceğine dair ustaca hazırlanmış bir kamu hizmeti duyurusu olduğu anlaşılmaktadır. BuzzFeed videosu, ışık hızında viral oldu. Hem de oldukça uygun bir tık tuzağı sloganı eşliğinde: "Obama'nın Bu Videoda Ne Söylediğine İnanamayacaksınız! ;)" Video Facebook'ta 5 milyon görüntülenme ve 83.000'den fazla paylaşım, YouTube'da 5 milyondan fazla görüntülenme ve Twitter'da 4,75 milyon görüntülenme ve yaklaşık 52.000 retweet aldı. (Facebook, 2018; Twitter, 2018; YouTube, 2018).

Deepfake'lerin Yükselişi

Haraketsiz görüntülerin "fotoşoplanması" uzun zamandır dijital kültürün temel dayanaklarından biri olmakla birlikte, insanların manipüle edilmiş görüntülerini içeren videoları giderek daha fazla internette yer almaktadır. BuzzFeed, videoyu "sentetik medya" (Witness, 2018) ya da "deepfake" olarak bilinen ve giderek yaygınlaşan teknikleri kullanarak oluşturmuştur. Makine öğrenimi algoritmalarına dayanan yazılım uygulamaları, bir kişinin ifadelerinin dikkatli bir şekilde başka bir kişinin kafasına bindirildiği, son derece ikna edici "yüz-grefti" videoları oluşturmaktadır (GitHub, 2019a, 2019b). Alternatif olarak, bir kişinin ağız hareketlerinin ve sesinin mevcut kayıtları, tersine mühendislik yapılarak herhangi bir cümleyi söylemelerini mümkün hale getirebilmektedir. Sonuçlar, özellikle internette yaygın olarak kullanılan düşük çözünürlüklü videolarda endişe uyandıracak kadar ikna edici olabilir.

Siyasi deepfake'ler, internetin görsel dönüşümünün önemli bir çıktısıdır. Çevrimiçi, video tabanlı dezenformasyonun öncüsüdürler ve göz yumulduğu takdirde gazetecilik, vatandaş yetkinliği ve demokrasinin niteliği üzerinde derin etkilere sahip olabilirler (Bennett & Livingston, 2018; Chadwick vd., 2018; Flynn vd., 2017; Rojecki & Meraz, 2016; Waisbord, 2018)[2]. Bu çalışma, deepfake'lerin yanıltıcılığına ilişkin ilk kanıtları sunmaktadır. Anekdotlara dayalı kanıtlar, kötü niyetli aktörler tarafından deepfake'lerin kitlesel olarak üretilmesi ve yayılması ihtimalinin, çevrimiçi siyasi söylemin

gerçekliğine yönelik şimdiye kadarki en ciddi zorluğu oluşturabileceğini göstermektedir. Görsellerin ikna gücü metinlerden daha fazladır ve vatandaşların bu tür görsel aldatmacalara karşı savunmaları nispeten daha zayıftır (Newman vd., 2015; Stenberg, 2006).

BuzzFeed Obama/Peele deepfake'indeki çok önemli eğitici "önemli açıklamanın" kurgulanmasının ne ölçüde bireylerin yanlış yönlendirilmesine yahut videonun doğru ya da yanlış mı olduğu konusunda emin olmamalarına yol açtığını belirlemek için temsili bir örneklem (N= 2,005) üzerinde çevrimiçi bir deney gerçekleştirdik. Başka bir ifadeyle, deneyimiz Hindistan'da viral olan kötü niyetli sahte kaçırma videosunun yarattığı sorunu yeniden üretmektedir.

Yanıltıcı siyasi deepfake'lerin katılımcılarımızı yanlış yönlendirdiğine dair bir kanıta ulaşamasak da, birçoğunu içeriğin doğruluğu konusunda belirsizliğe ittiğini bulduk. Buna karşılık, bu tür bir belirsizliğin sosyal medyadaki haberlere duyulan güvenin azalmasına neden olduğunu gösteriyoruz. Bu sonuçlara dayanarak, kontrol edilmediği takdirde, siyasi deepkfake'lerin yükselişinin, doğruluk ve yanlışlık konusunda bir belirsizlik iklimine katkıda bulunarak çevrimiçi yurttaşlık kültürüne zarar vereceğini ve bunun da çevrimiçi haberlere duyulan güveni azaltacağını öne sürüyoruz.

Görsel İletişimin Yenilenen Gücü

Görsel iletişimin gücü, siyasal iletişim araştırmalarında klasik bir inceleme konusu olmuştur. Kilometre taşı niteliğindeki deneyinde Graber (1990), televizyon izleyicilerinin görsel mesajları doğru bir şekilde anımsama olasılığının sözlü mesajlara oranla daha yüksek olduğunu bulmuştur. Grabe ve Bucy (2009); adayların gösterildiği ancak duyulmadığı klipleri içeren "görüntü bitlerinin", adayların konuşurken duyulduğu ancak konuşurken görüntülerinin olduğu ya da olmadığı "ses bitlerinden" daha güçlü olduğunu göstermiştir. Prior (2013), ankete katılanların, olgusal hatırlamayı sorgulayan sorularda hem görsel hem de sözel bilgilerin yer aldığı durumlarda daha yüksek bilgi düzeyleri gösterdiklerini bulmuştur.

Görseller, vatandaşların hafıza oluşturmalarına ve hafızayı geri getirmelerine yardımcı olarak bilgi aktarımını arttırmaktadır. Stenberg (2006), bireylerin görsel bilgiyi sözel bilgiye kıyasla daha doğrudan ve daha az çabayla işlediğini göstermektedir. Witten ve Knudsen (2005), algılanan "kesinliği" nedeniyle, görsel bilginin diğer duyusal veri türlerine göre daha etkili bir şekilde uyumlu hale geldiğini savunmaktadır. Yanıltıcı görsellerin yanlış algılar yaratma olasılığı, yanıltıcı sözel içeriklerden daha fazladır çünkü "gerçeklik etkisi" (Frenda vd., 2013; Sundar, 2008) temelinde, bireyler ses ve görüntülerin günlük deneyimin "gerçek dünyasına" benzeme olasılığını metne kıyasla daha yüksek olarak değerlendirmektedir.

Görüntülerin ve görsel işitsel içeriğin anlaşılması ve işlenmesi yazılı metinden daha kolay olduğunda, "üstbilişsel deneyim" (yeni enformasyonun işlenmesi gibi durumlara verdiğimiz tepkileri şekillendiren, deneyimsel olarak elde edilmiş düşüncelerimizle ilgili duygular) (Schwarz, vd., 2007) devreye girmektedir. Bu deneyimlerden biri olan "akıcılık", insanların yanlış bilgilere neden inandıklarını anlamak açısından özellikle önemlidir. İnsanların tanıdık olarak algıladıkları mesajları doğru olarak kabul etme olasılıkları daha yüksektir (Berinsky, 2017). Aşinalık, verinin özünsenmesini kolaylaştıran ve dolayısıyla daha inandırıcı kılan bir akıcılık hissi anlamına gelen "doğruluk etkisi"ni ortaya çıkarmaktadır (Newman vd., 2015). Teknik gerçeklikleri ve özellikle halihazırda tanınmış kamusal figürleri tasvir etmelerine bağlı olarak, deepfake siyasi videolar, videonun içeriğinin doğruluğuna bakılmaksızın aşinalık yoluyla akıcılık yaratabilir ve zaten ciddi oranda önemli olan bir sorunu potansiyel olarak arttırabilir.

Sosyal medya kullanıcılarının paylaşım davranışları da önemlidir. Videolar ve hareketsiz görüntülerin Twitter'da yayılma olasılığı haber ve çevrimiçi imza kampanyalarına kıyasla daha yüksektir (Goel vd., 2015; s. 186). 2016 ABD Başkanlık kampanyası sırasında Donald Trump ve Hillary Clinton'ın resim

veya video içeren tweetleri önemli ölçüde daha fazla beğeni ve retweet almıştır (Pancer & Poole, 2016).

Görsel Dezenformasyonun Kendine Özgü Bir Biçimi Olarak Siyasi 'Deepfake'ler

Deepfake'ler, Çekişmeli Üretici Ağ (Generative Adversarial Networks – GAN) (Goodfellow vd., 2014) adı verilen bir YZ teknolojisi sayesinde sentezlenebilmektedir. Ortalama bir insanın, kelimeleri oluştururken çıkardığı seslere karşılık gelen; tahmin edilebilir bir çene, dudak ve kafa hareketleri yelpazesi bulunmaktadır. GAN'lar eğitim seti olarak gerçek video görüntülerini kullanmakta ve iki yazılım sinir ağı arasında rekabet yaratmaktadır, böylece her biri diğerinin çıktısına göre gelişim göstermektedir. Bu tekniği kullanan Sawajanakorn ve arkadaşları (2017), konuşan insanların hem ses hem de video içeriklerini gerçekçi bir şekilde sentezlemiştir. Thies ve arkadaşları (2016), web kamerası olan herkesin diğer insanların yüz ifadelerinin kopyalarını oluşturmasını sağlayan bir yazılım geliştirmiştir. En güçlü teknik, bir konuşmacının yüz ifadelerini gerçek zamanlı olarak yeniden yapılandıran "kendi kendini yeniden canlandırma" videosunun üretilmesidir (Rössler vd., 2018). (Yine ücretsiz olarak erişilebilen bir yazılım tarafından çalıştırılan) GAN'lar için eğitim verisi olarak kullanıldığında, bu materyaller kullanıcıların kamuya mal olmuş kişilerin uydurma ama gerçekçi videolarını oluşturmasına olanak tanımaktadır ve bu videolar daha sonra gerçek görüntülerden ayırt edilebilecek herhangi bir belirli işaret olmaksızın çevrimiçi olarak paylaşılabilmektedir. YZ ayrıca insan sesini taklit eden yüksek kaliteli ses sentezlemek için de kullanılmaktadır (Baidu Research, 2017; Gault, 2016).

Çoğu insan, deepfake'ler tarafından ne zaman kandırıldıklarını ayırt edebilecek donanıma sahip olmayabilir. Rössler vd. (2018), insanların, vakaların yalnızca yaklaşık %50'si oranında sahtekarlıkları doğru bir şekilde tespit ettiğini bulmuştur – bu da istatistiksel olarak rastgele bir tahmin kadar iyidir. Algılama, özellikle sosyal medyada yaygın olarak kullanılan sıkıştırmanın neden olduğu bulanıklık ve blokluluğa sahip videoları değerlendirirken zayıftır. YZ tabanlı yöntemler insanlardan biraz daha iyidir, ancak video sıkıştırma kullanıldığında etkililiği de azalmaktadır.

Deepfake'lerin Etkisini Teorileştirmek: Yanıltma, Belirsizlik ve Güven

Deepfake'ler, video tabanlı görsel dezenformasyonun yeni ve benzersiz bir biçimidir. Bu yazı kaleme alındığı sırada, deepfake'lerin etkileri üzerine yapılmış akademik bir araştırma yoktu. Bu çalışmada, deepfake'lerin bireylerin doğruluk ve yanlışlık algılarını etkileyip etkilemediğini, ancak daha da önemlisi, aktardıkları bilgiler hakkında belirsizlik yaratıp yaratmadıklarını inceliyoruz. Son olarak, deepfake'lerin ortaya çıkardığı belirsizliğin insanların sosyal medyadaki haberlere duydukları güveni azaltıp azaltmadığını değerlendiriyoruz.

İlk olarak *bilişsel* çıktılara odaklanıyoruz. Sorunun can alıcı noktası, deepfake'lerin insanları yanıltabilmesidir. İzleyiciler deepfake tuzağına düşmeseler bile, içeriğin doğru mu yanlış mı olduğu konusunda belirsizliğe düşebilirler. Belirsizlik, kavramsal olarak kararsızlıktan farklıdır. Kararsızlık, bireylerin çatışan fikirleri içeren bir seçimle karşı karşıya kaldıklarında ortaya çıkar, böylece "ek bilgi yalnızca içselleştirilmiş çatışmayı arttırır" (Alvarez & Brehm, 1997, s.346). Buna karşın belirsizlik, bir seçim yapmak için yeterli bilgi mevcut olmadığında yaşanır ve bu nedenle yeni bilgilerin eklenmesiyle giderilebilir (Alvarez & Brehm, 1997). Downs'a göre (1957); doğru bilgi edinmenin maliyeti çok yüksek olduğu için vatandaşlar arasında belirsizlik ortaya çıkmaktadır. Deepfake'ler doğru bilgi edinme maliyetlerini yükseltebilir ve bunun sonucunda belirsizlik artabilir. Bu nedenle, yanıltıcı deepfake'lerin içerdikleri bilgi ile ilgili belirsizlik yaratıp yaratmadığına odaklanıyoruz.

Diğer dezenformasyon yöntemlerinin yanında, deepfake'ler de belirsizliği arttırmayı başarırsa, bunun başlıca sonuçlarından biri, deepfake'lerin en yaygın şekilde dolaşımda olduğu sosyal medyada yayınlanan haberlere duyulan güvenin azalması olabilir. Dolayısıyla, ikinci odak noktamız deepfake'lerin potansiyel *tutumsal* sonuçlarından

biri olan *sosyal medyadaki siyasi haberlere duyulan güven*dir. Haberlere duyulan güven dünya genelinde azalmaktadır (Hanitzsch vd., 2018) ve sosyal medyadaki haberlere duyulan güven, diğer kanallardan erişilen haberlere duyulan güvenden daha düşüktür (Newman vd., 2018).

Araştırmacılar güven ve belirsizlik arasındaki ilişkiyi farklı perspektiflerden incelemiştir. Bir taraftan, güven genellikle "toplumsal belirsizliğin yol açtığı sorunlara bir çözüm" olarak kavramsallaştırılmıştır (Yamagishi & Yamagishi, 1994, s. 131). Benzer şekilde Tsfati ve Cappella (2003, s. 505) "güvenin anlamlı olabilmesi için güvenen tarafında bir miktar belirsizlik olması gerektiğini" ileri sürmektedir. Bu yaklaşıma göre, belirsizlik güvenden önce gelmekte ve belirli koşullar altında ortaya çıkmaktadır. Öte taraftan, belirsizlik arttığında başkalarına güvenmek daha zor olabilir. Cook ve Gerbasi (2011, s. 219) insanların birbirlerine güvenmemelerinin nedenleri arasında "belirsizlik ve risk düzeyi gibi durumsal faktörleri" vurgulamaktadır. Artan belirsizliğin işle ilgili kararlara (Adobor, 2006), müzakere edilmiş ve karşılıklı alışverişlere (Molm vd., 2009), e-ticaret sitelerinin kullanımına (Anriawan ve Thakur, 2008) ve pazar araştırmalarına (Moorman vd. 1993) duyulan güveni azalttığı tespit edilmiştir. Çevrimiçi ortamdaki tartışmalı bilgilerle ilgili olarak artan belirsizlik, Van Duyn ve Collier'in (2018) insanları sahte haber sorunuyla ilgili seçkin tweet'lere maruz bırakmanın halkın haberlere olan güvenini azalttığını bulmasının nedenini açıklayabilir.

Dolayısıyla, çalışmamızın temelinde, diğer yanlış bilgi kaynaklarıyla (Örn; Vosoughi vd., 2018) ortak olarak, deepfake'lerin zaman içinde vatandaşlar arasında temel bir doğruluk zemininin oluşturulamayacağı varsayımını geliştirebileceği endişesi yatmaktadır. Araştırmalar, "kaos ihtiyacının" – sonuçlarını umursamadan "dünyanın yanışını izleme" arzusunun – internetteki sahte siyasi söylentilerin itici güçlerinden biri olduğunu göstermektedir (Petersen vd., 2018). Neyin doğru neyin yanlış olduğu konusunda belirsizlik tohumları ekmek, devlet destekli propagandanın temel stratejik hedeflerinden biri haline gelmiştir. Rus operasyonları hakkında yazan Pomerantsev (2015), "Amaç... bilgiye ayrılan yeri çöpe atmak ve böylece izleyicinin kaosun ortasında herhangi bir gerçeği aramaktan vazgeçmesini sağlamak" demektedir. Kötü niyetli aktörlerin dijital söyleme soktuğu çok sayıda çelişkili, anlamsız ve kafa karıştırıcı mesajın kümülatif etkisi (Chadwick vd., 2018; Phillips & Milner, 2017), sistemli bir belirsizlik durumu yaratabilir. Bu bağlamda, deepfake'lerin belirsizlik yaratıp yaratmadığına ve güveni azaltıp azaltmadığına odaklanmak özellikle önem kazanmaktadır.

Hipotezler

Bir deney yaparak, BuzzFeed Obama/Peele deepfake'inin iki yanıltıcı versiyonuna ve bir eğitici, düzenlenmemiş versiyonuna maruz kalan katılımcıların tepkilerini karşılaştırarak üç hipotezi test ediyoruz.

Öncelikle deepfake'lerin insanları ne ölçüde yanılttığı ve belirsizlik yarattığı ile ilgileniyoruz. Daha önce tartıştığımız gibi, birçok insanın deepfake videoları tespit etme konusunda yetkin olmadığını öne sürmek için geçerli nedenler bulunmaktadır. *(H1) Yalan ifadenin yalan olduğunun ortaya çıktığı bir deepfake siyasi videoyu izleyen kullanıcılarla kıyaslandığında; yalan olduğu ortaya çıkmayan yalan ifadeler içeren bir deepfake siyasi videoyu izleyen bireylerin kandırılma olasılığının daha yüksek olduğunu ve (H2) videonun içeriği ile ilgili belirsizlik yaşama olasılığının daha fazla* olduğunu düşünüyoruz.

Daha sonra, yanıltıcı deepfake'lere maruz kalma ile deepfake'in içeriğine ilişkin belirsizlik deneyiminin aracılık ettiği sosyal medyadaki haberlere duyulan güven arasındaki ilişkiye bakıyoruz. Önceki bölümde çerçevelenen argümanlara dayanarak, sosyal medya aracılığıyla bir deepfake'e maruz kalmanın içerik hakkında belirsizliğe yol açması halinde, bu artan belirsizliğin sosyal medyadaki haberlere duyulan güven düzeylerini azaltabileceğini düşünüyoruz. Başka bir deyişle, *bir deepfake'in içeriğine ilişkin belirsizlik, yanıltıcı deepfake'e maruz kalma ile sosyal medyadaki haberlere duyulan güven arasındaki ilişkiye aracılık etmektedir (H3)*[3].

Araştırma Tasarımı, Veriler ve Yöntem

Tasarım

Temsili bir örneklemin (*N*=2,005) BuzzFeed Obama/Peele videosunun iki yanıltıcı, bir eğitici olmak üzere üç versiyona nasıl tepki verdiğini değerlendirdik. Bu denekler-arası tasarım, yanıltıcı ve eğitici deepfake'e maruz kalmanın katılımcıların videoyu nasıl değerlendirdiklerini ve sosyal medyaya ne düzeyde güven duyduklarını etkileyip etkilemediğini ölçebilmemizi sağladı. Çalışmamız kontrol grubu içermemektedir; bu konuyu aşağıdaki "Sınırlılıklar" bölümünde ele alıyoruz.

Uygulamalar

İki sebepten dolayı mevcut bir siyasi deepfake'i kullanmayı seçtik. Birinci sebep; videonun bilinen bir viral başarıya ulaşmasının deneyimizin dış geçerliliğini arttırmasıdır. İkinci sebep ise, Buzzfeed videosunun kolayca farklı bölümlere ayrılması ve bu bölümlerin ayrı ayrı izlendiğinde izleyicileri çok farklı bilgilere maruz bırakmasıdır. Tamamlaycı Bilgiler bölümünde yer alan Ek 1'de transkiptlerin tamamı ve video indirme bağlantıları yer almaktadır.

Orijinal videoyu düzenleyerek üç ayrı video oluşturduk. İki video *yanıltıcıydı* çünkü orijinal videonun ikinci yarısı çıkarılmıştı; bu yarıda Obama'nın yüzü sentetik olarak yeniden yapılandırılmış ve sesi taklit edilmişti. Üçüncü video ise *eğiticiydi*: Deepfake'in bölünmüş ekranda gösterildiği ve konuşanın Obama değil Peele olduğunun gösterildiği videonun tamamını içermekteydi.

İlk yanıltıcı uygulama, sentetik Obama'nın "Başkan Trump tam bir ahmak" dediğini göstermektedir. Bu video dört saniye uzunluğundadır ve Obama'nın ifadesini bağlamsallaştıracak ya da yalan olabileceğini düşündürtecek herhangi bir ipucu vermemektedir. Bu mesajın uzunluğu, sosyal medyada sıklıkla paylaşılan kısa videolarla karşılaştırılabilir. Çalışmanın devamında bu uygulamayı "yanıltıcı 4 saniyelik klip" olarak adlandıracağız.

İkinci yanıltıcı uygulama BuzzFeed Obama/Peele videosunun ilk 26 saniyesini içermektedir, bu nedenle bu videoya "yanıltıcı 26 saniyelik klip" diyoruz. İlk uygulamada olduğu gibi, izleyicilere bu klibin gerçekliğini sorgulamalarına yol açabilecek herhangi açık bir bilgi verilmemiştir. Ancak, video sahte Obama'nın "Düşmanlarımızın bizi herhangi bir zamanda herhangi bir şeyi söylüyormuş gibi gösterebildiği bir çağda yaşıyoruz, bu şeyleri asla söylemeyecek olsak bile" sözleriyle başlamakta, dolayısıyla video izleyicileri sahteliği konusunda uyarabilecek bazı gizli sözlü ipuçları sunmaktadır. Videonun bu kesiti, en kısa videoyla aynı şekilde Obama'nın Trump'a ahmak demesiyle sona ermektedir. Bu daha uzun yanıltıcı videoyu sunduk, çünkü izleyicilerin sahte görüntülere, 4 saniyelik yanıltıcı klibe kıyasla daha uzun bir süre maruz kaldıkları için akıcılık ve dolayısıyla kabul edilebilirlik sağlayıp sağlamayacaklarını ortaya çıkarmak istedik.

Son olarak, üçüncü uygulama eğitici formdadır ve 1 dakika 10 saniye süren ve iki bölümden oluşan orijinal videonun tamamını içermektedir – biri sentetik Obama'nın kamerada tek başına konuştuğu ve Trump'a ahmak dediği, diğeri ise bunun yapay bir yaratım olduğunun belirtildiği ve Jordan Peele'in Obama'yı taklit ettiği bölümdür. Video, Obama'nın sesini taklit eden Peele tarafından söylenen ancak Obama'nın sentetik yüzü ve Peele'in gerçek yüzü kullanılarak görsel olarak temsil edilen deepfake'lerle ilgili bir uyarıyla sona ermektedir. Bu uygulamayı "Eğitici açıklamalı tam video" olarak adlandırıyoruz.

Bağımlı Değişkenlerin Ölçümü

Bağımlı değişkenlerimiz, katılımcıların deepfake'in doğruluğuna ilişkin değerlendirmeleri ve sosyal medyadaki haberlere duydukları güvendir. Bunları, katılımcıları uygulamaya maruz bırakmamızın ardından ölçtük.

Katılımcıların deepfake'in doğru olup olmadığına inanıp inanmadıklarını görmek için, sentetik Obama tarafından söylenen en uç ve olası olmayan cümleye odaklandık. "Barack Obama, Donald Trump'a hiç 'ahmak' dedi mi?" diye sorduk.

Katılımcılar, "Evet", "Hayır" ya da "Bilmiyorum" yanıtlarını verebilirdi. Böylesine doğrudan ve spesifik olgusal bir soru sormak, katılımcıların deepfake'in en az inandırıcı kısmına inanıp inanmadıklarını tespit etmemizi sağladı; bu da hipotezlerimizi videodaki daha makul bir ifadeye odaklanmamızdan daha güçlü bir şekilde test etmemizi mümkün kıldı. Ayrıca, bu soruda "hiç" kelimesini kullanarak, katılımcıların yalnızca izledikleri videoyu gerçek anlamda hatırlamalarını ve videonun doğruluğu hakkındaki inançlarını düşünmemelerini önlemeyi amaçladık.

"Evet" yanıtlarını deepfake'in katılımcıları yanılttığının göstergesi olarak ele aldık (H1). "Bilmiyorum" yanıtlarını ise deepfake'in belirsizliğe neden olduğunun göstergesi olarak kullanmayı tercih ettik (H2). Araştırmacılar genellikle "Bilmiyorum"u kayıp veri olarak ele alırken, Berinsky (2004) "Bilmiyorum"ların önemli bir anlam taşıdığını ve günlük sosyal etkileşimin doğasında bulunan ancak çoğunlukla bireylerin risklerden ve küçük düşmekten korunma veya belirsizlik ya da kararsızlığı ifade etme ihtiyacı gibi görüşmenin yapay bağlamında olmayan faktörlerle açıklanabileceğini ileri sürmektedir. Çevrimiçi olarak problemli olan bilgileri araştıranların Berinsky'nin argümanını verimli bir şekilde ileri taşıyabileceklerine inanıyoruz. Siyaset psikolojisinde belirsizlik üzerine yapılan çalışmaların çoğu, "bilmiyorum" yanıtlarının çelişkili görüşler nedeniyle kararsızlık göstergesi olarak görülebileceği konunun pozisyonlarına ve aday özelliklerine ilişkin algılara odaklanmıştır (Alvarez, 1997). Buna karşılık, yanlış anlamalarla ilgilenen araştırmacılar genellikle, Obama'nın Trump'a "ahmak" deyip demediğini sorarken yaptığımız gibi, katılımcının bilginin doğruluğuna ilişkin değerlendirmesine dokunan sorular sormaktadır. Burada, "Bilmiyorum" yanıtları belirsizliğin basit göstergeleri olarak önemli anlamlara sahip olabilir. Bir "Bilmiyorum" yanıtının, doğrudan belirsizlikle ilgili sorulara verilen yanıtları etkileyebilecek sosyal istenirlik yanlılığı veya utanç faktöründen nispeten uzak olması avantajı da vardır. Ayrıca, belirsizliğin ifadesi olarak "fikrim yok"u veya yanıt vermeyi reddetmeyi değil, kasıtlı olarak "Bilmiyorum"u belirledik, çünkü reddetmeler "Bilmiyorum"lardan belirgin ve sistematik yollarla farklıdır. Bir soruda kişisel olarak hassas bilgiler sorulduğunda reddetme olasılığı daha yüksektir; "Bilmiyorum"lar ise bilişsel çaba ve belirsizlikle ilişkilidir (Shoemaker vd., 2002).

Son olarak, sosyal medyadaki haberlere duyulan güveni ölçmek için (H3), "Sosyal medyada gördüğünüz siyaset ve kamu ilişkileriyle ilgili haber ve bilgilere ne kadar güveniyorsunuz?" diye sorduk. Yanıtlar şunlardı: "Çok", "Biraz", "Çok az", "Hiç" ve "Bilmiyorum". Bu soru bir olgu ifadesinden ziyade bir tutuma işaret etitğinden, "Bilmiyorum" cevaplarının belirsizlik değil kararsızlığa gönderme yapması daha mümkündür. Bu nedenle, katılımcılar arasından bu soruya "Bilmiyorum" cevabı veren %6,5'lik bir kesim analiz dışında bırakılmıştır.

Katılımcılar

Önde gelen bir anket şirketi olan Opinium Research'ün yaptığı panelde gerçekleştirdiğimiz çevrimiçi ankete katılan İngiliz katılımcılardan rastgele olarak seçilen üç alt örnekleme uygulamalarımızı yaptık[4]. %32,8'lik bir katılım oranı elde ettik ve 2,005 katılımcı anketi tamamladı. Örneklemimizin özellikliklerine ilişkin bilgiler Tamamlayıcı Bilgiler dosyasında raporlanmıştır. Laboratuvar temelli deneylerle karşılaştırıldığında, çevrimiçi anketlere yerleştirilmiş deneyler daha fazla temsil kabiliyeti sunmakta ve bu çalışmada kullanılanlar gibi zengin ve gerçekçi uygulamalara olanak sağlamaktadır (bkz. Iyengar & Vavreck, 2012). Ayrıca, kendi kendine uygulanan çevrimçi bir anket kullandığımız için yanıtlar sosyal istenirlik yanlılığından (Kreuter vd., 2008) daha az etkilenmiştir. Bu durum, sosyal istenirliğin eksik raporlamaya yol açabileceği dezenformasyon çalışmaları için özellikle önemlidir.

Prosedür

Ankette standart sosyo-demografik özellikleri ölçen 8 soru ve siyasi tutumları, sosyal medya kullanımını ve sosyal medyada haberlere erişim ve haberleri paylaşımı ölçen 21 soru yer almıştır. Bu soruları yanıtladıktan sonra katılımcılar

rastgele olarak üç uygulamadan birini izlemeye yönlendirildi ve ilk izlemeden sonra bir kez daha tekrar izleme imkanına sahiptiler. Daha sonra bağımlı değişkenlerimizi ölçen soruları ve bazı yanıt kalitesi kontrollerini yanıtladılar. Deney son olarak bir bilgilendirme notu ile sona erdi[5].

Karıştırıcı Faktörler

Üç durum için rastgele atama etkili olmuştur. 2,005 katılımcının 653'ü (%32,5) yanıltıcı 4 saniyeli klibi, 683'ü (%34,1) yanıltıcı 26 saniyelik klibi ve 669'u (%33,4'ü) eğitici açıklamalı tam videoyu izlemiştir. Rastgele dağıtım kontrolleri, hepsi deneyden önce ölçülen; üç alt örneklemin demografik özellikleri, siyasi tutumlar, dijital medya kullanımı, sosyal medyada siyasi konuşmalar ve sosyal medyada haberlere duyulan güven ile eşit derecede dengeli olduğunu doğrulamaktadır[6]. Bu nedenle sonraki analizlerimizde bu faktörleri kontrol etmiyoruz: rastgele atama, ilgilendiğimiz ilişkiler üzerinde bu faktörlerin etkilerini nötralize etmiştir.

Bununla beraber, sosyal medyadaki haberlere duyulan güven üzerindeki etkilere yönelik hipotezimizi test ederken, daha yüksek güven düzeyine sahip bireylerin ilk etapta belirsizliği ifade etme olasılıklarının daha düşük olması nedeniyle, tahminlerimizin yanlı olmadığından emin olmak için sosyal medyadaki haberlere duyulan güvenin uygulama öncesi bir ölçeğini kontrol ediyoruz[7].

Yanıt Kalite Kontrolleri

Videoyu izledikten sonra katılımcılara "Lütfen yukarıdaki videoyu sorunsuz bir şekilde izleyebildiğinizi onaylayın" sorusu yöneltilmiş ve tüm katılımcılar bu soruya olumlu yanıt vermiştir. Arayüzümüz, katılımcıların uygulamaların bulunduğu sayfada ne kadar zaman geçirdiklerini ölçmüştür. Hiçbir katılımcı kendilerine verilen videonun süresinden daha kısa bir süre sayfada kalmamıştır. Videoyu gösterdikten sonra katılımcılara "Bu videoyu daha önce hiç izlediniz mi?" diye sorduk ve 83 katılımcı (%4,1) "Evet" yanıtını verdi. Aynı zamanda "Videoyu izlemenizin hemen ardından, video hakkında daha fazla bilgi edinmek için herhangi bir araştırma (Örneğin; Google araması) yaptınız mı?" diye sorduk. 35 katılımcı (%1,7) araştırma yaptıklarını belirtti. Videoyu daha önce gördüklerini söyleyen ya da videoyu gördükten sonra araştırma yaptıklarını belirten katılımcıları araştırmanın dışında tutmadık çünkü bu sorular uygulama yapıldıktan *sonra* soruldu. Montgomery vd.'ne göre (2018); verilerin uygulama sonrası değişkenlere göre alt kümelere ayrılması nedensel tahminleri istatistiksel olarak saptırabilir ve rastgele atamanın avantajlarını geçersiz kılabilir.

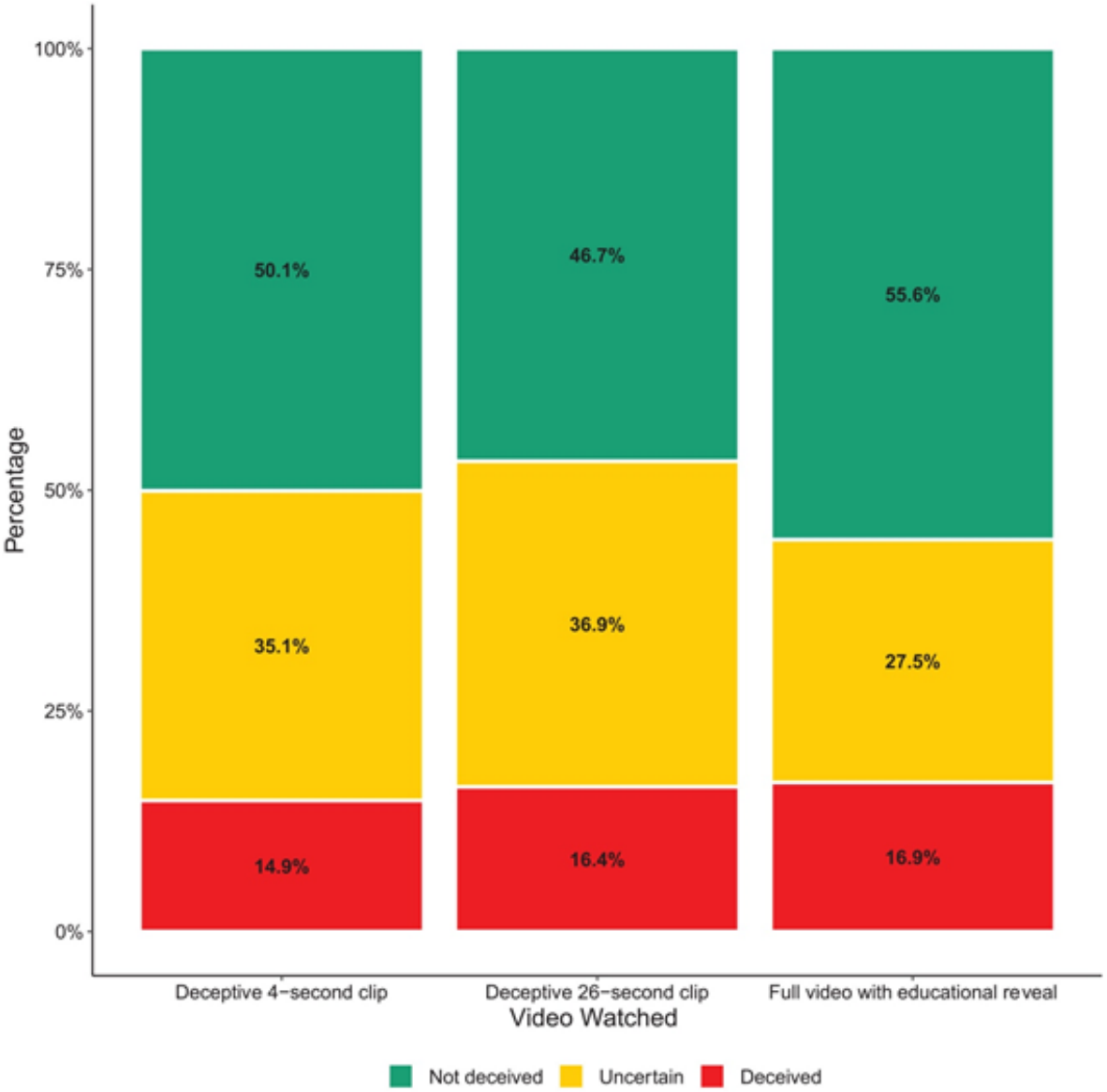
Analiz

İlk olarak, üç videodan birini izleyen katılımcıların Obama'nın Trump'a "ahmak" deyip demediğini sorduğumuz soruya ne ölçüde yanlış olarak "Evet" cevabı verdiklerini ve dolayısıyla yanıltıldıklarını (H1) ve katılımcıların ne ölçüde "Bilmiyorum" cevabı verdiklerini ve dolayısıyla içeriği hakında emin olmadıklarını (H2) test ediyoruz. Ki-kare bağımsızlık testleri ve lojistik regresyonlar kullanarak, iki yanıltıcı videoyu ve eğitici açıklamalı videonun tamamını izleyen katılımcılar arasındaki yanıtları karşılaştırıyoruz.

Genel olarak, katılımcıların yalnızca %50,8'i deepfake tarafından yanıltılmamıştır. İfadenin son derece olasılık dışı olduğu düşünüldüğü bu bulgu şaşırtıcıdır. Daha küçük bir grup (%16) yanılırken, %33,2'lik bir kısım kararsız kalmıştır. Ancak yanıtlar katılımcıların izledikleri videoya göre farklılık göstermiştir. Ki-Kare katsayısı (16.1, $df = 4$, $p = .003$), farklı uygulamalara katılanlar tarafından verilen yanıtlardaki farklılıkların istatistiksel olarak anlamlı olduğunu göstermektedir. İkili karşılaştırmalar, eğitici açıklamalı videonun tamamını izleyenlerin yanıtlarının, iki yanıltıcı deepfake videosunu izleyen katılımcıların yanıtlarından önemli ölçüde farklı olduğunu doğrulamaktadır. Buna karşın, iki yanıltıcı videonun ortaya çıkardığı yanıtlar birbirlerinden önemli ölçüde farklılık göstermemiştir[8].

Şekil 1'de görüldüğü üzere, 4 saniyelik ya da 26 saniyelik yanıltıcı deepfake kliplere maruz kalan katılımcıların yanılma olasılığı, eğitici açıklamalı videonun tamamına maruz kalanlardan çok fazla *değildir*. 4 saniyelik yanıltıcı videonun katılımcıları yanıltma olasılığı aslında en düşükken (%14,9), bunu 26 saniyelik yanıltıcı video (%16,4) ve eğitici

Şekil 1. Uygulamaya göre videonun doğruluğunun değerlendirilmesi.



açıklamalı tam video (%16,9) takip etmiştir. Ancak bu farklar çok küçüktür ve katılımcıların "Evet" yanıtlarını bağımsız değişkenler olarak lojistik regresyonda modellediğimizde anlamlı değildir[9]. Referans kategori olarak eğitici açıklamalı tam video kullanıldığında, yanıltıcı 4 saniyelik video için katsayı −0.152 ($SE = 0.151$, $p = .311$); yanıltıcı 26 saniyelik video için katsayı −0.035 ($SE = 0.146$, $p = .807$) olmuştur. Sonuç olarak, H1 – yalan olduğu ortaya çıkmayan yalan bir ifadeyi içeren deepfake siyasi bir video izleyen bireylerin yanlış ifadeye inanma olasılığının daha yüksek olduğu – doğrulanmamıştır.

Bununla beraber daha önemlisi, sonuçlar H2'yi – yalan olduğu ortaya çıkmayan yalan bir ifade içeren bir deepfake izlemenin belirsizliğe neden olma olasılığı daha yüksektir – desteklemektedir. Yanıltıcı videolardan herhangi birine maruz kalmak (4 saniyelik versiyonu izleyenler arasında %35,1 ve 26 saniyelik versiyonu izleyenler arasında %36,9); eğitici açıklamalı tam videoya maruz kalmaktan (%27,5) daha yüksek düzeyde belirsizlikle sonuçlanmıştır. Yanıltıcı videoların önemli ölçüde daha yüksek belirsizlik seviyeleri ortaya çıkarıp çıkarmadığını değerlendirmek için, gerçekleştirilen uygulamanın bir fonksiyonu olarak "Bilmiyorum" yanıtlarını temel alan bir lojistik regresyon gerçekleştirdik[10]. Eğitici açıklamalı tam video ile karşılaştırıldığında her iki yanıltıcı video için de pozitif ve anlamlı katsayılar elde ettik. Yanıltıcı 4 saniyelik video için katsayı 0.353'tür ($SE = 0.119$, $p = .003$, Holm-yöntemi $p = .003$, Bonferroni-yöntemi $p = .006$. Yanıltıcı 26 saniyelik

Tablo 1. Tablo 1. Sosyal Medyadaki Haberlere Duyulan Temel Güven Düzeylerini Kontrol Eden, Yanıltıcı Deepfake'e Maruz Kalmanın (X) ve Videonun Doğruluğuna İlişkin Belirsizliğin (M) Bir Fonksiyonu Olarak Sosyal Medyadaki Haberlere Duyulan Güveni (Y) Öngören En Küçük Kareler Regresyon Aracı Modeli

	Consequent							
Antecedent		Uncertainty (M)				Trust in news on social media (Y)		
		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Exposure to deceptive deepfake (X)	*a*	0.085***	0.022	.000	*c'*	0.005	0.034	.887
Uncertainty (M)		–	–	–	*b*	−0.175***	0.034	.000
Baseline trust in news on social media	z_1	−0.003	0.037	.925	z_2	0.661***	0.057	.000
N		1,763				1,763		
R^2		0.061				0.075		
F		57.9 (2, 1,760)				54.1 (3, 2,001)		
p		.000				.000		

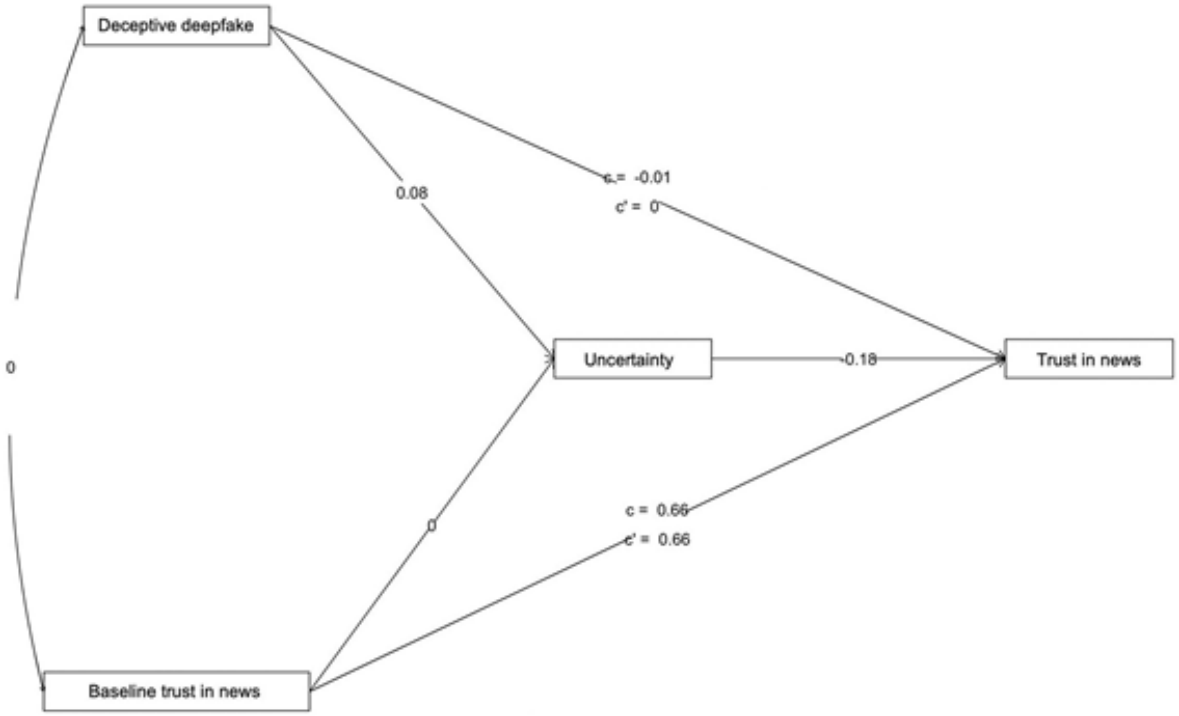
* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

video için katsayı 0.432'dir ($SE = 0.117$, $p = .000$, Holm-yöntemi $p = .000$, Bonferroni-yöntemi $p = .000$). Dolayısıyla, eğitim videosu ile kıyaslandığında, her iki yanıltıcı videonun belirsizliği ortaya çıkarma olasılığı önemli ölçüde daha yüksektir.

Son olarak, yanıltıcı bir deepfake'e maruz kalmanın artan belirsizlik seviyeleri aracılığıyla sosyal medyadaki haberlere duyulan güven oranını azaltıp azaltmadığını (H3) test ediyoruz. En küçük kareler regresyonu yöntemine (Hayes, 2013) dayalı basit bir aracı analiz gerçekleştirdik[11]. Eğitici açıklamanın yer aldığı tam videoya maruz kalmakla karşılaştırıldığında bağımsız değişken, yanıltıcı deepfake'lerden herhangi birine maruz kalmak olarak ele alındı. Belirsizlik (yani Obama'nın Trump'a 'ahmak' deyip demediğini soran soruya "Bilmiyorum" cevabı verilmesi) aracı değişken, sosyal medyadaki haberlere duyulan güven ise bağımlı değişkendir[12]. Aracı model aynı zamanda sosyal medyadaki haberlere duyulan güvenin maruz kalmadan önce ölçülen temel seviyelerini de kontrol etmektedir. Tablo 1 ve Şekil 2'de sonuçlar sunulmaktadır.

Yanıltıcı deepfake'lerden birine maruz kalan katılımcıların (eğitici açıklamalı tam videonun aksine) videonun içeriği hakkında belirsizlik ifade etme olasılığı önemli ölçüde daha yüksektir ($a = 0.085$), bu da H1'i doğrulamaktadır. Buna karşılık, videoyla ilgili belirsizlik ifade eden katılımcıların sosyal medyadaki haberlere duydukları güven düzeyleri ($b = -0.175$), uygulama öncesi güven düzeyleri kontrol edildikten sonra bile önemli ölçüde düşük çıkmıştır. Artan belirsizlik sonucu yanıltıcı bir deepfake'e maruz kalmanın sosyal medyadaki haberlere duyulan güven üzerindeki dolaylı etkisi bu iki katsayının çarpımıdır ($ab = -0.015$, $SE = 0.005$). Dolayısıyla, yanıltıcı deepfake'leri izlemenin bir sonucu olarak sosyal medyadaki haberlere duyulan güven azalmakta ve bu etkiye uygulamadan kaynaklanan

Şekil 2. Tablo 1'deki aracı modelin grafiksel gösterimi.



artan belirsizlik aracılık etmektedir. %95 güven aralıkları sıfırı içermemektedir (−0.026 to −0.007, 5,000 yeniden örneklenmiş örneklem ile). Buna karşılık, yanıltıcı bir deepfake'e maruz kalmak sosyal medyadaki haberlere duyulan güveni doğrudan ve belirsizlik üzerindeki etkisinden bağımsız olarak etkilememiştir ($c' = 0.005$, anlamlı değil). Daha da önemlisi, sosyal medyadaki haberlere duyulan güvenin başlangıç seviyeleri ile belirsizlik arasında anlamlı bir korelasyon bulamadık ($z_1 = -0.003$, anlamlı değil). Dolayısıyla, belirsizlik ve güven arasında bulduğumuz ilişki, uygulama öncesi güven düzeyleri daha düşük olan katılımcıların belirsizliği ifade etme olasılıklarının daha yüksek olmasıyla karıştırılmamıştır. Bunun yerine, sosyal medyadaki haberlere duyulan başlangıç düzeyindeki güven, beklendiği gibi, uygulama sonrası güven düzeylerini güçlü ve anlamlı bir şekilde yordamıştır. Böylece H3 doğrulanmıştır. Eğitici video ile kıyaslandığında, yanıltıcı deepfake'lerden birine maruz kalmak; başlangıçtaki güven düzeyleri kontrol edildikten sonra dahi, daha yüksek belirsizlik düzeyleri ortaya çıkararak sosyal medyadaki haberlere duyulan güveni dolaylı olarak azaltmıştır. Sonuç olarak, yanıltıcı deepfake'ler ile karşılaştırıldığında, eğitici videonun tamamı, belirsizliği azaltarak sosyal medyadaki haberlere duyulan güveni arttırmıştır[13].

Sınırlılıklar

Bulgularımızın sonuçlarını değerlendirmeden önce, bu çalışmanın bazı sınırlıklarını ele alıyoruz.

Öncelikle, deneyimiz sahada değil çevrimiçi bir anket kapsamında uygulandığından, dış geçerliliği konusunda tedbirli olmamız gerektiğini biliyoruz. Tüm anket çalışmalarında olduğu gibi, bu çalışma da sadece tek bir deepfake videoya maruz kalmanın olası etkilerini tek bir anda belirleme imkanı sunmaktadır. Bir deneyin kapalı ortamı dışında deepfake'lerin desteklenmesinde veya çürütülmesinde rol oynaması muhtemel faktörler olan kişiler arası ağlar, algoritmik filtreleme ve diğer mesajlarla rekabet hesaba katılmamıştır. Bununla birlikte, bulgularımızın geçerliliği deneysel yöntemlerin yaygın olarak kabul edilen güçlü yönleri ile artış göstermektedir. Dikkatle tasarlanmış uygulamaları büyük ve temsili bir örnekleme rastgele olarak atadık ve sıkı bir şekilde kontrol edilen bir ortamda insanların maruz kalma öncesi ve sonrası aşamalarındaki tutumlarını değerlendirdik. Bu nedenle bulgularımız, bireylerin yanıltıcı deepfake'lere maruz kaldıklarında bunun belirsizlik ve güvenin yayılmasında

daha geniş bir sosyal etkiye sahip olabileceğini göstermektedir; ancak bu süreçlerde kaçınılmaz olarak rol oynayacak bağlamsal koşulların çeşitliliği konusunda daha fazla araştırmaya ihtiyaç vardır.

İkincisi, kendi uygulamalarımızı üretmek yerine, mevcut bir deepfake'in farklı versiyonlarını oluşturmak için temel düzenleme teknolojisini kullandık ve böylece seçici olarak belirli bir kesiti almak dışında içeriğini değiştirmedik. Obama/Peele deepfake'indeki kilit ifade, genellikle soğukkanlılığıyla bilinen eski bir Başkan'ın, görevdeki bir Başkan'a argo bir küfürle hakaret etmesiydi. Aynı yapay zeka araçları kullanılarak oluşturulan bir video ile verilen daha ince bir mesaj kulağa daha inandırıcı gelebilirdi. Aynı şekilde, 26 saniyelik yanıltıcı videonun Obama'nın "düşmanlarımızın" "bizi herhangi bir zamanda herhangi bir şey söylüyormuş gibi gösterme" kabiliyeti hakkında uyarıda bulunmasıyla başlaması da katılımcıları uyarmış ve uygulamanın yanıltma potansiyelini azaltmış olabilir. Her iki sınırlık da en azından H1 açısından Tip II (yalancı-negatif) hatayı daha olası hale getirmiş olabilir. Deepfake üretme teknolojisi geliştikçe, akademisyenler özel yapım uygulamalar geliştirebilir, ancak riskler bağlamında etik sonuçların çok dikkatli bir şekilde tartışılması gerekecektir.

Üçüncüsü, deneyimizde bir kontrol grubu yer almamıştır. Bu tasarım, bir deepfake videoya maruz kalan kişileri bir "plasebo" videoya maruz kalanlarla karşılaştırmamızı engellemekle birlikte, hipotezlerimizin odak noktası olan aynı deepfake'in farklı (yanıltıcı veya eğitici) versiyonlarına verilen tepkileri karşılaştırmamıza olanak tanımaktadır.

Dördüncüsü, katılımcıların deepfake'teki sentetik Obama tarafından yapılan kilit açıklamaya inanıp inanmadıklarını ölçen sorumuz – "Barack Obama Donald Trump'a hiç 'ahmak' dedi mi?" – *tüm* uygulamalar tarafından ortaya çıkarılan belirsizlik seviyelerini arttırmamıza sebep olmuş olabilir. Bazı katılımcılar soruyu Obama'nın hem kamuya açık hem de özel konuşmalarını kapsayacak şekilde yorumlamış ve dolayısıyla Obama'nın özel olarak ne söylemiş olabileceğini bilemeyeceklerini düşündükleri için bilmedikleri yönünde cevap vermeye yönlendirilmiş olabilirler. Ancak, en önemlisi, belirsizliğin bu şekilde aşırı oranda çıkması tüm uygulamalara maruz kalan katılımcılar arasında eşit olarak dağılmış olacaktır ve dolayısıyla farklı deepfake'ler gören katılımcılar arasındaki yanıtları karşılaştırırken sonuçlarımızı etkilememelidir.

Beşinci olarak, sosyal medyadaki haberlere duyulan güven ölçütümüz geneldir, ancak sosyal medyadaki haberlere güven tartışmalı bir şekilde platforma özgüdür. Fazla uzun süren bir anketten kaçınmak ve katılımcılara bu sonuçla özellikle ilgilendiğimizi hissettirmemek için sosyal medyaya ilişkin her şeyi kapsayan bir soru sormayı tercih ettik. Gelecekteki araştırmalar, kanıtladığımız etki türlerinin platformlar arasında farklılık gösterip göstermediğini ele alabilir.

Altıncısı, çeşitli yanıt kalitesi ölçümleri kullanmamıza rağmen, katılımcıların deepfake'leri yanıltıcı veya eğitici olarak algıladıklarını doğrulamak için manipülasyon kontrolleri yapmadık. Bu tür soruların sorulması deneysel araştırmalarda iyi bir uygulama şekli olmakla beraber (Thorson vd., 2012), biz iki nedenden ötürü bu tür soruları kullanmadık. Birincisi, Rössler ve diğerleri (2018); çoğu kullanıcının deepfake ile gerçek videoyu ayırt etme becerisinin sınırlı olduğunu göstermiştir, dolayısıyla manipülasyon kontrolleri yüksek derecede tahmine dayalı olacaktır. İkincisi, birçok katılımcının izledikleri deepfake'in gerçekliği hakkında doğrudan bir soru sorduğumuzda yanıldıklarını kabul etmek istemeyeceklerini düşündük. Manipülasyon kontrolleri bulgularımızın geçerliliğini güçlendirecek olsa bile, çalışmamızın özel amacı bunların kullanımını sorunlu hale getirmiştir.

Son olarak, katılımcılarımızı seçmek için çevrimiçi panel tabanlı bir örneklemden yararlandık ve bu; bulgularımızı elverişli bir örneklemden elde ettiğimizden daha genellenebilir kılsa da, olasılıklı olmayan örneklemlerden elde edilen sonuçlar otomatik olarak nüfusa genellenemez (Pasek, 2015). Bununla birlikte, örneklemimiz cinsiyet,

yaş ve eğitim açısından yetişkin İngiliz nüfusuna oldukça benzemektedir[14].

Sonuç

Siyasi deepfake'lerin bireyleri mutlaka yanıltamayabileceğini, ancak belirsizlik yaratabileceğini ve bunun da sosyal medyada haberlere duyulan güveni azaltabileceğini gösterdik.

Uzun vadede bu etkiler çevrimiçi sivil kültüre de yansıyarak sorunlu norm ve davranışların ortaya çıkmasına neden olabilir. Güvenin düşük olduğu durumlarda bireylerin işbirliği yapma olasılığı daha düşüktür ve bu durum özellikle günümüzün kutuplaşmış siyaseti gibi yüksek çatışmalı durumlarda geçerlidir (Balliet & Van Lange, 2013). Sosyal medya kullanıcıları internette buldukları haberlere daha az güvenirlerse, kendileri haber paylaşırken diğer kullanıcılara karşı işbirlikçi ve sorumlu davranma olasılıkları da azalabilir. Uzun vadede, çevrimiçi ortamda yer alan içeriklerin çok azına güvenilebileceği yönündeki genel beklenti, çevrimiçi ortamda "her şeyin mümkün olduğu" yönündeki tutumsal sarmala daha çok katkıda bulunabilir. Bu da bireylerin paylaştıkları bilgilere yönelik sorumluluk duygusunu azaltabilir (Chedwick & Vaccari, 2019). Ayrıca vatandaşların belirsizlikten kaynaklanan stresten kaçınmak için haberlerden tamamen kaçmasına da sebep olabilir (Wenzel, 2019).

Vatandaşlar görsel içeriğe inanma eğilimi ile manipülatif deepfake'lere karşı uyanık olma ihtiyacını uzlaştırmaya çalışırken, bu senaryo ile anlamlı bir kamusal tartışma daha zor hale gelecektir. Aynı derecede endişe verici şekilde, seçkinci bir düzeyde; bu çevrimiçi bağlam, ifade özgürlüğünü ve diğer sivil hakları kısıtlayan liberal olmayan politikalar yoluyla "düzeni" ve "kesinliği" yeniden tesis etme vaatleri üzerine kampanya yürütmek için yeni fırsatlar yaratabilir (Arendt, 1951). Hannah Arendt'in (1978) belirttiği gibi;

> Artık hiçbir şeye inanmayan bir halk kendi kararını veremez. Sadece hareket etme kapasitesinden değil, aynı zamanda düşünme ve yargılama kapasitesinden de mahrumdur. Ve böyle bir halkla istediğinizi yapabilirsiniz.

Yaygın belirsizlik, sahtekar politikacıların hiçbir şeyin kanıtlanamayacağını ve hiçbir şeye inanılamayacağını iddia ederek ve yalan söyleyerek suçlamaları saptırmalarını da sağlayabilir.

Çevrimiçi dezenformasyona karşı geleneksel tepkiler bu bağlamda sınırlı bir etkiye sahip olabilir. Medya okuryazarlığını destekleyenler, halkı alternatif bilgi kaynakları aramaya ve bunları yetkili olduğunu iddia eden herhangi bir ifade veya kaynakla yan yana getirmeye teşvik etmeye odaklanmıştır (Örn; Aufderheide, 1992). Ancak bu amaç, siyasi bir ifadenin açık ve gözlemlenebilir bir şekilde gerçekleştiği ve gerekli olanın yalnızca bağlamsallaştırma olduğu varsayımına dayanır. Bu model, (hepsinin olmasa da) birçok doğruluk kontrol kuruluşunun merkezinde yer almaktadır. Deepfake'ler bu model için iki nedenden dolayı belirgin bir sorun teşkil etmektedir. Birincisi, birçok doğruluk kontrol merkezi, kamuya açık olarak söylenen her ne olursa olsun, ifade yanlış bile olsa, gerçek bir kişinin bunu söylediği temelinde çalışmaktadır. Bir deepfake söz konusu olduğunda durum böyle olmayacaktır. İkincisi ve daha önemlisi, deepfake'lerin bolca bulunduğu bir ortamda videoların doğruluğunu kontrol etmek için bir videonun gerçek olduğunu tespit etmek gerekir ancak bunu yapmak, deepfake'lerin teknik yeterliliği ve kısmen zaten kamuya açık olan videolardan üretilemeleri sebebiyle nispeten zordur.

Bireylerin deepfake'leri tespit etmesini sağlayacak yan yana getirme ve bağlamsallaştırma türlerinin kurumsallaştırılması da zor olabilir. Politikacılar, deepfake bir videoda söyledikleri gösterilen şeyleri söylediklerini inkar eden açıklamalar yapmakta hızlı davranacaklardır. Profesyonel gazeteciler önünde sonunda gerçeği ortaya çıkarabilir. Teknolojik açıdan yeterli küçük topluluklar GAN yazılımı tarafından ortaya çıkarılan hataları fark edebilir ve sahtekarlığı çevrimiçi olarak raporlayabilir, ancak uzun vadede yapay zeka tabanlı tespit yöntemlerinin kendi başarılarının kurbanı olması sorunu da vardır, çünkü eğitim veri kümeleri kötü niyetli aktörler tarafından deepfake üretimini daha da iyileştirmek için kullanılacaktır. Asıl soru

şudur: Deepfake'ler yoluyla dezenformasyonla mücadeleye yönelik tüm bu çabalar zamanında yapılacak ve gerektiği kadar geniş kapsamlı olacak mıdır? Yanıltıcı deepfake'lerin -belirsizliğin yayılması ve güvenin azalması gibi – kanıtladığımız olumsuz etkilerinin tamamını ya da çoğunu azaltabilecekler midir?

Daha iyimser bir bakış açısıyla, siyasi deepfake'lerle ilgili eğitici bir videonun, belirsizliği azaltmada başarılı olabileceğini ve böylece yanıltıcı deepfake'lere oranla sosyal medyadaki haberlere duyulan güveni arttırabileceğini gösterdik. Bununla birlikte, eğitici video doğrudan yanıltmayı azaltmamıştır; bu bulgu doğruluk kontrolünün sınırlı etkilerini gösteren önemli bir araştırma dizisiyle örtüşmektedir (örn; Garrett vd., 2013).

Yanıltıcı deepfake'lerin yarattığı belirsizlik nedeniyle sosyal medyadaki haberlere duyulan güvenin azalmasının sinizm ve yabancılaşma değil, şüphecilik yaratması da mümkündür (Cappella & Jamieson, 1996). Lewandowsky vd.'nin (2012, s. 12) öne sürdüğü gibi; "*şüphecilik*, insanları daha sonra yanlış olduğu ortaya çıkabilecek bilgilerin kaynağını sorgulamaya sevk ederse, yanlış bilginin etkilerine karşı duyarlılığı azaltabilir" ve aynı zamanda doğru bilginin tanınmasını ve değer görmesini sağlayabilir. Şüphecilik her şeyin ilacı olmasa da (Green & Donahue, 2011), demokrasi için sinizmden çok daha az sorunludur ve sağlıklı bir şeklide eleştirel ama ilgili bir çevrimiçi yurttaşlık kültürünün bir işareti, hatta bir bileşeni olabilir. Gelecekteki araştırmalar, sosyal medyadaki haberlere duyulan düşük güven düzeyinin sinizme mi yoksa şüpheciliğe mi yol açtığını ve hangi koşullar altında ortaya çıktığını dikkatlice birbirinden ayırmalıdır.

Siyasi deepfake'lerin gelecekte kamusal söylemde oynayacağı rol, nihayetinde bir dizi farklı aktörün onlara nasıl yaklaşacağına bağlı olacaktır. Teknoloji şirketlerinin insanların sentetik temsillerini üreten yapay zeka araçlarını daha fazla geliştirmeleri muhtemel görünmektedir, ancak aynı şirketlerin yapay zekalarını siyasi deepfake'lerin tespit edilmesine yardımcı olarak gerçekliğin demokratik faydasını korumak için de kullanmaya çalışacaklarını ümit ediyoruz. Sosyal medya platformları, otomatik ve insan kaynaklı sertifikasyon ve kontrol biçimlerinin depfake'lerin yayınlanmasını ve paylaşılmasını kolaylaştırıp kolaylaştırmayacağını belirleyecektir. Yerel ve uluslararası politika aktörleri, kamu hizmeti sunan sohbet robotları gibi nispeten zararsız olanlardan, muhaliflerin sahte videolarını oluşturmak ve yaymak gibi tehlikeli olanlara kadar farklı şekillerde deepfake'leri kullanacaktır. Gazetecilerin ve doğruluk kontrolörlerinin siyasi deepfake'lerin doğruluğunu sürekli olarak değerlendirmeleri, kötü niyetli kullanımları tespit etmeleri ve halkı tehlikeye karşı uyarıp uyarmama ya da nasıl uyaracakları konularında mantıklı seçimler yapmaları gerekecektir. Vatandaşlar; üreticiler, izleyiciler, yorumlayanlar ve paylaşanlar olarak sentetik medyada gezinmeye çalışacak ve bu davranışları benimserken uydukları normlar çok önemli olacaktır. Son olarak, siyasi deepfake'ler sosyal bilimciler için önemli ampirik zorluklar ve çözümü zor normatif bulmacalar yaratmaya devam edecektir. Deepfake'leri sadece teknolojik meraklar olarak ele almak akıllıca değildir. Riskler çok yüksektir ve siyasal iletişim akademisyenleri, kamusal tartışmanın niteliği ve kamuoyu oluşumu için siyasi deepfake'lerin etkilerini anlamak açısından çok önemli bir konuma sahiptir.

Çıkar Çatışması Beyanı

Yazar(lar) bu makalenin araştırması, yazarlığı ve/veya yayını ile ilgili olarak herhangi bir potansiyel çıkar çatışması beyan etmemiştir.

Finansman

Yazar(lar) bu makalenin araştırması, yazarlığı ve/veya yayını için herhangi bir mali destek almamıştır.

Tamamlayıcı Materyal

Bu makalenin ek materyali çevrimiçi olarak mevcuttur.

Notlar

1. Bu makalenin ilk taslağını yazdığımızda, video

Avustralya Yayın Kurumu'nun web sitesinde mevcuttu, ancak sonrasında kaldırıldı.

2. Çevrimiçi dezenformasyonu niyetli olarak yanlış yönlendiren kasıtlı davranış, çevrimiçi mezenformasyonu ise sehven yanlış yönlendiren kasıtsız davranış olarak tanımlayanları temel alıyoruz. Bkz. Jack'ten (2017) akt. Chadwick vd. (2018). Deepfake'ler dezenformasyondur çünkü kasıtlı eylemlerle (deepfake videonun oluşturulması) ortaya çıkarlar. Ancak, doğru temsiller olduğuna istemeden inanan kişiler tarafından çevrimiçi olarak dolaşıma sokulduklarında ise mezenformasyon haline gelirler. Bu çalışmanın amaçları doğrultusunda, bir deepfake paylaşma kararını şekillendiren faktörleri açıklamaya çalışmadığımız için bu ayrım önem arz etmemektedir.

3. Deepfake'in sosyal medyada haberlere duyulan güven üzerinde doğrudan bir etkisi olmasını beklemiyoruz. Hayes (2013, s. 88)'in belirttiği üzere, aracı bir modelin test edilmesi, bağımsız değişkenin bağımlı değişken üzerinde doğrudan bir etkisi olduğunu varsaymayı ve göstermeyi gerektirmez.

4. Loughborough Üniversitesi'ndeki Çevrimiçi Yurttaşlık Kültürü Merkezi'nin faaliyetlerini desteklemek amacıyla bu anketi ücretsiz olarak gerçekleştiren Opinium Research'e teşekkür ederiz. Anketler İngiliz halkının %99'unun hem Obama'yı hem de Trump'ı tanıdığını göstermektedir (YouGov, 2019a, 2019b).

5. Bkz. Tamamlayıcı Bilgiler, Ek 3.

6. Bkz. Tamamlayıcı Bilgiler, Ek 4.

7. Bu ölçümle ile ilgili bilgi için bkz. Tamamlayıcı Bilgiler, Ek 5.

8. Eğitici gösterimi içeren tam video ve yanıltıcı 4 saniyelik klip karşılaştırıldığında Ki-Kare = 8.8, $df = 2$, $p = .012$, p (Holm) = .024, p (Bonferroni) = .036; eğitici gösterimi içeren tam video ve yanıltıcı 26 saniyelik klip karşılaştırıldığında Ki-Kare=15, $df = 2$, $p = .000$, p (Holm) = .002, p (Bonferroni) = .002; yanıltıcı 4 saniyelik ve 26 saniyelik klipler karşılaştırıldığında, Ki-Kare =1.6, $df = 2$, $p = .572$, p (Holm) = .448, p (Bonferroni) = 1.000.

9. Bu regresyonun tam sonuçları için bkz. Tamamlayıcı Bilgiler, Ekler 6.

10. Bu regresyonun tam sonuçları için bkz. Tamamlayıcı Bilgiler, Ekler 7.

11. Modeli R'daki "*psych*" paketini kullanarak çalıştırdık (Revelle, 2018).

12. Test ettiğimiz aracı model, uygulamadan sonra ölçülen aracıyı – deepfake hakkındaki belirsizliği – içermektedir. Montgomery vd. (2018); bunun rastgele atamayı tehlikeye atabileceğini ve nedensel çıkarımları saptırabileceğini göstermektedir. Bununla birlikte, "Buradan çıkarılacak ders, mekanizmaları incelemenin imkansız olduğu ya da araştırmacıların nedensel yolları anlamaya çalışmaktan vazgeçmeleri gerektiği değildir". Olası çözümler olarak "aracı unsurları etkileyen ancak sonucu etkilemeyen bir uygulama" içeren tasarımlardan bahsetmektedirler (Montgomery vd., 2018, s. 772). Benzer şekilde Pearl (2014, s. 4), "yardımcı değişkenlerin [aracılar dahil] uygulamadan nedensel olarak etkilenmedikleri sürece, uygulama öncesinde olması gerekmediğini" savunmaktadır. Modelimiz bu kriterleri karşılamaktadır çünkü uygulamalarımız, H2 ile ilgili tartışmamızda gösterildiği gibi aracı değişkeni (belirsizlik) etkilemiş ancak aracı modelimizin sonucu olan sosyal medyadaki haberlere duyulan güveni etkilememiştir. Yanıltıcı 4 saniyelik klibi izleyen katılımcıların uygulama sonrası sosyal medyadaki haberlere duydukları ortalama güven düzeyleri 0,673, yanıltıcı 26 saniyelik klip izleyenlerin 0,711 ve eğitici açıklamalı tam videoyu izleyenlerin 0,707'dir. ANOVA F katsayısının 0.467 (p= .627) olması, izlenen video ile sosyal medyadaki haberlere duyulan güven arasında anlamlı bir ilişki olmadığını göstermektedir. Bu durum, H3'ü test etmek için yürüttüğümüz aracılık regresyonu (Tablo 1) tarafından da doğrulanmaktadır; bu regresyon,

uygulamanın sosyal medyadaki haberlere duyulan güven üzerinde önemli oranda doğrudan bir etkisi olmadığını göstermektedir (Katsayı= 0.005, *SE*= 0.034, *p*= .887).

13. Bu regresyon aracılık modelinin tam sonuçları için bkz. Tamamlayıcı Bilgiler, Ek 8. Dolaylı etki için ab katsayısı Tablo 1'deki modelle aynıdır, ancak işareti pozitiftir: −0.085 × −0.175 = 0.015 (95% CI = [0.007, 0.026]).

14. Bkz. Tamamlayıcı Bilgiler, Ek 2.

Kaynaklar

Adobor H. (2006). Optimal trust? Uncertainty as a determinant and limit to trust in inter-firm alliances. *Leadership & Organization Development Journal*, 27(7), 537–553.

Alvarez R. M. (1997). *Information and elections*. University of Michigan Press.

Alvarez R. M., Brehm J. (1997). Are Americans ambivalent towards racial policies? *American Journal of Political Science*, 41(2), 345–374.

Angriawan A., Thakur R. (2008). A parsimonious model of the antecedents and consequence of online trust: An uncertainty perspective. *Journal of Internet Commerce*, 7(1), 74–94.

Arendt H. (1951). *The origins of totalitarianism*. Harcourt Brace.

Arendt H. (1978, October 26). Hannah Arendt: From an Interview. *The* New York Review of Books https://www.nybooks.com/articles/1978/10/26/hannah-arendt-from-an-interview/

Aufderheide P. (Ed.). (1992). *Media literacy: A report of the national leadership conference on media literacy*. Aspen Institute.

Baidu Research. (2017). *Deep voice 3: 2000-speaker neural text-to-speech*. http://research.baidu.com/Blog/index-view?id=91

Balliet D., Van Lange P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, 139(5), 1090–1112.

BBC News. (2018, June 11). *India WhatsApp "child kidnap" rumours claim two more victims*. https://www.bbc.co.uk/news/world-asia-india-44435127

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.

Berinsky A. J. (2004). *Silent voices: Public opinion and political participation in America*. Princeton University Press.

Berinsky A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47(2), 241–262.

Cappella J. N., Jamieson K. H. (1996). News frames, political cynicism, and media cynicism. *The Annals of the American Academy of Political and Social Science*, 546(1), 71–84.

Chadwick A., Vaccari C. (2019). *News sharing on UK social media: Misinformation, disinformation, and correction*.

https://www.lboro.ac.uk/media/media/research/o3c/Chadwick%20Vaccari%20O3C-1%20News%20Sharing%20on%20UK%20Social%20Media.pdf

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.

Cook K. S., Gerbasi A. (2011). Trust. In Hedström P., Bearman P., Bearman P. S. (Eds.), *The Oxford handbook of analytical sociology*. Oxford University Press.

Downs A. (1957). *An economic theory of democracy*.

Harper.

Facebook. (2018, April 17). *You won't believe what Obama says in this video!* https://www.facebook.com/watch/?v=10157675129905329

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.

Garrett R. K., Nisbet E. C., Lynch E. K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication*, 63(4), 617–637.

Gault M. (2016, November 6). After 20 minutes of listening, new Adobe tool can make you say anything. *Motherboard*. https://www.vice.com/en_us/article/jpgkxp/after-20-minutes-of-listening-new-adobe-tool-can-make-you-say-anything

GitHub. (2019a). *Faceswap*. https://github.com/deepfakes/faceswap

GitHub.(2019b). *DeepFaceLab*. https://github.com/iperov/DeepFaceLab#Where_can_I_get_the_FakeApp

Goel S., Anderson A., Hofman J., Watts D. J. (2015). The structural virality of online diffusion. *Management Science*, 62(1), 180–196.

Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., . . . Bengio Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 3, 2672–2680.

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.

Green M. C., Donahue J. K. (2011). Persistence of belief change in the face of deception: The effect of factual stories revealed to be false. *Media Psychology*, 14(3), 312–331.

Hanitzsch T., Van Dalen A., Steindl N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *The International Journal of Press/Politics*, 23(1), 3–23.

Hayes A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Jack C. (2017) *Lexicon of Lies: Terms for Problematic Information*. Data & Society Research Institute.

Iyengar S., Vavreck L. (2012). Online panels and the future of political communication research. In Semetko H., Scammell M. (Eds.), *The SAGE handbook of political communication* (pp. 225–240). SAGE.

Kreuter F., Presser S., Tourangeau R. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.

Lewandowsky S., Ecker U. K. H., Seifert C. M., Schwarz N., Cook J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.

Molm L. D., Schaefer D. R., Collett J. L. (2009). Fragile and resilient trust: Risk and uncertainty in negotiated and reciprocal exchange. *Sociological Theory*, 27(1), 1–32.

Montgomery J. M., Nyhan B., Torres M. (2018). How

conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760–775.

Moorman C., Deshpande R., Zaltman G. (1993). Factors affecting trust in market research relationships. *Journal of Marketing*, 57(1), 81–101.

Newman E. J., Garry M., Unkelbach C., 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.

Newman N., Fletcher R., Kalogeropoulos A., Levy D. A. L., Nielsen R. K. (2018). *Reuters Institute digital news report 2018*. Reuters Institute for the Study of Journalism. http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf

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.

Pasek J. (2015). When will nonprobability surveys mirror probability surveys? Considering types of inference and weighting strategies as criteria for correspondence. *International Journal of Public Opinion Research*, 28(2), 269–291.

Pearl J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, 19(4), 459–481.

Petersen M. B., Osmundsen M., Arceneaux K. (2018, September 1). A "need for chaos" and the sharing of hostile political rumours in advanced democracies. *PsyArXiv Preprints*. https://psyarxiv.com/6m4ts/

Phillips W., Milner R. M. (2017). *The ambivalent internet: Mischief, oddity, and antagonism online*. Polity.

Pomerantsev P. (2015, January 4). Inside Putin's information war. *Politico*. https://www.politico.com/magazine/story/2015/01/putin-russia-tv-113960

Prior M. (2013). Visual political knowledge: A different road to competence? *Journal of Politics*, 76(1), 41–57.

Revelle W. (2018). *psych: Procedures for personality and psychological research*. Northwestern University. https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research

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.

Rössler A., Cozzolino D., Verdoliva L., Riess C., Thies J., Nießner M. (2018). FaceForensics: A large-scale video dataset for forgery detection in human faces. https://arxiv.org/pdf/1803.09179.pdf

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.

Shoemaker P. J., Eichholz M., Skewes E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, 14(2), 193–201.

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 Metzger M., Flanagin A.

(Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Suwajanakorn S., Seitz S. M., Kemelmacher-Shlizerman I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, 36(4), Article 95.

Thies J., Zollhofer M., Stamminger M., Theobalt C., Nießner M. (2016). Face2face: Real-time face capture and reenactment of RGB videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2387–2395).

Thorson E., Wicks R., Leshner G. (2012). Experimental methodology in journalism and mass communication research. *Journalism & Mass Communication Quarterly*, 89(1), 112–124.

Tsfati Y., Cappella J. N. (2003). Do people watch what they do not trust? Exploring the association between news media skepticism and exposure. *Communication Research*, 30(5), 504–529.

Twitter. (2018, April 17). *You won't believe what Obama says in this video!* https://twitter.com/BuzzFeed/status/986257991799222272

Van Duyn E., Collier J. (2018). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication & Society*, 22(1), 29–48. https://doi.org/10.1080/15205436.2018.1511807

Vosoughi S., Roy D., Aral S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.

Waisbord S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism Studies*, 19(13), 1866–1878.

Wenzel A. (2019). To verify or to disengage: Coping with "fake news" and ambiguity. *International Journal of Communication*, 13, Article 19.

Witness. (2018, June 11). *Mal-uses of AI-generated synthetic media and deepfakes: Pragmatic solutions discovery convening*. http://witness.mediafire.com/file/q5juw7dc3a2w8p7/Deepfakes_Final.pdf/file

Witten I. B., Knudsen E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, 48(3), 489–496.

Yamagishi T., Yamagishi M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, 18(2), 129–166.

YouGov. (2019a, November 1). *Public figure—Barack Obama*. https://yougov.co.uk/topics/politics/explore/public_figure/Barack_Obama

YouGov. (2019b, November 1). *Public figure—Donald Trump*. https://yougov.co.uk/topics/politics/explore/public_figure/Donald_Trump

YouTube. (2018, April 17). *You won't believe what Obama says in this video!* https://www.youtube.com/watch?v=cQ54GDm1eL0

Yazar Bilgileri

Author details

Doç. Dr., Ankara Hacı Bayram Veli Üniversitesi İletişim Fakültesi, seyda.kocak@hbv.edu.tr.

Kaynak Göstermek İçin

To Cite This Article

Vaccari, C., ve Chadwick, A. (2020). Deepfake ve dezenformasyon: Sentetik siyasi videoların yanıltma, belirsizlik ve haberlere duyulan güven üzerindeki etkisini araştırma. (Çev. Şeyda Koçak Kurt), *Yeni Medya (17)*, 384-398.