DEEP FAKES ACCOUNTABILITY ACT: OVERBROAD AND INEFFECTIVE

ZACHARY SCHAPIRO[*]

Abstract: Improvements in technology have made it easy to create deep fakes: realistic false digital records depicting a person performing actions that did not occur. Although there are legitimate uses for deep fakes, like parodies or finishing a movie where a principal performer has died, they can inflict harm upon the individual depicted or the general public if, for example, used to influence the outcome of an election. Considering these detrimental uses, Congress has proposed the Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019 (DEEP FAKES Act) to impose civil and criminal liability upon people who create or share deep fakes without an appropriate disclaimer. Despite its purported benefits, this legislation would actually burden legitimate users of deep fake technology and encumber courts with litigation due to its overbroad definition of "deep fake." Instead, the abuse of deep fake technology should be addressed judicially through a combination of copyright, criminal, and state tort law to impose liability on bad actors.

INTRODUCTION

Although Hollywood special effect artists have been manipulating audio and video for decades, now, anyone can download software that employs machine learning algorithms to impose their face upon a photo or video to create content that appears legitimate.[1] This type of artificially generated content, known as a "deep fake," came to social prominence in November 2017 when a Reddit user began posting pornographic clips featuring celebrity faces on the performers' bodies.[2] Since then, deep fakes have garnered legitimate uses in comedy and entertainment, appearing in viral mobile applications and videos.[3] For example, a viral video

[*] J.D. Candidate, Boston College Law School (expected 2022) B.S., Applied Economics and Management, Cornell University (2017).

[1] @iperov, *DeepFaceLab,* GITHUB (June 4, 2018), https://github.com/iperov/DeepFaceLab (distributing free software that lets users create their own deep fakes); J.M. Porup, *How and Why Deepfake Videos Work — and What Is at Risk*, CSO (Apr. 10, 2019, 3:00 AM), https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html (providing background on audio and video editing and the emergence of deep fakes).

[2] Tom Simonite, *Most Deepfakes Are Porn, and They're Multiplying Fast*, WIRED (Oct. 7, 2019, 10:00 AM), https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/ (discussing the rise of deep fakes).

[3] Colum Murphy & Zheping Huang, *China's Red-Hot Face-Swapping App Provokes Privacy Concern*, BLOOMBERG (Sept. 1, 2019, 9:47 PM), https://www.bloomberg.com/news/articles/2019-09-02/china-s-red-hot-face-swapping-app-provokes-privacy-concern (discussing Zao, a viral mobile app allowing users to impose their faces on famous movie and television scenes); David

depicting President Obama warning the public about the dangers of misinformation and deep fakes was actually a deep fake video of Jordan Peele, a comedian, performing a voice impression of the former President.[4]

Despite their potential for positive applications, deep fakes pose a legitimate risk to the democratic process.[5] An ill-timed release of a deep fake depicting a candidate could affect the outcome of an election before the video can be definitively refuted.[6] To combat the potential harm deep fakes pose to the democratic process, Texas and California have recently passed legislation to impose liability on those who create and/or distribute digitally manipulated media depicting political candidates.[7] Now, the House of Representatives has introduced the Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019 (DEEP FAKES Act) to impose federal restrictions on the use of harmful deep fake content and liability on those who create and distribute such content.[8] Part I of this Essay explains what deep fakes are, the problems they pose to online media consumption, and how the DEEP FAKES Act would combat these problems.[9] Part II discusses the scope of the DEEP FAKES Act.[10] Lastly, Part III argues that existing liability in copyright, state tort, and criminal law better protect against harmful deep fakes than the DEEP FAKES Act.[11]

I. DEEP FAKES AND THE END OF TRUTH ONLINE

This Part discusses what deep fakes are, the problems they pose to media consumption online, and how the DEEP FAKES Act would combat these problems.[12] Section A discusses how deep fakes are created and the potential harm they can cause.[13] Section B outlines how the proposed DEEP FAKES Act seeks to

Mack, *This PSA About Fake News from Barack Obama Is Not What It Appears*, BUZZFEED NEWS (Apr. 17, 2018, 11:26 AM), https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed#.gcxNolpGL (sharing a viral deep fake video that imposed President Barack Obama's face over comedian Jordan Peele's body and warned viewers of the dangers of "fake news").

[4] Mack, *supra* note 3 (warning viewers of the dangers of misinformation online using a deep fake video that imposed President Barack Obama's face over comedian Jordan Peele's body).

[5] *See* Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1753–54, 1768, 1778 (2019) (describing how a deep fake can quickly misinform the public and threaten the democratic process, especially in the age of instantaneous social media sharing).

[6] *See id.* at 1778 (describing how deep fakes can quickly misinform the public and undermine the democratic process).

[7] *See, e.g.*, TEX. ELEC. CODE ANN. § 255.004 (West 2019) (classifying the creation, publication, and distribution of a harmful deep fake video depicting a candidate as a Class A misdemeanor); CAL. ELEC. CODE § 20010 (West 2019) (allowing for equitable relief and damages for the distribution of "materially deceptive audio or visual media" depicting a candidate).

[8] Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019).

[9] *See infra* notes 12–42 and accompanying text.

[10] *See infra* notes 43–120 and accompanying text.

[11] *See infra* notes 121–151 and accompanying text.

[12] *See infra* notes 13–42 and accompanying text.

[13] *See infra* notes 15–25 and accompanying text.

address these problems by imposing disclosure requirements on deep fake content.[14]

A. What Is a Deep Fake?

Originally, the term "deep fake" referred only to video forgeries created through generative adversarial networks.[15] A generative adversarial network employs two machine learning models to create the forgery.[16] The first machine learning model iteratively attempts to superimpose the face of one person onto a video of another.[17] The second machine learning model then attempts to detect the forgery.[18] The final output is a deep fake created entirely through artificial intelligence that the second model is not able to detect.[19] Today, the term "deep fake" extends beyond video to include audio created using similar machine learning techniques.[20] Such deep fakes portray individuals speaking or acting in ways that did not actually occur.[21]

Deep fakes depicting politicians threaten to undermine the democratic process because the reputational damage is caused before the falsified videos can be definitively refuted.[22] Because people tend to perceive online content as true or false in accordance with their prior affiliations, even if a deep fake is eventually refuted, people might believe the version of events most in-line with their previously held opinions.[23] This problem is amplified as people increasingly consume a majority of their news on social media.[24] In fact, people are more likely to share and engage with false news and misinformation, such as deep fakes, on social media than they are the truth, thus exacerbating the harm deep fakes can cause.[25]

[14] *See infra* notes 26–42 and accompanying text.

[15] Porup, *supra* note 1.

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *Id.*

[20] *See* Catherine Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, WALL ST. J. (Aug. 30, 2019), https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (using the term "deep fake" to describe audio created using machine learning techniques).

[21] *See* Porup, *supra* note 1 (describing what deep fakes are and the damage they can cause). Specifically, deep fakes can damage an individual's reputation and misinform the public. *Id.*

[22] *See* Chesney, *supra* note 5 (describing how deep fakes can misinform the public and influence elections by spreading rapidly on social media).

[23] *See* Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned'*, WASH. POST (June 12, 2019, 4:44 PM), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ (describing how the party viewing altered photos or video is likely to perceive it in accordance with their previous affiliations).

[24] Nicole Martin, *How Social Media Has Changed How We Consume News*, FORBES (Nov. 30, 2018, 4:26 PM) https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/#3a493bc63c3c.

[25] *See id.* (describing how a majority of people consume news on social media); Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCI. 1146 (2018) (describing how people are more likely to share and engage with false news because of its novelty).

B. Combatting Deep Fakes: The DEEP FAKES Act

With the 2020 election on the horizon, Congress introduced the DEEP FAKES Act to mandate disclosure of deep fakes to ensure the electorate knows when online content is real or fake.[26] The DEEP FAKES Act mandates, with narrow exceptions, that any "advanced technological false personation record" shared or intended to be shared must contain a disclaimer that the content depicted is false.[27] The Act defines an "advanced technological false personation record" as any deep fake that a reasonable person would believe shows "material activity" of a living or deceased person that the person did not in fact perform.[28] Additionally, a "deep fake" is defined to include all forms of media, such as videos, sound recordings, photos, and any other "technological representation[s] of speech or conduct" that were created through substantially technical means (as opposed to through physical or verbal impersonation).[29] Under the Act, "material activity" is any falsified technological representation that either directly causes, or a reasonable person believes could cause, individual or societal harm.[30] Taken together, the DEEP FAKES Act concerns any media that depicts a person doing something that did not occur and that causes, or could be perceived to cause, harm to either society or the individual depicted.[31]

Depending on the medium of the record—audio, photo, or video—the DEEP FAKES Act outlines the specific method of disclosure or watermarking necessary to convey to viewers or listeners that the content is not real.[32] For example, any deep fake video must contain a clear verbal statement that the video is fake, a written statement identifying the video as fake, and a statement describing the alterations made that is visible throughout the video.[33]

The DEEP FAKES Act imposes liability on anyone who shares or intends to share a harmful deep fake that fails to meet the disclosure requirements.[34] If someone knowingly fails to disclose the deep fake or removes the disclosures, they are subject to criminal liability if the deep fake is pornographic, intends to incite violence, interferes with an official proceeding, including an election, amounts to foreign interference with domestic affairs, or facilitates fraud.[35] In addition, anyone who fails to disclose the deep fake or removes disclosures is subject to civil liability.[36] In such civil proceedings, the person depicted is entitled to monetary damages of at least $50,000 per deep fake for each record shared or altered.[37] In

[26] Mary Frost, *Clarke Introduces Bill to Combat High-Tech Altered Videos*, BROOKLYN DAILY EAGLE (June 14, 2019), https://brooklyneagle.com/articles/2019/06/14/clarke-introduces-bill-to-combat-high-tech-altered-videos/.
[27] H.R. 3230.
[28] *Id.*
[29] *Id.*
[30] *Id.*
[31] *See id.* (imposing criminal and civil liability on digital false personation records).
[32] *Id.*
[33] *Id.*
[34] *Id.*
[35] *Id.*
[36] *Id.*
[37] *Id.*

addition, the court may issue an injunction requiring that disclosures to be added to the offending content.[38]

Nonetheless, the DEEP FAKES Act recognizes that there are some legitimate uses of deep fakes that should not be subject to the disclosure requirements.[39] Specifically, the Act provides exceptions to the requirements for government-created records, parodies, and movies, television or music where the person represented provided consent.[40] The DEEP FAKES Act also creates a taskforce within the Department of Homeland Security dedicated to developing deep fake detection technologies to share with private Internet platforms.[41] Overall, the DEEP FAKES Act creates a broad regulatory scheme to prevent the dissemination of false and harmful information.[42]

II. Benefits and Drawbacks of the DEEP FAKES Act

This Part discusses the scope of the Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act (DEEP FAKES Act).[43] Section A examines the intended benefits of the DEEP FAKES Act.[44] Section B discusses unintended drawbacks of the DEEP FAKES Act.[45]

A. Benefits of the DEEP FAKES Act: Curbs the Dissemination of Misinformation and Allows for in Rem Civil Actions

The DEEP FAKES Act was introduced to address the disruption deep fakes could cause to elections.[46] Specifically, a deep fake has the potential to interfere with the political process when it is depicts a candidate and is released close to an election.[47] Before the deep fake can be refuted, voters may have already cast their vote, relying on false information portrayed in the deep fake.[48] Rapid technological change has made it difficult for regulators to ensure fairness and protect human dignity.[49] Deep fake technology has only recently achieved social prominence and become accessible to the general public.[50] With the 2020 election on the horizon, the DEEP FAKES Act represents the rare exception where the legislature, by

[38] *Id.*

[39] *Id.*

[40] *Id.*

[41] *Id.*

[42] *See id.* (creating civil and criminal liability for sharing deep fakes without appropriate watermarks and disclosures).

[43] *See infra* notes 46–120 and accompanying text.

[44] *See infra* notes 46–63 and accompanying text.

[45] *See infra* notes 64–120 and accompanying text.

[46] Frost, *supra* note 26.

[47] *Id.*

[48] *See* Chesney, *supra* note 5 (describing how deep fakes shared online misinform the public and can affect elections).

[49] Daniel Malan, *The Law Can't Keep Up With New Tech. Here's How To Close The Gap*, World Econ. F. (June 21, 2018), https://www.weforum.org/agenda/2018/06/law-too-slow-for-new-tech-how-keep-up/.

[50] *See* Simonite, *supra* note 2 (describing how deep fakes came to social prominence on Reddit in 2017).

criminalizing those who fail to mark their videos as deep fakes, is promptly responding to an emerging technology before it hurts the public.[51]

The DEEP FAKES Act also provides benefits beyond protecting the public from misinformation in elections.[52] In the context of deep fake pornography,[53] the Act serves to impose criminal liability on deep fake pornography creators.[54] Additionally, for anonymously shared deep fakes, the Act allows for an in rem civil action to be filed against the content itself.[55] In such situations, the court may issue an "order declaring there to be a substantial likelihood that the material activity depicted in such [deep fake] is false and lacking the watermarks and disclosures required under [the Act]."[56] The court may also order the forfeiture of profits derived from the deep fake, if any.[57]

The Act's creation of an in rem civil action offers clarity at a time where the Internet has posed significant challenges to traditional exercises of personal jurisdiction.[58] This is because the Internet operates across territorial boundaries, which a key factor in determining whether there is personal jurisdiction for a court to hear a case.[59] To determine whether a court can exercise personal jurisdiction over a website, courts have applied the *Zippo* test, which focuses on the interactive nature of the website and the commercial nature of the exchange of information that occurs on it.[60] This analysis breaks down in the context of social media because an individual poster is not the owner or operator of the website and the user posts are seen as shared information instead of interactive, so the exercise of personal jurisdiction is inappropriate.[61] The DEEP FAKES Act avoids these issues of personal jurisdiction for online action by allowing for an in rem action against the offending content when the court does not have personal jurisdiction over the deep fake creator.[62] Despite these benefits, however, critics argue that the Act sweeps in too much online content as deep fakes.[63]

[51] *See id.* (describing how deep fakes became more prominent in 2017); Frost, *supra* note 26 (stating the DEEP FAKES Act was introduced over fears of deep fake election interference in 2020); Malan, *supra* note 49 (discussing how the law is normally not responsive to rapid technological change).

[52] *See infra* notes 53–63 and accompanying text (describing additional benefits of H.R. 3230).

[53] Drew Harwell, *Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: 'Everybody Is a Potential Target'*, WASH. POST (Dec. 30, 2018, 10:00 AM), https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/. Deep fake pornography describes the situation where anonymous users create and share pornography that has, for example, a celebrity's face superimposed on the body of a pornographic film actor, to make people believe that the celebrity is actually the one participating in the video. *Id.*

[54] Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019).

[55] *Id.*

[56] *Id.*

[57] *Id.*

[58] Zoe Niesel, *#PersonalJurisdiction: A New Age of Internet Contacts*, 94 IND. L. J. 103, 103 (2019).

[59] *Id.*

[60] *Zippo Mfg. Co. v. Zippo Dot Com, Inc.*, 952 F. Supp. 1119, 1124 (W.D. Pa. 1997).

[61] Niesel, *supra* note 59, at 130.

[62] H.R. 3230. 119, 67-83, 84-94, 95-119

[63] *See infra* notes 68–84 and accompanying text.

B. Drawbacks of the DEEP FAKES Act: Its Overbroad Statutory Definition of "Deep Fake," Its Potential to Impede Legitimate Use, and Issues in Detecting Infringing Deep Fake Content

This Section examines some of the unintended drawbacks of the DEEP FAKES Act.[64] Subsection 1 examines the statutory definition of deep fake and its effects.[65] Subsection 2 discusses how the language of the Act might hinder legitimate users of deep fakes.[66] Finally, Subsection 3 explores the difficulty in detecting deep fake content and the role of social media companies in their detection.[67]

1. The Statutory Definition of Deep Fake

According to the commonly accepted definition of deep fake, the audio or video forgery must be created using automated machine learning algorithms.[68] The DEEP FAKES Act, however, broadly defines a "deep fake" as one that is created using predominantly technical means.[69] The term "technical means" can encompass any creation method using a computer, as opposed to traditional methods of physical and verbal impersonation.[70] For example, in May 2019, a video circulated online that appeared to show Nancy Pelosi, Speaker of the House of Representatives, drunkenly slurring her words while giving a speech.[71] The video went viral, quickly spreading across social media.[72] The video was doctored using conventional video and audio editing software to slow the video and adjust the pitch of Pelosi's voice.[73] Similarly, a slowed down video of President Donald Trump was shared online making him appear as if he was delivering a speech drunk.[74] As defined under the Act, deep fakes are seemingly authentic depictions of speech or conduct that did not actually occur created using technical means.[75] Even though

[64] *See infra* notes 68–120 and accompanying text.

[65] *See infra* notes 68–84 and accompanying text.

[66] *See infra* notes 85–95 and accompanying text.

[67] *See infra* notes 96–120 and accompanying text.

[68] *See* H.R. 3230 (defining deep fake as a technological false personation record created using technical means); Harwell, *infra* note 71 (outlining how the Nancy Pelosi video was created using traditional audio and video editing software); Kimmel, *infra* note 74 (describing how the Donald Trump video was created by slowing down existing footage); Porup, *supra* note 15 (defining deep fake as a video forgery created with machine learning algorithms without human intervention); Stupp, *supra* note 20 (using "deep fake" to describe audio forgeries created using machine learning algorithms).

[69] H.R. 3230.

[70] *See id.* (defining deep fake as a technological false personation record created using technical means).

[71] Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 24 2019, 4:41 PM), https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/.

[72] *Id.*

[73] *Id.*

[74] Jimmy Kimmel Live, *Drunk Donald Trump – Why He Got Elected*, YOUTUBE (Feb. 9, 2017), https://www.youtube.com/watch?v=dK3LbVFgyqQ.

[75] H.R. 3230.

Nancy Pelosi and Donald Trump actually gave those speeches, they would still be considered deep fakes under the Act.[76] This is because the seemingly authentic conduct—delivering speeches drunk—did not actually occur and was produced using technical means—audio and video editing software.[77]

The DEEP FAKES Act requires the deep fake "to cause perceptible individual or societal harm," which includes reputational damage.[78] This additional harm requirement of the DEEP FAKES Act is easily met because, like defamation at common law, any publicly circulated deep fake, which is inherently false, could be perceived as harmful to the individual's reputation.[79] The aforementioned Pelosi and Trump videos illustrate how the same creation method (slowing down a video) can be perceived as either harmful or humorous depending on one's political tendencies.[80] Additionally, deep fake detection methods currently struggle to distinguish between satirical and malicious deep fakes.[81]

Any record created using conventional audio, photo, and video editing software—of which there are many—can be easily and quickly debunked through the circulation of the unedited source materials.[82] Therefore, excessive litigation might arise concerning these types of records because of the presumption towards individual reputational harm, the ability to easily prove the record is false, and the significant civil liability imposed by the Act.[83] Because the Act sweeps in more than just deep fakes created using automated machine learning algorithms, which are more difficult to refute, the DEEP FAKES Act will likely encumber the courts with excessive litigation over easily-refuted deep fake content that causes minimal harm.[84]

[76] Jimmy Kimmel Live, *supra* note 74; Harwell, *supra* note 71; *see* H.R. 3230 (defining deep fake as a technological false personation record created using technical means).

[77] Jimmy Kimmel Live, *supra* note 74; Harwell, *supra* note 71; *see* H.R. 3230 (defining deep fake as a technological false personation record created using technical means).

[78] H.R. 3230.

[79] *See id.* (defining deep fake as a technological false personation record that causes harm to the individual depicted or society as a whole); RESTATEMENT (SECOND) OF TORTS § 621 (AM. LAW INST. 1965) (describing the common law presumption of actual harm from defamation).

[80] *See* Harwell, *supra* note 71 (describing how the Nancy Pelosi video harmed her reputation); Kimmel, *supra* note 74 (showing that slowing down footage of Donald Trump is humorous).

[81] Hayley Tsukayama et al., *Congress Should Not Rush to Regulate Deepfakes*, ELEC. FRONTIER FOUND. (June 24, 2019), https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes.

[82] *See* Harwell, *supra* note 71 (describing how the Nancy Pelosi video was debunked using the original, unedited video footage).

[83] *See* H.R. 3230 (defining deep fake as a technological false personation record created using technical means and imposing a minimum $50,000 civil liability for violation of the Act); RESTATEMENT (SECOND) OF TORTS § 621 (AM. LAW INST. 1977) (discussing defamation at common law and its presumption of actual harm); Harwell, *supra* note 71 (detailing how an altered video being shared online was quickly debunked using the original, unedited footage).

[84] *See infra* notes 109–120 and accompanying text (describing the difficulty in detecting deep fakes); H.R. 3230 (defining deep fake as a technological false personation record created using technical means and imposing a minimum $50,000 civil liability for violation of the Act); RESTATEMENT (SECOND) OF TORTS § 621 (describing the common law presumption of actual harm from defamation); Harwell, *supra* note 71 (describing how the Nancy Pelosi video was debunked using the original, unedited video footage).

2. Impeding Legitimate "Deep Fakes"

The number of deep fake videos shared online is growing rapidly, but the vast majority (ninety-six percent) of deep fakes shared online are pornographic.[85] These deep fake creator communities are hosted on deep fake porn websites or forum sites like 4chan and 8chan, which are known for hosting unethical and illegal activity.[86] Although the Act seeks to combat misinformation online, creators of these pornographic deep fakes are unlikely to be deterred by the Act.[87] This is because these creators probably have no intent to adhere to the transparency requirements imposed by the DEEP FAKES Act.[88] Assuming this is true, and malicious deep fake creators will create anyway, the Act would primarily hinder only legitimate deep fake users.[89] A legitimate deep fake might be one that discloses that is purposefully and creatively designed to promote public discourse. [90]

For example, a recent deep fake shared online depicted Mark Zuckerberg, creator of Facebook, boasting about the misuse of individuals' data.[91] This deep fake, which falls outside of the DEEP FAKES Act's listed exceptions, was meant to spur conversation regarding Facebook's data collection policies and the posting of fake news and deep fakes.[92] Under the DEEP FAKES Act, one could argue that the Zuckerberg deep fake materially harmed the reputation of Mark Zuckerberg and Facebook and that a reasonable person would believe the video to be true.[93] Although in this case the creator disclosed the record's false origins, this may not always be the case, and a deep fake created to promote discourse would be subject to liability.[94] Therefore, the Act may lead to fewer creative, legitimate users of deep fake technology for fear of inadvertently violating the Act.[95]

3. Issues in Detecting Infringing Deep Fake Content

[85] HENRY AJDER ET AL., THE STATE OF DEEPFAKES: LANDSCAPE THREATS AND IMPACT 1 (2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

[86] *Id.* at 4.

[87] *See* H.R. 3230 (imposing disclosure and watermarking requirements on deep fake content); AJDER ET AL, *supra* note 85, at 4 (stating that most deep fake creation communities are hosted on deep fake pornography websites and other forum-based sites like 4chan and 8chan, which are known for hosting illegal and unethical activity).

[88] *See* H.R. 3230 (imposing disclosure and watermarking requirements on deep fake content); AJDER ET AL, supra note 85, at 4 (discussing how most deep fake creation communities are located on deep fake porn websites and other forum communities known for illegal and unethical activity).

[89] *See* H.R. 3230 (imposing disclosure and watermarking requirements on deep fake content).

[90] E.g., Samantha Cole, *This Deepfake of Mark Zuckerberg Tests Facebook's Fake Video Policies*, VICE (June 11, 2019, 12:25 PM).

[91] *Id.*

[92] *Id.*; *see* H.R. 3230 (listing exceptions to the disclosure requirements imposed by the DEEP FAKES Act).

[93] *See* H.R. 3230 (imposing liability for failure to disclose that a record is a deep fake); Cole, *supra* note 90 (describing a deep fake of Mark Zuckerberg).

[94] *See* H.R. 3230 (imposing liability for failure to disclose that a record is a deep fake even if it is meant to promote discourse); Cole, *supra* note 90 (disclosing that the Mark Zuckerberg video is fake).

[95] *See* H.R. 3230 (imposing disclosure and watermarking requirements on all deep fake content).

The DEEP FAKES Act, if enacted, will create a taskforce within the Department of Homeland Security dedicated to developing deep fake detection technologies to share with private Internet platforms.[96] In other words, the Act encourages online sites, like Facebook, to take an active role in identifying deep fakes.[97] This additional responsibility for Internet service providers makes sense only if these sites are held liable for failing to remove deep fakes.[98] But neither the Act nor any other federal law currently imposes liability on these platforms for failure to remove harmful deep fakes.[99] Moreover, even if the DEEP FAKES Act imposed such liability, platforms would likely remove all deep fake content, harmful or not, to avoid all potential liability.[100]

Furthermore, Congress has explicitly protected websites from liability for the content posted by their users.[101] Section 230 of the Communications Decency Act of 1996 gives websites the ability to moderate content posted by their users without fear of liability.[102] Section 230 gives websites broad immunity from claims against them so long as they are not the publisher or speaker of the challenged content.[103] Social media and other websites predicated upon user-generated content have thrived due to Section 230 immunity.[104] Because of this immunity, social media companies and other websites have no current obligation to remove deep fake content because they serve as passive entities distributing the content, and not creating it.[105] Even so, Facebook announced that it will remove all deep fakes except those that are parodic or satirical.[106] Under the DEEP FAKES Act, social media providers and other websites would have no statutory obligation to remove

[96] *Id.*

[97] *See id.* (creating a Department of Homeland Security taskforce to develop deep fake detection technology).

[98] *See id.* (encouraging the development of deep fake detection technologies and sharing it with private companies).

[99] *See id.* (creating a taskforce to develop deep fake detection methods); 47 U.S.C. § 230 (2018) (providing broad protections to websites for the content posted by their users).

[100] *But see* Harwell, *supra* note 23 (explaining that whether a deep fake is considered harmful depends on how the party viewing the deep fake is likely to perceive it).

[101] *See* 47 U.S.C. § 230 (providing websites with broad immunity for the content posted by their users).

[102]*CDA 230: Legislative History*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230/legislative-history (last visited Jan. 6, 2020). Congress passed Section 230 in response to the New York State Supreme Court's controversial 1995 decision in *Stratton Oakmont, Inc. v. Prodigy Services Co.*, where an online service provider was liable for defamation due to content posted by its users. *See Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995) (finding defendant liable for defamation because they served as a moderator for content posted by third party users).

[103] 47 U.S.C. § 230; Amy J. Tindell, *"Indecent" Deception: The Role of Communications Decency Act § 230 in Balancing Consumer and Marketer Interests Online*, B.C. INTELL. PROP. & TECH. F. (July 19, 2009), http://bciptf.org/wp-content/uploads/2011/07/9-INDECENT.pdf; *CDA 230: The Most Important Law Protecting Internet Speech*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230 (last visited Jan. 6, 2020) [hereinafter *CDA 230*].

[104] *CDA 230*, *supra* note 102.

[105] *See id.* (describing the broad protections available to websites under 47 U.S.C. § 230); Cole, *supra* note 90 (discussing how Instagram will not remove deep fake content).

[106] Monika Bickert, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020), https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

the offending content because the Act fails to carve out an exception to the immunity provided by Section 230 of the Communications Decency Act.[107] Such an exception to Section 230, coupled with the broad definition of a deep fake in the DEEP FAKES Act may cause websites to censor all deep fakes to avoid incurring liability.[108]

Despite the lack of current legal obligations for websites to remove deep fake content, technology companies—in conjunction with academics and independent computer scientists—have started developing automated techniques to detect deep fake content.[109] Nevertheless, as deep fake detection methods improve, so do methods of generating deep fakes.[110] Any novel detection method informs the next iteration of generation methods of a new way to avoid detection.[111] For example, researchers identified that deep fake videos exhibited unhuman blinking patterns, but once this information was made public, the next iteration of deep fake generation algorithms took blinking patterns into account.[112] This never-ending game of cat and mouse allows bad actors to exploit software at the expense of either an individual person, a corporation, or the general public, while good actors attempt to stop them.[113] With this in mind, if the DEEP FAKES Act imposed liability on websites for failure to remove infringing deep fakes, they would be open to excessive litigation because it is impossible for them to detect all deep fake content.[114]

Even if all deep fake content could be algorithmically identified, in seeking to enforce the liabilities imposed by the DEEP FAKES Act, there persists the issue of discerning between an prohibited deep fake and a legitimate use.[115] A legitimate

[107] H.R. 3230.

[108] *See* H.R. 3230 (defining deep fake as any false media depiction created using technical means); Harwell, *supra* note 23 (discussing how people perceive manipulated photos or videos according to their previous affiliations).

[109] *See, e.g.*, Nick Dufour & Andrew Gully, *Contributing Data to Deepfake Detection Research*, GOOGLE AI BLOG (Sept. 24, 2019), https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html (outlining Google's efforts to develop automated deep fake detection methods, including releasing databases of deep fakes to help train detection algorithms); Mike Schroepfer, *Creating a Data Set and a Challenge for Deepfakes*, FACEBOOK ARTIFICIAL INTELLIGENCE (Sept. 5, 2019), https://ai.facebook.com/blog/deepfake-detection-challenge/ (announcing a Facebook-sponsored competition for detecting deep fake content).

[110] *See* Shruti Agarwal et al., *Protecting World Leaders Against Deep Fakes*, IEEE CONF. ON COMPUT. VISION AND PATTERN RECOGNITION WORKSHOPS (2019), http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf (describing how deep fake generation has changed over time).

[111] *See id.* (explaining how deep fake generation has evolved).

[112] *Id.*

[113] *See* Colin Barker, *Hackers and Defenders Continue Cybersecurity Game of Cat and Mouse*, ZDNET (Feb. 3, 2016, 2:00 PM), https://www.zdnet.com/article/hackers-and-defenders-continue-cyber-security-game-of-cat-and-mouse/ (describing the cat and mouse relationship between hackers and developers).

[114] *See id.* (describing how hackers continually identify new vulnerabilities after developers patch old ones); Harwell, *supra* note 23 (detailing how far more researchers are focusing on deep fake creation than detection).

[115] *See* H.R. 3230 (describing the types of deep fake content subject to the liability imposed by the proposed legislation).

use would be any of the enumerated exceptions listed in the DEEP FAKES Act—such as parody or government created deep fakes—or any digital record that fails to meet the definition of deep fake, most likely by failing to cause harm to the individual falsely depicted or the public.[116] Whether a deep fake depicting a politician is simply a parody or is instead a piece of propaganda used to negatively influence an election largely depends on the party consuming the deep fake content.[117] Courts then, must establish the "reasonable person," and as with any reasonableness test, this gives unguided discretion to the courts.[118] Furthermore, the DEEP FAKES Act mentions neither the burden of proof necessary to establish the answer to this question nor which party bears such a burden, which only muddies its enforcement.[119] Therefore, it is best left to the court to hold creators of malicious deep fakes accountable through the application of extant copyright, criminal, and state tort claims instead of imposing additional liability through the Act.[120]

III. EXISTING LAW BETTER ADDRESSES DEEP FAKES THAN THE DEEP FAKE ACT

This Part discusses why existing liability in federal copyright and state tort and criminal law better protect against harmful deep fakes than the Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act (DEEP FAKES Act).[121] Section A discusses how copyright law applies to deep fakes.[122] Section B explains how criminal law can apply to deep fakes.[123] Section C examines how state tort law would provide remedies to parties hurt by deep fake content.[124]

A. Copyright Law and Deep Fakes

Copyright law applies to deep fakes in narrow circumstances due to the doctrine of fair use and the nature of one's own image.[125] An individual's

[116] *See id.* (listing the types of deep fakes subject to the disclosure requirements imposed by the DEEP FAKES Act).

[117] *See* Harwell, *supra* note 23 (describing how people perceive manipulated media according to their previous beliefs).

[118] *See id.* (discussing how individuals interpret altered photos or video according to their previous affiliations); H.R. 3230 (imposing liability for failure to disclose content as false if it can be perceived to cause harm to the individual depicted or society at large).

[119] H.R. 3230.

[120] *See* David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELEC. FRONTIER FOUND. (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them (describing how existing tort, criminal, and copyright law apply to harmful deep fakes).

[121] *See infra* notes 122–152 and accompanying text.

[122] *See infra* notes 125–133 and accompanying text.

[123] *See infra* notes 134–140 and accompanying text.

[124] *See infra* notes 142–152 and accompanying text.

[125] *See* 17 U.S.C. § 102 (2018) (describing how copyrights only apply to works of authorship); Tom Kulik, *Faking It: Why Deepfakes Pose Specific Challenges Under Copyright & Privacy Laws*, ABOVE THE L. (July 15, 2019, 12:47 PM), https://abovethelaw.com/2019/07/faking-it-why-deepfakes-pose-specific-challenges-under-copyright-privacy-laws/ (describing how copyright law could be applied to deep fakes).

appearance cannot be copyrighted because an appearance is not created like a work of authorship—it simply exists.[126] Nonetheless, images and videos used to create a deep fake are subject to copyright protection as works of authorship of the original author.[127] But, by default, the copyright owner is the one who created content, so, unless stipulated in a contract with the photographer or videographer, the person depicted in the deep fake has no claim for copyright infringement since they would not own the copyright of the source material used to create the deep fake.[128]

Even if an individual depicted in the deep fake owns all images and video used to create the deep fake, a reviewing court must apply the doctrine of fair use to determine whether the deep fake infringes the copyright.[129] There are four factors for consideration in determining fair use: (1) the purpose and character of the use, (2) the nature of the copyrighted work, (3) the substantiality of the portion used in relation to the work as a whole, and (4) the effect of the use upon the potential market.[130] Applying these factors to deep fakes, courts would likely find that the use of deep fakes for scholarship, comedy, criticism, and reporting outweighs any other factors weighing against fair use.[131] If a deep fake portrays a candidate in a compromising position or is used to promote a product or cause, however, a court would likely find it to be infringing because of the purpose and character of the use and the effect of the use on the potential market.[132] Therefore, even though copyright protection extends to the owners of images used to create deep fakes, the outcome of the fair use analysis depends on the nature of the deep fake.[133]

B. Criminal Law and Deep Fakes

In cases of malicious use of deep fakes, some criminal laws may apply.[134] With the creation of pornography as the most prominent use of deep fake technology, remedies may be available via state revenge porn laws.[135] As of 2019,

[126] *See* 17 U.S.C. § 102 (granting copyright protections to works of authorship).

[127] *See id.* (listing pictorial and audiovisual works as works of authorship subject to copyright protection).

[128] *See* 17 U.S.C. § 106 (2018) (describing how the copyright owner is the only party with legal rights to the work).

[129] *See id.* (granting the copyright owner the exclusive legal rights to the work); 17 U.S.C. § 107 (2018) (describing the factors used to determine whether use of a copyrighted work violates copyright law).

[130] 17 U.S.C. § 107.

[131] *See id.* (describing the factors used to determine whether use of a copyrighted work violates copyright law); Kulik, *supra* note 125 (describing how, under limited circumstances, the nature and character of the deep fake makes it a fair use).

[132] *See* 17 U.S.C. § 107 (listing the factors used to determine whether a particular use of a copyrighted work violates the owner's copyright); Kulik, *supra* note 125 (describing how the effect of the use on the potential market can weigh against deep fake content).

[133] *See* Kulik, *supra* note 125 (describing how a deep fake's effect on the potential market may weigh against the deep fake in a fair use analysis); 17 U.S.C. § 106 (stating that the copyright owner is the exclusive owner of legal rights to the work).

[134] *See* Greene, *supra* note 120 (describing how different criminal laws apply depending on the context in which deep fake is used).

[135] Megan Farokhmanesh, *Is it Legal to Swap Someone's Face Into Porn Without Consent?*, THE VERGE (Jan. 30, 2018, 3:29 PM), https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal; *see* AJDER ET AL, *supra* note 85 (describing how pornography is the most common

forty-one states and the District of Columbia have passed revenge porn laws.[136] Looking at Alaska's revenge porn statute, deep fake pornography violates the law because it depicts a person engaged in a sexual act.[137] Other criminal laws may also apply depending on the purpose for which the deep fake is used.[138] For example, if someone threatens to publish a damaging deep fake depicting another unless they receive payment, then extortion laws would apply.[139] If the deep fake is used to obtain another's property through fraudulent means, then state and federal law prohibiting fraud would apply.[140] These are just some examples of how existing criminal laws may apply to deep fakes.[141]

C. State Tort Law and Deep Fakes

In cases of malicious use of deep fakes, state tort laws may apply.[142] Because deep fakes are false depictions, the tort of defamation may apply because it concerns, in part, communications of false information to the public.[143] At common law, anyone knowingly publishing a false record that harms an individual's reputation is subject to liability for either slander or false light.[144] Therefore, anyone who publishes a deep fake that harms an individual's reputation would be liable.[145] Courts have already shown a willingness to apply defamation

use of deep fakes). Some revenge porn statutes have language that may prevent them from applying to revenge porn. For example, Tennessee's revenge porn law requires that the "image was photographed or recorded under circumstances where the parties agreed or understood that the image would remain private." TENN. CODE ANN. § 39-17-318 (2020). Sharing a deep fake would not satisfy this law because the deep fake was neither photographed nor recorded. *See* TENN. CODE ANN. § 39-17-318.

[136] *State Revenge Porn Laws*, FINDLAW (last visited Jan. 6, 2020), https://criminal.findlaw.com/criminal-charges/revenge-porn-laws-by-state.html.

[137] *See* ALASKA STAT. § 11.61.120 (2019) (creating a Class B misdemeanor for sharing sexual content online without consent).

[138] *See* Greene, *supra* note 120 (discussing how different criminal laws may apply depending on the nature of the deep fake).

[139] *See, e.g.*, N.Y. PENAL LAW § 135.60 (McKinney 2018) (imposing criminal liability on anyone that induces another to do something under threat of reputational harm).

[140] *See, e.g.*, Stupp, *supra* note 20 (describing how deep fakes can be used to commit fraud).

[141] *See supra* notes 134–140 and accompanying text (providing examples of criminal laws that may apply to deep fakes).

[142] *See* Greene, *supra* note 120 (describing how different tort laws apply depending on the context in which the deep fake is used).

[143] *See* RESTATEMENT (SECOND) OF TORTS § 558 (AM. LAW INST. 1977) (listing the elements to prove defamation, including that the published record is actually false).

[144] *See id.* (listing the elements to prove defamation); RESTATEMENT (SECOND) OF TORTS § 568 (AM. LAW INST. 1977) (defining slander as defamation occurring in any way but the written word, including audio, pictures, or video); RESTATEMENT (SECOND) OF TORTS § 652E (AM. LAW INST. 1977) (defining the tort of false light as any act which portrays a third party in a false, harmful manner to the public).

[145] *See* RESTATEMENT (SECOND) OF TORTS § 558 (stating the elements of defamation); RESTATEMENT (SECOND) OF TORTS § 568 (defining slander as defamation in any manner but the written word, which includes audio, pictures, or video); RESTATEMENT (SECOND) OF TORTS § 652E (defining the tort of false light).

law to online statements, so it would reasonable to apply it to a malicious deep fake as well.[146]

Looking specifically at deep fake pornography, it is a form of non-consensual pornography. [147] Non-consensual pornography is shared with the intent to harm the individual by inflicting emotional distress and violating their privacy.[148] The tort of emotional distress imposes liability when an individual's actions either intentionally or recklessly inflict emotional distress on the individual.[149] Any individual sharing deep fake pornography would meet this standard because even if they do not intend to cause emotional distress, their reckless actions meet the common law test.[150] Because deep fake pornography also causes reputational damage to the individual, the individual also can assert a defamation claim against the creator.[151] Rather than impose additional liability with the DEEP FAKES Act, the court should apply extant copyright, criminal, and state tort laws to hold malicious deep fake creators accountable.[152]

CONCLUSION

Technological advancements in machine learning have enabled the automated creation of realistic digital depictions of acts that did not actually occur, or deep fakes. Deep fakes can have both beneficial purposes, such as the creation of parody content, and nefarious purposes, such as reputational damage influencing an election. Due to the potential harm deep fakes pose, the House of Representatives has introduced the DEEP FAKES Act to impose civil and criminal liability upon people that create or share deep fakes without an appropriate warning that the content depicted or expressed did not actually occur. Nevertheless, the proposed legislation would not change the behavior of malicious deep fake creators and would, in effect, only burden legitimate creators of deep fake technology. The overbroad definition of "deep fake" in the proposed legislation could reasonably extend to any false depiction created using a computer, effectively imposing criminal and civil liability across a large swath of digital content. Rather than imposing new legislation, the courts should apply existing copyright, criminal, and state tort law to malicious deep fake content. Applying this existing law rather than

[146] *See, e.g.*, *Bock v. Scheff*, 991 So.2d 1043, 1044 (Fla. Dist. Ct. App. 2008) (affirming trial court decision that held individual liable for defamation for false statements posted online); *Matter of Cohen v. Google, Inc.*, 887 N.Y.S.2d 424, 425 (N.Y. Sup. Ct. 2009) (requiring Google to disclose the identity of a blogger because the plaintiff's defamation claim was based upon the blogger's defamatory statements).

[147] Farokhmanesh, *supra* note 135.

[148] *Id.*

[149] RESTATEMENT (SECOND) OF TORTS § 46 (AM. LAW INST. 1965) (discussing the tort of emotional distress).

[150] *See id.* (describing how the tort of emotional distress is available for both intentional and reckless actions).

[151] *See* RESTATEMENT (SECOND) OF TORTS § 558 (listing the elements of defamation); RESTATEMENT (SECOND) OF TORTS § 568 (defining slander as defamation occurring through any medium except the written word); RESTATEMENT (SECOND) OF TORTS § 652E (defining false light as any act that falsely, harmfully portrays a third party to the public).

[152] *See* Greene, *supra* note 117 (describing how existing tort, criminal, and copyright law apply to harmful deep fakes).

passing the DEEP FAKES Act avoids the issue of the overly broad definition of "deep fake" within the Act because courts would have to fit the offending deep fake content within the existing copyright, criminal, and state tort frameworks. Therefore, instead of creating another basis for liability, courts should use existing copyright, criminal, and state tort liability to prevent the harmful use of deep fakes.

Recommended Citation: Zachary Schapiro, *DEEP FAKES Accountability Act: Overbroad and Ineffective*, B.C. INTELL. PROP. & TECH. F. (Apr. 27, 2020), http://bciptf.org/2020/04/deepfakes-accountability-act/.