Emily Ji

Professor Feigenbaum

CPSC 610

14 May 2021

Regulating Deepfakes in the United States

I. Introduction

Over the past few years, the public has grown increasingly concerned about deepfakes and the significant damage that they can inflict upon society. In ninety to ninety-five percent of deepfake videos found online, faces of celebrities or regular individuals are superimposed on porn stars.[1] Victims of deepfake pornography suffer from great trauma, and some have had to erase their presence on the Internet or change their names.[2] In 2018, Jordan Peele collaborated with *BuzzFeed* to create a deepfake of Barack Obama cursing Donald Trump.[3] This video indicates that deepfakes depicting politicians could threaten democracy and national security in the future. For instance, a few days before an election, a candidate could create a fake video that shows their opponent engaging in criminal activity. From these few examples, it is clear that regulating the creation and distribution of certain types of deepfakes would be beneficial. However, how exactly should the government regulate deepfakes? How can it protect individuals and society as a whole without restricting the beneficial applications of deepfakes?

This paper seeks to answer the above questions by presenting a comprehensive survey of deepfake regulation in the U.S. The majority of deepfakes are deepfake porn. Additionally, election misinformation has become a major issue for social media sites in recent years.

[1] Karen Hao, “Deepfake porn is ruining women’s lives. Now the law may finally ban it,” *MIT Technology Review,* February 12, 2021, https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/
[2] Ibid.
[3] James Vincent, “Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news,” *The Verge*, April 17, 2018, https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed

Therefore, this paper will mainly focus on pornographic deepfakes—deepfakes that depict sexually explicit conduct without the individual's consent—and political deepfakes—deepfakes that depict candidates prior to an election. After analyzing the recommendations of various legal scholars and lawmakers, this paper will conclude by proposing the best ways of combining technology and law, in order to reduce the negative effects of deepfakes.

II. What Are Deepfakes?

Deepfakes are photos, videos, or audio recordings that were manipulated or generated using machine learning. They appear to be authentic and have the potential to deceive someone.

The above definition is the synthesis of various definitions by computer scientists and legal scholars.[4] These experts interpret the term "deepfake" in slightly different ways, but they all highlight the use of machine learning or artificial intelligence (AI) to create the content.[5] An image that was manually edited in Adobe Photoshop, for example, is unlikely to be mistaken as an accurate depiction. Machine learning, however, is powerful enough to alter an image in ways that cannot be detected by the human eye. This feature of deepfakes is what makes them potentially dangerous. If deepfakes become ubiquitous, then people will no longer be able to trust their own senses; they will no longer be able to trust what they see and what they hear.

There are several different types of deepfakes, including, but not limited to, face-swapping, puppeteering, lip-synching, and voice cloning.[6] Robert Chesney and Danielle Citron, two legal scholars who have studied deepfakes in great detail, describe the numerous positive and negative applications of deepfakes in their paper: "Deep Fakes: A Looming

[4] Jessica Ice, "Defamatory Political Deepfakes and the First Amendment," *Case Western Reserve Law Review* 70, no. 2 (2019): 426 – 427, https://scholarlycommons.law.case.edu/caselrev/vol70/iss2/12/ ; James Vincent, "Why we need a better definition of 'deepfake'," *The Verge*, May 22, 2018, https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news

[5] Ibid.

[6] Ashish Jaiman, "AI Generated Synthetic Media, Aka Deepfakes." *Medium*, September 28, 2020, https://towardsdatascience.com/ai-generated-synthetic-media-aka-deepfakes-7c021dea40e1

Challenge for Privacy, Democracy, and National Security."[7] Banning all deepfakes would be irrational because some are innocuous, entertaining, or even beneficial. For example, deepfake videos of historical figures could make history class more engaging for students.[8] Paralyzed individuals could regain their mobility via a deepfake video.[9] However, it is also feasible to use deepfakes in many harmful ways, and the consequences would be dire. Deepfakes could "manipulate elections," "erode trust in institutions," "undermine public safety," "jeopardize national security," and more.[10] For instance, a fake video of a government official announcing an impending missile strike would cause great panic and chaos. As of 2020, there was no evidence of deepfakes being created for a political campaign, but that scenario is certainly within the realm of possibility.[11] The most pressing issue, however, is deepfake porn, which is the most popular form of deepfakes found on the Internet. These negative applications of deepfakes warrant government attention and action.

III. Existing Deepfake Regulation

Federal Laws

Currently, there are no federal laws that regulate the creation or distribution of deepfakes. The National Defense Authorization Act (NDAA) for Fiscal Year 2020, the NDAA for Fiscal Year 2021, and the Identifying Outputs of Generative Adversarial Networks (IOGAN) Act are three laws that mention deepfakes, but they only include research and reporting requirements. For example, the 2020 NDAA mandates that the Director of National Intelligence (DNI) provide Congress with a comprehensive report about the foreign weaponization of deepfakes.[12] The

[7] Robert Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review* 107, no. 1753 (2019): 1768 – 1785, http://dx.doi.org/10.2139/ssrn.3213954

[8] Chesney and Citron, 1769.

[9] Chesney and Citron, 1771.

[10] Chesney and Citron, 1771 – 1785.

[11] Nick Statt, "TikTok is banning deepfakes to better protect against misinformation," *The Verge*, August 5, 2020, https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference

[12] National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116 – 92, 133 Stat. 1198 (2019).

report should consist of multiple components, including an assessment of foreign entities' technical capabilities and the ways that foreign entities use or could use deepfakes to harm U.S. national security.[13] The 2020 NDAA also requires the DNI to establish a competition that would "stimulate the research, development, or commercialization" of deepfake-detection technologies.[14] For the 2021 fiscal year, however, the NDAA does not mention a "deepfakes prize competition."[15] Congress still requests an assessment of the foreign weaponization of deepfakes, but the report should specifically be about deepfakes that depict members of the armed forces or their families.[16] Finally, the IOGAN Act directs the National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST) to support research on manipulated media, such as deepfakes.[17] This research should focus on deepfake-detection tools and the "development of measurements and standards" for analyzing GANs, among other topics.[18] Ultimately, the two NDAAs and the IOGAN Act demonstrate that Congress is beginning to view deepfakes as a national security threat. It is possible that Congress wants to gain a more thorough understanding of deepfakes before implementing any specific regulations.

State Laws

There has been slightly more interest in regulating depfakes at the state level compared to the federal level. As of 2021, six states have enacted bills that target two harmful applications of deepfakes. The first category is deepfake porn. New York S.B. S5959D, which will take effect in 2023, and California A.B. 602 allow private parties to file lawsuits against creators or distributors of deepfake porn; the requirement is that the defendant knew or reasonably should

[13] National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116 – 92, 133 Stat. 1198 (2019).
[14] Ibid.
[15] National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116 – 283 (2021).
[16] Ibid.
[17] Identifying Outputs of Generative Adversarial Networks Act, Pub. L. No. 116 – 258, 134 Stat. 1150 (2020).
[18] Ibid.

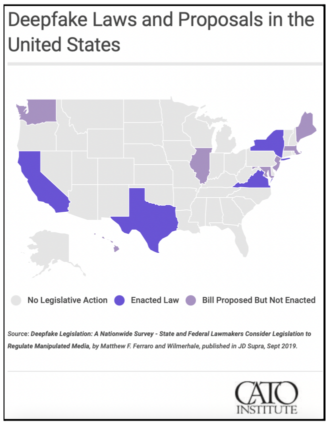
have known that the depicted individual did not approve the deepfake's creation.[19] Virginia H.B. 2678 expanded the state's existing ban on nonconsensual porn to explicitly cover deepfake porn.[20] Similarly, Maryland H.B. 1027 amended an existing child pornography law so that it now applies to any computer-generated image of a minor engaging in sexual conduct.[21] Finally, Georgia S.B. 337 prohibits a person from distributing deepfake porn online if the action "is harassment or causes financial loss to the depicted person and serves no legitimate purpose to the depicted person."[22] Other deepfake legislation relates to counterfeit and deceptive campaign material. California A.B. 730 and Texas S.B. 751 make it a crime to produce, publish, or distribute deepfakes of a political candidate within sixty and thirty days of an election respectively.[23] Both laws require proof of actual malice. In other words, the perpetrator must have intended to "injure a candidate or influence the result of an election."[24] California A.B. 730 also includes a few exceptions. For example, the conduct would be legal if the deepfake is accompanied by a disclosure of its inauthenticity.[25] The law also does not apply to deepfakes that are clearly satire or parody.[26] The differences between California A.B. 730 and Texas S.B. 751 demonstrate that even when states attempt to address the same issue, they might do so in different ways. For similar legislation, the regulation window, exceptions, and punishments may vary across states.

As of 2021, a few states have pending deepfake bills. New Jersey A.3006 requires that all deepfake producers add a digital watermark to their creation, in order to indicate that the content

[19] Assembly Bill 602, 2019 – 2020 Leg., Reg. Sess. (California 2019); Senate Bill S5959, 602, 2019 – 2020 Leg., Reg. Sess. (New York 2019).
[20] House Bill 2678, 2019 – 2020 Leg., Reg. Sess. (Virginia 2019).
[21] House Bill 1027, 2019 – 2020 Leg., Reg. Sess. (Maryland 2019).
[22] Senate Bill 337, 2019 – 2020 Leg., Reg. Sess. (Georgia 2020).
[23] Assembly Bill 730, 2019 – 2020 Leg., Reg. Sess. (California 2019); Senate Bill 751, 86th Leg., (Texas 2019).
[24] Senate Bill 751, 86th Leg., (Texas 2019).
[25] Assembly Bill 730, 2019 – 2020 Leg., Reg. Sess. (California 2019).
[26] Ibid.

was manipulated by technology.[27] Depending on the type of media, a written or verbal statement that explains the alterations must also accompany the deepfake.[28] Hawaii S.B. 309 makes it a crime to create, distribute, or threaten to distribute deepfake porn.[29] New Jersey A.4985 regulates deepfakes that depict candidates in the days leading up to an election, and it is almost an exact copy of California A.B. 730.[30] Florida S.B. 658 mandates that any deepfake with "a manipulation of a candidate's likeness," if it is used for a political purpose, include a conspicuous disclaimer regarding its inauthenticity.[31] Based on these pieces of proposed legislation, pornographic deepfakes and political deepfakes seem to be the main sources of concern for lawmakers.

The passage of the above bills, however, is uncertain. Although six states successfully enacted laws to protect society from certain harmful deepfakes, a comparable number of states failed to accomplish the same task. In 2019, Rachel Chiu of the Cato Institute created a map of U.S. state laws and pending legislation that mention deepfakes (fig. 1).[32]


Deepfake Laws and Proposals in the United States
● No Legislative Action ● Enacted Law ● Bill Proposed But Not Enacted
CATO INSTITUTE

Source: *Deepfake Legislation: A Nationwide Survey - State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*, by Matthew F. Ferraro and Wilmerhale, published in JD Supra, Sept 2019.

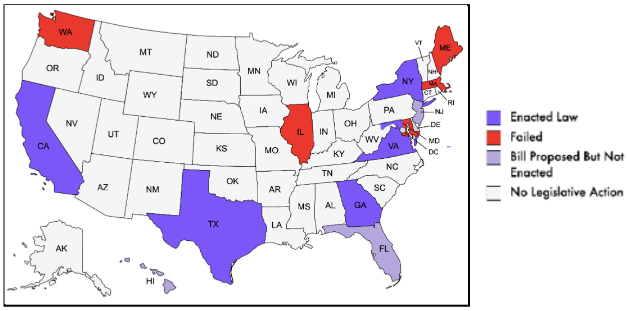


Figure 1 (left). A map of deepfake regulation in the U.S. (created by Rachel Chiu in 2019)
Figure 2 (right). An updated map of deepfake regulation in the U.S. (created by Emily Ji in 2021)

[27] Assembly Bill 3006, 2020 – 2021 Leg., Reg. Sess. (New Jersey 2021).
[28] Ibid.
[29] Senate Bill 309, 2020 – 2021 Leg., Reg. Sess. (Hawaii 2021).
[30] Assembly Bill 4985, 2020 – 2021 Leg., Reg. Sess. (New Jersey 2020).
[31] Senate Bill 658, 2020 – 2021 Leg., Reg. Sess. (Florida 2021).
[32] Feeney, Matthew. *Deepfake Laws Risk Creating More Problems Than They Solve*, 7, figure 1. Regulatory Transparency Project, 2021. https://regproject.org/wp-content/uploads/Paper-Deepfake-Laws-Risk-Creating-More-Problems-Than-They-Solve.pdf

The map shows Maryland H.B. 198, Washington S.B. 6513, Maine L.D. 1988, Illinois S.B. 3171, and Massachusetts H.B. 3366 as proposed bills, but in the end, none passed the state legislature (fig. 2). It is important to note that the first four of those bills are similar to California A.B. 730 and Texas S.B. 751, which did become laws. They all regulate deepfakes that could impact an election. For example, Senator Rebecca Millet (D-ME) actually used California A.B. 730 as a model for Maine L.D. 1988.[33] Both bills prohibit a person from distributing, with actual malice and within sixty days of an election, deepfakes that depict a political candidate.[34] Maine L.D. 1988 also includes the same exceptions as California A.B. 730.[35] However, Republican legislators in Maine worried that the bill would suppress freedom of speech, which may be the main reason why the bill failed.[36] In general, there is very little information to explain the lack of support for any of the five bills. Lawmakers, specifically those who are seeking re-election, should personally benefit from the regulation of deepfakes that could damage a candidate's reputation. Therefore, it is possible that lawmakers understood the dangers of deepfakes, but they simply viewed the proposed legislation as overly restrictive. Ultimately, state-level deepfake regulation is currently a patchwork of inconsistent laws; only a small percentage of the fifty states have either passed or proposed legislation that addresses any type of deepfake use.

Online Platform Policies

In recent years, several of the largest social media sites have announced that they will ban deepfakes. Facebook's ban drew a lot of public attention, given its enormous influence and user base.[37] Deepfakes will be removed from Facebook if two requirements are met: the media is a

[33] Scott Thistle, "Maine Lawmakers Aim to Ban Deepfake Political Ads With Bill," *GovTech*, January 29, 2020, https://www.govtech.com/computing/maine-lawmakers-aim-to-ban-deepfake-political-ads-in-new-bill.html

[34] Legislative Document 1988, 129th Leg., (Maine 2020).

[35] Ibid.

[36] Thistle, "Maine Lawmakers Aim to Ban Deepfake Political Ads With Bill."

[37] David McCabe and Davey Alba, "Facebook Says It Will Ban 'Deepfakes," *New York Times*, January 7, 2020, https://www.nytimes.com/2020/01/07/technology/facebook-says-it-will-ban-deepfakes.html

creation of machine learning, and the media is "likely to mislead someone into thinking that a subject of the video said words that they did not actually say."[38] The rule does not apply to parody, satire, videos that were only altered to remove or reorder words, and deepfakes of false actions.[39] Reddit's policies also have an exception for parodic or satirical deepfakes.[40] Beyond this similarity, however, Reddit's deepfake ban is more extensive than Facebook's. Any manipulated media that is "presented to mislead, or falsely attributed to an individual or entity" is not allowed on the site.[41] Therefore, the policy extends beyond AI-generated content and content that presents false, verbal statements. Reddit also prohibits deepfake porn.[42] Facebook does not specifically mention pornographic deepfakes in its ban, but its Community Standards already forbid explicit images of sexual activity, such as pornography.[43] Facebook and Reddit's policies demonstrate that there is no standard approach to regulating deepfake distribution on online platforms.

Twitter's synthetic and manipulated media policy is also unique in that it includes several components. There are four possible outcomes that correspond to four levels of severity. If Twitter determines that the deceptive content is "likely to impact public safety or cause serious harm" *and* it was "shared in a deceptive manner," then there is a high likelihood of removal.[44] If the former condition is not met, then it is likely that Twitter will label the Tweet.[45] If, instead, the

[38] Monika Bickert, "Enforcing Against Manipulated Media," About Facebook, last modified January 6, 2020, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

[39] Ibid.

[40] "Do not impersonate an individual or entity," Reddit Help, last modified July 2020, https://www.reddithelp.com/hc/en-us/articles/360043075032

[41] Ibid.

[42] "Do Not Post Involuntary Pornography," Reddit Help, last modified July 2020, https://www.reddithelp.com/hc/en-us/articles/360043513411

[43] "Community Standards," Facebook, accessed May 11, 2021, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity

[44] Yoel Roth and Ashita Achuthan, "Building rules in public: Our approach to synthetic & manipulated media," Twitter Blog, last modified February 4, 2020, https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html

[45] Ibid.

latter condition is not met, then Twitter will likely add a label and may end up removing the content.[46] If neither condition is met, then the Tweet "may be labeled."[47] The purpose of the label is to inform the viewer about the content's inauthentic elements. For the most part, Twitter will reduce the visibility of labeled Tweets, and users will receive warnings when they try to share or like them.[48] The use of the words "likely" and "may" suggests that Twitter prefers to review deepfakes on a case-by-case basis rather than impose a strict ban.

Other large online platforms have also chosen to focus on deepfakes that can cause harm. In this regard, their policies are a bit more narrowly tailored than that of Facebook and Reddit. For instance, TikTok prohibits users from posting synthetic or manipulated media that "distort[s] the truth of events" *and* "cause[s] harm to the subject of the video, other persons, or society."[49] Although the definition of "harm" is very broad, TikTok's main goal was to protect its users from election misinformation; it implemented the ban a few months before the 2020 U.S. presidential election.[50] YouTube has a similar approach to regulating deceptive content, but its rules will only apply to election-related deepfakes.[51] This type of content will be removed if it "may pose a serious risk of egregious harm."[52] Deepfakes that show a candidate committing a crime that never happened or deepfakes that discuss inaccurate voting information would fall under the category of "election-related." Ultimately, online platforms' deepfake policies differ significantly in their focus and scope.

[46] Yoel Roth and Ashita Achuthan, "Building rules in public: Our approach to synthetic & manipulated media."
[47] Ibid.
[48] Ibid.
[49] "Community Guidelines," TikTok, accessed May 11, 2021. https://www.tiktok.com/community-guidelines?lang=en#37
[50] Vanessa Pappas, "Combating misinformation and election interference on TikTok," TikTok Newsroom, last modified August 5, 2020, https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok
[51] Leslie Miller, "How YouTube supports elections," YouTube Official Blog, last modified February 3, 2020, https://blog.youtube/news-and-events/how-youtube-supports-elections
[52] Ibid.

IV. Other Legal Remedies

Given the limited amount of government deepfake regulation, legal scholars have explored how to use other existing laws to protect the rights of deepfake victims. This section highlights the shortcomings of these approaches, in order to demonstrate why there is a need for new regulation targeted specifically at harmful deepfakes.

Criminal Law and its Limitations

The creation and distribution of pornographic deepfakes might be illegal under revenge porn or cyberstalking laws. However, since these laws were not created with deepfakes in mind, they will only be applicable in certain scenarios. Revenge porn, for example, is viewed as "a violation of the victim's right to sexual privacy."[53] In other words, revenge porn statutes are mainly for situations in which the perpetrator publicizes a real, unedited image that the victim intended to keep private.[54] Although forty-eight states currently criminalize revenge porn, many of the statutes require proof that the deepfake creator or distributor intended to harm an individual.[55] An important question then arises: What if someone simply creates deepfake porn of person A for their own enjoyment and then shares it with others, but there is no intention for person A to ever see it? According to lawyer Anne Gieseke, this situation often occurs, and as a result, it is difficult to meet the law's intent requirement.[56] There is also a similar issue with cyberstalking laws. For example, the federal cyberstalking statute punishes those with "the intent to kill, injure, harass, intimidate, or place under surveillance."[57] Many deepfake porn creators and

[53] Rebecca Delfino, "Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act," *Fordham Law Review* 88, no. 3 (2019): 897, http://dx.doi.org/10.2139/ssrn.3341593

[54] Anne Pechenik Gieseke, ""The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography," *Vanderbilt Law Review* 73, no. 5 (2020): 1488, https://scholarship.law.vanderbilt.edu/vlr/vol73/iss5/4/

[55] "48 STATES + DC + ONE TERRITORY NOW HAVE REVENGE PORN LAWS," Cyber Civil Rights Initiative, accessed May 9, 2021, https://www.cybercivilrights.org/revenge-porn-laws/

[56] Gieseke, ""The New Weapon of Choice"," 1499.

[57] "18 U.S. Code § 2261A - Stalking," LII / Legal Information Institute, accessed May 8, 2021. https://www.law.cornell.edu/uscode/text/18/2261A

distributors will not fall under this description. Additionally, some state stalking statutes, such as those of Texas and Florida, can only be applied after multiple harassment incidents occur.[58] Therefore, creating or distributing a single pornographic deepfake is still legal, even if it causes great emotional distress to the depicted individual.[59] Victims of deepfake porn will only have sufficient protection if existing revenge porn laws are updated or if a new law that specifically regulates pornographic deepfakes is created.

Many deepfakes depict someone saying or doing something that they did not actually say or do, so one might also assume that online impersonation laws are always relevant. However, there are situations in which those statutes would fail to regulate both pornographic and political deepfakes. Several states, including three of the largest states by population (New York, Texas, and California), have laws that prohibit online impersonation, but they all require intent.[60] The accused must have intended to threaten, intimidate, harm, or defraud an individual.[61] As explained above, establishing the mens rea of a pornographic deepfake creator or distributor is a difficult task. Those who create or distribute political deepfakes, on the other hand, likely aim to damage someone's reputation. The issue is that online impersonation laws apply "where the depiction at issue was shared or posted in a way that made it seem like the victim was the poster."[62] For example, assuming that the intent requirement is met, Texas Penal Code section 33.07 makes it a crime to create a web page using the "name or persona of another person" without their consent.[63] It is also illegal to include identifying information in electronic communication without a person's permission, such that the recipient would "reasonably

[58] Florida Statute § 784.048 (1992) (amended 2019) ; Texas Penal Code § 42.072 (1997) (amended 2013).
[59] Delfino, "Pornographic Deepfakes," 897.
[60] Chesney and Citron, "Deep Fakes: A Looming Challenge," 1802.
[61] Ibid.
[62] Delfino, "Pornographic Deepfakes," 919.
[63] Texas Penal Code § 33.07 (2009) (amended 2011).

believe" that the person sent the message.[64] If someone maliciously creates a deepfake of a political candidate committing a crime, it may be illogical or ineffective to pretend that the candidate posted it online. Instead, the creator could simply share the deepfake via their own social media account. Texas' online impersonation law would not cover this conduct, even though the perpetrator used technology to fake a candidate's identity and actions. Ultimately, if a humiliating or scandalous deepfake becomes viral, it will negatively impact the depicted individual regardless of who distributed it and whether or not there was an intent to harm.

Tort Law and its Limitations

Deepfake victims could also use tort law to receive compensatory damages from the deepfake creator or distributor. This means that the perpetrator would be subject to civil liability instead of criminal liability. In total, there are four applicable torts—defamation, intentional infliction of emotional distress (IIED), appropriation of name or likeness, and false light—but they each have their own limitations. Deepfakes that cause reputational harm, specifically deepfake porn, would likely be considered libel.[65] As a result, individuals could file a defamation lawsuit. The issue is that a significant number of states first require the plaintiff to prove that the defendant intended to cause emotional distress.[66] Individuals who use the IIED tort would face the same obstacle. The tort of wrongful appropriation "require[s] a tortfeasor to have benefitted commercially from unlawfully using a victim's likeness."[67] Those who create deepfake porn, however, may do so for their own enjoyment. Those who create embarrassing deepfakes of political candidates may be attempting to manipulate an election. In other words, some malicious

[64] Ibid.
[65] Russell Spivak, ""Deepfakes": The Newest Way To Commit One of the Oldest Crimes," *Georgetown Law Technology Review* 3, no. 2 (2019): 373 – 374, https://georgetownlawtechreview.org/deepfakes-the-newest-way-to-commit-one-of-the-oldest-crimes/GLTR-05-2019/ ; Gieseke, ""The New Weapon of Choice"," 1500.
[66] Gieseke, ""The New Weapon of Choice"," 1500.
[67] Gieseke, 1497.

deepfake producers will not have a profit motive, and the tort of wrongful appropriation will be unusable. Finally, the tort of false light seems to be the most relevant tort; as legal scholar Russell Spivak explains, "Deepfakes, by definition, place an individual before the public in a false light."[68] Gieseke, however, doubts the usefulness of this tort. Her concern is that deepfake creators could simply add a disclaimer about the content's inauthenticity, in order to avoid liability.[69] If audiences know that the media has been altered, then the portrayal of the individual may not be considered misleading. In the end, plaintiffs may not even have the resources to file a civil lawsuit. Even if they do, the available tort laws are not sufficiently comprehensive. This suggests that new regulations are necessary and that creating or distributing a harmful deepfake should be considered a criminal offense rather than a civil one.

V. Proposed Regulations

In recent years, several lawmakers and legal scholars have also concluded that existing laws do not adequately address deepfakes. This section discusses the various benefits and drawbacks of their ideas for new regulation.

Malicious Deep Fake Prohibition Act

In 2018, Senator Ben Sasse (R-NE) presented the first federal bill that targets deepfake creation and distribution.[70] The Malicious Deep Fake Prohibition Act forbids an individual from creating a deepfake if there is an intent to distribute it and if the subsequent distribution would "facilitate criminal or tortious conduct."[71] Online platforms that knowingly disseminate this type of harmful deepfake would be implicated as well.[72] Therefore, the main benefit of this bill is that it regulates the behavior of individuals *and* online platforms. It is unclear, however, if they would

[68] Spivak, "“Deepfakes”: The Newest Way To Commit One of the Oldest Crimes," 380.
[69] Gieseke, "“The New Weapon of Choice”," 1498.
[70] Nina Iacono Brown, "Congress Wants to Solve Deepfakes by 2020," *Slate*, July 15, 2019, https://slate.com/technology/2019/07/congress-deepfake-regulation-230-2020.html
[71] Malicious Deep Fake Prohibition Act of 2018, S.3805, 115th Congress (2018).
[72] Delfino, "Pornographic Deepfakes," 909.

receive different punishments. The bill simply states that "any person" who violates the law will be fined, imprisoned, or both.[73] Additionally, the punishment will increase in severity if the deepfake interferes with the actions of a government agency or encourages violence.[74] Senator Sasse introduced the bill at the very end of 2018, and it did not gain any co-sponsors before it expired a few days later.[75]

Although the short time frame is likely the main reason why the bill failed to pass Congress, legal scholars have retrospectively pointed out several of the bill's shortcomings. First of all, the Malicious Deep Fake Prohibition Act only regulates conduct that is already illegal: the facilitation of a crime. Orin Kerr, a professor at the University of Berkeley School of Law, argues that this actually makes the legislation overly broad. He gives the example of an individual who creates a funny deepfake video, invites their friends to a party so that they can watch the video together, and then violates a state law because the party is a public nuisance.[76] According to Kerr, this is an example of deepfake distribution that facilitates tortious conduct; therefore, this innocuous event becomes a federal crime under the Malicious Deep Fake Prohibition Act.[77] Kerr's hypothetical scenario is unlikely to happen. However, it still supports his argument that the bill should be more specific about the types of deepfakes that it regulates. Danielle Citron also believes that the Malicious Deep Fake Prohibition Act is overly broad, but for a different reason. Her concern is that the bill could lead to first amendment violations because it places too much liability on online platforms.[78] If platforms are punished for knowingly distributing

[73] Malicious Deep Fake Prohibition Act of 2018, S.3805, 115th Congress (2018).
[74] Ibid.
[75] "US - S3805," BillTrack50, accessed May 6, 2021, https://www.billtrack50.com/billdetail/1000397.
[76] Orin Kerr, "Should Congress Pass a "Deep Fakes" Law?" *Reason*, January 31, 2019, https://reason.com/volokh/2019/01/31/should-congress-pass-a-deep-fakes-law/
[77] Ibid.
[78] Kaveh Waddel, "Lawmakers plunge into the "deepfake" war," *Axios*, January 31, 2019, https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html

harmful deepfakes, then they will likely impose very strict deepfake removal policies.[79] Many *harmless* deepfakes would be taken down as a result. Ultimately, it seems that the Malicious Deep Fake Prohibition Act, if passed by Congress, would have more negative effects than positive ones.

DEEPFAKES Accountability Act

In 2019, Representative Yvette Clark (D-NY) introduced the Defending Each and Every Person from False Appearances by Keeping Exploitation Subject (DEEPFAKES) Accountability Act. This bill requires that all deepfake producers add a digital watermark to their creation, in order to indicate that the content is false.[80] The deepfake must also include a statement that identifies the media as having manipulated elements and describes how the media was altered.[81] This statement should be verbal, written, or both, depending on the format of the deepfake.[82] Anyone who fails to comply with the law would be subject to a civil penalty, which takes the form of fines.[83] This could escalate to a criminal penalty if, for example, the perpetrator intended to "humiliate or otherwise harass the person falsely exhibited" or "cause violence and physical harm."[84] The bill also gives deepfake victims a private right of action against creators who do not add the mandated watermark and statement.[85] Although the DEEPFAKES Accountability Act had almost thirty sponsors in total, it did not end up passing the House, and it expired in 2019.[86]

The exact reason for the bill's failure is unclear, but similar to the Malicious Deep Fake Prohibition Act, it has its strengths and weaknesses. Soon after the bill was introduced, members

[79] Ibid.
[80] Defending Each and Every Person from False Appearances by Keeping Exploitation Subject Accountability Act of 2019, H.R. 3230, 116th Congress (2019).
[81] Ibid.
[82] Ibid.
[83] Ibid.
[84] Ibid.
[85] Ibid.
[86] "US - HR3230." BillTrack50. Accessed May 6, 2021.https://www.billtrack50.com/BillDetail/1132741

of the Electronic Frontier Foundation expressed concern about its insufficient first amendment protections. They claim that in general, the first amendment will only allow the government to impose criminal penalties if there is proof of an identifiable harm.[87] The DEEPFAKES Accountability Act, however, simply requires an intent to harm.[88] Devin Coldeway, a reporter for TechCrunch, believes that this bill is not very effective. He argues that those who create harmful deepfakes, such as deepfake revenge porn, are probably acting illegally already; therefore, they will have no incentive to comply with this law.[89] Furthermore, one could argue that placing watermarks on pornographic deepfakes does nothing to mitigate their negative effects; just by existing, those deepfakes cause great emotional distress to the depicted individuals. It is important to note, however, that this bill would give the court system more guidance on how to address deepfake cases, which is useful. Mutale Nkonde, a fellow at the Berkman Klein Center at Harvard who helped draft the bill, explains that the bill adds deepfakes to the category of "misappropriation of information."[90] Additionally, at the very least, this bill gives deepfake victims another method for receiving compensatory damages. The DEEPFAKES Accountability Act is ultimately insufficient as it is currently drafted, and it would need to be paired with more regulation. Nevertheless, it underscores the importance of labeling deepfakes, which cannot be easily detected by the human eye.

Federal Criminalization of Deepfake Porn

Rebecca Delfino, a professor at Loyola Law School, has focused her efforts on regulating

[87] Hayley Tsukayama, India McKinney, and Jamie Williams, "Congress Should Not Rush to Regulate Deepfakes," *Electronic Frontier Foundation*, June 24, 2019, https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes

[88] DEEPFAKES Accountability Act of 2019, H.R. 3230, 116th Congress (2019).

[89] Devin Coldewey, "DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start," *TechCrunch*, June 13, 2019, https://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/

[90] Mutale Nkonde, "Congress Must Act on Regulating Deepfakes," *Medium*, June 17, 2019, https://onezero.medium.com/congress-must-act-on-regulating-deepfakes-1e7e94783be8

a specific type of harmful deepfake. In 2019, she proposed a new piece of federal legislation in her paper, “Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn’s Next Tragic Act.”[91] The Pornographic Deepfake Criminalization Act makes it a crime to create or distribute deepfake porn “with knowledge or reckless disregard for” the depicted individual’s lack of consent.[92] Similar to the Malicious Deep Fake Prohibition Act, anyone who violates the law will be fined, imprisoned, or both.[93] The proposed statute also suggests remedies for the deepfake porn victim. The remedy could be ordering an online service provider to remove the deepfake from its platform or providing the victim with monetary damages.[94] The service provider will only be held liable if it distributes deepfake porn “knowingly or with reckless disregard.”[95]

Delfino drafted the Pornographic Deepfake Criminalization Act with the shortcomings of the Malicious Deep Fake Prohibition Act in mind. Therefore, it seems that Delfino’s bill would face less criticism if it were introduced to Congress. The Pornographic Deepfake Criminalization Act is specifically targeted at “deepfakes portraying “sexually explicit conduct”,” so it should not be viewed as overly broad.[96] Online platforms will still need to self-regulate, in order to avoid liability, but the bill only requires that they screen for deepfake porn. In other words, it is unlikely that harmless deepfakes will be taken down at the same time. A first amendment violation may arise, however, because the bill does not require proof of an identifiable harm or even an intent to harm. Delfino addresses this concern in her paper. She chose the phrase “with knowledge or reckless disregard for the lack of consent for the individual” because many

[91] Delfino, “Pornographic Deepfakes,” 928.
[92] Delfino, 929.
[93] Ibid.
[94] Delfino, 930.
[95] Ibid.
[96] Delfino, 931.

deepfake producers may simply create deepfake porn for their own entertainment.[97] Therefore, including an intent requirement in the legislation would be too restrictive. According to Delfino, "recklessness as an alternative mens rea, is a reasonable compromise."[98] Although lawmakers and legal scholars may point out additional shortcomings in the future, the Pornographic Deepfake Criminalization Act appears to be a promising option for deepfake regulation.

VI. Recommendations

Deepfakes have the potential to significantly damage an individual's reputation. This is especially true for pornographic deepfakes, which are currently a major threat, and political deepfakes, which could become more common in the future. Given the limited amount of existing deepfake regulation and the inadequacy of other laws, it is essential to develop new regulation that specifically targets harmful deepfake creation and distribution. This section presents recommendations for the government and for online platforms. It will also explain how various technologies could be used to successfully implement these recommendations.

Federal and State Laws

Congress should enact legislation that criminalizes the creation and distribution of deepfake porn. The Pornographic Deepfake Criminalization Act, proposed by Rebecca Delfino, serves as a good model. There are several advantages to this approach, beyond the ones listed in section V of this paper. The analysis of state-level regulation in section III shows that a few states have passed laws that prohibit pornographic deepfake creation and distribution. However, the number of states is very small, and there is no guarantee that other states will take appropriate action. With a federal law, all deepfake porn victims, regardless of where they live, will be able to seek legal redress. Delfino also argues that "criminalization should be national, uniform, and

[97] Delfino, 931 - 932.

[98] Delfino, "Pornographic Deepfakes," 932.

consistent everywhere"; after all, if deepfake porn spreads on the Internet, then the harmful conduct is not confined to a single state.[99] Section III demonstrates that even when states pass similar laws, there are still differences regarding the regulation window, exceptions, and punishments. Therefore, another benefit of federal criminalization is that it will standardize the punishment for creating or distributing deepfake porn. Most importantly, Delfino's proposed bill will encourage online platforms to self-regulate. A pornographic deepfake causes reputational harm as soon as it is posted online and other people see it. It is therefore essential for service providers to remove this type of content as quickly as possible. As Delfino explains, Section 230 immunity will not apply to violations of federal law.[100] Online platforms will need to routinely screen for deepfake porn under the Pornographic Deepfake Criminalization Act. However, since this legislation only targets a specific type of deepfake, the content removal should not be excessive. Although deepfakes can be fun and entertaining, deepfake porn is one type of entertainment that should not be tolerated.

Deepfakes that show a political candidate saying or doing something that they did not actually say or do should also be prohibited during election season. As legal scholar Rebecca Green explains, "counterfeit campaign speech deprives voters of agency, the ability to absorb "real" information from candidates to make informed decisions about who should represent them in government."[101] Deepfakes can be weaponized, used as election misinformation, and hurt democracy. However, it seems that regulating political deepfakes is not as straightforward as regulating pornographic ones. The different outcomes of Maine L.D. 1988 and California A.B. 730 demonstrate that some states will view this type of regulation as too restrictive. Even though

[99] Delfino, 927.
[100] Delfino, 927.
[101] Rebecca Green, "Counterfeit Campaign Speech," *Hastings Law Journal* 70, no. 6 (2019): 1458, https://www.hastingslawjournal.org/counterfeit-campaign-speech/

California A.B. 730 was signed into law, it still raised concerns about first amendment violations from groups such as the American Civil Liberties Union.[102] Therefore, letting states regulate political deepfakes will likely lead to an inconsistent patchwork of laws. At the same time, however, each state has its own unique election process and rules; it may be inappropriate for the federal government to regulate political deepfakes, unless the regulation is for federal elections. Green argues that "a narrow law targeting counterfeited candidate speech produced and distributed with knowing intent to confuse voters and disrupt elections should survive First Amendment scrutiny."[103] She concludes that such a law fits the exception from *U.S. v. Alvarez*, a Supreme Court case which determined that false speech is constitutional.[104] Therefore, legislation will likely have a better chance of success if it targets deepfakes of fake candidate speech rather than all deepfakes that depict a political candidate. States that failed to pass political deepfake regulations, such as Maryland, Washington, Maine, and Illinois, should try to narrow the scope of their proposed bills in this way.

Online Platform Policies

Facebook should also apply its deepfake ban to media that is likely to mislead someone into thinking that a subject of the video did something that they did not do. This assumes that the media meets Facebook's first requirement: it is the creation of machine learning. As it is currently written, Facebook's policy only covers deepfakes of false speech. However, these are not the only types of deepfakes that can cause harm. A deepfake of a government official committing a crime, for example, would erode the public's trust in government. Deepfake porn is another type of video that depicts false actions and leads to emotional distress. Even though

[102] Colin Lecher, "California has banned political deepfakes during election season," *The Verge*, October 7, 2019, https://www.theverge.com/2019/10/7/20902884/california-deepfake-political-ban-election-2020
[103] Green, "Counterfeit Campaign Speech," 1486.
[104] Ibid.

Facebook’s Community Standards should already prohibit deepfake porn, this change to Facebook’s policy would undeniably ban deepfake porn from the platform.

Twitter, another influential social networking service, should have a harsher policy for deepfakes that were “shared in a deceptive manner” and are “likely to impact public safety or cause serious harm.”[105] Based on Twitter’s standards, these deepfakes warrant the most regulation. It is unclear why Twitter would not definitely remove this type of manipulated media, especially since the content may incite violence. Twitter’s deepfake ban is already more lenient than Facebook’s, which does not consider a deepfake’s potential to cause harm. Additionally, other online platforms such as TikTok and YouTube already remove deepfakes if they pose a risk to society. Twitter would not be alone in making this change to its policies.

Finally, all online platforms should model Twitter and apply labels to deepfakes. As technology improves, it will be impossible for humans to detect deepfakes on their own and discern truth from falsehood. Therefore, even if the manipulated media is harmless, it will be informative to know that it was altered by machine learning. If the content has the potential to cause harm, then online platforms could look to Twitter’s other strategies, such as reducing the deepfake’s visibility or showing a warning when users try to like or share it.

Technologies

Although laws can prohibit the creation and distribution of certain deepfakes, they will not be effective if the deepfakes remain unnoticed online. Therefore, service providers should leverage existing deepfake-detection tools and routinely screen for illegal content.

One promising technology is Microsoft’s Video Authenticator. It searches for “subtle

[105] Yoel Roth and Ashita Achuthan, “Building rules in public: Our approach to synthetic & manipulated media.”

fading or grayscale elements" that will appear at the blending boundary of the deepfake.[106] It then returns a numerical value that describes the likelihood that the content was altered.[107] Since this technology can be used on both photos and videos, it would be specifically helpful for detecting deepfake porn. The majority of pornographic deepfakes are face-swapping videos, in which person A's face was pasted onto person B's body.[108] It is likely that the Video Authenticator could detect the boundary of person A's face in the video and determine that the pornography is fake. Another method for detecting deepfakes is to analyze the blinking patterns in a video. Researchers at the University of Albany, SUNY discovered that deepfakes often depict individuals who do not blink.[109] This is because many training datasets, which are used for creating deepfakes, do not contain images of people with closed eyes.[110] The researchers' deepfake-detection algorithm is relatively robust, and there is room for improvement as well; they plan to analyze abnormal blinking rates in a video, in addition to a lack of blinking, which would be further proof that the video is a deepfake.[111]

Ultimately, as more deepfake-detection tools are developed, it will be easier to discover and eliminate harmful deepfakes on the Internet. Online platforms should use these tools to check if a photo or video is a deepfake before a user is even allowed to post the content. This preliminary screening will help prevent malicious deepfakes from becoming viral and inflicting great harm.

VII. Conclusion

[106] Tom Burt and Eric Horvitz, "New Steps to Combat Disinformation," Microsoft On the Issues, last modified September 1, 2020, https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/

[107] Ibid.

[108] Hao, "Deepfake porn is ruining women's lives. Now the law may finally ban it."

[109] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking," *ArXiv:1806.02877 [Cs]*, June 11, 2018. http://arxiv.org/abs/1806.02877.

[110] Ibid.

[111] Ibid.

Deepfakes, like any new technology, are a double-edged sword; they have the potential to bring benefits to society and cause irreversible harm. As of 2021, the amount of existing deepfake regulation at the federal and state levels is severely lacking. Online platforms could also improve their policies to prevent the spread of malicious deepfakes online. Therefore, in the next few years, it is essential for the government to collaborate with legal scholars, technologists, and service providers to mitigate the negative effects of deepfakes.

Bibliography

"18 U.S. Code § 2261A - Stalking." LII / Legal Information Institute. Accessed May 8, 2021. https://www.law.cornell.edu/uscode/text/18/2261A

"48 STATES + DC + ONE TERRITORY NOW HAVE REVENGE PORN LAWS." Cyber Civil Rights Initiative. Accessed May 9, 2021. https://www.cybercivilrights.org/revenge-porn-laws/

Assembly Bill 602, 2019 – 2020 Legislature, Regular Session (California 2019). https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602

Assembly Bill 730, 2019 – 2020 Legislature, Regular Session (California 2019). https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730

Assembly Bill 3006, 2020 – 2021 Legislature, Regular Session (New Jersey 2021). https://legiscan.com/NJ/text/A3006/2020

Assembly Bill 4985, 2020 – 2021 Legislature, Regular Session (New Jersey 2020). https://legiscan.com/NJ/text/A4985/2020

Bickert, Monika. "Enforcing Against Manipulated Media." About Facebook. Last modified January 6, 2020. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

Brown, Nina Iacono. "Congress Wants to Solve Deepfakes by 2020." *Slate*, July 15, 2019. https://slate.com/technology/2019/07/congress-deepfake-regulation-230-2020.html

Burt, Tom and Eric Horvitz. "New Steps to Combat Disinformation." Microsoft On the Issues. Last modified September 1, 2020. https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/

Chesney, Robert and Danielle Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107, no. 1753 (2019): 1753 – 1820. http://dx.doi.org/10.2139/ssrn.3213954

Chesney, Robert and Danielle Citron. "Deepfakes and the New Disinformation War." *Foreign Affairs*, February, 2019. https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war

Coldewey, Devin. "DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start." *TechCrunch*, June 13, 2019. https://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforc eable-rules-but-its-a-start/

"Community Guidelines." TikTok. Accessed May 11, 2021. https://www.tiktok.com/community-guidelines?lang=en#37

"Community Standards." Facebook. Accessed May 11, 2021. https://www.facebook.com/communitystandards/adult_nudity_sexual_activity

Dack, Sean. "Deep Fakes, Fake News, and What Comes Next." *The Henry M. Jackson School of International Studies*, March 20, 2019. https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next/

Defending Each and Every Person from False Appearances by Keeping Exploitation Subject Accountability Act of 2019, H.R. 3230, 116th Congress (2019). https://www.congress.gov/bill/116th-congress/house-bill/3230/text

Delfino, Rebecca. "Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act." *Fordham Law Review* 88, no. 3 (2019): 887 – 938. http://dx.doi.org/10.2139/ssrn.3341593

"Do not impersonate an individual or entity." Reddit Help. Last modified July 2020. https://www.reddithelp.com/hc/en-us/articles/360043075032

"Do Not Post Involuntary Pornography." Reddit Help. Last modified July 2020. https://www.reddithelp.com/hc/en-us/articles/360043513411

Feeney, Matthew. *Deepfake Laws Risk Creating More Problems Than They Solve*, 7, figure 1. Regulatory Transparency Project, 2021. https://regproject.org/wp-content/uploads/Paper-Deepfake-Laws-Risk-Creating-More-Problems-Than-They-Solve.pdf

Florida Statute § 784.048 (1992) (amended 2019). https://www.flsenate.gov/Laws/Statutes/2018/784.048

Gieseke, Anne Pechenik. ""The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography." *Vanderbilt Law Review* 73, no. 5 (2020): 1479 – 1515. https://scholarship.law.vanderbilt.edu/vlr/vol73/iss5/4/

Green, Rebecca. "Counterfeit Campaign Speech." *Hastings Law Journal* 70, no. 6 (2019): 1445 – 1490. https://www.hastingslawjournal.org/counterfeit-campaign-speech/

Hao, Karen. "Deepfake porn is ruining women's lives. Now the law may finally ban it." *MIT Technology Review,* February 12, 2021. https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/

House Bill 1027, 2019 – 2020 Legislature, Regular Session (Maryland 2019). https://legiscan.com/MD/text/HB1027/2019

House Bill 2678, 2019 – 2020 Legislature, Regular Session (Virginia 2019). https://lis.virginia.gov/cgi-bin/legp604.exe?191+ful+HB2678S1+hil

Ice, Jessica. "Defamatory Political Deepfakes and the First Amendment." *Case Western Reserve Law Review* 70, no. 2 (2019): 417 – 455. https://scholarlycommons.law.case.edu/caselrev/vol70/iss2/12/

Identifying Outputs of Generative Adversarial Networks Act, Pub. L. No. 116 – 258, 134 Stat. 1150 (2020). https://www.congress.gov/116/plaws/publ258/PLAW-116publ258.pdf

Jaiman, Ashish. "AI Generated Synthetic Media, Aka Deepfakes." *Medium*, September 28, 2020. https://towardsdatascience.com/ai-generated-synthetic-media-aka-deepfakes-7c021dea40e1

Kerr, Orin. "Should Congress Pass a "Deep Fakes" Law?" *Reason*, January 31, 2019. https://reason.com/volokh/2019/01/31/should-congress-pass-a-deep-fakes-law/

Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking." *ArXiv:1806.02877 [Cs]*, June 11, 2018. http://arxiv.org/abs/1806.02877.

Lecher, Colin. "California has banned political deepfakes during election season." *The Verge*, October 7, 2019. https://www.theverge.com/2019/10/7/20902884/california-deepfake-political-ban-election-2020

Legislative Document 1988, 129th Legislature, (Maine 2020). https://mainelegislature.org/legis/bills/bills_129th/billtexts/SP069001.asp

Malicious Deep Fake Prohibition Act of 2018, S.3805, 115th Congress (2018). https://www.congress.gov/bill/115th-congress/senate-bill/3805/text

McCabe, David and Davey Alba. "Facebook Says It Will Ban 'Deepfakes.'" *New York Times*, January 7, 2020. https://www.nytimes.com/2020/01/07/technology/facebook-says-it-will-ban-deepfakes.html

Miller, Leslie. "How YouTube supports elections." YouTube Official Blog. Last modified February 3, 2020. https://blog.youtube/news-and-events/how-youtube-supports-elections

National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116 – 92, 133 Stat. 1198 (2019). https://www.congress.gov/116/plaws/publ92/PLAW-116publ92.pdf

National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116 – 283 (2021). https://www.congress.gov/bill/116th-congress/house-bill/6395/text

Nkonde, Mutale. “Congress Must Act on Regulating Deepfakes.” *Medium*, June 17, 2019. https://onezero.medium.com/congress-must-act-on-regulating-deepfakes-1e7e94783be8

“Our Synthetic and Manipulated Media Policy.” Twitter Help. Accessed May 11, 2021. https://help.twitter.com/en/rules-and-policies/manipulated-media

Pappas, Vanessa. “Combating misinformation and election interference on TikTok.” TikTok Newsroom. Last modified August 5, 2020. https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok

Roth, Yoel and Ashita Achuthan. “Building rules in public: Our approach to synthetic & manipulated media.” Twitter Blog. Last modified February 4, 2020. https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html

Senate Bill 309, 2020 – 2021 Legislature, Regular Session (Hawaii 2021). https://legiscan.com/HI/text/SB309/id/2362095

Senate Bill 337, 2019 – 2020 Legislature, Regular Session (Georgia 2020). https://www.billtrack50.com/BillDetail/1189798

Senate Bill 658, 2020 – 2021 Legislature, Regular Session (Florida 2021). https://www.flsenate.gov/Session/Bill/2021/658/BillText/Filed/HTML

Senate Bill 751, 86th Legislature, (Texas 2019). https://capitol.texas.gov/BillLookup/History.aspx?LegSess=86R&Bill=SB751

Senate Bill S5959, 2019 – 2020 Legislature, Regular Session (New York 2019). https://legislation.nysenate.gov/pdf/bills/2019/S5959D

Spivak, Russell. ““Deepfakes”: The Newest Way To Commit One of the Oldest Crimes.” *Georgetown Law Technology Review* 3, no. 2 (2019): 339 – 400. https://georgetownlawtechreview.org/deepfakes-the-newest-way-to-commit-one-of-the-oldest-crimes/GLTR-05-2019/

Statt, Nick. “TikTok is banning deepfakes to better protect against misinformation.” *The Verge*, August 5, 2020. https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference

Texas Penal Code § 33.07 (2009) (amended 2011).

https://statutes.capitol.texas.gov/Docs/PE/htm/PE.33.htm

Texas Penal Code § 42.072 (1997) (amended 2013). https://statutes.capitol.texas.gov/Docs/PE/htm/PE.42.htm

Thistle, Scott. “Maine Lawmakers Aim to Ban Deepfake Political Ads With Bill.” *GovTech*, January 29, 2020. https://www.govtech.com/computing/maine-lawmakers-aim-to-ban-deepfake-political-ads-in-new-bill.html

Tsukayama, Hayley, India McKinney, and Jamie Williams. “Congress Should Not Rush to Regulate Deepfakes.” *Electronic Frontier Foundation*, June 24, 2019. https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes

“US - HR3230.” BillTrack50. Accessed May 6, 2021.https://www.billtrack50.com/BillDetail/1132741

“US - S3805.” BillTrack50. Accessed May 6, 2021. https://www.billtrack50.com/billdetail/1000397.

Vincent, James. “Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news.” *The Verge*, April 17, 2018. https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed

Vincent, James. “Why we need a better definition of ‘deepfake’.” *The Verge*, May 22, 2018. https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news

Waddel, Kaveh. “Lawmakers plunge into the “deepfake” war.” *Axios*, January 31, 2019. https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html