Idaho Law Review

Volume 58 | Number 2 Article 6

January 2022

NONCONSENSUAL DEEPFAKES: DETECTING AND REGULATING THE RISING THREAT TO PRIVACY

Natalie Lussier

NONCONSENSUAL DEEPFAKES: DETECTING AND REGULATING THIS RISING THREAT TO PRIVACY

NATALIE LUSSIER*

ABSTRACT

This paper surveys the emerging threat of deepfake technology, largely in relation to nonconsensual deepfake pornography. Part I of this Article provides an understanding of deepfake technology and its increasing threat to privacy. Part II then canvases the steps that public and private entities are taking to combat these threats. Lastly, Part III explores legal avenues for victims and engages potential legislative solutions.

TABLE OF CONTENTS

* J.D. Candidate, University of Idaho College of Law, 2022.

I. INTRODUCTION

Deepfake technology uses artificial intelligence to manipulate human images, yielding fabricated images and videos that appear strikingly authentic. Deepfakes are a rapidly increasing presence on the internet: in 2020, the quantity of deepfake videos increased to six times that of the year prior.[1] This number will continue to climb with the increasing availability of deepfake technology and ease in which these images and videos are created.

The concern surrounding deepfake technology focuses on cybercrime in relation to the use of one's images and privacy, and this relatively new area of crime incites a new spin on a subject that many women are too familiar with—nonconsensual pornography. Cybercriminals can and are using deepfake technology to create sexually explicit photos and videos of individuals, predominantly women, and in turn using these videos to threaten, extort, and humiliate.[2]

With the lack of knowledge about deepfakes, it is unsurprising that those in the legislature are only recently starting to understand and care about the effects of both nonconsensual pornography and deepfake technology.[3] Consequently, legislation that is an amalgamation of the two is scarce: nonconsensual pornography proposals overlook the artificial intelligence subset and deepfake technology proposals are typically tailored to election AI. Neither proposal considers the most prominent category of deepfakes: nonconsensual deepfake pornography.

Companies, researchers, and organizations are also recognizing this threat to privacy. For example, Amsterdam-based intelligence company, Sensity, has developed tools to help the public determine the authenticity of images and videos.[4] But what about the websites that hold this media—what is their burden? Researchers seek to develop and provide tools for these sites because the solution is greater than just public knowledge or offender liability: websites that house these

---

1. In 2019 the internet held an estimated 15,000 deepfake videos. And of these videos, 96% were pornographic in nature, with exclusively female targets. *See* HENRY AJDER ET AL., THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1–2 (2019), https://sensity.ai/reports/. This number grew to 85,000 by December of 2020, almost six times that of the year before. *See How to Detect a Deepfake Online: Image Forensics and Analysis of Deepfake Videos*, SENSITY: DEEPFAKE DETECTION (Feb. 8, 2021), [hereinafter *How to Detect a Deepfake Online*], https://sensity.ai/how-to-detect-a-deepfake/.

2. *Deepfakes and Cheapfakes: The Biggest Threat is Not What You Think*, TRTWORLD (Jan. 7, 2021), https://www.trtworld.com/magazine/deepfakes-and-cheap-fakes-the-biggest-threat-is-not-what-you-think-43046.

3. One expert opined that "80% [of people] have no idea what a deepfake is." Karen Hao, *Deepfake Porn Is Ruining Women's Lives. Now the Law May Finally Ban it*, MIT TECH. REV. (Feb. 12, 2021), https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/.

4. *Forensic Deepfakes Detection*, SENSITY, https://sensity.ai/deepfakes-detection/ (last visited Oct. 15, 2021).

altered images and videos need to take action.[5] This Article seeks to address website liability, specifically relating to nonconsensual pornography, as well as explore current legislation, propose new legislation, and assess the impact of sitting idle.

II. UNDERSTANDING DEEPFAKE TECHNOLOGY AND ITS INCREASING THREAT TO PRIVACY

A. Deepfake Technology: A Rising Risk

For decades people have relied on video and audio recording, but deepfake technology now casts doubt on the ideology that seeing is believing. There is no reliable way to currently detect deepfakes, and researchers have a catch-22 in exploring solutions—the technology utilized for detection can also be used for creation.[6] "As of now, we lack automated ways to detect Deepfakes in a reliable and scalable fashion," Dawn Song, Professor at the University of California Berkeley said, "[i]t will be an arms race between those that create Deepfakes and those [sic] seek to detect them."[7]

Deepfake technology uses the likeness of others to manipulate human images, producing fabricated images and videos that appear strikingly authentic to the average viewer.[8] In the most modern fashion, the term "deepfake" was first coined in 2017 by a Reddit user of the same name who shared pornographic videos that used face-swapping technology.[9] With the simple creation of a username and a handful of posts, this rapidly expanding technological advancement was titled.

In 2019, Sensity, a company dedicated to researching deepfakes and their evolving threats, released a report titled *The State of Deepfakes: Landscape, Threats, and Impact*.[10] This report found that the number of deepfake videos on the internet doubled from 2018 to 2019.[11] And of these videos, ninety-six percent were pornographic in nature and almost exclusively targeted women.[12] The number of fake online videos has grown drastically since this report, roughly doubling every

5. Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned'*, THE 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/.

6. Chenxi Wang, *Deepfakes, Revenge Porn, and the Impact on Women*, FORBES (Nov. 1, 2019, 7:39 PM), https://www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and-the-impact-on-women/?sh=64ed80a71f53.

7. *Id.*

8. Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 39, 39 (2019).

9. *Id.*

10. AJDER ET AL., *supra* note 1, at foreword.

11. *Id.* at 1.

12. *Id.* at 1–2.

six months.[13] As of December 2020, Sensity has detected about 85,000 fake videos, almost six times what the number was the year prior.[14]

Deepfakes are quickly integrating into the mainstream internet. Most recently, a conversation about deepfakes infiltrating social media was sparked by a series of videos posted on TikTok by Tom Cruise, except it was not Tom Cruise.[15] It was an actor that looked similar to Cruise, filmed by creator Chris Ume, and then manipulated by Ume to appear to be Cruise.[16] The intent behind these videos was not malicious, as demonstrated by the account's username "deeptomcruise."[17] Ume said that the goal of these videos was to draw attention to deepfakes and advocate for their regulation.[18] One of many reasons that a need for regulation exists is due to the ease by which these videos can be created—anyone can do it.

i. GAN Technology: Anyone Can Do It

Deepfake algorithms are open source, which makes them easy to access for anyone with rudimentary programming skills.[19] In short, no expertise is needed. The most popular algorithms available are Generative Adversarial Networks ("GAN"), originally proposed by Ian Goodfellow.[20] GAN's make up includes two neural networks that contest with each other, with one network operating as a "generative" model that is trained to generate new examples—similar to a counterfeiter producing false currency.[21] The other network is the "discriminatory" model, which is designed to classify whether data is synthetic or original, similar to police attempting to detect the counterfeit currency.[22]

So, say a user wants their GAN to draw a cat, the user would give the GAN various images of cats in order to teach it what cats look like.[23] The GAN would then compare its generated cat to the images, and in time, the algorithm would learn

13. *How to Detect a Deepfake Online*, *supra* note 1.

14. *Id.*

15. Tom Knowles, *Deepfakes Are Risky Business, Warns Creator of Viral Tom Cruise*, THE TIMES (Mar. 5, 2021, 5:00 PM), https://www.thetimes.co.uk/article/deepfakes-are-risky-business-warns-creator-of-viral-tom-cruise-x2zmkqc8h?utm_medium=Social&utm_source=Twitter#Echobox=1614964116.

16. *Id.*

17. *Id.*

18. *Id.*

19. Wang, *supra* note 6.

20. Ian J. Goodfellow et al., *Generative Adversarial Nets*, *in* 1 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 27 (2014).

21. *Id.* at 1.

22. *Id.*

23.Donovan Alexander, *Artificial Intelligence Creates Better Art Than You (Sometimes),* INTERESTING ENG'G (April 11, 2021), https://interestingengineering.com/artificial-intelligence-creates-better-art-than-you-sometimes.

how to create a cat.[24] In fact, this has already been done, and a site houses endless GAN-generated cat images.[25] With each refresh of the page comes a different cat, and none of these cats actually exist.[26]

GAN has also been used less innocuously. In late 2018, a GAN-generated painting was sold at a fine auction house for $432,500—forty-two times the initial estimates.[27] This "painting" replicated a nineteenth-century portrait of a man and was created by being fed 15,000 portraits between the fourteenth and twentieth centuries.[28] The painting seems authentic until closer inspection, where the work appears unfinished and the faces unclear.[29]

Additionally, researchers from the University of Helsinki and Copenhagen used a GAN to generate images of false faces that it knew users would find attractive.[30] They attempted this objective by feeding the technology 200,000 images of celebrities.[31] This AI then produced hundreds of images of imaginary people, which were in turn shown to a group of study participants while researchers monitored their brain activity.[32] As predicted, brain activity increased when participants viewed an image of a face they were attracted to.[33] Unlike the study itself, the drive behind it was far from surface-level beauty standards.[34] Tuukka Ruotsalo, an associate professor at the University of Helsinki, stated, "[t]his could help us to understand the kind of features and their combinations that respond to cognitive functions, such as biases, stereotypes, but also preferences and individual differences."[35] As shown, deepfakes can help advance psychological and social sciences, but what is the psychological effect of knowing these deepfakes exist at all?

ii. The Liar's Dividend

The average person cannot discern deepfake from real. One 2018 study published by six professors in Germany, Italy, and France found that people only

---

24.*Id.*

25. Lindsey Romain, *This Website Creates Photos of Cats That Don't Really Exist,* NERDIST (Apr. 13, 2021, 7:53 AM), https://nerdist.com/article/this-cat-does-not-exist-website-ai-fake-cats/.

26. *Id.*

27. Alexander, *supra* note 23.

28. *Id.*

29. *Id.*

30. Vanessa Bates Ramirez, *This AI Uses Your Brain Activity to Create Fake Faces It Knows You'll Find Attractive*, SINGULARITY HUB (Mar. 18, 2021), https://singularityhub.com/2021/03/18/this-ai-uses-your-brain-activity-to-create-fake-faces-it-knows-youll-find-attractive/.

31. *Id.*

32. *Id.*

33. *Id.*

34. *Id.*

35. *Id.*

correctly identify fakes in about fifty percent of cases, which is no better than simply guessing.[36]

This gives way to the "liar's dividend," the idea that the existence of deepfakes creates distrust, casting doubt on what could very possibly be real.[37] With mainstream deepfakes, anyone can claim a released tape of them displaying prejudice or stating anything unfavorable was fake, and it would be difficult to know the difference. This is particularly concerning for public officials. Recall the infamous Hollywood Access tape where former president Donald Trump was recorded disparaging women, in turn causing a significant hit to his campaign.[38] One may argue that the impact of this recording would have been mitigated if voters or the former president questioned the authenticity of the video.

iii. Deepfake Concerns with Elections, Stock Markets, and Courtrooms

Continuing with political AI, fake videos of politicians displaying prejudice or accepting bribes can sabotage elections.[39] This should concern public officials given the creation of these videos are not difficult due to the vast amount of media that exists of public officials on the internet.[40]

There have also been concerns that deepfakes will be used to manipulate the stock market through deepfake material events, such as false mergers or catastrophic financial losses that never happened.[41] Stock markets are volatile by nature, and the existence of deepfakes aggravate this base volatility. This was demonstrated in 2011, when the Associated Press' Twitter was targeted and hacked, resulting in a tweet stating that there was an explosion in the White House and former President Barack Obama was injured.[42] The Dow Jones immediately plummeted and the S&P 500 lost $136.5 billion in market capitalization.[43]

Additionally, in the legal context, deepfakes can call into question courtroom evidence. The New Evidence Rule 902(13) allows authentication of records

36. Rössler et al., *FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces*, ARXIV, Mar. 24, 2018, https://arxiv.org/abs/1803.09179.pdf.

37. STEPHEN PROCHASKA ET AL., CTR. FOR AN INFORMED PUB., UNIV. OF WASH., DEEPFAKES IN THE 2020 ELECTIONS AND BEYOND: LESSONS FROM 2020 WORKSHOP SERIES (2020), https://cpb-us-e1.wpmucdn.com/sites.uw.edu/dist/6/4560/files/2020/10/CIP_Deepfakes_Report_Extended.pdf.

38. Lawrence Goodman, *How the Access Hollywood Tape Affected the 2016 Election*, BRANDEISNOW (Sept. 30, 2020), https://www.brandeis.edu/now/2020/september/access-hollywood-greenlee.html.

39. Westerlund, *supra* note 8, at 39–40.

40. *Id.*

41. Prajakta Pradhan, *AI Deepfakes: The Goose Is Cooked?,* U. ILL. L. REV. BLOG (Oct. 4, 2020), https://illinoislawreview.org/blog/ai-deepfakes/.

42. Shawn Langlois, *This Day in History: Hacked AP Tweet About White House Explosions Triggers Panic*, MARKETWATCH (Apr. 23, 2018, 2:08 PM), https://www.marketwatch.com/story/this-day-in-history-hacked-ap-tweet-about-white-house-explosions-triggers-panic-2018-04-23.

43. *Id.*

"generated by an electronic process or system that produces an accurate result" if "shown by the certification of a qualified person."[44] So, a video can be authenticated by having the person who took the video, or was present for the video, testify to the accuracy and validity of this video.[45] An individual would have to perjure himself or herself if there was knowledge that the video was inaccurate or a deepfake.[46]

What is concerning about this Rule is it allows video evidence to be authenticated by a witness who is familiar with the suspect in the video, and this witness can mistakenly testify that the suspect in the video is who he or she appears to be.[47] Since deepfakes are difficult to detect, even a close relative who knows the suspect in question would be at risk of false authentication.[48] Many courts also use the "silent witness" approach to authentication, which allows for evidence when no human witnessed the incident, such as security footage of a break-in.[49] This also opens the door in a dangerous way for deepfakes.

Again, this is not a future issue—this is happening now. Already, people accused of possessing child pornography often claim that it's computer-generated, says Hany Farid, a digital forensics expert at UC Berkeley.[50] "I expect that in this and other realms, the rise of AI-synthesized content will increase the likelihood and efficacy of those claiming that real content is fake."[51] Once a video is seen, its effects on the jury can be irreversible. Conversely, juries could fall victim to the liar's dividend if lawyers erroneously claim that media against their client is fake.

As illustrated, deepfake technology has found its way into various areas, including, but not limited to, elections, market manipulation, and the legal system. However, the most prominent subset remains undiscussed: nonconsensual deepfake pornography.

B. Where Deepfake Technology Meets Nonconsensual Pornography

Deepfake incitement has recently emerged with election tampering, but for women, the trauma of deepfakes has been present for years. Danielle Citron, Professor of Law at Boston University and author of *Hate Crimes in Cyberspace*, told Sensity that deepfakes are especially harmful to women:

> Deepfake technology is being weaponized against women by inserting their faces into porn. It is terrifying, embarrassing, demeaning, and

---

44. FED. R. EVID. 902(13).

45. *Id.*

46. *Id.*

47. Theodore F. Claypoole, *AI and Evidence: Let's Start to Worry*, IX THE NAT'L L. REV. 1 (2019), https://www.natlawreview.com/article/ai-and-evidence-let-s-start-to-worry.

48. Wang, *supra* note 6.

49. *Id.*

50. Kaveh Waddell, *The Deepfake Threat to Evidence*, AXIOS (Oct. 12, 2019), https://www.axios.com/deepfakes-evidence-law-f36e6538-f075-496d-bb56-64fcc29f21ef.html.

51. *Id.*

> silencing. Deepfake sex videos say to individuals that their bodies are not their own and can make it difficult to stay online, get or keep a job, and feel safe.[52]

Ninety-six percent of online deepfake videos are pornographic in nature and almost exclusively targeted women.[53]

This issue may be almost completely gender exclusive, but it is not region exclusive. As of 2019, ninety percent of general deepfake videos on YouTube specifically featured Western subjects.[54] However, non-Western subjects were featured in almost one-third of videos on specifically deepfake pornography websites, indicating that deepfake pornography is not restricted to the West, but quickly becoming a global issue.[55]

Unfortunately, this type of AI targets victims of all ages. In 2019, a deepfake bot was uncovered that used GAN technology to "undress" women, and many of these images were in reality underage children.[56] For no fee, someone could upload an image of a clothed women and receive a photo back of her with the clothes seemingly removed.[57] As of July 2020, over 100,000 women's images were used.[58] The victims come from a broad range of countries, including Russia, the United States, Argentina, and Italy.[59] Most of the images uploaded were provided from someone the user knows in real life or from Instagram.[60]

Another prominent threat is utilizing deepfakes as harassment to silence female journalists. For one journalist,[61] what started with a series of fake xenophobic tweets in retaliation of a speech on child sex abusers quickly evolved to a nonconsensual deepfake pornographic video.[62] "Within hours, I was receiving screenshots of the video on my WhatsApp, Twitter and Facebook," the journalist said in an interview with The World public radio: "I felt like I was naked for the

---

52. AJDER ET AL., *supra* note 1, at 6.

53. *Id.* at 1.

54. *Id.* at 2.

55. *Id.*

56. Karen Hao, *A Deepfake Bot is Being Used to "Undress" Underage Girls*, MIT TECH. REV. (Oct. 20, 2020), https://www.technologyreview.com/2020/10/20/1010789/ai-deepfake-bot-undresses-women-and-underage-girls/.

57. *Id.*

58. *Id.*

59. *Id.*

60. *Id.*

61. The names of the victims in this section have been left out due to the sensitive nature of the content discussed.

62. *See Internet 'Deepfakes' Threaten Truth and Reality*, THE WORLD (June 13, 2019, 5:15 PM), https://www.pri.org/stories/2019-06-13/internet-deepfakes-threaten-truth-and-reality; *see also* Rana Ayyub, *I Was the Victim of a Deepfake Porn Plot Intended to Silence Me*, HUFFPOST: LIFE LESS ORDINARY (Nov. 21, 2018), https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316.

world. I was throwing up, I was in the hospital, I had palpitations for two days and my blood pressure shot up."[63] When talking about lasting effects, the journalist stated that the video "broke" her, concluding that the experience was "scarring."[64]

In 2018, a high school student searched her name on Google and discovered that her face had been inserted into a pornographic video and various photos.[65] The image count was in the hundreds, with her face doctored onto bodies of women in numerous sexual poses and situations.[66] Further, and perhaps even more disturbing, many of these images contained identifying information, such as her name, her residence, and her school.[67] The impact was both devasting and lasting: five years after she discovered the deepfakes, she is still being targeted.[68]

Deepfake videos hosted on three of the world's largest porn websites alone have been viewed millions of times.[69] Just one deepfake video, a video using Emma Watson's face, has been viewed more than 23 million times.[70] And for these websites, views beget profit. Each video contains ad attachments that generate revenue with each click, providing little incentive for sites to remove videos such as these.[71] This lack of incentive is dangerous because, as illustrated, nonconsensual deepfake pornography can have lasting emotional and psychological harms on women, including violence, harassment, blackmail, and reputational harm.

III. COMPARING PRIVATE TO PUBLIC ACTION

A. Company Action

Companies, researchers, and organizations are recognizing this threat as well. Cyber-companies such as Truepic and Sensity have created software to detect deepfakes, with Truepic partnering with social media giant Twitter, software company Adobe, and mass media company The New York Times to create the

63. *Internet 'Deepfakes' Threaten Truth and Reality*, *supra* note 62.

64. *Id.*

65. Pradhan, *supra* note 41 (citing Ally Foster, *Picture Reveals Sickening Online Secret*, NEWS.COM.AU (June 30, 2018, 7:33 AM), https://www.news.com.au/technology/online/security/teens-google-search-reveals-sickening-online-secret-about-herself/news-story/ee9d26010989c4b9a5c6333013ebbef2).

66. Ally Foster, *Picture Reveals Sickening Online Secret*, NEWS.COM.AU (June 30, 2018, 7:33 AM), https://www.news.com.au/technology/online/security/teens-google-search-reveals-sickening-online-secret-about-herself/news-story/ee9d26010989c4b9a5c6333013ebbef2. The article cited details of the crime—one example being that she, a high school girl, was put on the cover of two adult movies, with one movie titled "Treat me like a whore" on the front. These details are uncomfortable to read but traumatizing for the victim to experience—especially when that victim is a teenager.

67. *Id.*

68. *Id.*

69. Matt Burgess, *Deepfake Porn Is Now Mainstream. And Major Sites Are Cashing in*, WIRED (Aug. 27, 2020, 6:00 AM), https://www.wired.co.uk/article/deepfake-porn-websites-videos-law.

70. *Id.*

71. *Id.*

Content Authenticity Initiative, which implements Truepic's deepfake technology detection to inform and create a standard for digital content provenance.

With the recent developments in deepfake detection technology, do platforms have an excuse for turning a blind eye? Researchers are desperately working on tools for these sites because the solution to the threat of deepfakes is greater than just public knowledge or offender liability: sites that house these altered images and videos need to mitigate. And these technological developments, as well as public pressure, are forcing companies to pay attention. Facebook has evolved from a deepfake policy of deflect and ignore, to a deepfake policy of blanket removal with limited exceptions.[72] Further, other social media sites such as YouTube, TikTok, and Twitter have followed suit.[73]

i. Sensity

Sensity is a visual threat intelligence company based in Amsterdam that is responsible for the deepfake statistics referenced in Part I of this article, as well as numerous law review papers and journalism articles on the subject.[74] Since 2019, Sensity has released three reports on the state of deepfake technology, discussing a broad range of subjects in the area.[75]

In 2021, Sensity released a "detection platform" that monitors over 500 sources that commonly host "malicious deepfakes."[76] A user can upload their file onto the platform, or copy and paste a URL to verify its presence in their record of authenticated videos.[77] Sensity promises a detection confidence of "95-99.9%."[78] Currently, the detection platform can run authenticity tests on both images and videos, support facial manipulation analysis for images and videos, and "GAN-generated faces" analysis for solely images.[79]

This resource is valuable because, of the companies discussed, Sensity is the only company that has developed technology that allows a user to check for authenticity in real time, with any type of device that has access to the internet.[80]

72. *See* discussion *infra* Section III.A.iii.

73. *See* discussion *infra* Section III.A.iii. Though there are many companies taking action to combat the risk of deepfakes, this Article touches on the companies that are taking the greatest steps.

74. AJDER ET AL., *supra* note 1 (finding that the number of deepfake videos on the internet doubled from 2018 to 2019 and at the time of its release, the web held an estimated 15,000 deepfake videos).

75. *Reports*, SENSITY, https://sensity.ai/reports/ (last visited Apr. 30, 2021).

76. *How to Detect a Deepfake Online*, *supra* note 1.

77. *Id.*

78. *Id.*

79. *Id.*

80. *Id.*

ii. Truepic

Another company working to combat the harms of deepfakes is Truepic. Truepic, founded by Craig Stack, a former Goldman Sachs employee, creates camera technology for mobile devices.[81] At the outset, Truepic was formed to combat fraud, such as Craigslist scammers and dating-site predators.[82] Truepic's biggest clients evolved to insurance companies, who used its technology to verify that policyholders' photographs of their flooded homes or broken windshields were real.[83] The company then sought to expand to industries where there is a "trust gap" and integrate the software into cameras so that "verification can begin the moment photons enter the lens."[84] Today, the company's primary mission is to authenticate digital photos and videos.[85]

Integrating verification software in cameras is no small step to combat deepfakes. Since humans are expected to take over 1.4 trillion photos in 2021, with 93.1% of these photos taken using a mobile phone or tablet,[86] the ability to install software into these devices could prevent smartphones from contributing to deepfake attacks and hacking. Further, adoption can build a foundation for the receding trust many currently experience with visual media.

Truepic has already tested its software in a prototype mobile device, which captures photos or videos with cryptographically-sealed provenance data, creating what is almost a digital fingerprint.[87] The authenticity of these photos or videos can be verified by recipients and installation does not require the download of a third-party app—the code is integrated into the device processor in a secure area that is home to sensitive tasks, such as fingerprint scanning and mobile payments.[88] By engaging the "secure" camera mode, every photo or video taken yields a digital

---

81. J.J. McCorvey, *This Image-Authentication Startup Is Combating Faux Social Media Accounts, Doctored Photos, Deep Fakes, and More*, FAST COMPANY (Feb. 19, 2019), https://www.fastcompany.com/90299000/truepic-most-innovative-companies-2019.

82. *Id. See also About Us: We're on a Mission to Restore Trust to the Internet*, TRUEPIC, [hereinafter *About Us*], https://truepic.com/about-us/ (last visited Jan. 26, 2022).

83. Joshua Rothman, *In the Age of A.I., Is Seeing Still Believing?,* The NEW YORKER (Nov. 5, 2018), https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing.

84. *Id.*

85. *See About Us*, *supra* note 82.

86. Nina Pantic, *How Many Photos Will be Taken in 2021?*, MYLIO, https://focus.mylio.com/tech-today/how-many-photos-will-be-taken-in-2021 (last visited Jan. 26, 2022).

87. *See Pioneering Provenance-Based Media Authentication*, TRUEPIC, https://truepic.com/technology/ (last visited Sept. 23, 2021); *see also* Truepic, *Truepic Breakthrough Charts a Path for Restoring Trust in Photos and Videos at Internet Scale*, CISION (Oct. 15, 2020, 8:00 PM), [hereinafter *Truepic Breakthrough*], https://www.prnewswire.com/news-releases/truepic-breakthrough-charts-a-path-for-restoring-trust-in-photos-and-videos-at-internet-scale-301152998.html.

88. *See Truepic Breakthrough*, *supra* note 87.

photo that contains cryptographic data, tags your photo or video, and authenticates various features, such as the date and geolocation.[89]

Truepic has partnered with tech giant Qualcomm to embed its technology into Qualcomm's smartphone chip, the Snapdragon 888.[90] The launch for these new chips was in 2021 and are currently available in certain Android devices, such as the popular Samsung Galaxy S21 models.[91]

Twitter, Adobe, and The New York Times are also creating a system for discerning authentic digital photography and videography from manipulated, named the Content Authenticity Initiative.[92] Truepic's technology will work with this system to tag and cryptographically encode important distinguishing information.[93] Adobe has already released this technology in Adobe Creative Cloud, which uses a version of the open standard created that will provide the author's name, location, and edit history.[94]

iii. Facebook

Companies such as Facebook have been criticized for their unwillingness to do more to temper deepfakes. To circle back to the story of the journalist from Part I, who had a retaliatory nonconsensual deepfake pornographic video go viral, she asserts that when she spoke with Facebook and Twitter about removing the video, they refused.[95] "The problem was that they were not willing to concede that there was a problem on their part that their platform was being used to disseminate this video," the journalist stated in an interview with The World public radio.[96]

Months after the interview, in September of 2019, Facebook partnered with AWS, Microsoft, and others to create a "Deepfake Detection Challenge" to measure the progress on deepfake detection technology.[97] In this challenge, Facebook

---

89. *See Truepic Breakthrough, supra* note 87.

90. *About Us*, *supra* note 82.

91. *See id.*; *see also* Tuan Do, *List of Phones Powered by Qualcomm Snapdragon 888 Processor Released in 2021*, TECHWALLS, https://www.techwalls.com/snapdragon-888-smartphone-list/ (Sept. 21, 2021).

92. Arooj Ahmed, *The Truepic Technology and Joint Venture with QUALCOMM May Give You the Exact Time and Location of a Photo or Video Taken from Your Phone*, DIGIT. INFO. WORLD (Oct. 18, 2020, 11:00 PM), https://www.digitalinformationworld.com/2020/10/the-truepic-technology-and-joint.html.

93. *Id.*

94. Will Allen*, The Content Authenticity Initiative unveils content attribution tool within Photoshop and Behance*, ADOBE BLOG (Oct. 20, 2020), https://blog.adobe.com/en/publish/2020/10/20/content-authenticity-initiative-unveils-content-attribution-tool-within-photoshop-behance.html#gs.z3shja; *see also* Ahmed, *supra* note 92.

95. *Internet 'Deepfakes' Threaten Truth and Reality*, *supra* note 62.

96. *Id.*

97. *See Deepfake Detection Challenge*, KAGGLE, https://www.kaggle.com/c/deepfake-detection-challenge (last visited Apr. 19, 2021); *see also Deepfake Detection Challenge Dataset*, META AI (June 25, 2020), https://ai.facebook.com/datasets/dfdc/.

created a dataset with individuals who agreed to the "use and manipulation of their likenesses."[98] Facebook released this dataset both as part of the challenge and to the general public in order to "accelerate progress on detecting harmful manipulated media."[99] This challenge released a public preview dataset, which consisted of 5,000 videos and two facial modification algorithms, and the black box dataset, which consisted of 124,000 videos and eight facial modification algorithms.[100] The evaluation of the algorithms created were on an unseen black box dataset that was home to both organic content found on the internet and new videos created for the project, consisting of largely tough to classify videos such as makeup tutorials and paintings.[101]

This competition was hosted by Kaggle, a subsidiary of Google, and included a $1 million prize.[102] The top-performing model on average achieved 82.56% precision with the public dataset.[103] However, with the black box dataset, the highest-performing model on average achieved just 65.18% precision.[104] This discrepancy verifies that a key challenge for detecting deepfakes is generalizing detection techniques to both known and unforeseen synthetic examples.

In October of 2019, shortly after the creation of the Deepfake Detection Challenge, the co-founder and CEO of Facebook testified in a House Financial Services Committee hearing.[105] Mark Zuckerberg answered questions on data and election security issues related to Facebook posts and ads during testimony before the House Financial Services Committee.[106] When criticized for allowing the spread of a deepfake video containing manipulated media of House Speaker Nancy Pelosi appearing to be inebriated, Zuckerberg conceded that it was a company failure to not "fact check" the video and that he played a role in the decision not to remove the video due to company policy.[107] During this testimony, Representative Jennifer Wexton asked a poignant question: "Do you understand there's a difference between misinformation and disinformation?"[108] Zuckerberg responded that he did but that it is hard to determine intent.[109]

In January of 2020, the social media giant announced its firmest stance against manipulated media to date, revealing that Facebook will remove manipulated

---

98. *Deepfake Detection Challenge Dataset*, *supra* note 97.

99. *Id.*

100. *Id.*

101. *Id.*

102. *Deepfake Detection Challenge*, *supra* note 97.

103. *Deepfake Detection Challenge Dataset*, *supra* note 97.

104. *Id.*

105. *Facebook CEO Testimony before House Financial Services Committee*, C-SPAN (Oct. 23, 2019), https://www.c-span.org/video/?465293-1/facebook-ceo-testimony-house-financial-services-committee#.

106. *Id.*

107. *Id.*

108. *Id.*

109. *Id.*

media that meet two criteria.[110] First, manipulated media will be removed if the media has been edited or synthesized in ways that an average person would not recognize or would mislead someone into thinking that a subject of the video actually said something that, in reality, they did not.[111] Second, manipulated media will be removed if it is the product of AI or machine learning that "merges, replaces or superimposes content onto a video" to make it appear to be authentic.[112]

There are exceptions to this new policy: parody, satire, or videos that have been edited solely to omit or change the order of the words will remain.[113] Further, videos that may not fall under the criteria for removal may still be reviewed by Facebook's numerous independent third-party fact-checkers.[114] If the image or video is deemed false, distribution will be reduced in users' News Feed and rejected if an ad.[115] Additionally, those who see it, attempt to share it, or already shared it will receive a warning that it is false.[116]

Other social media sites have also recently taken steps to mitigate deepfakes. In February of 2020, YouTube announced that they will disallow deepfake "election-related content."[117] And later that year, TikTok announced that they were removing all deepfakes "which prohibits synthetic or manipulated content that misleads users by distorting the truth of events in a way that could cause harm."[118] Finally, in 2021, Twitter implemented a policy that allows the site to "label Tweets containing synthetic and manipulated media to help people understand their authenticity and to provide additional context."[119]

B. Federal Action

Legislators are only recently starting to understand and care about the effects of both revenge porn and deepfake technology. Former President Donald Trump implemented unprecedented legislation to combat deepfakes in response to

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

111. *Id.*

112. *Id.*

113. *Id.*

114. *Id.*

115. *Id.*

116. Bickert, *supra* note 110.

117. Leslie Miller, *How YouTube Supports Elections*, YOUTUBE OFFICIAL BLOG: NEWS & EVENTS (Feb. 3, 2020), https://blog.youtube/news-and-events/how-youtube-supports-elections?m=1.

118. Vanessa Pappas, *Combating Misinformation and Election Interference on TikTok*, TIKTOK: COMPANY (Aug. 5, 2020), https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok.

119. *Synthetic and Manipulated Media Policy*, TWITTER: HELP CENTER, https://help.twitter.com/en/rules-and-policies/manipulated-media (last visited Apr. 29, 2021).

election interference,[120] and it appears that current President Joe Biden will be furthering the fight against deepfakes. In 2021, Representative Yvett Clarke addressed the Biden administration shortly upon introduction of the DEEPFAKES Accountability Act: "We're in a new Congress," Clarke said about the Biden administration.[121] "There are members in the Congress, both on the Senate and House side, who recognize what this threat is to our way of life, and how it has already been used to abuse women."[122]

i. The National Defense Authorization Act for Fiscal Year 2020

Former President Donald Trump signed the first federal law regarding deepfakes in response to election interference by foreign entities: The National Defense Authorization Act for Fiscal Year 2020.[123] This Act directed the Director of National Intelligence to produce a comprehensive report on the foreign weaponization of deepfakes and even created a "deepfakes prize competition" to encourage research and development of technology to detect deepfakes.[124] The competition provided a $5 million prize for one or more winners.[125]

This Act detailed reporting requirements for the Director of National Intelligence, stating that within six months of enactment, the Director is to submit a report on the potential national security impacts of deepfakes and how foreign governments are or might use them "to spread disinformation or engage in other malign activities."[126]

For election deepfakes specifically, this Act required the Director to notify the congressional Intelligence Committees whenever he or she ascertained that there is credible information or intelligence that a foreign entity has or is utilizing deepfakes aimed at elections or the political processes of the U.S.[127]

ii. IOGAN Act

In one of former President Donald Trump's last acts as president, he signed into law the Identifying Outputs of Generative Adversarial Networks Act.[128] With this act, the Director of the National Science Foundation ("NSF") must support research on manipulated media and information authenticity, as well as support

120. *See* National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, § 5724, 133 Stat. 1198, 2177-78 (2019) (codified as 50 U.S.C.A. § 3024 (West 2019)).

121. Hao, *supra* note 3.

122. *Id.*

123. *See* National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, § 5724, 133 Stat. 1198, 2177-78 (2019) (codified as 50 U.S.C.A. § 3024 (West 2019)).

124. *Id.* at §§ 5709, 5724.

125. *Id.* at § 5724.

126. 50 U.S.C.A. § 3369a(a)(1)(B) (West 2019).

127. *Id.*

128. Identifying Outputs of Generative Adversarial Networks Act, Pub. L. 116-258, 134 Stat. 1150, 1150–52 (2019) (codified as 15 U.S.C.A. § 9202 (West 2020)).

research for developing standards to accelerate the development of technology regarding GANs or other similar media manipulation technology.[129]

When describing the bill, the House committee report explained that "the intent of this legislation is to accelerate the progress of research and the development of measurements, standards, and tools to combat manipulated media content, including the outputs of generative adversarial networks."[130] The committee encouraged NSF "to continue to fund cross-directorate research through these programs, and others, to achieve the purposes" of the legislation, "including social and behavioral research on the ethics of these technologies and human interaction with the content generated by these technologies."[131]

iii. Non-Deepfake Federal Liability

For a nonconsensual deepfake pornography victim, there are many potential avenues, but they are not promising. Currently, no federal law criminalizes nonconsensual deepfake pornography and the criminal options are significantly less than civil. Civil avenues for nonconsensual deepfake pornography are scarce and only allow for filing suit against the creator of the deepfake, not the platform that holds or distributes it.

There is a cyberstalking statute that could be effective, but only if there is an element of intent.[132] A victim may have tort options, but many are written too narrowly to encompass all victims. A victim may also have a copyright claim if the deepfake photo or video is from photos or videos the victim took personally, but questions exist for deepfakes that need to be hashed out, such as exploring who owns the IP of the synthetic data—the algorithm or writer. Most of these current legal avenues lack the nuances of AI necessary to be effective.

Law surrounding cyberstalking could provide an avenue to some. 18 U.S.C. § 2261A(1)(B) provides liability if someone uses "any interactive computer service or electronic communication service or electronic communication system" to engage in conduct that causes or attempts to cause "substantial emotional distress."[133] However, there is always the argument that the creator lacked the intent needed to file suit under this law.

The tort of intentional infliction of emotional distress provides a seemingly simple avenue, but upon further glance is also too narrow to encompass all victims. This tort provides liability for a victim if "[a]n actor . . . by extreme and outrageous conduct intentionally or recklessly causes severe emotional harm to another."[134] The intent requirement can leave out a variety of victims if the creator did not intend to create the harm. For instance, perhaps the creator never thought the

129. 15 U.S.C.A. § 9202 (West 2020).

130. H.R. REP. NO. 116-268, at 6 (2019).

131. *Id.*

132. 18 U.S.C.A. § 2261A(1)(B) (West 2020).

133. *Id.*

134. RESTATEMENT (THIRD) OF TORTS § 46 (AM. L. INST. 2012).

victim would see the deepfake or created the deepfake solely for his or another's pleasure.

Privacy-based torts also at first glance seem to provide legal avenues for victims, but upon further reading require elements that are too narrow for most cases. The right of privacy is defined as "the right to be let alone,"[135] which certainly can transfer to nonconsensual deepfake pornography; however, it is clear that the torts described were not written with artificial intelligence in mind. There are three possible privacy-based torts that one could cite when pursing a tort claim.[136]

First, the appropriation of name or likeness tort grants an avenue for a legal remedy if someone "appropriates to his own use or benefit the name or likeness of another."[137] However, the use must "be of benefit" to the individual using the likeness.[138]

Next, the publicity given to private life tort imposes liability on one who "gives publicity to a matter concerning the private life of another" if the matter publicized "would be highly offensive to a reasonable person" and "is not of legitimate concern to the public."[139] However, this tort requires that the publicity concern the *private* life of the individual.[140] There is no liability when the information is already public.[141] So, in many cases when images of someone's face are taken from their public social media account, this tort would not be a viable option.

Lastly, the publicity placing person in false light tort punishes "one who gives publicity to a matter concerning another that places the other before the public in a false light" if the false light would be "highly offensive to a reasonable person" and "the actor had knowledge of or acted in reckless disregard as to the falsity of the publicized matter and the false light in which the other would be placed."[142] While this tort could work in some instances, if the video or image is noticeably manipulated or has a label stating that it has been manipulated, this tort would not be of use.

As seen, these privacy-based torts at first glance seem to provide legal remedies to the victim, but further reading informs that they were drafted too narrowly to encompass many of the issues of AI. The elements for the three torts listed—such as necessary benefit, that the image taken to create the deepfake be "private," and that the deepfake must appear to be true—provide too narrow of a road for many victims to explore.

---

135. RESTATEMENT (SECOND) OF TORTS § 652A cmt. a (AM. L. INST. 1977).

136. *See id.* § 652A.

137. *Id.* § 652C.

138. *Id.*

139. *Id.* § 652D.

140. *See* RESTATEMENT (SECOND) OF TORTS § 652D cmt. B (AM. L. INST. 1977).

141. *Id.*

142. *Id.* § 652E.

Copyright claims may also exist, but only if the image used to make the deepfake was taken by the victim.[143] However, if the created material was "transformative," then this option disappears.[144] The U.S. Copyright Office of Fair Use stated that "[t]ransformative uses are those that add something new, with a further purpose or different character, and do not substitute for the original use of the work."[145] A creator could easily argue that taking a photograph and manipulating it into a nonconsensual deepfake pornography video is indeed transformative. One would be hard-pressed to find a deepfake that does not add something new, which is why it is called manipulated media.

iv. Pending Legislation

In 2019, Representative Yvett Clarke introduced the DEEPFAKES Accountability Act.[146] This Act would require creators of synthetic media that contain the likeness of a person to disclose that the video has been manipulated, using "irremovable digital watermarks, as well as textual descriptions."[147] Importantly, this act would allow victims to sue the creators and "vindicate their reputations" in court.[148] However, this Act, while a step in the right direction, contains loopholes that must be addressed.[149] Creators of harmful deepfakes usually preserve anonymity, so harmful media with anonymous creators will likely disregard the watermark requirement.[150] Further, even if there is a watermark attached, they are relatively easy to remove.[151] As one writer explained, "[t]ext can be cropped, logos removed (via more smart algorithms), and even a sophisticated whole-frame watermark might be eliminated simply by being re-encoded for distribution on Instagram or YouTube."[152]

Next, in March of 2021, the U.S. House of Representatives voted to reauthorize the Violence Against Women Act, a bill designed to protect victims of

143. *More Information on Fair Use*, COPYRIGHT.GOV, https://www.copyright.gov/fair-use/more-info.html (May 2021).

144. *Id.*

145. *Id.*

146. Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act, H.R. 3230, 116th Cong. (2019) (this bill has not yet been brought for a vote).

147. *Id.* § 1041(a)(1).

148. *Id.* § 1041(f).

149. *Id.*

150. Ashley Dean, *Deepfakes, Pose Detection, and the Death of "Seeing is Believing,"* LAW TECH. TODAY (Aug. 6, 2020), https://www.lawtechnologytoday.org/2020/08/deepfakes-pose-detection-and-the-death-of-seeing-is-believing/.

151. Devin Coldewey, *DEEPFAKES Accountability Act Would Impose Unenforceable Rules—But It's a Start*, TECHCRUNCH (June 13, 2019, 1:25 PM), https://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/.

152. *Id.*

domestic violence and sexual assault.[153] This amendment included a ban on knowingly or recklessly distributing "intimate visual depictions" of individuals without their consent.[154] Offenders would face up to two years in prison for each individual victim depicted.[155] This would be the first U.S. federal law that attempts to seriously address online nonconsensual pornography.[156] "Writing and passing VAWA is one of the legislative accomplishments of which I'm most proud," Biden said in the statement.[157] "VAWA has transformed the way our country responds to violence against women."[158] This bill, while a positive step forward for nonconsensual pornography, is a missed opportunity to provide for the deepfake subset of this crime.

C. State Action

In the United States, forty-eight states, Washington D.C., and Guam have some variant of a ban on nonconsensual pornography, and only Virginia and California have a nonconsensual pornography variant that includes faked and deepfaked media.[159] Approximately a dozen states have legislation pending, though nearly all legislation is civil and addresses "actual malice related to the intent to deceive and the knowledge of deceivery."[160] So, to file a suit against a deepfake creator, the deepfake would have to not only be deceptive, but the creator must have intended to deceive.

i. Virginia

To address the disproportionate number of deepfakes in the form of non-consensual pornography with female victims, Virginia expanded its ban on non-consensual pornography to images of people "whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person and who is recognizable as an actual person by the person's face,

153. Violence Against Women Reauthorization Act of 2021, H.R. 1620, 117th Cong. (2021).

154. *Id.*

155. *Id.*

156. Adi Robertson, *A Federal 'Revenge Porn' Ban Could Transform Online Harassment Laws*, THE VERGE (Apr. 15, 2021, 10:00 AM), https://www.theverge.com/2021/4/15/22340260/vawa-shield-act-revenge-porn-first-amendment-questions.

157. Statement by President Biden on the Passage of the Violence Against Women Reauthorization Act of 2021 in the House of Representatives (Mar. 17, 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/17/statement-by-president-biden-on-the-passage-of-the-violence-against-women-reauthorization-act-of-2021-in-the-house-of-representatives/.

158. *Id.*

159. *See 46 States + DC + One Territory Now Have Revenge Porn Laws*, CYBER CIVIL RIGHTS INITIATIVE, https://web.archive.org/web/20200722180154/https://www.cybercivilrights.org/revenge-porn-laws/ (last visited Apr. 19, 2021); *see also* PROCHASKA ET AL., *supra* note 37.

160. PROCHASKA ET AL., *supra* note 37.

likeness, or other distinguishing characteristic."[161] With the passage of this legislation, Virginia became the first state to impose criminal penalties for the distribution of nonconsensual deepfake pornography.[162]

ii. California

On October 3, 2019, California enacted two new laws that regulate the distribution of deepfakes: A.B. 602 and 730.[163] A.B. 602 is tailored to nonconsensual deepfake pornography, and broadly allows victims to prosecute creators for damages.[164] A plaintiff can recover disgorgement of profits, economic and noneconomic damages, and statutory damages up to $150,000 if the act was "committed with malice."[165] However, this is a small sum for what could be a life-altering event.

As noted, a study from Sensity found that ninety-six percent of deepfakes online are sexually explicit, but what has not been discussed is that ninety-nine percent of these women were or are currently entertainment professionals.[166] The Screen Actors Guild, a union that represents film, TV, and other media professionals, praised Governor Newsom for the bill, citing a step towards protecting women. "We are absolutely thrilled that Governor Newsom stood by the victims, most of whom are women, of non-consensual pornography by signing A.B. 602 into law," Gabrielle Carteris, president of the union, told Deadline.[167] "Every person deserves the basic human right to live free from image-based sexual abuse."[168]

Next, A.B. 730 addresses elections, and grants political candidates for public office an avenue to prosecute individuals or organizations that maliciously distribute "materially deceptive" media about any candidate within sixty days of an

161. *See* H.D. 2678, 2019 Gen. Assemb., (Va. 2018) (amending VA. CODE. ANN. § 18.2-386.2 (2019)); *see also* Matthew F. Ferraro, *Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*, WILMERHALE (Sept. 25, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey.

162. Ferraro, *supra* note 161.

163. *See* Assemb. 602, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. CODE CIV. PRO. § 1708.85–86 (2020)); *see also* Assemb. 730, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. ELEC. CODE § 20010(a) (2020)).

164. *See* Assemb. 602, (amending CAL. CODE CIV. PRO. § 1708.85–86 (2020).

165. *Id.*

166. AJDER ET AL., *supra* note 1, at 1–2.

167. David Robb, *SAG-AFTRA Commends Gov. Newsom for Signing "DeepFakes" Bill*, DEADLINE (Oct. 3, 2019), https://deadline.com/2019/10/sag-aftra-commends-gov-newsom-for-signing-deepfakes-bill-1202752095/.

168. *Id.*

election.[169] Victims can pursue both damages and equitable relief, such as an injunction to dissemination of the deepfake.[170] There is no liability for print of online media paid to disseminate the deepfake as long as the site provides a disclosure with the media stating that it has been manipulated.[171] The bill also contains exceptions for deceptive media that can be considered satire.[172] The bill is set to sunset on January 1, 2023.[173]

Both of these laws have exceptions in place to alleviate First Amendment infringements. A.B. 730 does not alter protections under Section 230 of the Communications Decency Act, nor does it apply to satire, and it allows for websites or radio and television stations to circumvent liability if there is a disclaimer informing the reader of the manipulation.[174] Similarly, A.B. 602 asserts that a creator cannot be held liable for creating or publishing manipulated content that is a "matter of legitimate public concern," if the media is of "political or newsworthy value," or within state or federal protections for "commentary or criticism."[175]

Nevertheless, this bill has faced criticism for its restriction on free speech from organizations like the ACLU of California and the California Broadcasters Association.[176] Mark Powers, Vice President of the California Broadcasters Association, said the bill would be impossible to comply with.[177] "By passing this bill, you put your broadcasters in jeopardy," Powers told the Senate Elections and Constitutional Amendments Committee.[178] This puts broadcasters in a tough position where if fact-checking a political ad proves to be too costly, they may choose not to run ads at all to avoid potential liability.[179] California News Publishers Association Staff Attorney Whitney Prout stated the bill was "an ineffective and frankly unconstitutional solution that causes more problems than it solves," noting that A.B. 730 would impose restrictions on speech that must survive strict scrutiny.[180]

There is truth to these concerns. A.B. 730 would likely ban altering content to reenact true events that were not filmed or recorded and could prohibit a

---

169. *See* Assemb. 730, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. ELEC. CODE § 20010(a) (2020)).

170. *Id.*

171. Assemb. 730, 2019-2020 Reg. Sess. (Cal. 2019).

172. *Id.*

173. *Id.*

174. *Id.*

175. *See* Assemb. 602, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. CIV. PROC. CODE § 1708.86 (2020)).

176. Nick Cahill, *Bill to Fight 'Deepfake' Videos Advances in California, Despite Free-Speech Fears*, COURTHOUSE NEWS SERV. (July 2, 2019), https://www.courthousenews.com/bill-to-fight-deepfake-videos-advances-in-california-despite-free-speech-fears/.

177. *Id.*

178. *Id.*

179. *Id.*

180. *Id.*

candidate's use of altered videos of him or herself.[181] Additionally, A.B. 602 could impose liability for content viewed by only the creator and lacks clarification for when consent is revoked after creation or distribution.[182]

Marc Berman, the California assembly member who introduced AB. 730, addressed concerns about A.B. 730, while also offering a judicial counterpoint.[183] "I understand that there are significant First Amendment concerns with the bill as it is currently drafted, and I'm committed to working through these issues," Berman said.[184] "I would note however, that I haven't seen a court determine that the First Amendment grants someone the right to literally put their words into my mouth, which is what this technology does."[185]

iii. Texas

Texas has deepfake legislation, but only narrowly tailored to elections.[186] Texas imposes penalties on the creation or distribution of deepfake videos intended to harm candidates for public office or influence elections.[187] S.B. 751 makes it a Class A misdemeanor, punishable by up to a year in the county jail and a fine of $4,000, for a person to "create[] a deepfake video" and "cause[] the deepfake video to be published or distributed within 30 days of an election," if the person does so with the "intent to injure a candidate or influence the result of an election."[188]

The Texas Senate Research Center released an analysis on the bill, acknowledging that deepfake technology "likely cannot be constitutionally banned altogether," but concluded that "it can be narrowly limited to avoid what may be its greatest potential threat: the electoral process."[189] This statement is valid. Though banning deepfakes would solve many of the issues discussed, it would be a blatant disregard for First Amendment rights, as well as the benefits that deepfakes provide. A legislative solution must be drafted that is not too broad as to cause

181. *See* Assemb. 730, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. ELEC. CODE § 20010(a) (West 2020)); *see also* K.C. Halm et al*., Two New California Laws Tackle Deepfake Videos in Politics and Porn*, DAVIS WRIGHT TREMAINE LLP (Oct. 14, 2019), https://www.dwt.com/insights/2019/10/california-deepfakes-law.

182. *See* Assemb. 602, 2019-2020 Reg. Sess. (Cal. 2019) (amending CAL. CIV. PROC. CODE § 1708.86 (West 2020)); *see also* K.C. Halm et al*., supra* note 181.

183. Cahill, *supra* note 176.

184. *Id.*

185. *Id.*

186. S. 751, 86th Legis., Reg. Sess. (Tex. 2019) (amending TEX. ELEC. CODE ANN. § 255.004 (West 2019)). Maryland, Maine, and Washington also have proposed deepfake election legislation. Matthew Feeney, *Deepfake Laws Risk Creating More Problems Than They Solve*, REGUL. TRANSPARENCY PROJECT (Mar. 1, 2021), https://regproject.org/paper/deepfake-laws-risk-creating-more-problems-than-they-solve/.

187. S. 751, 86th Legis., Reg. Sess. (Tex. 2019); Feeney, *supra* note 186, at 6.

188. TEX. ELEC. CODE ANN. § 255.004 (West 2019); TEX. PENAL CODE ANN. § 12.21 (West 2021).

189. Hughes, *Bill Analysis*, SENATE RSCH. CTR. (June 12, 2019) https://capitol.texas.gov/tlodocs/86R/analysis/html/SB00751F.htm.

constitutional distress, but not too narrow, such as this Texas legislation, as to bar a remedy for a large subset of victims.

IV. EXAMINING POTENTIAL SOLUTIONS

The current roadmap for victims of nonconsensual deepfake pornography is cloudy at best. A nationwide ban on deepfakes would be the most effective solution to the issues discussed, but this is not possible and would create more issues than remedies. Implementing a ban such as that would be a suffocation of the freedom of expression that Americans have a fundamental right to. Similarly, an injunction against deepfakes likely infringes on the Constitution's First Amendment.[190] But what if the legislation drafted was not a blanket ban on deepfakes, but solely for nonconsensual deepfake pornography, with the possible extension for other pressing issues, such as intentional election interference?

As seen above, current legislation targeting deepfakes is too narrow, with most efforts focusing solely on election issues, and current legislation for platform protections which is too broad.[191] Future legislation must balance the interests of free speech and encouragement for technological advancements with the interests of privacy. To find this balance and address the necessary issues of manipulated media separate from the narrow lens of elections, legislators must work with experts in the technology industry, the field of cybersecurity law, and academia to address all of the necessary nuances, so that no victim falls through the cracks. This solution must focus on ensuring companies are making the detection of deepfakes a priority, as well as removing harmful manipulated videos from their platform and holding the creators accountable.

State laws alone are not reliable nor effective due to the nature of the internet as a national and global force.[192] Federal legislation must be drafted to truly be

---

190. For a comprehensive analysis of potential permissible injunctions regarding deepfakes, see Jessica Ice, *Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RSRV. L. REV. 417 (2019).

191. *See supra* notes 123–89.

192. The rationale behind why a federal law would be the most effective solution rather than state specific legislation mirrors the rationale for why an international standard would be more effective than a nation-wide standard. Simply put, the internet is not confined to just one state or, in most cases, one country. For an example of how the wide reach of a current international privacy standard can be utilized to regulate deepfakes, see Martijn van der Helm, Harmful Deepfakes and the GDPR (Dec. 10, 2021) (M.A. thesis, Tilburg University), http://arno.uvt.nl/show.cgi?fid=156861.

Currently, there is no global standard, though there is some international liability. Former President Barack Obama amended Executive Order 13964 to forbid foreign entities from using cyber-enabled means to "[t]amper with, alter, or cause a misappropriation of information with the purpose or effect of interfering with or undermining election processes or institutions." Obama used this Order to impose sanctions against Russia, as did former President Donald Trump. Press Release, Office of the Press Secretary, Actions in Response to Russian Malicious Cyber Activity and Harassment (Dec. 29, 2016), https://obamawhitehouse.archives.gov/the-press-office/2016/12/29/fact-sheet-actions-response-russian-malicious-cyber-activity-and.

effective, and tech giants must step up to mitigate. If a solution can be found for this issue of nonconsensual deepfake pornography, which encompasses the vast majority of deepfakes, then this will open doors to extend legislation or technological advances to other harmful deepfake subsets, such as election interference.

A. Section 230 of the Communications Decency Act

The Communications Decency Act ("CDA") shaped the internet into what it is today. The CDA is a federal law that was passed in 1996 and regulates pornography on the internet, as well as protects websites from liability for content posted by their users.[193] Under Section 230 of the CDA, the owners of internet services and websites are not regarded as the publishers of the content that their users post.[194] As such, there is no legal obligation to remove nonconsensual pornography, or nonconsensual deepfake pornography, unless it violates copyright or federal criminal laws.[195]

The purpose behind Section 230 was to allow website owners to moderate their websites without concern for legal liability, as part of a larger initiative to enable growth of the internet.[196] However, the landscape of the internet has changed, and this concern is no longer relevant. Section 230 as it reads currently is overly broad, protecting both good and bad actors.[197] Consequently, aside from

---

While this Article agrees that international liability is needed, the United States must also work with other countries to impose standards and regulations regarding deepfakes. The United States is not the only country feeling the effects—other countries are hearing outcries from their citizens and implementing legislation of their own. *See e.g.*, *#MyImageMyChoice is A Coalition of Survivors and Advocates Calling for Reform in Law and Government Policy on Image-Based Sexual Abuse*, ELIZABETH WOODWARD, https://elizabethwoodward.com/work/myimagemychoice/ (last visited Feb. 28, 2022); *see also, e.g.*, Julian Ryall, *'Deepfakes' Rattle South Korea's Tech Culture*, DW (Jan. 22, 2021), https://www.dw.com/en/deepfakes-rattle-south-koreas-tech-culture/a-56310213. The United States should work with these countries to adopt a more uniform approach to deepfakes because the internet is global, with the need for global solutions.

Further analysis of international liability is beyond the scope of this Article.

193. *See* 47 U.S.C. § 230*; see also Frequently Asked Questions*, CYBER CIV. RTS. INITIATIVE, https://www.cybercivilrights.org/faqs/ (last visited Apr. 21, 2021) (choose "[a]ren't civil and copyright remedies adequate to address this conduct?" to expand text).

194. 47 U.S.C § 230.

195. *Id.*

196. *See* Alexander F. Magee, *Back Against the Wall: Are Section 230's Days Numbered?,* WAKE FOREST L. REV.: CURRENT ISSUES BLOG, http://www.wakeforestlawreview.com/2020/10/back-against-the-wall-are-section-230s-days-numbered/ (last visited Apr. 26, 2021); *see also* Ashley Johnson & Daniel Castro, *Overview of Section 230: What it is, Why it Was Created, and What it Has Achieved*, INFO. TECH. & INNOVATION FOUND. (Feb. 22, 2021), https://itif.org/publications/2021/02/22/overview-section-230-what-it-why-it-was-created-and-what-it-has-achieved.

197. *See* 47 U.S.C. § 230.

public disapproval, there is little incentive for online platforms to monitor or remove harmful media. An amendment would modernize the statute and mitigate the harm that is a consequence of the current lack of platform incentive.

i. Amending Section 230

The strength in Section 230 derives from its treatment as an absolute exemption from almost every civil federal law. Currently, Section 230 shields sites against civil suits and state laws.[198] Section 230 does not explicitly prohibit states from enforcing their own "consistent" laws but does not allow enforcement if the laws are "inconsistent" with Section 230.[199] In both *Voicenet Communications, Inc. v. Corbett* and *Backpage v. McKenna*, the courts reinforced that online platforms are not liable under state criminal laws deemed inconsistent.[200] Similarly, in *Perfect 10, Inc. v. CCBil LLC*, the court held that online platforms are not liable under state intellectual property laws deemed inconsistent.[201] And recently, the second circuit in *Domen v. Vimeo* applied Section 230 immunity in the context of removing content on an online platform, as opposed to simply allowing the content on the site.[202] This is the first circuit court of appeal to make a decision applying Section 230 as a basis for immunity during the pleading stage, denying opportunity for discovery.[203]

As illustrated, Section 230 grants great power to platforms. The reason this statute in particular is detrimental to deepfake victims is that in many instances, the creators of deepfakes are hard to find.[204] This leaves limited avenues for a victim to pursue legal remedies through other means, such as suing the platform holding or disseminating the media. Utilizing or amending Section 230 to impose platform liability is arguably the most efficient and effective solution to this issue.

The arguments against Section 230 have included both an amendment and an abolishment all together.[205] In December of 2019, then former Vice-President and

198. *See* 47 U.S.C. § 230*; see also Frequently Asked Questions*, *supra* note 193.

199. 47 U.S.C. § 230(e).

200. Voicenet Commc'ns, Inc. v. Corbett, No. 04-1318, 2006 WL 2506318 (E.D. Pa. Aug. 30, 2006); Backpage.com, LLC v. McKenna, 881 F. Supp. 2d 1262, 1273 (W.D. Wash. 2012).

201. Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1118–19 (9th Cir. 2007).

202. Domen v. Vimeo, Inc., 6 F.4th 245, 253 (2d Cir. 2021).

203. Fenwick & West LLP, *Second Circuit Affirms Video Sharing Site's Immunity from Suit Under CDA Section 230 for Removal of User Content*, LEXOLOGY (Mar. 18, 2021), https://www.lexology.com/library/detail.aspx?g=2279bc78-8906-44bc-8371-388c1ff46ee2.

204. Robert Chesney & Danielle K. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1792 (2019) (discussing a lack of metadata available to identify a creator, technology available to ensure anonymity, and jurisdictional restraints).

205. The New York Times Editorial Board*, Joe Biden: Former Vice President of the United States*, N.Y. TIMES: OPINION (Jan. 17, 2020), https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html?smid=nytcore-ios-share (Interview conducted on Dec. 16, 2019) (Former Vice President Biden, discussing C.D.A. 230, stated that "[i]t should be revoked because it is not merely

now current President Joe Biden stated that Section 230 should be abolished completely, explaining that for Facebook and other platforms, Section 230 "should be revoked because it is not merely an internet company. It is propagating falsehoods they know to be false."[206]

However, an abolishment would negate the valid uses of Section 230. For instance, sites such as Wikipedia, which host user contributions and volunteer editors, as well as blogs, and online business reviews, would not exist without Section 230's liability shield.[207] For some of these sites, such as Facebook where 1.91 billion people log in daily, it can be nearly impossible to regulate with 100% accuracy.[208] Without Section 230's shield, companies, especially start-ups, would falter under legal expenses. As the Ninth Circuit has explained, without Section 230, start-ups would face "death by ten thousand duck-bites" defending lawsuits.[209] For context, for each case that reaches the discovery stage, a start-up is forced to pay anywhere from $100,000 to $500,000 defending against the lawsuit.[210]

An amendment to Section 230 is not impossible. Section 230 has previously been amended due to public interest and a concern for safety, opening the door for an amendment regarding nonconsensual deepfake pornography. In 2018, the statute was amended to address sex trafficking by broadly expanding prosecutorial power over companies used by sex workers for their business.[211] Unlike the new legislation drafted in and contested in *Ashcroft*, this amendment to Section 230 was successful and passed both the House and Senate with overwhelming bipartisan support.[212] As a result, platforms can no longer be used for assisting, supporting, or facilitating sex trafficking and are thus held accountable for third-party activity regarding sex trafficking or prostitution.[213] The amendment also contains

---

an internet company."); *see also* Rachel Lerman, *Social Media Liability Law is Likely to Be Reviewed Under Biden*, WASH. POST (Jan. 18, 2021, 8:00 AM), https://www.washingtonpost.com/politics/2021/01/18/biden-section-230/ ("Democrats also think the law should be amended . . . .").

206. The New York Times Editorial Board, *supra* note 205.

207. *See generally* 47 U.S.C. § 230.

208. Dan Noyes, *The Top 20 Valuable Facebook Statistics – Q2 2021*, ZEPHORIA, https://zephoria.com/top-15-valuable-facebook-statistics/ (July 28, 2021) ("1.91 billion people on average log onto Facebook daily and are considered daily active users . . . .").

209. Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157, 1174 (9th Cir. 2008).

210. *Section 230: Cost Report*, ENGINE, https://static1.squarespace.com/static/571681753c44d835a440c8b5/t/5c8168cae5e5f04b9a30e84e/1551984843007/Engine_Primer_230cost2019.pdf (last visited Jan. 28, 2022).

211. Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115-164, 132 Stat. 1253 (2018) (codified as amended at 47 U.S.C. § 230(d) (2018)).

212. *See generally* Ashcroft v. Free Speech Coal., 535 U.S. 234 (2002); Allow States and Victims to Fight Online Sex Trafficking Act of 2017.

213. Allow States and Victims to Fight Online Sex Trafficking Act of 2017.

retroactive provisions, allowing prosecutors to pursue posts or other means deemed to assist, support, or facilitate sex trafficking or prostitution that were created before the law was passed.[214]

Amending the statute provided platforms an incentive to take a closer look at their sites and self-censor to ensure that they were abiding by the law. For example, shortly after the legislation passed through Congress, Craigslist preemptively removed their Personals section,[215] Reddit removed several forums related to sex work,[216] and Tumblr banned "adult content" within months of the passage of FOSTA and just days after the app was removed from Apple's iOS App Store over a child pornography incident.[217]

It can be argued that the Fight Online Sex Trafficking Act ("FOSTA") contains First Amendment concerns that are tougher to navigate than the Section 230 amendment this Article advocates for. Namely, FOSTA is tailored to discriminate against specific content and a specific viewpoint: the promotion of sex trafficking and prostitution.[218] So, it prohibits speech promoting prostitution and sex work views. Conversely, an amendment to protect victims from having their images manipulated into deepfake pornography is not tailored to discriminate against a viewpoint. There could be a counterargument that the amendment proposed concerns a restriction on subject-matter, but the broader subject of deepfakes will not be barred. Further, the Supreme Court has held that content-based restrictions will nevertheless be treated as content neutral if they are designed to prevent the speech's adverse secondary effects, which is the case here.[219]

This amendment could start with targeting nonconsensual deepfake pornography, with the goal to expand to all nonconsensual deepfakes. The amendment should ensure that platforms are no longer shielded unless they take reasonable measures to prevent and remove nonconsensual deepfake pornography, as well as utilize emerging technology. Similar to how companies must take reasonable measures to keep customer information safe, they should take reasonable measures to prevent and remove nonconsensual deepfakes. Additionally, these platforms should hold the creators of these deepfakes accountable, such as a suspension of their account or an outright ban.

---

214. *Id.*

215. *About: FOSTA*, CRAIGSLIST, http://www.craigslist.org/about/FOSTA (last visited Apr. 27, 2021).

216. Annamarya Scaccia, S*ESTA Is Already Having Devastating Impacts on Sex Workers—Just Like They Predicted (Updated)*, REWIRE NEWS (Apr. 2, 2018, 5:33 PM), https://rewirenewsgroup.com/article/2018/04/02/sesta-already-devastating-impacts-sex-workers-just-like-predicted/.

217. Shannon Liao, *Tumblr Will Ban All Adult Content on December 17th*, THE VERGE (Dec. 3, 2018, 12:26 PM), https://www.theverge.com/2018/12/3/18123752/tumblr-adult-content-porn-ban-date-explicit-changes-why-safe-mode.

218. Allow States and Victims to Fight Online Sex Trafficking Act of 2017.

219. *See* City of Renton v. Playtime Theatres, 475 U.S. 41, 49 (1986).

B. New Legislation

Currently, Section 230 shields sites against civil suits and state laws but does not shield from federal criminal charges.[220] So, in theory the federal government could simply pass legislation making nonconsensual deepfake pornography a criminal offense. However, this could prove difficult, as shown in *Ashcroft* and how that holding related to the CPPA.[221]

In 1996, Congress passed the Child Pornography Prevention Act ("CPPA").[222] This Act prohibited actual child pornography, as well as virtual, defining "child pornography" as including a "computer-generated image or picture" that "appears to be" or "conveys the impression that the material is . . . of a minor engaging in sexually explicit conduct."[223] The lawmakers behind this Act believed that by passing this law and cutting off access to child focused pornographic material, even if not actual children, would decrease incidents of child abuse, as well as real child pornography.[224] However, this was contested in *Ashcroft v. Free Speech Coalition*, with the court finding the statute overbroad, explaining that the government cannot prohibit speech simply because the speech "increases the chance an unlawful act will be committed 'at some indefinite future time.'"[225]

Next, the government argued that virtual images must be prohibited in order to eliminate the market for child pornography because they are indistinguishable from real images.[226] The court found this argument "implausible," reasoning that "[i]f virtual images were identical to illegal child pornography, the illegal images would be driven from the market by the indistinguishable substitutes. Few pornographers would risk prosecution by abusing real children if fictional, computerized images would suffice."[227] The court also explained that courts have banned child pornography because there was a "proximate link to the crime from which it came," elaborating that "the CPPA prohibits speech that records no crime and creates no victims by its production."[228]

This holding, however, arose before deepfakes increased in prominence and technology advanced. If this case were decided today, it may produce a different result due to the indistinguishable comparison of deepfakes today from real media, and the fact that they are of real people. And if the coin was flipped, the logic used

---

220. *See* 47 U.S.C. § 230*; see also Frequently Asked Questions*, *supra* note 193.

221. *See* Ashcroft v. Free Speech Coal., 535 U.S. 234, 236 (2002).

222. H.R. 4123, 104th Cong. (1996) (codified as 18 U.S.C.A. 2252A).

223. *Id.*

224. *See Ashcroft*, 535 U.S. at 241–42 (2002); *see also* John Schwartz, *New Law Expanding Legal Definition of Child Pornography Definition of Child Pornography Draws Fir*e, WASH. POST (Oct. 4, 1996), https://www.washingtonpost.com/archive/politics/1996/10/04/new-law-expanding-legal-definition-of-child-pornography-draws-fire/3259a7d5-3349-4b4b-a49a-f2d1ba534019/.

225. *Ashcroft* 535 U.S. at 253 (quoting Hess v. Indiana, 414 U.S. 105, 108 (1973)).

226. *Id.* at 254.

227. *Id.*

228. *Id.* at 236.

by lawmakers in *Ashcroft* could be mirrored concerning nonconsensual deepfake pornography with adult victims as well, amplified by the reality that *real* nonconsensual pornography[229] is also a pressing issue.[230] The court's reasoning in *Ashcroft* simply would not translate for nonconsensual deepfake pornography, because the creator is using someone's actual images, their likeness, to create this pornography. A real victim exists.

Much can be learned from the passage and abolishment of this Act—it is smart to be mindful that solutions must be tailored as to not opine on future harms, but instead focus on present threats, and to stress that there are real victims in the world of nonconsensual deepfake pornography, with lasting impacts and few remedies. If litigation was passed, the most effective solution is a federal statute, as states are constrained by the restrictions imposed by Section 230.[231] And Section 230 aside, if new proposed legislation is contested, courts will have to weigh preventing harm against free speech concerns, while acknowledging a public interest in protecting people's privacy, and chilling speech.

A completely new law could be created, yes, or existing tort law be modified with AI in mind. As described in Part II, there are tort avenues currently available—but as they currently read, they are detrimentally narrow.[232] One solution could be the removal of the intent requirement for the intentional infliction of emotional distress tort, or the benefit element for appropriation of name or likeness, specifically when dealing with pornographic deepfakes. Amending tort law may be perfect timing. The "Defamation and Privacy" section of the Restatement Third of Torts is currently being drafted.[233] However, to avoid a clash with Section 230 as it currently reads, an amendment to the Act would nevertheless have to accompany a new law.

C. Regulation

There is no agency that deepfakes cleanly fall under, though debates tend to be divided between the three options briefly discussed below. Scholars writing in this subject matter have already begun the debate concerning which agency should regulate deepfakes.[234]

The first agency that could potentially regulate deepfakes is The Federal Communications Commission ("FCC"). The FCC regulates broadcasting

---

229. Also known as "revenge porn."

230. Ashcroft v. Free Speech Coal., 535 U.S. 234, 241–42 (2002).

231. 47 U.S.C. § 230.

232. *See supra* Part II.

233. Richard Revesz, *Completing the Restatement Third of Torts*, THE AM. L. INST. (Apr. 4, 2019), https://www.ali.org/news/articles/completing-restatement-third-torts/.

234. Proposed agencies for regulation of deepfakes warrant additional exploration and are beyond the scope of this Article. This topic is subject to thoughtful scholarship by others. *See, e.g.*, Anne P. Gieseke, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 VAND. L. REV. 1479 (2020); *see also* Robert Chesney & Danielle K. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753 (2019).

communications and is the "primary authority" on issues including "communications law, regulation and technological innovation."[235] The FCC has rules in place for false information for television and radio, but the jury is out on whether they can or are willing to extend these regulations to the internet.[236] Notably, the FCC has never said that platforms like Facebook should be included in telecommunications.[237] To muddle the issue further, the passage of the Restoring Internet Freedom Order, which removed net neutrality protections, was the FCC stepping back from regulating internet, suggesting they do not desire to embrace the internet under their umbrella of authority.[238]

The Federal Trade Commission ("FTC") is another option. The FTC protects consumers and competition by preventing "anticompetitive, deceptive, and unfair business practices."[239] Deepfakes certainly can be deceptive, so there is an argument here for the FTC to take these on. Additionally, the FTC hosts an online complaint process where individuals can report "fraud, scams, and bad business practices."[240] The agency lists "identity theft," as well as "computers, the internet and online privacy" as topics able to be reported.[241] Nonconsensual deepfake pornography is arguably a mix of the two topics.

Lastly, this problem is incredibly complex, which could warrant a new agency, specifically tailored to regulate artificial intelligence or for the internet as a whole.

V. CONCLUSION

The threat from deepfakes is not hypothetical. Deepfakes are here to stay, and the impact is global. As illustrated, nonconsensual deepfake pornography can have lasting emotional and psychological harms, including violence, harassment, blackmail, and harm to one's reputation.

The current legal landscape for victims of deepfake videos has not kept up with technology, and what few laws have been passed largely focus on political concerns without consideration for the most prominent deepfake category: nonconsensual deepfake pornography. Legislation must be drafted that balances preventing and punishing serious harms and limiting free expression. This solution must focus on ensuring companies are making the detection of deepfakes a priority, as well as removing harmful manipulated videos from their platform and holding the creators accountable. There must also be an increase in how quickly nonconsensual pornography deepfakes can be detected and removed across

235. *What We Do*, FED. COMMC'NS COMM'N, https://www.fcc.gov/about-fcc/what-we-do (last visited Apr. 27, 2021).

236. *See generally* 47 C.F.R.

237. *Id.*

238. Restoring Internet Freedom, 33 FCC Rcd. 311 (2018).

239. *About the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc (last visited Apr. 27, 2021).

240. *Report Fraud to the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/faq/consumer-protection/submit-consumer-complaint-ftc (last visited Apr. 27, 2021).

241. *Id.*

varying platforms. With how quickly an image can travel, the damage is usually done before sites can remove them.

The only way to combat deepfakes is by bringing together experts in varying fields to create technology to combat this threat, as well as the government utilizing its influence to spread public awareness for deepfakes and pass legislation for nonconsensual deepfake mitigation. Platforms must be aligned regarding methods of removal and prevention and ensuring they are staying informed on technological advancements for deepfake detection. Deepfake technology is continually evolving, so there will be a need for constant updates. Legislators, companies, and experts in various fields necessary for deepfake advisement and prevention must work together to address this issue before the harm and erosion of public trust in images and videos is irreversible.