DON'T ALWAYS BELIEVE WHAT YOU SEE: SHALLOWFAKE AND DEEPFAKE MEDIA HAS ALTERED THE PERCEPTION OF REALITY

*Samuel D. Hodge, Jr.**

We live in a fantasy world, a world of illusion. The great task in life is to find reality.

Iris Murdoch.

I. INTRODUCTION

A picture may be worth a thousand words, but they are valueless if their authenticity must be questioned. This danger is exemplified by the manipulated videos that showed Nancy Pelosi slurring her words and appearing intoxicated[1] or the clip of Barack Obama calling Donald Trump a "total and complete dipshit."[2] These tapes were widely circulated on social media platforms and energized segments of the population who perceived them as real.[3] However, they are examples of

* Samuel D. Hodge, Jr. is a Legal Studies Professor at Temple University, where he teaches law, anatomy, and forensics. He is also a member of the Dispute Resolution Institute, where he serves as a mediator and neutral arbitrator. Professor Hodge has authored more than 140 articles in medical or legal journals and has written ten books. He is also a national public speaker and has participated in over 500 continuing legal education programs.

1. Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 24, 2019), https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media (stating that distorted videos of Nancy Pelosi were "altered to make her sound as if she's drunkenly slurring her words," and how this "video's dissemination highlights the subtle way that viral misinformation could shape public perceptions in the run-up to the 2020 election"); Hannah Denham, *Another Fake Video of Pelosi Goes Viral on Facebook*, WASH. POST (Aug. 3, 2020), https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook (stating that "a manipulated and widely shared video" of Nancy Pelosi that depicted her "slurring her speech and appearing intoxicated was labeled 'partly false' by Facebook").

2. Gabe Worgaftik, *Jordan Peele Makes Obama Call Trump A "Complete Dipshit" in PSA About Fake-Video Technology*, AV CLUB (Apr. 17, 2018, 3:15 PM) https://news.avclub.com/jordan-peele-makes-obama-call-trump-a-complete-dipshit-1825333067 (internal quotation marks omitted).

3. *See id.*

new technology that uses artificial intelligence to create convincing fake videos of third parties.[4] This ability to replicate and distort the words or images of another is not limited to celebrities. If some reprehensible individual has pictures of your face, you can appear as the next star of a pornographic video. If someone has a recording of your voice, your speech can be simulated. You can then be heard calling the office manager and authorizing a money transfer because the employee recognizes your voice. These examples may seem unnerving, but they are today's reality.[5]

These techniques are known as shallowfake[6] and deepfake,[7] and they are an ever-present danger because of their potential to deceive. In a legal proceeding, they pose an appreciable threat to the authenticity of crucial demonstrative evidence in a courtroom. This risk is real because the average person cannot distinguish between a legitimate video or picture, and a manipulated one.[8] This Article will discuss the growing phenomenon of using artificial intelligence ("AI") to create digital content of people voicing things that they never said or participating in events that never took place.[9] Following a discussion of how the technology works, this Article will explore the legal theories employed by an aggrieved party against the creators or distributors of these manipulations, and whether those remedies are sufficient to curtail the problem.[10]

II. THE TECHNOLOGY

Computer technology has become so sophisticated that it is common to think that AI is on the verge of worldwide adoption, creating a torrent of fake images and videos.[11] However, many of the digitally altered depictions are generated manually by selective editing.[12] Shallowfake and deepfake media are not alike, so it is important to

4. *Id.*

5. Sharon D. Nelson et al., *Detecting Deepfakes Deepfake Videos Are Becoming Harder to Identify and May Threaten the 2020 Election*, LAW PRAC., Jan./Feb. 2020, at 42, 44.

6. *See infra* Part II.A.

7. *See infra* Part II.B.

8. Prajakta Pradhan, *AI Deepfakes*, U. ILL. L. REV.: BLOG (Oct. 4, 2020), https://www.illinoislawreview.org/blog/ai-deepfakes.

9. *See infra* Part II.

10. *See infra* Part IV.

11. Kalev Leetaru, *The Real Danger Today Is Shallow Fakes and Selective Editing Not Deep Fakes*, FORBES (Aug. 26, 2019, 1:11 PM) https://www.forbes.com/sites/kalevleetaru/2019/08/26/the-real-danger-today-is-shallow-fakes-and-selective-editing-not-deep-fakes/?sh=485acbc74ea0.

12. *Id.*

understand the differences between the two. This comprehension is the only way a person can understand the problems with AI-manipulated materials.[13]

A. Shallowfakes

Shallowfakes reference videos or pictures that have been "manually altered or selectively revised."[14] The technique gained popularity in early 2018 when video content first appeared on the website Reddit.[15] The footage purported to show celebrities engaged in various sexual acts. In reality, the videos were pornographic films modified by hand to show the heads of famous personalities on the bodies of adult film actors who originally appeared in the footage.[16]

Shallowfakes do not involve deep-learning systems, which make them vastly different from their deepfake counterpart.[17] These reproductions utilize standard editing software and use pre-existing media. The creator will expend a significant amount of time on the alteration and use software that allows the person to generate the falsified content.[18] The video can also be manipulated by purposely slowing down or accelerating the film to depict the subject in a false light. This alteration does not change the content in any manner. Instead, by reframing how the viewer sees the film, the modifications can provide a new meaning to the previously innocent appearing video.[19] This is how Nancy Pelosi appeared intoxicated and slurring her words.[20] Another method of changing a video is by splicing together unaltered snippets of a talk. For instance, a politician's speech can be changed by removing the word "never" from a sentence, thereby wholly altering the meaning of a statement.[21] This selective editing practice is common in campaign advertising and is considered a dirty trick.[22]

---

13. *See* Ashley Stoll, *Shallowfakes and Their Potential for Fake News*, WASH. J.L. TECH. & ARTS (Jan. 13, 2020), https://wjlta.com/2020/01/13/shallowfakes-and-their-potential-for-fake-news.
14. *Id.*
15. *Id.*
16. *Id.*
17. Arnold, *What Is the Difference Between a Deepfake and Shallowfake?*, DEEPFAKENOW, (Apr. 21, 2020) https://deepfakenow.com/what-is-the-difference-between-a-deepfake-and-shallowfake/#:~:text=While%20shallowfakes%20apply%20general%20editing,data%20to%20a%20computer%20program.
18. *Id.*
19. Leetaru, *supra* note 11.
20. *Id.*; *see supra* note 1 and accompanying text.
21. Leetaru, *supra* note 11.
22. *Id.*

B. Deepfakes

"Deepfake" is a combination of the terms "deep learning" and "fake."[23] The word signifies sound or visual media that has been altered or created using deep learning.[24] However, the depiction has been manipulated using AI. The technology can change faces, control facial expressions, and synthesize faces and speech.[25] Deepfakes are still in their infancy, but the know-how is quickly becoming more convincing.[26] Despite the mantra of "fake news," deepfake videos present a substantially different problem.[27] Even if a news account uses fake facts or statements, video evidence enjoys a ring of truth. Regardless of what a person says, the ability to visualize something is uniquely believable.[28]

The images or audio recordings are created by using "generative adversarial networks."[29] This technique uses software that captures facial images and then "maps" them to show how those faces would move based upon certain audio clues.[30] Two machine learning models or algorithms work contemporaneously to produce the content. One device reviews a data set and makes video forgeries, while the other tries to identify the fabrication.[31] The forger persists in creating the imposter video until the other algorithm can't detect the fake.[32]

Most deepfake videos are created on high-end desktops with strong graphics cards or with computing abilities in the cloud.[33] This shortens the production period from days to hours. However, it also takes much expertise, such as touching up the final product to decrease flicker and

---

23. RAINA DAVIS ET AL., DEEPFAKES 2 (Amritha Jayanti ed., 2020), https://www.belfercenter.org/sites/default/files/2020-10/tappfactsheets/Deepfakes.pdf.

24. *Id.*

25. *Deconstructing Deepfakes—How Do They Work and What Are the Risks?*, WATCHBLOG (Oct. 20, 2020), https://blog.gao.gov/2020/10/20/deconstructing-deepfakes-how-do-they-work-and-what-are-the-risks/#:~:text=Deepfakes%20rely%20on%20artificial%20neural,and%20reconstruct%20patterns%E2%80%94usually%20faces.

26. Nicholas O'Donnell, Note, *Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos*, 2021 U. ILL. L. REV. 701, 703 (2021).

27. *Id.*

28. *Id.*

29. Stoll, *supra* note 13.

30. heyjuliesmith, *Shallow Fakes & Deep Fakes: The Next #digcit Frontier*, HEYJULIESSMITH.COM (Sept. 16, 2019), http://heyjuliesmith.com/2019/09/16/shallow-fakes-deep-fakes-the-next-digcit-frontier.

31. Stoll, *supra* note 13.

32. *Id.*

33. Ian Sample, *What Are Deepfakes – and How Can You Spot Them?*, GUARDIAN (Jan. 13, 2020, 5:00 PM), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

other visual flaws.[34] Some firms will create deepfakes for others and use the cloud for processing. Even the mobile phone application, Zao, can be employed by consumers to replace their faces for those from an inventory of television and movie celebrities on which the system has been taught.[35] Recently, a single picture can be used with an iPhone app, called Avatarify, to control a person's face like a puppet. By using the phone's selfie camera, whatever a person does with his or her face will be duplicated on the other.[36] The result is not particularly sophisticated, but it has been downloaded more than six million times in its two months of availability.[37] Another app for the phone, Wombo, transforms a straight-on picture into a humorous lip-synced music video. This app has generated more than 100 million clips in its first two weeks of availability.[38]

Apprehension about the growth of the technology has proven justified as the number of deepfake videos on the Internet doubled between December of 2018 and July of 2019.[39] This number will only increase as more people learn about the technology.

C. Disadvantages

A significant risk with deepfakes is that the manipulation is so exacting that it is almost impossible to differentiate them from real videos. With its growing use, there is increasing apprehension that they will be weaponized to bolster political campaigns and be abused by undemocratic administrations.[40] Several factors exist that increase the risk of their employment in a political setting. These influences include the propensity of individuals to be drawn to scandalous items frequently contained in deepfakes. This human propensity only creates wider audiences and fosters increased dissemination.[41] There is also concern that hysteria over these fraudulent videos could cause individuals to

---

34. *Id.*

35. *Id.*

36. Geoffrey A. Fowler, *Anyone with an iPhone Can Now Make Deepfakes. We Aren't Ready for What Happens Next.*, WASH. POST (Mar. 25, 2021, 8:00 AM), https://www.washingtonpost.com/technology/2021/03/25/deepfake-video-apps.

37. *Id.*

38. *Id.*

39. O'Donnell, *supra* note 26 at 706-07.

40. Deeptesh Sen, *Explained: Why Is It Becoming More Difficult to Detect Deepfake Videos, and What Are the Implications?*, INDIAN EXPRESS, https://indianexpress.com/article/explained/explained-deepfake-video-detection-implications-7247635 (Apr. 3, 2021, 10:25 AM).

41. Shannon Reid, Comment, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, 23 U. PA. J. CONST. L. 209, 211 (2021).

repudiate valid video evidence or overpower people to the stage of "reality apathy."[42] The problem is that this indifference will cause a person to reject all video evidence as untrustworthy, so that they will stand by their prior position or affiliation regardless of what they see.[43] This is similar to the Trump supporters who steadfastly believe that the Presidential election was stolen despite the overwhelming evidence to the contrary.

Another major concern is the creation of nonconsensual pornographic materials. This apprehension is demonstrated by the deepfake pornographic videos of celebrities, such as Scarlett Johansson, Taylor Swift, and Gal Gadot, posted on the Internet.[44] It is estimated that about ninety-six percent of these videos are nonconsensual pornography, usually showing a computer-produced face of a famous individual superimposed over that of the original actor in a sexually explicit scene.[45] These videos will have a disparate effect on women and marginalized groups. These deepfake-sex videos diminish women to genitalia, breasts, and buttocks, establishing a sexual character, not of the person's creation.[46] If the video then appears in an Internet search of the individual's name, it may be impossible to obtain employment or keep a job. It can also cause havoc to the victim's social life and perception of security.[47]

Malevolent manipulators employ these videos to defame people, disseminate falsehoods, affect elections, and polarize citizens.[48] Furthermore, the more challenging it is to discover the deception, the greater the danger it presents to pass off the video as genuine and cause untold difficulties.[49] Whether the video is genuine becomes largely irrelevant. The more important message is that this technology will only make it more challenging to differentiate between what is genuine and what is a sham, a reality that malicious actors will utilize—with possibly destructive results.[50]

---

42. *Id.*
43. *Id.*
44. Sen, *supra* note 40.
45. Pradhan, *supra* note 8.
46. Robert Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 886 (2019).
47. *Id.*
48. Sen, *supra* note 40.
49. *See id.*
50. Rob Toews, *Deepfakes Are Going to Wreak Havoc on Society. We Are Not Prepared.*, FORBES (Mar. 25, 2020, 11:54 PM), https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=6225fd6e7494.

The government has subsidized several projects on how to detect these false narratives. However, scientists remain "vastly overwhelmed by a technology that they fear could herald a damaging new wave of disinformation campaigns[.]"[51] Nevertheless, researchers continue to examine soft biometrics, including how an individual talks and other attributes in the footage to help detect this fake media. These traits are noteworthy because they allow a person to look for these revealing characteristics on their own.[52] Helpful signs for detecting a fake video include unnatural eye movements, abnormal facial expressions, a lack of emotion, unnatural body movement, fake-looking hair or teeth, blurred edges of the video, and hashtag discrepancies.[53]

D. Advantages

Deepfake technology has several beneficial uses. One of its most valuable contributions is to the world of parody or satire.[54] Many content creators have posted these videos on social media platforms, like YouTube, that feature such applications.[55] For instance, the creators of *South Park* posted a satirical video called "Sassy Justice," which was devoted to the growing use of deepfakes.[56] The presentation used generated images of Donald Trump, Mark Zuckerberg, and Julie Andrews.[57] Educators have also employed manipulations to generate digital recreations of dead historical figures. This process permits students and the general public to review content, such as hearing former President John F. Kennedy give the speech he was going to deliver on the day of his assassination. It can also bring the *Mona Lisa* or Albert Einstein to life.[58] Documentary filmmakers can use the technology to narrate a story or to present a particular point of view. For instance, in the documentary *Welcome to Chechnya*, the director used deepfakes to safeguard the names of gay Chechens whose sexual orientation could

51. Reid, *supra* note 41, at 213 (alteration in original) (internal quotation marks omitted).

52. Alison Grace Johansen, *How to Spot Deepfake Videos — 15 Signs to Watch for*, NORTON (Aug. 13, 2020), https://us.norton.com/internetsecurity-emerging-threats-how-to-spot-deepfakes.html.

53. *See id.*

54. 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.

55. *Id.*

56. *Id.*; *see* Sassy Justice, *Sassy Justice with Fred Sassy (Full Episode) | from Trey Parker, Matt Stone, and Peter Serafinowicz*, YOUTUBE (Oct. 26, 2020), https://www.youtube.com/watch?v=9WfZuNceFDM.

57. Sassy Justice, *supra* note 56.

58. Feeney, *supra* note 54.

cause serious consequences in their native country.[59] This application permitted the filmmakers to protect their speakers' identities without showing them in a silhouette form with distorted voices.[60] Cloning of one's voice can even restore the speech of those who lose the ability to talk because of disease.[61]

III. REMEDIES

The threats posed by shallowfake and deepfake creations are real and deeply concerning. However, the solution is not easily discernable.[62] Current and proposed answers endeavor to apply civil and criminal remedies to the creators of these false presentations without changing the immunity provided by Section 230 of the Communication Decency Act ("CDA").[63] Remedial remedies have also emerged in both existing and proposed legislation. Some scholars have suggested that existing civil liability theories should be expanded to inculpate creators of these false narratives. In contrast, others assert that the immunity provided to online intermediaries under the CDA should be abolished.[64] Regardless of the thought process, specific theories of liability can be advanced under common law and by statute.

A. Federal Initiatives

Congress first addressed the issue of these false narratives in the National Defense Authorization Act for Fiscal Year 2020 ("NDAA").[65] The provisions that relate to this technology mandate a comprehensive report on foreign weaponization of deepfakes and requires the government to inform Congress of foreign deepfake disinformation actions affecting U.S. elections.[66] The second requirement involves creating a "Deepfake Prize" competition to support the research or commercialization of detection technologies.[67] As noted by

59. *Id.*

60. *Id.*

61. Sample, *supra* note 33.

62. O'Donnell, *supra* note 26, at 711.

63. *Id.* 47 U.S.C. § 230(c)(1) provides that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." 47 U.S.C. § 230(c)(1) (2018). This section provides online intermediaries that host internet content with immunity. *Id.* § 230(c)(2).

64. O'Donnell, *supra* note 26, at 711-12; *see infra* Parts III.C, IV.A.

65. Matthew F. Ferraro et al., *First Federal Legislation on Deepfakes Signed into Law*, WILMERHALE (Dec. 23, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20191223-first-federal-legislation-on-deepfakes-signed-into-law.

66. *Id.*

67. *Id.*

Representative Jennifer Wexton, "[d]eepfakes pose a serious threat to our national security, and there are significant challenges in our ability to effectively identify this manipulated content."[68]

According to Section 5724 of the NDAA, the Director of National Intelligence is to provide up to five million dollars to one or more winners in a competition to foster "the research, development, or commercialization of technologies to automatically detect machine-manipulated media."[69]

The NDAA also mandates an updated analysis of how foreign governments could employ or are using machine-manipulated media and machine-generated text to damage the national security interests of the United States.[70] This includes an evaluation of the historic, present, or possible future attempts by China and Russia to use deepfake media "to . . . overseas or domestic dissemination of misinformation; the attempted discrediting of political opponents or disfavored populations; and intelligence or influence operations directed against the United States, allies, or partners of the United States, or other jurisdictions believed to be subject to Chinese or Russian interference."[71]

The Identifying Outputs of Generative Adversarial Networks Act is another 2020 piece of legislation involving this manipulative technology.[72] This law directs the National Science Foundation to investigate deepfake usage.[73] It also orders the National Institute of Standards and Technology to support the creation of standards related to the media.[74] Both agencies are further mandated to work with the private sector on deepfake identification abilities.[75]

In 2021, Congress passed another initiative addressing the problem.[76] This law requires the Department of Homeland Security ("DHS") to provide an annual report over the next five years on deepfakes.[77] This account should include all of the ways that the

---

68. Anthony Kimery, *US Defense Bill Requires Comprehensive Deepfake Weaponization, Countermeasures Initiative*, BIOMETRICUPDATE.COM (Dec. 31, 2019), https://www.biometricupdate.com/201912/us-defense-bill-requires-comprehensive-deepfake-weaponization-countermeasures-initiative (internal quotation marks omitted).

69. *Id.* (internal quotation marks omitted).

70. *Id.*

71. *Id.* (alteration in original) (internal quotation marks omitted).

72. Scott Briscoe, *U.S. Laws Address Deepfakes*, SEC. MGMT. (Jan. 12, 2021), https://www.asisonline.org/security-management-magazine/latest-news/today-in-security/2021/january/U-S-Laws-Address-Deepfakes.

73. *Id.*

74. *Id.*

75. *Id.*

76. *Id.*

77. *Id.* President Trump vetoed this law, but Congress overrode that action. *Id.*

technology can injure election campaigns to fraud against specific population segments.[78] The mandate also requires DHS to investigate the technology with a focus on detection and mitigation solutions.[79] This analysis requires the Department of Defense to examine the likelihood of adversaries making deepfake videos of this country's military personnel, or their families, and making policy changes to address the technique.[80] While these efforts are admirable, they do not authorize any type of legal action against the creator or distributor of this false information. Instead, the laws are focused on information gathering.

B. State Initiatives

Most of the initiatives to regulate the technology have been done on the state level. For the most part, these efforts are narrowly crafted.[81] In this regard, Texas, in 2019, became the first state to prohibit political deepfakes.[82] The state amended its election code to criminalize deepfakes generated "with intent to injure a candidate or influence the result of an election."[83] The one caveat is that the video must be "published and distributed within [thirty] days of an election."[84] A violation of the law is a misdemeanor and perpetrators can be sentenced to a year in jail and fined up to $4,000.[85] However, at least one scholar believes that this law will be declared unconstitutional. Jared Schroeder, a journalism professor specializing in freedom of the press matters, noted that the way the courts interpret First Amendment safeguards for free expression severely restricts legislators' options to prevent emerging technologies like deepfakes.[86] In this regard, the Supreme Court has safeguarded intentionally false speech.[87]

Later the same year, Virginia became the first state to impose criminal penalties[88] on the circulation of nonconsensual deepfake

78. *Id.*
79. *Id.*
80. *Id.*
81. *See* Stoll, *supra* note 13.
82. Pradhan, *supra* note 8.
83. Kenneth Artz, *Texas Outlaws 'Deepfakes'—but the Legal System May Not Be Able to Stop Them*, LAW.COM (Oct. 11, 2019, 1:20 PM), https://www.law.com/texaslawyer/2019/10/11/texas-outlaws-deepfakes-but-the-legal-system-may-not-be-able-to-stop-them (internal quotation marks omitted).
84. *Id.* (internal quotation marks omitted).
85. *Id.*
86. Jared Schroeder, *Texas Deepfake Law Unlikely to Survive Scrutiny of the Courts*, SMU, https://blog.smu.edu/opinions/2019/09/25/texas-deepfake-law-unlikely-to-survive-scrutiny-of-the-courts (last visited Oct. 13, 2021).
87. *Id.*
88. *See* VA. CODE ANN. § 18.2-386.2 (2019).

pornography.[89] This prohibition was accomplished by amending the state's law on revenge pornography. The substance of the law provides:

> Any person who, with the intent to coerce, harass, or intimidate, maliciously disseminates or sells any videographic or still image *created by any means whatsoever* that depicts another person who is totally nude, or in a state of undress so as to expose the genitals, pubic area, buttocks, or female breast, where such person knows or has reason to know that he is not licensed or authorized to disseminate or sell such videographic or still image is guilty of a Class 1 misdemeanor.[90]

It is important to note that this measure is narrowly confined to those who use deepfakes to "coerce, harass, or intimidate" another.[91] The penalty for distributing deepfakes of a person without permission is incarceration of up to twelve months and a $2,500 fine.[92]

California has enacted the most sweeping of the remedial legislation. Civil Code Section 1708.86 creates a private cause of action against a person who does either of the following:

(1) Creates and intentionally discloses sexually explicit material and the person knows or reasonably should have known the depicted individual in that material did not consent to its creation or disclosure.
(2) Intentionally discloses sexually explicit material that the person did not create and the person knows the depicted individual in that material did not consent to the creation of the sexually explicit material.[93]

There are several exceptions to this law, such as if the material is disclosed as part of the reporting of unlawful activity; the person reveals the sexually explicit material in the course of reporting unlawful activity; it is revealed while exercising the individual's law enforcement duties; or if the matter is of legitimate public concern, has political or newsworthy value or is protected by the California Constitution or the U.S. Constitution.[94]

---

89. Ferraro et al., *supra* note 65.
90. § 18.2-386.2(A) (emphasis added).
91. Feeney, *supra* note 54.
92. Michael Grothaus, *Virginia Updates Its Revenge Porn Laws to Include Deepfakes*, FAST CO. (July 2, 2019), https://www.fastcompany.com/90372079/virginia-updates-its-revenge-porn-laws-to-include-deepfakes.
93. CAL. CIV. CODE § 1708.86(b)(1)–(2) (West 2019).
94. *Id.* § 1708.86(c)(1).

The statute permits the recovery of no less than $1,500, but not more than $30,000.[95] However, if the act was done with malice, the amount of damages may be increased to $150,000.[96] Under the appropriate circumstances, punitive damages and reasonable attorney's fees may be obtained.[97]

California's second initiative deals with using deepfake images or videos in connection with political campaigns. This measure makes it illegal for any entity to maliciously distribute or create "materially deceptive" media pertaining to a political candidate within sixty days of an election.[98] As noted by the legislator who introduced the bill, "[d]eepfakes are a powerful and dangerous new technology that can be weaponized to sow misinformation and discord among an already hyper-partisan electorate."[99]

New York took a different approach. In 2020, the state enacted legislation that addresses synthetic or digitally manipulated media.[100] The law has two main components. It creates a postmortem right of publicity to safeguard performers' likenesses, including digitally altered images, from unapproved commercial use for forty years after death.[101] This includes a safeguard pertaining only to professional artists and performers who were domiciled in New York at the time of their death and whose digital replica is being used in a scripted audiovisual work or for the live performance of a musical work.[102] The law also prohibits nonconsensual, computer-generated pornography.[103] An aggrieved person can sue any individual who distributes or publishes the sexually explicit material and knows, or should have known, that the person shown in the media did not consent to its creation or publication.[104] However, the sexually explicit materials shall not be deemed

95. *Id.* § 1708.86(e)(1)(B)(ii)(I).

96. *Id.* § 1708.86(e)(1)(B)(ii)(II).

97. *Id.* § 1708.86(e)(1)(C)–(D).

98. State Assemb. 730, 2019 Leg. (Cal. 2019).

99. Kari Paul, *California Makes 'Deepfake' Videos Illegal, but Law May Be Hard to Enforce*, GUARDIAN (Oct. 7, 2019, 6:42 PM), https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce (internal quotation marks omitted).

100. S.B. 5959D, 2019-20 Legis., Reg. Sess. (N.Y. 2019).

101. Matthew F. Ferraro & Louis W. Tompros, *New York's Right to Publicity and Deepfakes Law Breaks New Ground*, WILMERHALE (Dec. 17, 2020), https://www.wilmerhale.com/en/insights/client-alerts/20201217-new-yorks-right-to-publicity-and-deepfakes-law-breaks-new-ground.

102. *Id.*

103. *Id.*

104. Andrea L. Calvaruso & Taraneh J. Marciano, *New Year Brings Expanded Protections for Publicity and Privacy Rights Under New York Law*, INTELL. PROP. & TECH. L.J., Feb. 2021, at 12, 13.

newsworthy merely because they involve a public figure.[105] Remedies include injunctive relief, compensatory and punitive damages, and counsel fees.[106]

Following these initiatives, the momentum to pass legislation regulating deepfakes seems to have lost momentum. A legislative search on Westlaw of the term "deepfake" failed to find any new statutes on the issue.[107] An internet search did disclose proposed legislation in Maryland, Massachusetts, Pennsylvania, and Florida.[108] For instance, the Florida Senate, in 2021, introduced a bill that would make it illegal to manipulate a candidate's likeness or message for political purposes.[109] Likewise, the Pennsylvania House introduced a bill in 2021 that prohibits anyone from disseminating a deepfake with the intent to harass, annoy, or alarm a candidate for political office.[110]

The future of new legislative initiatives remains to be seen. Critics have attacked these remedial measures as "overbroad, uninformed, and, in their attempt to regulate one problem, actually trample on the protected rights of Americans."[111] It is also believed that these statutes will face First Amendment challenges.[112] For instance, some of the laws do not provide exemptions for satire or parody in a political context. It is also claimed that the legislation has the potential of repressing helpful speech, and the efforts are duplicative because broadcasters and others already reject political advertising and comparable content to mitigate their potential liability.[113]

The First Amendment continues to create an ever-present roadblock. For example, in *United States v. Alvarez*,[114] a divided Supreme Court found that the First Amendment bars the government from regulating speech simply because it is a lie.[115] Alvarez was a retired

105. *Id.*

106. *Id.*

107. Westlaw legislative search of "deepfake," WESTLAW EDGE, http://1.next.westlaw.com (follow "search" hyperlink; then sign-in to Westlaw; select "Content Types;" then select "Proposed & Enacted Legislation;" search "deepfake") (last visited Oct. 13, 2021).

108. Google search of "deepfake," GOOGLE, http://www.google.com (follow "search" hyperlink; search "deepfake") (last visited Oct. 13, 2021).

109. *See* S.B. 658, 2021 Legis., Reg. Sess. (Fla. 2021).

110. *See* H.B. 1942, 2021 Gen. Assemb., Reg. Sess. (Pa. 2021). This bill was referred to the Committee on the Judiciary on September 30, 2021, but no vote has been taken on the proposed legislation. *See id.*

111. David Ruiz, *Deepfakes Laws and Proposals Flood US*, MALWAREBYTES LABS (Jan. 23, 2020) https://blog.malwarebytes.com/artificial-intelligence/2020/01/deepfakes-laws-and-proposals-flood-us.

112. Pradhan, *supra* note 8.

113. Feeney, *supra* note 54.

114. 567 U.S. 709 (2012).

115. *See id.* at 712-30.

marine who falsely represented that he had been awarded the Congressional Medal of Honor.[116] He was charged with violating the Stolen Valor Act for lying about the award.[117] The defendant challenged the statute as a "content-based suppression of pure speech."[118] In siding with the defendant, the Court opined that the constitutional guarantee means that the government has no power to limit speech because of its message, ideas, or subject matter.[119] As noted:

> Permitting the government to decree this speech to be a criminal offense . . . would endorse government authority to compile a list of subjects about which false statements are punishable. That governmental power has no clear limiting principle. . . . Were this law to be sustained, there could be an endless list of subjects the National Government or the States could single out.[120]

The Court stated that if it ruled that the interest in truthful dialog alone is adequate to uphold a ban on speech, it would offer the government a broad censorial authority unparalleled in its cases or the country's constitutional tradition.[121] The mere possibility of implementing that power throws a chill against the First Amendment, which the Court cannot permit if free speech and discourse are to persist as a foundation of our freedoms.[122] This broad protective language suggests that these remedial statutes may not survive a constitutional challenge.

C. Social Media

The most immediate way to stop deepfake technology could come from social media platforms, like Facebook, Google, and Twitter. These tech giants can take immediate action to restrict the distribution of this harmful media.[123] For instance, Facebook has already removed deepfake and other altered media from its platform.[124] However, these online vendors currently do not have the incentive to become actively involved in this controversy since they have immunity. Section 230 of the CDA provides online platforms with immunity for content posted on their

116. *Id.* at 713-14.
117. *Id.*
118. *Id.* at 716.
119. *Id.*
120. *Id.* at 723.
121. *Id.*
122. *Id.*
123. Pradhan, *supra* note 8.
124. *Id.*

sites.[125] To be precise, the statute provides: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[126] This protection is afforded not only to traditional Internet service providers, but to other "interactive computer service providers," such as any online service that publishes third-party content.[127] This section also extends its legal protections to bloggers who act as intermediaries by presenting comments on their platforms. Therefore, bloggers are not responsible for remarks left by readers, the postings of guest bloggers, or tips sent via email. This immunity applies even if a blogger knows about the objectionable content or exercises editorial judgment.[128]

This blanket immunity has been criticized by some scholars who maintain that the Court should remove these legal protections for bad actors.[129] Perhaps the immunity provided by the CDA should even be amended to premise it upon reasonable moderation practices rather than the blanket immunity that currently exists.[130]

IV. THEORIES OF LIABILITY

Because the technology is relatively new, there is little court guidance on the topic. Very few cases even mention the terms shallowfake and deepfake. A Westlaw search reveals a mere five cases which mention "deepfake" and no cases in which the word "shallowfake" arises.[131] Therefore, one is left to predict the remedies a person aggrieved by this technology may pursue. The most logical theories would include claims for defamation, infliction of emotional distress, placing a person in a false light, cyberstalking, cyberbullying, a copyright violation, and "revenge porn," or nonconsensual pornography statutes.

Regardless of the harm caused by this manipulated media and the reprehensible nature of the actions, shallowfake and deepfake victims face unique legal challenges. For instance, a person whose face has been superimposed on the body of another will be confronted with whether

125. 47 U.S.C. § 230(c)(1) (2018).

126. *Id.*

127. *Section 230 of the Communications Decency Act*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230 (last visited Oct. 13, 2021).

128. *Id.*

129. O'Donnell, *supra* note 26, at 713-14.

130. Chesney & Citron, *supra* note 46, at 890.

131. Westlaw case search of "deepfake," WESTLAW EDGE, http://1.next.westlaw.com (follow "search" hyperlink; then sign-in to Westlaw; select "Content Types;" then "Cases;" select "All State" and "All Federal;" search "deepfake;" search "shallowfake") (last visited on Oct. 13, 2021).

they can pursue compensation for exposure of personal details that do not show their intimate body parts.[132] Equally, the individual whose body is depicted in a manipulated image may have difficulty demonstrating that their form is sufficiently recognizable to qualify as an identifiable misrepresentation.[133] Assuming that these hurdles can be overcome, both the person whose face is used and the original actor in the video are victims. Therefore, they must each establish the harm caused by the manipulated media. This is challenging because the depictions are not revealing the intimate particulars of any one victim.[134]

A. Civil Liability

The purpose of the law of torts is to offer relief to an aggrieved person for the harm caused by another, impose liability on those responsible for the offending conduct, and discourage others from perpetrating these harmful acts.[135] Several theories can be advanced in an attempt to recover damages for the creation or distribution of manipulated media. Most often, these remedies are dictated by state and not federal law.[136]

The most logical causes of action for this false exposure would be for false light publicity under the tort of invasion of privacy and infliction of emotional distress.[137] This is because a tortfeasor who desires to control, expose, and damage the identity of another regularly does so by invading their sexual privacy.[138] This form of confidentiality is at the top of society's privacy values because of its significance to sexual agency, intimacy, and equality.[139] Without this confidentiality, individuals would have trouble forming intimate associations. These relationships of love and caring occur through a progression of mutual self-disclosure and vulnerability.[140] Partners disclose their innermost secrets to one another with the anticipation that they will safeguard each

132. Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887, 902 (2019).

133. *Id.*

134. *Id.*

135. *Tort*, LEGAL INFO. INST., https://www.law.cornell.edu/wex/tort#:~:text=The%20primary%20aims%20of%20tort,others%20from%20committing%20harmful%20acts.&text=Typically%2C%20a%20party%20seeking%20redress,the%20form%20of%20monetary%20compensation (last visited Oct. 13, 2021).

136. *See* Reid, *supra* note 41, at 214.

137. *Id.* at 215.

138. Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1870 (2019).

139. *Id.* at 1874.

140. *Id.* at 1875.

other's intimate disclosures. When that confidence is violated, it can be challenging to trust others in the future.[141]

Deepfake sex videos are not the same as the nonconsensual publication of intimate images because the video does not show a victim's actual naked form. While these manipulations do not show the actual genitals, breasts, or buttocks, they appropriate the victim's sexual and intimate identities.[142] In turn, these films generate a sexual personality, not of the person's creation. These depictions are an insult to the idea that an individual's intimate characteristics are their own to disclose or keep confidential.[143]

Considering the likelihood of abuse, and the fundamental harm to the creation and distribution of digitally manipulated media, there is minimal precedent on the issue.[144] One possible explanation for the death of litigation is the Supreme Court's position on civil suits involving free speech. Landmark cases, such as *New York Times Co. v. Sullivan*[145] and *Hustler Magazine, Inc. v. Falwell*,[146] have created high benchmarks to satisfy a tort claim against another's speech.[147] Fortunately, this precedent requiring a showing of reckless disregard for the truth and actual malice is limited to public figures. There is a more liberal rule for "private individuals," as set forth in *Gertz v. Robert Welch, Inc.*[148] and *Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc*.[149] These decisions involved claims by private citizens against other private entities, which did not implicate matters of public concern. Therefore, the claims have increased viability.[150] As the *Gertz* Court noted, a publisher of a defamatory statement about a person who is neither a public official nor a public figure is unable to claim the protections afforded by *Sullivan*. Because a private person has not willingly subjected themselves to increased risk of harm from defamatory falsehoods, they are more worthy of recovery.[151]

---

141. *Id.*
142. *Id.* at 1921.
143. *Id.*
144. Michael Scott Henderson, Note, *Applying Tort Law to Fabricated Digital Content*, 2018 UTAH L. REV. 1145, 1152 (2018).
145. 376 U.S. 254 (1964).
146. 485 U.S. 46 (1988).
147. Henderson, *supra* note 144, at 1152.
148. 418 U.S. 323 (1974).
149. 472 U.S. 749 (1985).
150. Henderson, *supra* note 144, at 1153.
151. *See Gertz*, 418 U.S. at 339-48.

1. False Light

Generally, an action for an invasion of privacy's false light may be instituted against a person who is responsible, either by writing or speaking, for a false account that positioned the claimant in an untrue light and the publisher of the representation is someone other than the writer or speaker.[152] It is also permissible to make a false light claim against those who were complicit in creating or communicating the false representation.[153] According to the *Restatement (Second) of Torts*, liability will attach if the victim was placed in a position that would be highly offensive to a reasonable person, and the perpetrator had awareness of or acted in reckless disregard as to the falsity of the publicized material and false light in which the victim was presented.[154] The cause of action is viable only if there is a major misrepresentation of the claimant's character, history, activities, or beliefs and that much offense may be taken by a reasonable person in her position.[155] It would certainly seem logical that placing someone else's face on a different person's body engaging in a sexual act would be the type of highly offensive conduct contemplated by this tort.[156]

False light is comparable to defamation, but the jurisdictions that recognize this cause of action acknowledge that the torts are different. However, they also overlap to some extent.[157] Defamation provides compensation for damage to a person's reputation, while false light offers money damages for being subject to offensiveness.[158] Also, only about two-thirds of the states recognize the tort of false light.[159]

2. Intentional Infliction of Emotional Distress

Intentional infliction of emotional distress customarily requires some type of conduct that is so extreme and outrageous that it causes

---

152. Richard E. Kaye, *Cause of Action for False Light Invasion of Privacy*, *in* 33 CAUSES OF ACTION SECOND SERIES 1, at § 3 (2007) (Thomson Reuters) (database updated July 2021).

153. *Id.*

154. RESTATEMENT (SECOND) OF TORTS § 652(E) (AM. L. INST. 1977).

155. Kaye, *supra* note 152, at § 4.

156. David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELEC. FRONTIER FOUND. (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them#:~:text=February%2013%2C%202018-,We%20Don%E2%80%99t%20Need%20New%20Laws%20for,Videos%2C%20We%20Already%20Have%20Them&text=Video%20editing%20technology%20hit%20a%20milestone%20this%20month.&text=As%20Samantha%20Cole%20at%20Motherboard,(non%2Dpornography)%20actors.

157. *False Light*, DIGIT. MEDIA L. PROJECT (Jan. 22, 2021), https://www.dmlp.org/legal-guide/false-light.

158. Greene, *supra* note 156.

159. *Id.*

severe emotional trauma to another.[160] Behavior "is 'extreme and outrageous' only if it is 'so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious and utterly intolerable in a civilized community.'"[161]

Scholars have opined that creating and distributing pornographic deepfakes would fulfill this high burden of proof because they fall "outside the norms of decency."[162] A victim would feel great humiliation and anguish from such a depiction of themselves. However, when the false portrayal falls within the ambit of subjects of public concern, the claimant will have to overcome the requirement of actual malice mandated to be successful in an infliction claim.[163] Actual malice is not the same as malicious intent. Instead, it is knowing that a matter is untrue or proceeding with reckless disregard for the matter's falsity.[164]

In *Ault v. Hustler Magazine, Inc.*,[165] a woman who was an anti-pornography lobbyist was interviewed by a newspaper about her stance on the subject.[166] The defendant then ran an article featuring Ms. Ault as the "Asshole of the Month."[167] The story portrayed the plaintiff "as a 'tightassed housewife,' 'frustrated,' . . . 'crackpot,' and a 'deluded busybody.'"[168] The account also contained a photograph of the plaintiff superimposed over the buttocks area of a bent-over naked man.[169] The woman sued the magazine for a variety of torts, including intentional infliction of emotional distress. The magazine countered by filing a motion to dismiss for the failure to state a cause of action, which the court granted.[170]

This decision was sustained on appeal.[171] The court felt that the threshold inquiry is whether the article deals with an opinion or a factual

160. FindLaw, *Intentional Infliction of Emotional Distress*, FINDLAW, https://www.findlaw.com/injury/torts-and-personal-injuries/intentional-infliction-of-emotional-distress.html (Aug. 8, 2018).

161. Doe v. Brandeis Univ., 177 F. Supp. 3d 561, 617 (D. Mass. 2016) (quoting Foley v. Polaroid Corp., 508 N.E.2d 72, 82 (Mass. 1987)).

162. Kareem Gibson, Note, *Deepfakes and Involuntary Pornography: Can Our Current Legal Framework Address This Technology?*, 66 WAYNE L. REV. 259, 279 (2020).

163. *Id.* at 279-80.

164. *Id.* at 280.

165. 860 F.2d 877 (9th Cir. 1988).

166. *Id.* at 879.

167. *Id.*

168. *Id.*

169. *Id.*

170. *Id.* at 880.

171. *Id.* at 884.

statement.[172] This distinction is important because the First Amendment protects an opinion.[173] The court referenced *Hustler Magazine v. Falwell*,[174] where it was noted:

> [P]ublic figures and public officials may not recover for the tort of intentional infliction of emotional distress by reason of publication . . . without showing . . . that the publication contains a false statement of fact which was made with 'actual malice,' [for instance,] with knowledge that the statement was false or with reckless disregard as to whether or not it was true.[175]

Falwell acts as a bar for an emotional distress claim against a public figure or private person under the opinion provided.[176] Therefore, the plaintiff's claim in the *Ault* matter was precluded regardless of her status because of the First Amendment.[177] The *Ault* case involves pornography, a heated and spirited debate, of which the magazine article is a part. Therefore, epithets, fierce rhetoric, and exaggeration are foreseeable.[178] The derogatory article about the plaintiff is constitutionally protected opinion, which bars her claim.[179]

Hustler's action in superimposing the plaintiff's face on a nude man's body is an example of a shallowfake image. Because the court felt that a debate about pornography was a matter of public interest, it barred the claim despite the offensive nature of the comments and manipulated image. This case demonstrates the hurdles that a plaintiff must overcome when the actions are a matter of political, social, or other community concern.[180]

3. Defamation

The law has long acknowledged the worth of a person's reputation.[181] Therefore, this tort would appear to be a reasonable cause of action to combat shallowfake and deepfake media. This type of claim requires the publication of a false narrative that injures a person's

172. *Id.* at 880.
173. *Id.*
174. 485 U.S. 46 (1988).
175. *Ault*, 860 F.2d at 880 (quoting *Falwell*, 485 U.S. at 56).
176. *See Falwell*, 485 U.S. at 56.
177. *Id.*
178. *Id.* at 881.
179. *Id.*
180. Gibson, *supra* note 162, at 280-81.
181. Henderson, *supra* note 144, at 1157.

reputation.[182] As set forth in the *Restatement (Second) of Torts*, liability will be created when there is:

> (a) [A] false and defamatory statement concerning another; (b) an unprivileged publication to a third party; (c) fault amounting at least to negligence on the part of the publisher [with respect to the act of publication]; and (d) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.[183]

After all, the creator of the manipulated or edited material has passed off a fake video or picture as though it is real to the detriment of the person visualized.[184]

There is a lack of judicial precedent about such a claim resulting from an altered video. However, in some states, defamation explicitly applies to altered pictures.[185] For instance, in *Kiesau v. Bantz*,[186] the plaintiff was a law enforcement official, and a fellow officer altered a photograph that made it appear that she was standing in front of her police cruiser with her breasts exposed.[187] In concluding that the picture was libelous per se, the court noted that defamation diminishes a personal interest. It belittles the view which others maintain about a person and invades the claimant's interest in her reputation and good name.[188] The tort is premised upon the communication of derogatory words, not any physical or emotional harm to the plaintiff which may arise.[189] The defendant maintained that the altered image was not libelous per se and could be reasonably appreciated as a parody.[190] The court disagreed and noted that the altered picture was clear and precise in its presentation. The altered image was also not made in any political context.[191] Showing the plaintiff in uniform with her breasts exposed could be reasonably understood to attack her integrity and moral

182. Gibson, *supra* note 162, at 268, 271.

183. RESTATEMENT (SECOND) OF TORTS § 558 (AM. L. INST. 1977).

184. *See* Stoll, *supra* note 13.

185. Erik Gerstner, *Face/Off: "DeepFake" Face Swaps and Privacy Laws*, DEF. COUNS. J., Jan. 2020, at 1, 5, https://www.iadclaw.org/assets/1/17/Face_Off_-_DeepFake_Face_Swaps_and_Privacy_Laws.pdf?4179.

186. 686 N.W.2d 164 (Iowa 2004).

187. *Id.* at 170.

188. *Id.* at 175.

189. *Id.*

190. *Id.* at 176.

191. *Id.* at 177-78.

character.[192] Therefore, substantial evidence was presented to show that the altered photograph was libelous per se.[193]

Videos should be viewed in the same light as still images under the law. Therefore, a deepfake video should be considered defamatory.[194] The primary defense that a content creator will assert is one of a parody, and the result will be determined on a case-by-case basis.[195] Different fact-finders may reach significantly different conclusions based on the facts of each case.[196]

An interesting question arises when the creator of the scandalous fake video labels the film as fake. While the video creator may be able to overcome a defamation claim initially, the outcome is less certain if those who subsequently post the deepfake remove the fake label.[197] It seems fair to find the originator partially liable for the subsequent postings. The Internet makes the widespread distribution of materials very easy, and this is a foreseeable event.[198] An individual who disseminates a deepfake for sensationalism, or otherwise, probably understands that the creation will be reposted many times. Even though the film is "out of the creator's control," it is certainly foreseeable that the deepfake will achieve widespread circulation and that the "fake" label will be removed along the way.[199]

B. Criminal Liability

Civil liability has several imposing hurdles, as the creator of manipulated media may be judgment-proof, making an award of damages a hollow victory.[200] Those whose bodies have been presented in the video may not desire to have their names disseminated or known.[201] There is also no assurance that the manipulated media will be removed from the Internet, especially because of the immunity provided to the social media providers under the CDA.[202] A criminal remedy does not have these limitations, and criminalizing the misconduct may be the best deterrent.[203] Making these actions illegal also sends a clear message that

192. *Id.* at 178.
193. *Id.*
194. Gerstner, *supra* note 185, at 5.
195. *Id.*
196. *Id.*
197. Gibson, *supra* note 162, at 272.
198. *Id.* at 273.
199. *Id.* at 272-73.
200. Delfino, *supra* note 132, at 902.
201. *Id.*
202. *Id.*
203. *Id.*

society finds this behavior reprehensible. It lets the creators and distributors of the manipulated media know that their actions are not considered trivial and that it is harmful to those depicted in the images and offensive to society.[204] Considering that this technology disproportionately affects females, a societal action in the form of criminal punishment is in order.[205]

Various criminal offenses can be pursued to punish a wrongdoer who uses or distributes deepfake or shallowfake media with criminal intent. For instance, if the perpetrator employs the manipulated video to pressure a victim to pay money to withhold or destroy the material, extortion laws are a logical charge.[206] Likewise, if the materials are used to badger the person, the harassment laws would apply.[207] Several other offenses also come into play, such as revenge porn, cyberstalking, and cyberbullying.[208]

1. Federal Law

Stalking, threats, and harassment are traditionally within the province of local law enforcement officials.[209] However, with the expanded employment of technology and the multi-jurisdictional nature of many of these offenses, the federal government can offer supplementary resources to prosecute matters that may exceed the abilities of local law enforcement officials.[210] Undeniably, technology has eliminated conventional borders and made more difficult matters that would otherwise seem straightforward.[211] While no explicit federal statute criminalizes deepfake videos, the federal government may utilize the laws dealing with cyber exploitation to prosecute those who create and disseminate these fake materials.[212] These theories include cyberbullying, cyberthreats, cyberharassment, cyberstalking, sextortion, and nonconsensual pornography.[213]

---

204. *Id.* at 903.
205. *Id.*
206. Greene, *supra* note 156.
207. *Id.*
208. *See* Reid, *supra* note 41, at 224-28.
209. Monty Wilkinson, *Introduction*, U.S. ATT'YS' BULL. (U.S. Att'ys, Columbia, S.C.), May 2016, at 1, https://www.justice.gov/usao/file/851856/download.
210. *Id.*
211. *Id.*
212. Delfino, *supra* note 132, at 904.
213. *See* Joey L. Blanch & Wesley L. Hsu, *An Introduction to Violent Crime on the Internet*, U.S. ATT'YS' BULL. (U.S. Att'ys, Columbia, S.C.), May 2016, at 2, 3-7, https://www.justice.gov/usao/file/851856/download.

The law on interstate communications provides in part that it is a crime to convey "in interstate or foreign commerce any . . . threat to injure the person of another . . . ."[214] If one assumes that posting a deepfake or shallowfake constitutes a "communication" under this provision, and the other requirements of 18 U.S.C. § 875(c) are satisfied, the federal government may be able to charge the creator of the media with violating this provision.[215]

The cyberstalking laws[216] can also be utilized to pursue this offending conduct. The Department of Justice's Office of Victims of Crimes notes that stalking consists of "repeated and unwanted attention, harassment, contact, or any other course of conduct directed at a specific person that would cause a reasonable person to feel fear."[217] The term "cyberstalking" is both an informal word that can include a wide array of conduct and a legal term of art. It is generally appreciated to mean stalking that happens online and may be compatible with other terms covered by this article, such as cyberharassment, cyber threats, or revenge porn.[218] This conduct would include harassing a person through the Internet and posting information or spreading rumors about an individual on social media sites.[219] Therefore, if posting revenge media is linked with, or rises, to the level of cyberstalking, then a person may be pursued by the federal government.[220]

In *United States v. Cardozo*,[221] the court rejected several attacks to the cyberstalking statute under the First Amendment.[222] As the court noted:

> [W]e must read 'intent to . . . harass,' as referring to criminal harassment which is unprotected because it constitutes true threats or speech that is integral to proscribable criminal conduct. We think that this logic would also apply to the term 'intimidate' in the current version of the statute. Indeed, 'interpreting the statute to avoid a serious constitutional threat,' points to reading the statute as referring to '[i]ntimidation in the constitutionally proscribable sense of the word[, which] is a type of true threat.'[223]

---

214. 18 U.S.C. § 875(c) (2018).
215. Delfino, *supra* note 132, at 904-05.
216. *See* 18 U.S.C. § 2261A (2018).
217. Blanch & Hsu, *supra* note 213, at 4.
218. *Id.* at 5.
219. *Id.* at 4-5.
220. Delfino, *supra* note 132, at 905.
221. No. 1:18-CR-10251-ADB, 2019 WL 2603096 (D. Mass. June 24, 2019).
222. *See id.* at *3-5.
223. *Id.* at *3 (alterations in original) (quoting United States v. Ackell, 907 F.3d 67, 76 (1st Cir. 2018)).

This type of language would suggest that the court has opened the door for the cyberstalking statute to be applied to things such as revenge pornography.[224]

Another criminal option is cyberharassment. This crime differs from cyberstalking in that it is commonly defined as not pertaining to a credible threat.[225] Rather, it applies to menacing or annoying emails, instant messages, blog posts, or websites devoted to tormenting a victim.[226] Federal law makes no general reference to "cyberharassment," but an action that constitutes this offense might nevertheless be pursued under other statutes, contingent upon the specific facts.[227]

Sextortion happens when a perpetrator demands that the victim provide the offender with pictures of a sexual nature, sexual favors, or other matters of value. These requests also contain threats to harm or embarrass the victim if she fails to comply.[228] For instance, one can envisage a perpetrator threatening to create or distribute a pornographic deepfake video unless the victim supplies nude pictures of herself.

The National Center for Missing & Exploited Children has witnessed a mounting number of these kinds of cases, a comparatively novel form of online child sexual exploitation.[229] Sextortion occurs when non-physical forms of coercion are employed, such as blackmail, to obtain sexual content from children, including pictures and videos, extort money, or engage in sex with a child.[230] Threats are made to harm the child or their family, or to make sexual content of the child using digital-editing tools if the victim will not provide sexual images.[231]

2. State Laws

No state has yet to make deepfake misuse criminal.[232] However, fairly new nonconsensual pornography laws may be the most helpful way to prevent distribution of deepfake pornographic videos of nonconsenting victims.[233] Also known as revenge porn, this kind of online harassment takes place when an ex-partner or hacker publishes

---

224. *See* Delfino, *supra* note 132, at 905.
225. Blanch & Hsu, *supra* note 213, at 5.
226. *Id.*
227. *Id.*
228. *Id.* at 6.
229. John F. Clark, *Growing Threat: Sextortion*, U.S. ATT'YS' BULL. (U.S. Att'ys, Columbia, S.C.), May 2016, at 41, 42, https://www.justice.gov/usao/file/851856/download.
230. *Id.*
231. *Id.* at 43.
232. Delfino, *supra* note 132, at 909.
233. Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 119 (2019).

sexually explicit images of an individual online without their permission.[234] State law penalties vary from making a revenge pornography case a felony, to punishing the same offense as a misdemeanor, to not having any criminal charges for the offending conduct.[235]

As of 2021, forty-six states and the District of Columbia have specific laws prohibiting the distribution of revenge porn.[236] However, these laws are still in their infancy, and the statutes are continuing to evolve.[237] In most states, the crime requires the distributor to send out pictures or videos deemed sexual, such as depicting the victim's intimate body parts or performing a sexual act.[238] Merely posting an uncomplimentary image of an offender's ex in a bathing suit is not pornographic, in the absence of any other circumstances, such as the victim's breasts being observable.[239]

The exact wording of the laws vary by jurisdiction, with many states focusing on matters where former sexual partners post sexually explicit media to cause distress or embarrassment.[240] These statutes use phrases that are both relevant and inapplicable to personal deepfakes. For example, Maryland's law provides:

> A person may not intentionally cause serious emotional distress to another by intentionally placing on the Internet a photograph, film, videotape, recording, or any other reproduction of the image of the other person that reveals the identity of the other person with his or her intimate parts exposed or while engaged in an act of sexual contact: (1) knowing that the other person did not consent to the placement of the image on the Internet; and (2) under circumstances in which the other person had a reasonable expectation that the image would be kept private.[241]

The laws of twenty-four jurisdictions have a culpability mandate that the offender must have the intent to cause harm to the victim by

---

234. FindLaw, *State Revenge Porn Laws*, FINDLAW, https://www.findlaw.com/criminal/criminal-charges/revenge-porn-laws-by-state.html (Jan. 13, 2020).

235. Delfino, *supra* note 132, at 909.

236. *Nonconsensual Pornography (Revenge Porn) Laws in the United States*, BALLOTPEDIA, https://ballotpedia.org/Nonconsensual_pornography_(revenge_porn)_laws_in_the_United_States (last visited Oct. 13, 2021).

237. FindLaw, *supra* note 234.

238. *Id.*

239. *Id.*

240. Harris, *supra* note 233, at 120.

241. MD. CODE ANN., CRIM. LAW § 3-809 (2018).

posting or distributing the sexually explicit material.[242] In addition, sixteen of these states require the intent to "harass" while others employ like language, such as the objective to "intimidate," cause "emotional distress," or damage the victim's "health, safety, business, calling, career, financial condition, reputation, or personal relationships[.]"[243]

Despite these efforts, existing criminal laws are largely inadequate to penalize the makers and distributors, and to remedy the trauma suffered by the victims.[244] These shortcomings continue in the face of deepfake pornographic videos, which correspondingly do not fit within existing criminal law definitions, even those that make revenge pornography illegal.[245]

C. Copyright Infringement

A copyright is one of the various groups of intellectual property protections intended to safeguard the creator's or holder's sole ability to claim an original work as their own when the effort is fixed in a tangible medium.[246] As soon as the creation is reduced to a written form, recorded digitally, or typed electronically, the work is afforded copyright protection, for a specified time period.[247] Photographs are protected materials and include images created with a camera and captured in a digital file or other visual media, such as film. These protected images include color photos, black and white images, and similar types of pictures.[248]

A copyright infringement generally occurs when someone uses an individual's original creative or copyrighted work without permission.[249] Therefore, a person's image of their face or body enjoys copyright protection. Logically, if another takes that image without permission and uses it in a deepfake video, a copyright infringement has occurred. However, it is not that simple because of the fair use doctrine.

Fair use is the reproduction of a copyrighted work for comment, teaching, scholarship, criticism, or research.[250] This doctrine is the byproduct of multiple court decisions codified in Title 17 of the United

---

242. Harris, *supra* note 233, at 121.

243. *Id.* (alteration in original) (internal quotation marks omitted).

244. Delfino, *supra* note 132, at 918.

245. *Id.*

246. Jonathan Layton, *How to Avoid Copyright Infringement*, LEGALZOOM https://www.legalzoom.com/articles/how-to-avoid-copyright-infringement (Mar. 5, 2021).

247. *Id.*

248. *Photographs*, COPYRIGHT.GOV, https://www.copyright.gov/registration/photographs (last visited Oct. 13, 2021).

249. Layton, *supra* note 246.

250. *What Is Fair Use and What About Parodies?*, COTMAN IP, https://www.cotmanip.com/articles/fair-use-parody (last visited Oct. 13, 2021).

States Code within the Copyright Act.[251] The statute sets forth four elements in deciding whether something qualifies as fair use:

> (1) [T]he purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.[252]

The creator or distributor of shallowfake or deepfake media will assert that the fake work is a parody and exempt from the law. As noted by the Court, a parody is a legally established and approved form of fair use of an original copyrighted work.[253] It will qualify as a parody "if its aim is to comment upon or criticize a prior work by appropriating elements of the original in creating a new artistic, as opposed to scholarly or journalistic, work."[254] In making a parody, the new creation predictably uses parts of the copyrighted work to access the original for reasons of comment or critique.[255]

The use of copyrighted material in an offensive or pornographic manner has been permitted as fair use in various contexts. For example, in *Mattel, Inc. v. MCA Records, Inc.*,[256] the court noted that even if the song parody "Barbie Girl" blemishes the Barbie doll mark through its sexual or degrading lyrics, it is within the ambit of the "noncommercial use of a mark" and not prohibited.[257] Likewise, the animated film *Starballz*, a pornographic spin-off of the movie, *Star Wars*, was noted by the court to "likely survive Lucasfilm's copyright infringement claim because of the fair use doctrine."[258]

One scholar notes that the parody exemption does not protect a personal deepfake video because this type of distortion presupposes an original work that the parody is commenting on in some way.[259] The pornographic version is a new production that offers no commentary on a past work. The creator of the deepfake wants the viewer to believe that

---

251. *Id.*

252. Copyright Act, 17 U.S.C. § 107 (2018).

253. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

254. CCA and B, LLC v. F + W Media, Inc., 819 F. Supp. 2d 1310, 1318 (N.D. Ga. 2011) (quoting Suntrust Bank v. Houghton Mifflin Co., 268 F.3d 1257, 1268-69 (11th Cir. 2001)) (internal quotation marks omitted).

255. *Id.*

256. 28 F. Supp. 2d 1120 (C.D. Cal. 1998).

257. *Id.* at 1155.

258. Lucasfilm, Ltd. v. Media Mkt. Grp., Ltd., 182 F. Supp. 2d 897, 901 (N.D. Cal. 2002).

259. Harris, *supra* note 233, at 109.

the new version is that of the victim.[260] However, the victim's claim will likely fail because the deepfake is a transformative work.[261] Such a use takes the original copyrighted item and transforms its look or character to such a high degree that the use no longer qualifies as infringing.[262] A transformative use does not automatically mean that a fair use argument will be successful, but it will weigh heavily in favor of the alleged infringer.[263]

V. CONCLUSION

Computer technology has become so sophisticated that it is common to think that AI is on the verge of worldwide adoption. This widespread use has created a torrent of fake media known as shallowfake and deepfake technology.[264] Apprehension about the growth of these products has proven justified as the number of manipulated media has grown by leaps and bounds.[265] Unfortunately, this number will only increase as more people learn about the technology.

The threats posed by these false creations are real and deeply concerning. However, the solution is not easily discernable.[266] Current and proposed solutions endeavor to apply civil and criminal remedies, but the results of these efforts are unknown. The government has sponsored several endeavors on how to detect these false narratives. However, scientists remain "vastly overwhelmed by a technology that they fear could herald a damaging new wave of disinformation campaigns[.]"[267] Part of the problem is created by Section 230 of the CDA. This law provides immunity to online intermediaries who allow this false information to be posted to their social media sites.[268]

For the most part, there is little legislative or court guidance pertaining to this fake media. Criminal and civil remedies can be suggested, but there are obstacles to the various theories of responsibilities.[269] Paramount among them is the First Amendment safeguards related to freedom of speech. Until litigation filed by the

---

260. *Id.*

261. *Id.*

262. Richard Stim, *Fair Use: What Is Transformative?*, NOLO, https://www.nolo.com/legal-encyclopedia/fair-use-what-transformative.html (last visited Oct. 13, 2021).

263. *Transformative Use*, JUSTIA, https://www.justia.com/intellectual-property/copyright/fair-use/transformative-use (Oct. 2021).

264. Leetaru, *supra* note 11.

265. O'Donnell, *supra* note 26, at 706-07.

266. *Id.* at 711.

267. Reid, *supra* note 41, at 213 (alteration in original) (internal quotation marks omitted).

268. O'Donnell, *supra* note 26, at 711.

269. *See supra* Parts III–IV.

victims of this harmful technology works its way through the courts, the success of these proposed remedies remains to be seen. The problem is so widespread and devastating that Congress should pass legislation to remediate this problem that will overcome a constitutional attack.[270]

270. *See supra* Parts III, IV.B.