PennState Dickinson Law

DICKINSON LAW REVIEW
PUBLISHED SINCE 1897

Volume 126 | Issue 2

Winter 2022

# Deepfakes, Shallowfakes, and the Need for a Private Right of Action

Eric Kocsis
*Penn State Dickinson Law*

# Deepfakes, Shallowfakes, and the Need for a Private Right of Action

Eric Kocsis*

ABSTRACT

For nearly as long as there have been photographs and videos, people have been editing and manipulating them to make them appear to be something they are not. Usually edited or manipulated photographs are relatively easy to detect, but those days are numbered. Technology has no morality; as it advances, so do the ways it can be misused. The lack of morality is no clearer than with deepfake technology.

People create deepfakes by inputting data sets, most often pictures or videos into a computer. A series of neural networks attempt to mimic the original data set until they are nearly indistinguishable. The result is an ability to create pictures and videos entirely from data points.

There are many positive uses for deepfakes, such as in education, entertainment, and business, but the potential for misuse is high. People can create pornographic images of others and make it appear as if they are performing sexual acts on video that they had not. Deepfakes such as these often target women and celebrities. People also use deepfakes to target politicians, which has deeper implications for democracy and the electoral process.

Unfortunately, the legal system is currently unequipped to handle the problems that deepfakes are causing. In response, many lawmakers and policy experts are calling for legislation to protect people from these dangers. These proposals range from technological preventative measures to legal remedies. Many people are calling for criminal liability for those engaging in malicious deepfake activities, but there has been reluctance towards enacting a civil remedy. Malicious deepfakes overwhelmingly are nonconsensual porn that target women. Currently the law in most jurisdictions offers little to no legal recourse for those who

* J.D. Candidate, Pennsylvania State University Dickinson Law, 2022. Thank you to Jeremy Ulm for his countless hours of assistance throughout the process of writing this Comment. I would also like to thank Professor Peter Glenn for being a great mentor; I would not be here without you.

are targeted. Therefore, it is necessary that the federal government include a private right of action in any proposed deepfake legislation.

TABLE OF CONTENTS



## I. INTRODUCTION

"Your eyes can deceive you; don't trust them."[1] Although Obi Wan Kenobi's line in Star Wars: Episode IV – A New Hope encourages Luke Skywalker to use the force,[2] people should still heed this advice in the real world. People have been manipulating photo-

1. STAR WARS: EPISODE IV – A NEW HOPE (Lucasfilm Ltd. 1977).
2. *Id.*

graphs for nearly as long as photographs have existed.[3] People often use manipulated photographs in humorous or whimsical manners without intending to trick the viewer into believing them.[4] In other contexts, people use manipulated photographs to trick observers into believing that they are real, but they are often benign.[5] But, in some cases, people can harm one another using manipulated images.[6]

As technology advanced, so did the ability to manipulate media. Over 30 years ago, Adobe created Photoshop, which ushered in a new era of media manipulation.[7] Because of Photoshop, people were able to manipulate photographs quicker and more easily.[8] Photoshop had become so widespread that the word "photoshop" itself became a verb to describe the act of doctoring images.[9] Despite their convincing appearance, experts can quite successfully identify media manipulated by Photoshop.[10] Unfortunately, it becomes harder to detect manipulated media as technology becomes more advanced.[11]

Deepfakes are the newest link in the ever-growing chain of media manipulation. Deepfakes utilize artificial intelligence to, among

---

3. *See* Jocelyn Sears, *How Photo Retouching Worked Before Photoshop*, MENTAL FLOSS (July 28, 2016), https://bit.ly/2PBjw9B [https://perma.cc/LLE9-QZWW] (describing how Calvert Richard Jones performed the first known photo edit in 1846 by using photo negatives).

4. *See* Hann Brooks Olson, *Before Photoshop: A Brief History of Photo Manipulation*, CREATIVELIVE (Dec. 8, 2017), https://cr8.lv/3bdzP4Q [https://perma.cc/JNQ9-SJ8F]. Around the early 20th century, people would purchase "tall-tale" postcards that made it appear that animals and plants were giant. *See id.*

5. *See, e.g.*, *Cottingley Fairies: How Sherlock Holmes's Creator Was Fooled by Hoax*, BBC NEWS (Dec. 5, 2020), https://bbc.in/3rkfuQR [https://perma.cc/QB4H-627D]. In 1917, two young girls produced images of themselves with fairies, and they did not confess to them being fake until almost 70 years later. *See id.* Sir Arthur Conan Doyle got involved, likely believed the photos were real, and even published articles in support of their authenticity. *See id.*

6. *See* Megan Garber, *Oprah's Head, Ann-Margaret's Body: A Brief History of Pre-Photoshop Fakery*, ATL. (June 11, 2012), https://bit.ly/30dcf1M [https://perma.cc/VJH7-MNRT]. In 1989, *TV Guide* featured Oprah Winfrey on its cover, but instead of using a picture of Oprah Winfrey, *TV Guide* spliced her head onto the body of Ann-Margret. *See id.* This image was created and distributed without the consent of either woman. *See id.* Ann-Margret was supposedly "shocked," and a spokeswoman for Oprah Winfrey stated, "It's not something she would ever do." John Horn, *Oprah Winfrey Portrait Puts Talk Hostess' Head on Ann-Margret's Body*, AP NEWS (Aug. 28, 1989), https://bit.ly/3qjnIaF.

7. *See* Garber, *supra* note 6.

8. *See* Olson, *supra* note 4.

9. *See* Jenn Shreve, *Photoshop: It's All the Rage*, WIRED (Nov. 19, 2001, 2:00 AM), https://bit.ly/3sUBGS9 [https://perma.cc/KV9X-FUFL].

10. *See* Steven Melendez, *How DARPA's Fighting Deepfakes*, FAST COMPANY (Apr. 4, 2018), https://bit.ly/30cyK6T [https://perma.cc/8UH8-YDAE].

11. *See id.*

other things, replace one person's likeness with another.[12] Experts warn that because of deepfakes, manipulated media are becoming increasingly realistic and more difficult to identify.[13] Since deepfakes are so difficult to detect, they pose a significant threat to individuals, organizations, and society.[14]

As concerns grow over the looming threat of deepfakes, Congress has begun to take notice.[15] Lawmakers included requirements in the National Defense Authorization Act (NDAA) in both 2020 and 2021 for the Department of Defense and the Department of Homeland Security to research and issue reports on deepfakes.[16] In the 2021 NDAA, legislators specifically requested an "analysis of technical countermeasures to deepfakes and for detecting digital content forgeries."[17] Legislators are focusing heavily on technical countermeasures because of the effect that deepfakes have on national security.[18]

Although national security issues are of grave concern, legislators must not forget individuals who may be personally harmed by deepfakes. The overwhelming majority of deepfakes currently being produced are pornographic images of women.[19] Worse still, these deepfakes often target minors.[20] This Comment will show why a federal private right of action is necessary to remedy the harms caused by these malicious deepfakes.[21] A private right of action has been suggested previously in Congress and this Comment

12. *See* Dave Johnson, *What Is a Deepfake? Everything You Need to Know About the AI-Powered Fake Media*, INSIDER (Jan. 22, 2021, 11:46 AM), https://bit.ly/3bZYAR6 [https://perma.cc/9TUQ-SMBU].

13. *See* Melendez, *supra* note 10.

14. *See generally* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1771–86 (2019).

15. *See generally* Matthew F. Ferraro, *Congress's Deepening Interest in Deepfakes*, HILL (Dec. 29, 2020, 12:00 PM), https://bit.ly/3kLQq2F [https://perma.cc/C23H-TYHM] (explaining various actions Congress is taking relating to deepfakes, such as the FY 2021 NDAA and the IOGAN Act).

16. *See id.*

17. *See id.*

18. *See generally*, Anthony Kimery, *Deep Fake Technology Outpacing Security Countermeasures*, BIOMETRICUPDATE.COM (Dec. 11, 2018), https://bit.ly/30akZWm [https://perma.cc/FJ4G-LT48].

19. *See* HENRY AJDER, GIORGIO PATRINI, FRANCESCO CAVALLI, & LAURENCE CULLEN, THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1–2 (2019), https://bit.ly/30Qyk9T [https://perma.cc/W2YY-LH7N].

20. *See* EJ Dickson, *TikTok Stars Are Being Turned Into Deepfake Porn Without Their Consent*, ROLLING STONE (Oct. 26, 2020, 1:16 PM), https://bit.ly/2Je9sQw [https://perma.cc/5V68-P8LH].

21. *Infra* Part III.B.

will analyze that bill as well.[22] This Comment reaches this conclusion based on the analysis of why the current law is an insufficient remedy.[23]

## II. BACKGROUND

### *A. History*

In 2017, the term deepfake rose to public conscience.[24] A Reddit user by the same name created a subreddit[25] for users to post digitally altered pornographic pictures of celebrities using open source digital learning software.[26] Creators quickly filled this subreddit with fake, but convincing, nude and pornographic images of celebrities such as Gal Gadot, Daisy Ridley, and Maisie Williams.[27] This continued for several months until Reddit finally shut the subreddit down in February of 2018, after it had amassed nearly 90,000 followers.[28] By that point, the individuals making these fakes had moved to new communities on websites such as Discord.[29]

Unfortunately, people have been digitally altering images and videos for years, but people can use deepfake technology to much greater effect to cause the same harms.[30] People digitally alter photographs for a variety of purposes, both malicious and benign.[31] Even though altered media is not new, the recent commodification

22. *Infra* Part III.A.2.b.

23. *Infra* Part III.A.

24. The term "deepfake" has been used interchangeably with the term "deep fake." Both spellings refer to the same issue, but for the purpose of consistency, the author uses the spelling "deepfake" throughout this Comment. *Compare* Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339 (2019) (utilizing the "deepfake" spelling), *with* Chesney & Citron, *supra* note 14, at 1753 (2019) (utilizing the "deep fake" spelling).

25. A subreddit is an "individual message board[ ] devoted to one particular topic." Michael Franco, *The Beginner's Guide to Reddit*, LIFEHACKER (Sept. 5, 2017, 9:30 AM), https://bit.ly/33nQuhN [https://perma.cc/U3MH-UGGE].

26. *See* Meredith Somers, *Deepfakes, Explained*, MIT SLOAN SCH. OF MGMT. (July 21, 2020), https://bit.ly/34y58Ef [https://perma.cc/83DR-JGCF].

27. Spivak, *supra* note 24, at 339.

28. Samantha Cole, *Reddit Just Shut Down the Deepfakes Subreddit*, VICE (Feb. 7, 2018, 1:35 PM), https://bit.ly/2TqWe53 [https://perma.cc/QM43-TP9V].

29. *Id.*

30. Several years before the term "deepfake" was coined, people utilized other means to digitally alter images in harmful manners. *See generally* Ruby Harris, *How it Feels to Find Your Face Photoshopped Onto Internet Porn*, VICE (Apr. 17, 2019, 8:46 PM), https://bit.ly/35z7NN6 [https://perma.cc/85P8-54XL] (discussing how photoshopped images of a woman would colloquially be known as deepfakes today).

31. *See generally* Joshua Solomon Fischer, *Can a Photograph Lie? Remedies for an Age of Image Alteration*, SETON HALL L. SCH. STUDENT SCHOLARSHIP, May 2013, at 1 (discussing that some altered photographs are intended as harmless fun, while others are intended to embarrass or shame).

of deepfakes and the technology used to make them has created a dangerous marketplace where people can buy and sell potentially damaging deepfake content.[32]

### *B. What Is a Deepfake and How Are They Made?*

Policy makers need to accurately define what is and what is not a deepfake to properly remedy the injuries caused by deepfakes. The media often use "deepfake" as a buzzword to describe things that are not deepfakes.[33] Deepfakes are pieces of synthetic media that creators produce by utilizing various forms of machine learning.[34] Experts have also expanded the definition of deepfakes to include similar types of synthetic media that existed before the term was coined.[35]

Of these various machine learning techniques, Generative Adversarial Networks (GANs) are the most popular.[36] GANs use two neural networks simultaneously during a learning process, which are known as the generator and the discriminator.[37] The generator is utilized to create the fake and then the discriminator assesses the quality of the fake.[38] When using a GAN, the user inputs a dataset and the generator produces a sample to try and mimic the dataset.[39] In response, the discriminator analyzes the quality of the results; then, the generator uses those results to create a new sample.[40] In return, the generator iteratively creates increasingly more convincing outputs.[41]

---

32. *See* AJDER ET AL., *supra* note 19, at 5.

33. Charles Towers-Clark, *Mona Lisa and Nancy Pelosi: The Implications of Deepfakes*, FORBES (May 31, 2019, 10:32 AM), https://bit.ly/35Ecg11 [https://perma.cc/65NB-VMLK] (noting that several news outlets reported that an altered video of Nancy Pelosi was a deepfake when in fact it was not).

34. Chesney & Cintron, *supra* note 14, at 1759.

35. Somers, *supra* note 26.

36. AJDER ET AL., *supra* note 19, at 3.

37. Chesney & Cintron, *supra* note 14, at 1760. Neural networks are a way to conduct machine learning that is loosely modeled on the human brain, which teach a computer through the use of training examples. *See* Larry Hardesty, *Explained: Neural Networks*, MIT NEWS (April 14, 2017), https://bit.ly/30a9IFx. During this process, a computer will be given images of an object, and through pattern recognition and repetition, the computer will eventually learn to identify those objects on its own. *See id.*

38. *Id.*

39. *Id.*

40. *Id.*

41. Kyle Wiggers, *Carnegie Mellon Researchers Create the Most Convincing Deepfakes Yet*, VENTURE BEAT (Aug. 16, 2018, 8:12 AM), https://bit.ly/3jAEmiw [https://perma.cc/PG5T-HHQF].

This technology is only going to get more advanced and the content it produces is only going to get more realistic.[42] Utilizing GANs, researchers at Carnegie Mellon created deepfakes that were able to fool a group of people nearly 30 percent of the time.[43] Their deepfakes were not just limited to human recreations, but also non-human objects such as flowers and changing weather patterns.[44]

Not only are deepfakes getting more realistic, but the technology used to make them is also getting easier to access.[45] People can now make their own deepfakes by downloading the same code that researchers use, such as Carnegie Mellon, for free from the internet.[46] For individuals without the time or experience to work with the source code, push-button phone applications like Reface now allow anybody with a smartphone to make convincing deepfakes using GANs.[47] Although the creators of Reface say they are implementing counter-measures to prevent misuse, such as pornography detection and digital watermarking, not all application developers will implement the same security measures.[48]

While deepfakes pose a risk, they are not the only type of altered media that may do harm.[49] Another type of altered media that poses similar threats as deepfakes are colloquially known as shallowfakes.[50] A shallowfake is generally media that has been deceptively edited or is placed out of context.[51] For example, President Donald Trump retweeted a video of Speaker of the House Nancy Pelosi, in which the creator insinuated that she was drunk and slurring her words. However, in reality the video of Speaker Pelosi was slowed down to create the appearance of slurring.[52] Since this video was made using basic editing skills and not machine

42. Chesney & Cintron, *supra* note 14, at 1760.

43. *See* Wiggers, *supra* note 41.

44. *See id.*

45. *See generally* Natasha Lomas, *Deepfake Video App Reface Is Just Getting Started on Shapeshifting Selfie Culture*, TECHCRUNCH (Aug. 17, 2020, 1:35 PM), https://tcrn.ch/31JG3Ek [https://perma.cc/WV6Z-8Y6T] (discussing a phone application that allows users to make deepfakes easily on their phones).

46. *See* Wiggers, *supra* note 41.

47. *See* Lomas, *supra note* 45.

48. *See* Aaron Holmes, *Popular Deepfake Apps Are Making it Easier Than Ever to Make AI-Powered Manipulated Videos – Spawning New Memes, and an Increased Potential for Abuse*, INSIDER (Sept. 26, 2020, 9:00 AM), https://bit.ly/2G3U7kq [https://perma.cc/PX89-8GBC].

49. *See* AJDER ET. AL., *supra* note 19, at 11.

50. *Id.*

51. *Id.*

52. *See id.*

learning technology, it was technically a shallowfake and not a deepfake.[53]

Although the majority of deepfakes are pictures and videos, there is a growing concern over the use of this technology to create synthetic voice impersonations.[54] Around the same time as public awareness of deepfakes started to grow, Google released a system that used neural networks to generate synthetic human speech from text inputs.[55] While not discussed as much as altered photography and videos, maliciously altered and synthetic audio can pose just as serious of a threat as video.[56]

Regardless of whether the threat is from an altered recording or video, deepfakes pose a danger to public and private citizens alike.[57] Even less sophisticated shallowfakes pose tangible threats.[58] Whichever category the threat may fall under, federal law is not equipped to address the potential injuries to individual people that malicious actors may cause.[59]

### *C. Why People Make Deepfakes*

Deepfakes originally entered the public conscious in a negative light and were primarily seen as a way for people to create pornographic images of celebrities.[60] But, not all uses of deepfake technology are inherently negative or malicious.[61] In many cases, deepfake technology benefits education, entertainment, and business.[62]

---

53. *See id.*

54. *Id.* at 14.

55. *See* Jonathon Shen & Ruoming Pang, *Tacotron 2: Generating Human-Like Speech from Text*, GOOGLE AI BLOG (Dec. 19, 2017), https://bit.ly/3ec8BLS [https://perma.cc/72VF-FS2L].

56. *See, e.g.*, Catherine Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, WALL ST. J. (Aug. 30, 2019, 12:52 PM), https://on.wsj.com/3kGBAK4 [https://perma.cc/5HVZ-J9MR] (reporting on an incident where a company was scammed for nearly $250,000 because the scammers were able to recreate the voice of their parent company's CEO and convince the company to wire money to a supposed supplier).

57. *See* Chesney & Cintron, *supra* note 14, at 1757.

58. *See* AJDER ET. AL., *supra* note 19, at 9.

59. *See* Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 102 (2019).

60. *See* Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 339 (2019).

61. *See generally Yes, Positive Deepfake Examples Exist*, THINKAUTOMATION, https://bit.ly/3jJ05VE [https://perma.cc/39XG-9JFW] (last visited Dec. 10, 2021) (discussing how deepfakes have positive uses in education, art, and medicine).

62. *See id.*

### *1. Beneficial Uses*

Recently, there has been considerable discussions about how best to combat the threat that deepfake technology has posed.[63] Experts and lawmakers have proposed several remedies such as criminal liability, federal agency regulation, private detection technology, and civil liability.[64] In spite of the potential harms that this technology can cause, policymakers need to remember the benefits of deepfake technology and ensure that the solution does not inhibit those potential benefits.[65]

#### a. Education

Education has the potential to be changed for the better through the use of deepfake technology and other artificial learning and synthetic media.[66] This is especially true in the field of history, because educators can use deepfake technology to preserve past and present historical figures and events.[67] This type of preservation allows people to interact with and get immersed in history in new ways.[68]

By nature, history is more difficult to preserve over time, especially as those who lived it pass on. Deepfake and artificial learning technology now make it possible to preserve that history in new and interactive mediums.[69] Utilizing this technology, the Institute for Visual History and Education and the USC Shoah Foundation have been able to preserve the history of many survivors of the holocaust.[70] Their exhibition takes pre-recorded conversations with holocaust survivors and turns them into "high-definition projections," and visitors are then able to ask that person questions.[71] As more people ask questions, the technology utilizes machine learning and artificial intelligence that allows the exhibit to respond to the visi-

63. *See, e.g.*, Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 23–53 (2020).

64. *See id.*

65. *See* Chesney & Cintron, *supra* note 14, at 1788–89.

66. *See id.*

67. *See* Bernice Chen, *The Unique Potential of Deepfakes*, ASS'N OF INT'L RESEARCHERS FOR FUTURE EDUC. (July 2, 2020), https://bit.ly/2TA08sh [https://perma.cc/MD3E-6MAL].

68. *See id.*

69. *See Holocaust Museum Houston Opens Award-Winning 'Dimensions in Testimony' Exhibit Featuring Houston-Area Holocaust Survivor William J. Morgan*, USC SHOAH FOUND. (Jan. 11, 2019, 3:09 PM), https://bit.ly/2JmghzT [https://perma.cc/M28X-T2UM].

70. *See id.*

71. *See id.*

tors' questions, thus creating lifelike and real-time conversations.[72] This use of machine learning and deepfake-style technology allows important stories of history to stay alive for future generations and prevents them from being forgotten.[73]

In a more traditional use of deepfake technology, an organization recreated President John F. Kennedy's voice to deliver the speech that he was supposed to give the day he was assassinated.[74] The group used Deep Neural Networks to recreate the unique speaking style of President Kennedy.[75] This marked the first time that an entire speech was created using machine learning, completely from data.[76] This use of machine learning technology allows for a more immersive and unique approach to history and education.[77]

b. Entertainment

Similarly, the art and entertainment industries have also benefited from the development of deepfake technology.[78] Deepfake technology allows for faster and higher quality production in entertainment such as movies.[79] Deepfake technology also allows museum and exhibit patrons to better immerse themselves into art.[80]

Disney has already flirted with the idea of using deepfakes in their movies.[81] At the Eurographics Symposium on Rendering, Disney displayed their own proprietary technology and method for photo-realistic face swapping.[82] The deepfake content that Disney created was the first to be convincing at a megapixel resolution.[83] Disney may start to use deepfake technology in their movies, since

72. *See id.*

73. *See* Perlita Stroh, *How 3D Holograms and AI Are Preserving Holocaust Survivor's Stories*, CBC NEWS (Jan. 26, 2020, 4:00 PM), https://bit.ly/31Ungqa [https://perma.cc/42DZ-RX5V].

74. *JFK Unsilenced*, CEREPROC, https://bit.ly/3ecCxYb [https://perma.cc/B2ZZ-DQVD] (last visited Dec. 10, 2021).

75. *Id.*

76. *Id.*

77. *Id.*

78. Chen, *supra* note 67.

79. *See id.*

80. *See id.*

81. *See generally* James Vincent, *Disney's Deepfakes Are Getting Closer to a Big-Screen Debut*, VERGE (June 29, 2020, 11:53 AM), https://bit.ly/3eaTIJs [https://perma.cc/JXA3-FAKC] (discussing how Disney's internal researchers are studying deepfake technology to potentially use in the future).

82. Jacek Naruniec et al., *High-Resolution Neural Face Swapping for Visual Effects*, 39 COMPUT. GRAPHICS F. 173, 175 (2020).

83. *See id.* at 181.

they already use hyper-realistic facial swapping technology that could reasonably be described as shallowfakes.[84]

The Dalí Museum already has started to use this technology to enhance their exhibits.[85] Through the use of machine learning, the museum took archival footage and quotes and used it to train the AI, which then created an interactive version of the artist himself.[86] Now people can listen to Salvador Dalí describe his own work over three decades after he passed away.[87]

c. Business

Finally, businesses can use deepfake technology and machine learning in many unique ways to benefit their business, especially through marketing.[88] Businesses have started to take advantage of this technique due to the COVID-19 pandemic.[89] As a result of the COVID-19 pandemic, advertising companies lost their ability to produce new commercials because they were unable to film new footage.[90] To get around this challenge, companies such as ESPN have utilized various video manipulating techniques to create new advertisements out of old footage.[91] ESPN and State Farm Insurance utilized this technology to create an advertisement that took footage of reporter Kenny Mayne from 1998 and made it appear that he was talking about events in the future.[92] ESPN and other corporations have stated that they expect to continue to use this technology for their advertisements in the future.[93]

---

84. Dave Itzkoff, *How 'Rogue One' Brought Back Familiar Faces*, N.Y. TIMES (Dec. 27, 2016), https://nyti.ms/31Zv4GX [https://perma.cc/SQ4X-KASW] (reporting on how Lucasfilm, a subsidiary of Disney, utilized visual effects during the editing of *Rogue One: A Star Wars Story* to recreate the face of actor Peter Cushing who had passed away in 1994); *Disney Gallery: Star Wars: The Mandalorian: Making of S2 Finale* (Disney 2021) (discussing how Disney initially worked with deepfakes when de-aging Mark Hamill in *The Mandalorian* but ultimately decided against it).

85. *See generally Dalí lives: Museum Brings Artist Back to Life with AI*, DALI (Jan. 23, 2019), https://bit.ly/2HOqrZv [https://perma.cc/ZHM6-7ZE6].

86. *Id.*

87. Chen, *supra* note 67.

88. *See Somers*, *supra* note 26.

89. *See generally* Tiffany Hsu, *An ESPN Commercial Hints at Advertising's Deepfake Future*, N.Y. TIMES (Apr. 22, 2020), https://nyti.ms/3mLJAKy [https://perma.cc/Q32L-LUJH].

90. *See id.*

91. *See id.*

92. *See* NBA (@NBA), TWITTER (Apr. 18, 2020, 1:00 PM), https://bit.ly/2I0gV53 [https://perma.cc/W2V9-M34J].

93. *See* Hsu, *supra* note 89.

*2. Malicious Uses*

There are plenty of beneficial uses for deepfake and machine learning technology,[94] but those benefits do not eliminate the harms created when the technology is misused.[95] Historically, people have been able to differentiate real from fake, especially when looking at videos.[96] Since the machine learning technology behind these deepfakes continues to improve, it is becoming much harder to detect what is real and what is not.[97] Deepfake technology can be used for a variety of malicious purposes, such as undermining politicians and government officials, committing fraud, and creating nonconsensual pornographic images.[98]

a. Politics

Political deepfakes are an area of great concern for political scientists. Some worry that a well-timed deepfake or shallowfake could affect the outcome of an election.[99] Before the 2020 election, there was plenty of anxiety over the prospect that deepfakes would play a starring role on the campaign trail.[100] Fortunately, those anxieties were mostly unfounded, notwithstanding a few one-off examples.[101] Nevertheless, experts are still concerned about the future impact that deepfakes may play in the political arena.[102]

In 2018, a socialist political party in Belgium (SP.A) created a deepfake video of President Donald Trump calling climate change

94. *See generally Yes, Positive Deepfake Examples Exist*, THINKAUTOMATION, https://bit.ly/3jJ05VE [https://perma.cc/39XG-9JFW] (last visited Dec. 10, 2021).

95. *See* Harris, *supra* note 59, at 99–100.

96. *See* Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U. J. L. & TECH. 51, 61 (2018).

97. Chesney & Cintron, *supra* note 14, at 1753.

98. *See generally* Hall, *supra* note 96 at 57–58.

99. *See* William A. Galston, *Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics*, BROOKINGS (Jan. 8, 2020), https://brook.gs/3nYvA0q [https://perma.cc/ZDF3-YW8P].

100. *See* Gary Grossman, *Deepfakes May Not Have Upended the 2020 U.S. Election, but Their Day is Coming*, VENTURE BEAT (Nov. 1, 2020, 2:22 PM), https://bit.ly/3kMuxP8 [https://perma.cc/P9M2-898N].

101. *See id.* The author refers to an incident where President Donald Trump retweeted a video of Joe Biden sticking out his tongue. *Id. The Atlantic* reported that this was a deepfake mostly because the original poster labeled it as such. *See generally* David Frum, *The Very Real Threat of Trump's Deepfake*, ATL. (Apr. 27, 2020), https://bit.ly/2UH8fUl [https://perma.cc/U3EN-5UA2]. While still arguably inappropriate, it was in fact not a deepfake. Cole, *supra* note 28. Instead, this type of content was made using an app that lets users distort videos by dragging them with their finger, which is much more in line with the concept of a shallowfake. *See id.*

102. *See generally* Grossman, *supra* note 100 (arguing that deepfakes are going to play a large role in the future of politics).

fake.[103] The SP.A used deepfake technology to create the appearance that President Trump was giving a speech on the decision to withdraw the United States from the Paris Climate Agreement.[104] At the end of the video the "President" states that the video is fake and the SP.A states that it created the deepfake to "start a public debate" about climate change.[105] Even though the group stated in the video that it was fake, some still were convinced that it was real.[106] The SP.A's video, though clearly fake, demonstrates the dangers and potential injury that this technology poses.[107]

Shallowfakes can be just as dangerous as deepfakes in the political arena and should be addressed in any attempt to address and resolve the problem of deepfakes.[108] In 2019, a video of Nancy Pelosi was shared around social media that gave the appearance that she was giving a speech drunk and slurring her words.[109] These types of altered media are still as serious of a threat in the political arena as deepfakes, and they are easier to make than deepfakes.[110]

b. Fraud

Fraudsters can use deepfake technology to defraud victims ranging from individuals[111] to large corporations.[112] Audio deepfakes are often overlooked in the discussion of deepfakes but pose a serious threat when it comes to financial fraud.[113] Companies such as Google have created text-to-speech machine learning technology that allows users to recreate a person's voice.[114] Someone can use this type of text-to-speech technology to make it appear as

103. Hans Von Der Burchard, *Belgian Socialist Party Circulates 'Deep Fake' Donald Trump Video*, POLITICO (May 21, 2019, 2:13 PM), https://politi.co/3pIdgdO [https://perma.cc/3Y2J-8P32].

104. *A Faked Video of Donald Trump Points to a Worrying Future*, ECONOMIST (May 24, 2018), https://econ.st/2UBVOJy [https://perma.cc/884V-D3YH].

105. Von Der Bruchard, *supra* note 103.

106. *See id.*

107. *See A Faked Video of Donald Trump Points to a Worrying Future, supra* note 104.

108. *See* AJDER ET. AL., *supra* note 19, at 11.

109. *See* Towers-Clark, *supra* note 33.

110. *See* Tim Mak & Dina Temple-Raston, *Where Are The Deepfakes In This Presidential Election?*, NPR (Oct. 1, 2020, 5:05 AM), https://n.pr/2J12leF [https://perma.cc/CJX6-EVRN].

111. *See* Chesney & Cintron, *supra* note 14, at 1802.

112. *See* Hannah Murphy, *Cyber Security Companies Race to Combat 'Deepfake' Technology*, FINANCIAL TIMES (Aug. 16, 2019), https://on.ft.com/3703KtT [https://perma.cc/9QBZ-G5YG].

113. *See id.*

114. *See generally* Shen & Pang, *supra* note 55.

if someone is saying something that they never did, which is especially troubling when it is done without their consent.[115]

Audio deepfakes have the potential to cost corporations vast sums of money in simple fraud schemes.[116] In 2019, a United Kingdom company, which was a subsidiary of a German company, was defrauded for around $250,000 as a result of artificial-intelligence software.[117] The fraudsters utilized this technology to mimic the voice of the parent company's CEO and used that fake voice to request urgent payment to a supplier.[118] The CEO of the subsidiary complied with the request because he honestly believed he was speaking to the CEO of the parent company.[119] The fraud was only discovered when a promised reimbursement was not transferred back to the subsidiary.[120]

Fraud is not limited to attacks against large companies, it also poses a threat to individuals in unique ways.[121] During a custody hearing in 2019, a woman from the United Kingdom presented as evidence a recording of her husband threatening their children.[122] In reality, the recording was a shallowfake that was manipulated using publicly available technology and internet tutorials.[123] Her forgery was only discovered when the attorney for the husband reviewed the metadata of the evidence.[124] Had the attorney not recognized the fake, the client could have lost his children and his reputation could have been damaged.[125]

---

115. *See, e.g.*, Bill Hochberg, *YouTube Won't Take Down A Deepfake of Jay-Z Reading Hamlet—To Sue Or Not To Sue*, FORBES (May 18, 2020, 10:00 AM), https://bit.ly/36Zju09 [https://perma.cc/58HW-JD94]. In early 2020, the YouTube channel "Voice Synthesis" posted a recording that sounded like the rapper Jay-Z reciting Hamlet by Shakespeare. *Id.* The creator of the channel uses deepfake software to mimic the voices of celebrities, including Bob Dylan and several U.S. presidents. *Id.* Unlike the others, Jay-Z unsuccessfully filed take down notices with YouTube. *Id.* While clearly done in jest, this YouTube channel highlights the real threats that deepfakes can cause when used with malicious intent. *See id.*

116. *See generally* Murphy, *supra* note 112.

117. *See generally* Stupp, *supra* note 56.

118. *See id.*

119. *See id.*

120. *See id*.

121. *See* Chesney & Cintron, *supra* note 14, at 1772.

122. Matt Reynolds, *Courts and Lawyers Struggle with Growing Prevalence of Deepfakes*, A.B.A. J. (June 9, 2020, 9:29 AM), https://bit.ly/39f4L41 [https://perma.cc/5YTT-5TMF].

123. *See id.*

124. *See id.*

125. *See id*.

c. Pornography

Without a doubt, pornography is the most prevalent type of deepfake on the internet.[126] Although deepfakes often target celebrities, there is an increasing concern about deepfakes being used to target private citizens, such as ex-partners.[127] The director of the Artificial Intelligence Institute at the University of Buffalo, David Doermann, told *NPR*, "The place that we saw these fakes hurt people, initially, was at a very grassroots level. They were using it for revenge on a spouse or partner."[128] Currently, there is no sufficient remedy for the injuries caused by these deepfakes, especially for private citizens.[129]

Many of the women who are targeted by deepfake creators are young and some are even underage.[130] In the summer of 2020, a pornographic deepfake video of a TikTok influencer made its way onto social media without her consent.[131] While investigating an article about the subject, *Rolling Stone* discovered over two dozen incidences of deepfake pornography involving TikTok creators.[132] Many of these came from Discord servers[133] that allow users to request custom made deepfake pornography with many of the top requests being underage girls.[134]

Recently, people looking for custom deepfake pornography have turned to another messaging app, Telegram.[135] On Telegram, people can send pictures of clothed women through the app and a

126. AJDER ET. AL., *supra* note 19, at 1–2. As of 2019, Deepfake pornography made up 96 percent of all deepfakes on the internet, and 100 percent of the pornographic deepfakes targeted women. *Id.* at 1–2.

127. *See* Harris, *supra* note 59, at 101. A reddit user posted on the original deepfake Reddit page, "I want to make a porn video with my ex-girlfriend. But I don't have high-quality video with her, but I have a lot of good photos. If I can do something using only a photos [sic]?" *See* Cole, *supra* note 28.

128. Mak & Raston, *supra* note 110 (discussing how deepfakes could be used to disrupt elections, especially at the local level).

129. *See* Harris, *supra* note 59, at 102–03.

130. *See* Dickson, *supra* note 20.

131. *Id.*

132. *Id.*

133. Discord is an app through which people can converse with one another using voice or text in groups called servers. *See generally* Devon Delfino & Grace Dean, *'What is Discord?': Everything You Need to Know About the Popular Group-Chatting Platform*, INSIDER (Mar. 24, 2021, 3:58 PM), https://bit.ly/39ju0Ce [https://perma.cc/3ZQF-UGLB] (explaining what Discord is and how it is used).

134. *See* Dickson, *supra* note 20.

135. *See generally* HENRY AJDER, GIORGIO PATRINI, & FRANCESCO CAVALLI, AUTOMATING IMAGE ABUSE: DEEPFAKE BOTS ON TELEGRAM (2020), https://bit.ly/3bYsDZH [https://perma.cc/X7X5-DVCV] (analyzing the growing deepfake market on Telegram).

bot will digitally remove her clothes.[136] The bot used to make the deepfakes is free to use, but users can pay money to have their picture created faster and to remove the watermark.[137] Based on a self-reporting poll through Telegram, 63 percent of the bot's users were interested in it for the purpose of creating pornographic images of "[f]amiliar girls, whom I know in real life."[138] This statistic is even more concerning alongside the fact that the main channel for the bot has over 45,000 unique users.[139]

Terrifyingly, a developer[140] has created a web based application that allows users to create their own deepfake pornography.[141] This app allows users to create deepfakes by uploading pictures of unsuspecting victims and then selecting from a library of pornographic videos.[142] Within seconds, the application will generate a deepfake pornographic video of the victim without their consent.[143] For a fee, the perpetrator can download the video to keep or share.[144] While deepfakes overwhelmingly target women,[145] the site also provides a library of homosexual pornography.[146]

People create nonconsensual deepfake pornography for a variety of reasons, but all of them are deeply harmful to the subject of the deepfake content.[147] Some people believe it is a "victimless crime" while others believe that they can "do what they want" with the images of public figures for the sole reason that they are public figures.[148] Some have far more malicious intent and create these

136. *Id.* at 3.

137. *See id.*

138. *Id.* at 7.

139. *Id.* at 3.

140. To prevent further traffic to or promotion of the application, I have decided not to publish the name of the developer or application.

141. *See* Karen Hao, *A Horrifying New AI App Swaps Women into Porn Videos with a Click*, MIT TECH REV. (Sept. 13, 2021), https://bit.ly/2XJYhXm [https://perma.cc/LP2Q-G4UU].

142. *See id.*

143. *See id.*

144. *See id.*

145. *See* AJDER ET. AL., *supra* note 19 at 1–2.

146. *See* Hao *supra*, note 141. The author of the article notes how these deepfakes in particular could pose additional threats to people who live in areas of the world where homosexuality is not yet legal and even criminalized. *Id.* The rise in deepfakes that target the LGBT community is in contrast to just a few years ago when deepfakes exclusively targeted women. *See* AJDER ET. AL., *supra* note 19 at 2.

147. *See* Dickson, *supra* note 20.

148. *See id.*

deepfakes to extort or undermine women in positions of authority and power.[149]

## III. ANALYSIS

### *A. Current Legal Remedies and Where They Fall Short*

The introduction of deepfakes and other forms of digitally altered media have created various ways that malicious actors can injure individual citizens.[150] Recently, most efforts to combat malicious deepfakes have been focused on creating methods to detect deepfakes[151] and ways to combat misinformation on a larger scale as a result of deepfakes.[152] Although those are necessary and beneficial to society as a whole, individual victims of malicious deepfakes do not have adequate civil legal remedies to their injuries.[153]

#### *1. State Law*

In recent years various states have considered or enacted several bills aimed at either curbing the malicious use of deepfake technology or creating remedies for those who are harmed by it.[154] States such as Texas and Virginia impose criminal liability for certain uses of deepfake technology.[155] California is the only state that has created a private right of action for individuals harmed by mali-

149. *See generally* Rana Ayyub, *In India, Journalists Face Slut-Shaming and Rape Threats*, N.Y. TIMES (May 22, 2018), https://nyti.ms/37e1DTv [https://perma.cc/YVN6-8ZL7]. The author of this article is an investigative political journalist in India who wrote books and articles critical of the current party in power. *Id.* After a quote was falsely attributed to her on a parody account, she began to receive threatening and harassing messages. *See id.* The harassment culminated in a two-minute-long video that appeared to show her performing a sexual act. *Id.* The video was a deepfake that imposed her face on another woman's body and was quickly circulated around social media, thus resulting in more harassment. *See id.*

150. *See generally* Chesney & Cintron, *supra* note 14, at 1771–75.

151. *E.g.*, Tom Burt and Eric Horvitz, *New Steps to Combat Disinformation*, MICROSOFT: MICROSOFT ON THE ISSUES (Sept. 1, 2020), https://bit.ly/3q9c43k [https://perma.cc/8E2V-WNCP].

152. *See* Alex Engler, *Fighting Deepfakes When Detection Fails*, BROOKINGS (Nov. 14, 2019), https://brook.gs/39oPihO [https://perma.cc/8ZLE-6BJB].

153. *See* Harris, *supra* note 59, at 103.

154. *See generally* Matthew F. Ferraro, *Deepfake Legislation: A Nationwide Survey–State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*, WILMERHALE (Sept. 25, 2019), https://bit.ly/37gn9H1 [https://perma.cc/GFH8-RTBG].

155. . *See, e.g.*, TEX. ELEC. CODE ANN. § 255.004 (West 2021) (making it a Class A misdemeanor to use a deepfake video to injure a candidate or influence an election). *See also* VA. CODE ANN. § 18.2-386.2 (West 2021) (making it a Class 1 misdemeanor to distribute nonconsensual pornography, including those that are digitally made to depict another person).

cious deepfakes.[156] State tort laws, such as the right to privacy, defamation, or intentional infliction of emotional distress may offer some relief but ultimately are insufficient in remedying many deepfake cases.[157]

a. California AB-602 – A First Attempt at Private Right of Action

In 2019, the California Legislature passed AB-602, which created a civil private right of action for victims of nonconsensual deepfake pornography.[158] Instead of using the term deepfake explicitly, the statute uses the terms "altered depiction," "depicted individual," and "digitization."[159] For the purposes of the statute, "depicted individual" means an individual who appears, as a result of digitization, to be giving a performance they did not actually perform or to be performing in an altered depiction."[160] In respect to the definition of "depicted individual," the definition of 'digitization' is:

> [T]o realistically depict any of the following: (A) The nude body parts of another human being as the nude body parts of the depicted individual. (B) Computer-generated nude body parts as the nude body parts of the depicted individual. (C) The depicted individual engaging in sexual conduct in which the depicted individual did not engage.[161]

The definitions within this legislation appear to be inclusive of more than just deepfakes. Instead, this language likely falls in line with allowing a private right of action for shallowfakes as well.[162] The notion that the statute broadly covers both deepfakes and shallowfakes is supported by the word deepfake being included in the legislative history, but being omitted in the actual statute.[163] The law specifically avoids any language regarding machine learning or artificial intelligence and only talks in terms of digital "depictions"

156. CAL. CIV. CODE § 1708.86 (West 2021) (creating a civil private right of action against individuals who "create" or "intentionally disclose sexually explicit material . . . of a depicted individual").

157. *See generally* Harris, *supra* note 59, at 104–16 (explaining the various potential legal actions that can be used to combat deepfakes).

158. *See* Ferraro, *supra* note 154.

159. *See* CAL. CIV. CODE § 1708.86 (West 2021).

160. *Id.*

161. *Id.*

162. *See* AJDER ET. AL., AJDER ET. AL., at 11.

163. *See, e.g.*, STATE OF CAL. ASSEMB. COMM. ON JUDICIARY, DEPICTION OF INDIVIDUAL USING DIGITAL OR ELECTRONIC TECHNOLOGY: SEXUALLY EXPLICIT MATERIAL: CAUSE OF ACTION (2019).

of people.[164] This distinction is important and should be included in any legislation, because of the danger that shallowfakes can cause, regardless of quality.[165]

The California law also allows for various forms of damages, including reclaiming any monetary gain by the defendant, actual or statutory damages, punitive damages, and reasonable attorney's fees.[166] Most notable is the one-way shifting of reasonable attorney's fees and costs to the defendant, which could potentially allow for the better access to the courts for those harmed in personal injury cases, such as malicious deepfake use.[167]

Conversely, the law in California falls short in regard to remedying malicious deepfake injuries aside from nonconsensual pornography. Deepfake technology creates the potential for a wide variety of injuries, not just nonconsensual pornography.[168] Yet under the California law, people can only recover for nonconsensual pornography and no other potentially harmful uses, such as deepfake financial fraud.[169] While not directly aimed at deepfake technology, California passed a law allowing for civil recovery for victims of malicious deepfakes within 60 days of an election, but the law is only temporary and is set to expire on January 1, 2023.[170]

b. Tort Law

Even though California is the only state with specific laws allowing recovery for malicious deepfake technology, common law tort claims may allow recovery in some instances.[171] Plaintiffs in tort cases regarding deepfakes may potentially recover under theories of defamation, intentional infliction of emotional distress, or the right to privacy—specifically the tort of false light.[172]

---

164. *See* CAL. CIV. CODE § 1708.86 (West 2021).

165. *See* AJDER ET. AL., *supra* note 19.

166. *See* CAL. CIV. CODE § 1708.86 (West 2021).

167. *See* Gregory A. Hicks, *Statutory Damage Caps Are an Incomplete Reform: A Proposal for Attorney Fee Shifting in Tort Actions*, 49 LA. L REV. 763, 793–94 (1989).

168. *See generally* Chesney & Cintron, *supra* note 14, at 1771–75 (explaining the various malicious uses for deepfake technology).

169. *See* CAL. CIV. CODE § 1708.86 (West 2021) (allowing civil action and recovery against a person who within 60 days of an election distributes "materially deceptive audio or visual media . . . with the intent to injure a candidate's reputation or to deceive a voter into voting for or against a candidate"); Ferraro *supra* note 154.

170. *See* CAL. ELEC. CODE § 20010 (West).

171. *See generally* Erik Gerstner, *Face/off: "Deepfake" Face Swaps and Privacy Laws*, 87 DEF. COUNS. J. 1, 4–9 (2020) (explaining the various common law remedies available for victims of deepfakes).

172. *See* Chesney & Citron, *supra* note 14, at 1793–95.

Generally, "[t]he elements of a cause of action for defamation are a false statement, published without privilege or authorization to a third party, constituting fault as judged by, at a minimum, a negligence standard, and it must either cause special harm or constitute defamation per se."[173] While potentially plausible for private plaintiffs; public figures face a heightened burden of proving actual malice,[174] making it harder for public figures to recover under this theory.[175] Defamation claims, especially for public figures, are made even harder with the defense of parody.[176] If the creator can show that the deepfake cannot be perceived as real, then the plaintiff will lose the claim for defamation.[177]

Intentional infliction of emotional distress is even harder to prove than defamation.[178] While states have differing definitions, a plaintiff generally must show "that the defendant intentionally or recklessly engaged in extreme and outrageous conduct that caused the plaintiff to suffer severe emotional distress."[179] Deepfakes that harm individuals that are not particularly graphic, such as political deepfakes or deepfakes used to damage someone's reputation, are not likely to satisfy the emotional distress requirement of intentional infliction of emotional distress.[180] While people may be able to succeed on an intentional infliction of emotional distress claim for nonconsensual deepfake pornography, it is much less likely in other cases.[181]

Finally, a person may seek remedy under multiple tort privacy theories, but the tort of false light is most likely to succeed.[182] While false light is a potentially viable claim against malicious deepfakes, it is only available in 32 states and the District of Columbia.[183] Generally, to succeed on a false light claim, the person must be placed in a "highly offensive" false light and the creator must have "had

173. D'Amico v. Corr. Med. Care, Inc., 120 A.D.3d 956, 962 (N.Y App. Div. 2014).

174. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 283–84 (1964).

175. *See* Chesney & Citron, *supra* note 14, at 1793.

176. *See* Gerstner, *supra* note 171, at 5.

177. *See id.*

178. *See* Harris, *supra* note 59, at 112–14.

179. *E.g.*, Snyder v. Phelps, 562 U.S. 443, 451 (2011).

180. *See* Harris, *supra* note 59, at 112.

181. *See* Chesney & Cintron, *supra* note 14, at 1794.

182. *See* Harris, *supra* note 59, at 115.

183. *Id.* at 115–16. The states that allow a false light cause of action are: Alabama, Arizona, Arkansas, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Jersey, New Mexico, Oklahoma, Oregon, Pennsylvania, Rhode Island, Tennessee, Utah, Washington, and West Virginia. *Id.*

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."[184] In a majority of states that recognize the tort of false light, the creator of the content needs to publish it for public viewing, while in the minority of states it only needs to be viewed by one other person.[185] Therefore, in most states, this remedy would only be available to a victim of malicious deepfake content if a large portion of people actually viewed the deepfake.[186]

### *2. Federal Law*

Although most personal injury and tort law is handled at the state level, federal law provides several potential remedies. Most notably, federal copyright law offers some protections but also has some shortcomings.[187] Legislators introduced a bill in 2019 that included a provision that would create a private right of action for malicious deepfake injuries, but it garnered little traction.[188]

#### a. Federal Copyright Law

People harmed by the misuse of deepfake technology may seek remedies under The Copyright Act of 1976.[189] Generally, an individual owns the rights to their pictures.[190] Since the deepfake creator usually uses a series of photographs to create the deepfake,[191] they potentially infringe the rights of the owner of the photographs.[192] If the plaintiff could prove the requisite elements of copyright infringement, then they would be able to access certain remedies, such as injunctions, disposition of the infringing material, damages, profits, and attorney's fees.[193]

Although federal copyright law offers some recourse, the scope of protection is far too narrow. Depending on the content of the deepfake, the creation may be protected under the fair use doctrine.[194] Importantly, parody is a recognized purpose under the fair

184. RESTATEMENT (SECOND) OF TORTS § 652E (1977).
185. *See* Harris, *supra* note 59, at 116–17.
186. *See id.*
187. *See generally id.* at 107–11.
188. *See* Daniel Lipkowitz, *Manipulated Reality, Menaced Democracy: An Assessment of The Deep Fakes Accountability Act of 2019*, 2020 N.Y.U. J. LEGIS. & PUB. POL'Y QUORUM 30 (2020); *see also infra* Part III.A.2.b.
189. *See* Copyright Act of 1976, 17 U.S.C. §§ 101–1511.
190. *See* 17 U.S.C. § 102.
191. *See supra* Part II.B.
192. *See* 17 U.S.C. § 501.
193. *See id.* §§ 502–505.
194. *See* Harris, *supra* note 59, at 107–08.

use doctrine.[195] It is also worth mentioning that the parodied work does not need to be well-known,[196] which could potentially open up the defense to a plethora of deepfake situations, such as with smaller social media influencers.[197] While parody and fair use may not be as relevant in regard to nonconsensual deepfake pornography of non-celebrities, it may arise more often with celebrities.[198]

Since fair use is an affirmative defense, a defendant would need to show these factors to avoid liability: (1) the purpose and character of the use, (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.[199] Not everything would be considered fair use, especially many pornographic depictions, but it is easy to see how a deepfake of Jay-Z reciting Shakespeare would be protected by the fair use doctrine[200] While the ethics surrounding the legality of recreating someone else's likeness without their consent is outside the scope of this Comment, it is not difficult to see why someone would want a legal remedy. In that case, federal copyright law falls short.

b. The DEEPFAKES Accountability Act

The issue of malicious deepfakes has not gone unnoticed by Congress, and in 2019 Rep. Yvette Clark (D-NY) introduced the DEEPFAKES Accountability Act.[201] This bill addresses the danger of deepfakes by including a private right of action for individuals harmed by malicious deepfake use.[202] The DEEPFAKES Accountability Act ("Act") would allow an "individual . . . who has been exhibited as engaging in falsified material activity in an advanced technological false personation record" to receive either actual or statutory damages and injunctive relief.[203] For the purpose of this

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

196. *See* Northland Family Planning Clinic, Inc. v. Center for Bio-Ethical Reform, 868 F. Supp. 2d 962, 976 (C.D. Cal. 2012).

197. *See supra* notes 131–34 and accompanying text (discussing how people are often making deepfakes of young women from TikTok).

198. *See generally* Tiffany C. Li, *Kim Kardashian vs. Deepfakes*, SLATE (June 18, 2019, 8:34 PM), https://bit.ly/3qOQ9hm [https://perma.cc/2SMX-FQ8N] (discussing how most "high-profile" deepfakes likely fall within fair use).

199. 17 U.S.C. § 107.

200. *See generally* Hochberg, *supra* note 115 (discussing how a deepfake video of Jay-Z reading Hamlet would likely fall under the parody category of fair use).

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

202. H.R. 3230 § 1041(g).

203. *See* H.R. 3230 § 1041(g)(1).

legislation, the Act defines the term deep fake [sic] in regard to advanced technological false personation record as:

> [A]ny video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct substantially derivative thereof—(A) which appears to authentically depict any speech or conduct; and (B) the production of which was substantially dependent upon technical means, rather than the ability of another person to physically or verbally impersonate such person.[204]

This definition would likely satisfy the need for a legal definition that encompasses both the narrow technical definition of deepfake and the broader colloquial definition that includes shallowfakes.[205]

While the bill would likely have a positive effect, it is does have several shortcomings. First, the requirements under subsection (a) of the bill only require the creator to include disclosures that state the content is altered and a watermark.[206] Although the criminal portion of the bill specifically addresses sexual harassment, political interference, and fraud, the civil liability and private right of action portion only requires watermarks and disclosures that identify a piece of media as a deepfake.[207] Based on this reading of the law, it is conceivable that an individual could avoid civil liability for creating an embarrassing deepfake by including the required watermarks and disclosures.

Second, this bill could potentially harm innocent parties, specifically artists. Regardless of the law, those looking to harm others with deepfakes are unlikely to include watermarks or disclosures in their creation.[208] To combat this, the bill attempts to carve out exceptions for movies, television, music, parody, and historical reenactments.[209] Since the bill includes specific exceptions, many other equally innocent uses may get vilified if the courts apply the *expressio unuis* canon of construction.[210] While these detailed exceptions

204. H.R. 3230 § 1041(n)(3).

205. *See generally supra* Part II.B (discussing the difference between deepfakes and shallowfakes).

206. *See* H.R. 3230 § 1041(a).

207. *Compare* H.R. 3230 § 1041(f)(1) (including sexual harassment, political interference, and fraud in the criminal liability section), *with* H.R. 3230 § 1041(f)(2), *and* H.R. 3230 § 1041(g) (including reference to only subsection (a), which makes disclosures and watermarks a requirement).

208. *See* Matthew Ingram, *Legislation Aimed at Stopping Deepfakes Is a Bad Idea*, COLUM. JOURNALISM REV. (July 1, 2019), https://bit.ly/3a5yLxZ [https://perma.cc/9PZF-W7QQ].

209. *See* H.R. 3230 § 1041(j).

210. *See, e.g.*, Sebastian Peris, *Star Wars Deepfake Shows Stranger Things' Millie Bobby Brown As Leia*, GAMERANT (Jan. 23, 2021), https://bit.ly/2KOBWBA

were likely at the behest of those industries, it leaves the door open to a broad statute that could potentially harm innocent parties.[211] Although the DEEPFAKES Accountability Act includes some important provisions, and represents a step in the right direction, it still falls short of what is needed.

### *B. A Federal Private Right of Action*

As discussed over the course of this Comment, deepfakes pose a substantial threat across various facets of society. Both state and federal governments recognize the need for action and have begun to devise legislation to combat the rise in deepfake threats.[212] For several reasons, these attempts have come up short.[213] When the federal government takes up this issue again, they must include a private right of action in future deepfake legislation.

#### *1. Why Is a Private Right of Action Necessary?*

Although there is much talk about large-scale deepfake fake crimes that can disrupt elections or have large scale consequences, the vast majority of deepfakes are pornographic images of individuals.[214] In fact, 96 percent of all deepfakes on the internet are pornography and virtually entirely targeting women.[215] While other large-scale solutions to deepfake crimes are certainly important, a private right of action is necessary because the law fails to provide remedy and recourse to those currently being harmed.[216]

A private right of action is necessary because it would give more choice over the remedy process to those who are harmed. It has been argued that a civil litigation process would require the party bringing the suit to relive the trauma caused by the deepfake.[217] While this is an important consideration, that decision

---

[https://perma.cc/4D57-QREF]. Without getting into the underlying ethical concerns, would fan edits of movies, such as this one showing Mille Bobby Brown in the place of Carrie Fisher as Leia, be considered movies or would they require watermarks? *Id.* This uncertainty potentially causes issues for artists and entertainers.

211. *See* Kaveh Waddell, *A Shaky First Pass at Criminalizing Deepfakes*, AXIOS (Jul. 27, 2019), https://bit.ly/2YgvIgX [https://perma.cc/95AM-M5NF].

212. *See, e.g.*, *supra* Part III.A.1.a, Part III.A.2.b.

213. *See generally supra* Part III.A (discussing why the current state of the law is inadequate to provide remedies for victims of malicious deepfakes).

214. *See* AJDER ET. AL., *supra* note 19, at 6.

215. *See id.* at 1–2.

216. *See generally supra* Part III.A (discussing why the current state of the law is inadequate to provide remedies for victims of malicious deepfakes).

217. *See* Brown, *supra* note 63, at 41.

should be made by the victim.[218] The lack of a private right of action is actually drawing out the litigation process even more.[219] Similar to revenge porn cases, attorneys for individuals harmed by malicious deepfakes are forced to carve out new areas of common law and struggle to convince judges to accept that law.[220] Instead, a private right of action would give the legal system a statutory framework to work with when handling the inevitable rise in deepfake cases.[221]

Finally, it would allow for some legal recourse in the case of anonymous creators. One of the biggest criticisms of against a private right of action is the fact that many creators are anonymous and it is incredibly difficult to identify them.[222] To the contrary, a private right of action would allow courts to hear malicious deepfake cases without having first identified the creator of the deepfake.[223] In cases where the defendant is not identifiable or is otherwise outside the jurisdiction of the courts, the plaintiff could commence an in rem civil action.[224] This would at least provide some remedy through a court order declaring the material to be a deepfake and awarding any profits to the plaintiff, if the defendant is ever identified.[225] Although there are many other necessary steps to solve the deepfake problem, any proposed legislation must include a private right of action.

### *2. What Would the Law Look Like?*

Ultimately, a federal law creating a private right of action would need to include several components to ensure that individuals have their harms remedied, while not being overly restrictive. First, the law must create a proper definition for what will be considered a deepfake under the statute.[226] By technical definition, the

218. *See* Geoff Bartlett, *'Revenge Porn' Law Will Allow People to Sue Those Who Share Intimate Photos*, CBC NEWS (Apr. 16, 2018, 9:25 AM), https://bit.ly/39mVkz6 [https://perma.cc/5HKQ-YCUV] (discussing how a Canadian law that creates a civil action right against revenge porn gives victims "a voice").

219. *See* Cara Bayles, *With Online Revenge Porn, The Law Is Still Catching Up*, LAW360 (Mar. 1, 2020, 8:02 PM), https://bit.ly/3raI4E3 [https://perma.cc/PBV6-GHPU].

220. *See id*.

221. *See* Devin Coldeway, *DEEPFAKES Accountability Act Would Impose Unenforceable Rules—But It's a Start*, TECHCRUNCH (June 13, 2019, 3:25 PM), https://tcrn.ch/36gQH7V [https://perma.cc/2WM3-D4UG].

222. *See id*.

223. *See generally* H.R. 3230 § 4(a) (allowing in rem litigation against unknown defendants).

224. *See id.* § 4(a).

225. *See id.* § 4(d).

226. *See supra* Part II.B (discussing the technical definition of deepfake).

word deepfake is quite narrow, but for the purposes of the law, it should be broadened to encompass shallowfakes.[227] As previously discussed, the DEEPFAKES Accountability Act has several issues, but the way the bill defined deepfakes was not one of them.[228] Therefore, the government should adopt, for the purposes of civil liability, the deepfake definition from the DEEPFAKES Accountability Act.[229]

Second, the law must explicitly make it a violation to knowingly distribute or solicit deepfake pornography without the consent of the depicted party. This provision is necessary because the overwhelming majority of deepfakes made are nonconsensual pornography.[230] It is crucial to include the acts of distribution and solicitation because individuals are often not creating the deepfakes themselves; rather, they are purchasing custom deepfakes from others.[231] Also, a comprehensive ban of illicit deepfakes would almost certainly be unconstitutional.[232] Therefore, the law should focus on creating liability for distributing or soliciting malicious deepfakes.[233]

Lawmakers should also include a specific provision aimed at political deepfakes alongside nonconsensual pornographic deepfakes. Malicious political deepfakes have the potential to harm society as a whole by disrupting the electoral process.[234] Curtailing the societal harm from political deepfakes will not be accomplished through a private right of action but will require one of the various other approaches suggested by experts.[235] While the greater harm to society is important, politicians should still be able to seek individual remedy for the harms they may face.[236] To balance the relative harm to politicians with not being overly restrictive, the law

227. *See generally id*.

228. *See generally supra* Part III.A.2.b.

229. H.R. 3230 § 1041(n)(3); *see also supra* text accompanying note 196.

230. AJDER ET. AL., *supra* note 19, at 1–2.

231. *See* James Vincent, *Deepfake Bots on Telegram Make the Work of Creating Fake Nudes Dangerously Easy*, VERGE (Oct. 20, 2020, 10:00 AM), https://bit.ly/2M8znLC [https://perma.cc/WN9A-6LM3].

232. *See generally* Harris, *supra* note 59, at 125–28 (arguing that interpreting courts would likely find that some illicit deepfakes have societal value and that an outright ban would be overbroad).

233. *See id*.

234. *See* Lipkowitz, *supra* note 188, at 34.

235. *See, e.g.*, Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1483–86 (2019) (discussing a partial ban on counterfeit campaign speech).

236. *See id*. at 1465–66 (arguing in favor a law protecting the dignity of political candidates from "falsification of their identities").

should prohibit politically motivated deepfakes during election season.[237]

Although it would be tempting to create a comprehensive list of offending deepfake uses, any law should avoid doing so—this would allow the law to handle new and developing harms that are not yet identified.[238] To prevent the law being closed to future developments, lawmakers should draft legislation to prohibit the use of deepfakes during the commission of any crime, violation of law, or any other tort.[239] This framework would also allow for easier recovery based for malicious deepfake use in the context of fraud.[240]

Finally, lawmakers must include several procedural rules to help protect the victims and allow for efficient remedy, several of which are already included in the DEEPFAKES Accountability Act. While not a perfect solution, lawmakers should allow in rem litigation against unknown defendants, because it will allow a court to determine that a deepfake is in fact materially false.[241] This could be beneficial, especially in regard to the dignity of politicians who are harmed by malicious electoral deepfakes.[242] The law should also allow plaintiffs to have privacy protections during the proceeding by allowing the lawsuit to be filed under seal.[243] When combined, these provisions would create a law that allows individual citizens to seek remedies with dignity, is adaptable to unforeseen threats, and provides a statutory framework for addressing malicious deepfakes.

### *C. Deepfakes and the First Amendment*

Deepfakes are undeniably speech and any law that limits their creation would be subject to First Amendment scrutiny. However, the First Amendment would not bar the limitations outlined in this Comment. While the First Amendment protects free speech, "it is well understood that the right of free speech is not absolute."[244] Based on the exceptions of obscenity and defamation, the proposed legislative action would not violate the First Amendment.[245]

---

237. *See supra* note 169 (discussing California's prohibition of political deepfakes near an election).

238. *See* Lipkowitz, *supra* note 188, at 31.

239. *See id.* at 32.

240. *See supra* Part II.C.2.b.

241. *E.g.*, H.R. 3230 § 4(a).

242. *See supra* note 237 and accompanying text.

243. *See* H.R. 3230 § 1041(4)(2).

244. Chaplinsky v. State of New Hampshire, 315 U.S. 568, 571 (1942).

245. *See generally* Jessica Ice, Note, *Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RES. L. REV. 417, 440–43 (2019).

The First Amendment does not protect obscenity as a form of free speech.[246] The court in *Miller v. California* articulated a three-prong test to determine if the speech in question is obscene: [1] whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest; [2] whether the work depicts or describes [sexual conduct], in a patently offensive way . . .; [3] whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.[247]

Without a doubt, certain deepfakes could be permissibly regulated under this standard.[248] This standard would also safeguard against legislative overreach by protecting beneficial deepfake uses.[249]

Lawmakers would also likely be able to restrict political deepfakes as long as the First Amendment standard for defamation is satisfied. In *New York Times Co. v. Sullivan*, the Supreme Court held that "actual malice" is required to sustain a defamation case against a political figure.[250] The bar for actual malice is incredibly high, requiring "that the statement was made . . . with knowledge that it was false or with reckless disregard of whether it was false or not."[251] In political deepfake cases, this standard is likely not as hard to meet because of the nature of deepfakes.[252] Since the creator of a deepfake is in fact taking definitive steps to create a piece of content that did not previously exist, it is a near certainty that the creator knew that what they were publishing was false.[253] Therefore, it is incredibly likely that both restrictions on nonconsensual pornographic deepfakes and political deepfakes would be upheld under existing First Amendment standards.

## IV. CONCLUSION

Deepfakes are the newest rendition in a long line of media manipulation.[254] Unlike prior forms of manipulated media, deepfakes are easier to make and harder to detect.[255] Creators are able to use emerging technology in artificial intelligence to make not only near

246. *See* Miller v. California, 413 U.S. 15, 23 (1973).
247. *Id*. at 24.
248. *See* Ice, *supra* note 245, at 440.
249. *See generally supra* Part II.C.1 (discussing beneficial uses for deepfakes).
250. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).
251. *See id*.
252. *See* Ice, *supra* note 245 at 433–34.
253. *See id*.
254. *Supra* Part I.
255. *See supra* Part I.

realistic deepfake photographs, but video and audio as well.[256] As the technology improves, more people are gaining the access and ability to create deepfakes.[257]

People are rightfully concerned about the threat that deepfakes pose, but it is important to remember that there are many beneficial uses for deepfakes and the technology used to create them.[258] Deepfakes can be used in various fields, including education, entertainment, and business.[259] Lawmakers must keep these beneficial purposes in mind when drafting any deepfake legislation.[260]

Despite the possible benefits of deepfakes, deepfakes still pose a significant threat.[261] People are starting to use deepfakes to affect political discourse, create nonconsensual pornography, and even commit fraud.[262] Currently, nearly all of the deepfakes on the internet are fake pornographic images of women.[263] While it is crucial to develop detection software and national security plans for confronting the deepfake challenge, it is necessary that we create solutions for individuals that are harmed by this malicious deepfake use.[264]

As it stands, victims of malicious deepfake use have little to no sufficient remedy for their harms.[265] State tort law and federal copyright law provide potential remedies, but both prove challenging.[266] Some states, such as California, have started to provide specific remedies, but most have not.[267] At the federal level, legislators have introduced legislation to address the issue, but the bills have not been enacted.[268] By no means would a private right of action fix all the harm done by malicious deepfakes and many victims may choose not to pursue litigation for various reasons.[269] But, those harmed should be the ones to make that choice and if they chose to pursue litigation, a cause of action should be readily availa-

256. *See supra* Part II.B.
257. *See supra* Part II.B.
258. *See generally supra* Part II.C.1.
259. *See supra* Part II.C.1.
260. *See* Chesney & Cintron, *supra* note 14, at 1788–89.
261. *See generally supra* Part II.C.2.
262. *See supra* Part II.C.2.
263. *See* AJDER ET. AL., *supra* note 19, at 1–2.
264. *See supra* Part III.B.1.
265. *See generally supra* Part III.A.
266. *See supra* Part III.A.
267. *Supra* Part III.A.1.a.
268. *Supra* Part III.A.2.b.
269. *Supra* III.B.1.

ble for them.[270] Deepfakes pose a significant and dangerous challenge and no one solution will fix that problem outright, but any solution must consist of a private right of action so those harmed can have a legal remedy.

270. *See Supra* III.B.1.