DEEPFAKES AND INVOLUNTARY PORNOGRAPHY: CAN OUR CURRENT LEGAL FRAMEWORK ADDRESS THIS TECHNOLOGY?

KAREEM GIBSON†

I. INTRODUCTION

In a time where buzzwords such as "fake news" and "alternative facts" are ever-present, the "truthfulness" of the media we consume today has likely never been more in doubt. One new technology that has gained some attention is called deepfake. Deepfake is a computational algorithm for swapping faces on a picture or video.[1] The technology works off of an AI neural net technology.[2] In the past, tools such as

† B.S., 2014, University of South Carolina; J.D., 2020, Wayne State University Law School.

1. *Frequently Asked Questions*, DEEPFAKES.CLUB, https://www.deepfakes.club/faq/ [https://web.archive.org/web/20181015022418/https://www.deepfakes.club/faq/].

2. *Id.* Neural net technology is a type of machine learning that operates as a pseudo-brain system. The goal is to use this technology to identify patterns in a large amount of data much more quickly and more efficiently than humans can. The more data that is input into the network, the more it learns, and the more efficient the process becomes. The use of neural net technology is widespread with companies like Google using it to power its Photos app and YouTube using it to generate "watch next" recommendations. Facebook uses neural net technology for its face recognition algorithms that users utilize to tag their friends in photos. *See* Bernard Marr, *What Are Artificial Neural Networks - A Simple Explanation For Absolutely Anyone*, FORBES (Sep. 24, 2018, 12:46 AM),

Adobe Photoshop could be used to doctor photos one at a time.[3] Now, the deepfake process allows this to be done on a much larger scale, thereby processing hundreds or thousands of photos to create a video.[4] Unfortunately, there has been controversy over its use in superimposing images of people onto the bodies of adult performers in pornography.[5]

While the above-mentioned use of the deepfake technology has mostly been confined to celebrities and those in the public eye, it is likely only a matter of time before its use becomes more widespread. This technology has only recently developed enough to become workable on a consumer-level computer.[6] It will also likely be used in the future to spread misinformation in the "Fake News" context.[7] The use of deepfake technology to portray people as if they were acting in a pornographic film has been dubbed "involuntary pornography" and it currently operates in a legal grey area.[8] Although several popular online platforms, including some pornographic video sharing websites, have banned the sharing of involuntary pornography on their platforms, the legal implications of this media have only just begun to be explored.[9]

---

https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#1118f1dd1245 [http://web.archive.org/web/20200406164947/https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/].

3. *Id.*

4. *Id.*

5. Bloomberg, *How Faking Videos Became Easy — And Why That's so Scary,* FORTUNE (Sept. 11, 2018, 1:22 PM), http://fortune.com/2018/09/11/deep-fakes-obama-video/ [http://web.archive.org/web/20200406165126/https://fortune.com/2018/09/11/deep-fakes-obama-video/] (outlining other fears of potential uses of this technology, including inciting unrest through portraying a presidential candidate doing something criminal, or having a police chief depicted as being violent against minority groups).

6. *Id.*

7. *Id.* Fake news has been used as a catch-all term for propaganda, lies, and campaigns of misinformation to sway public opinion on certain issues. UM News*, What Is Fake News*, NEWS@THEU (Dec. 3, 2018), https://news.miami.edu/stories/2018/12/what-is-fake-news.html [http://web.archive.org/web/20200406165258/https://news.miami.edu/stories/2018/12/what-is-fake-news.html].

8. Gemma Askham, *Are Deepfakes the New Revenge Porn?*, BBC (Apr. 25, 2018), https://www.bbc.co.uk/bbcthree/article/779c940c-c6c3-4d6b-9104-bef9459cc8bd [http://web.archive.org/web/20200406165403/https://www.bbc.co.uk/bbcthree/article/779c940c-c6c3-4d6b-9104-bef9459cc8bd].

9. Samantha Cole, *Reddit Just Shut down the Deepfakes Subreddit*, VICE (Feb. 7, 2018, 1:35 PM), https://motherboard.vice.com/en_us/article/neqb98/reddit-shuts-down-deepfakes [http://web.archive.org/web/20200406170805/https://www.vice.com/en_us/article/neqb98/reddit-shuts-down-deepfakes].

This Note will scrutinize what legal framework is currently available to victims of involuntary pornography when the publisher has released it purely for pornographic purposes, and whether there is a legal cause of action suitable to address the harm caused. This Note will also examine the legal framework that is available to aid victims against creators of deepfakes and, to a lesser degree, individuals who disseminate the material. Section 230 of the Communications Decency Act, which protects online platforms from liability due to the publishing of third-party content, is outside the scope of this Note.[10] Whether there is an established tort, such as defamation or intentional infliction of emotional distress, suitable to address the harm suffered, or whether there needs to be a custom-made criminal cause of action tailored to address the harmful use of this media, will be examined in-depth. One complex issue that arises is what if the creator of a deepfake does not purport that the video is real footage? Much of the early use of this technology has been for purely pornographic purposes.[11] If a creator has labeled his pornographic deepfake video as just that, torts such as defamation and false light, which require a false statement of fact, will be ineffective.[12]

While there are inherent limitations in the application of the torts of defamation, false light, and intentional infliction of emotional distress (IIED) to deepfakes, this Note will argue that each of these torts has a place in addressing some of the harm that may result from this media.

---

10. 47 U.S.C. § 230(c)(1) (2018).

> (c) Protection for "Good Samaritan" blocking and screening of offensive material
>
> > (1) Treatment of publisher or speaker. 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.

Id.

> In other words, online intermediaries that host or republish speech are protected against a range of laws that might otherwise be used to hold them legally responsible for what others say and do. The protected intermediaries include not only regular Internet Service Providers (ISPs), but also a range of "interactive computer service providers," including basically any online service that publishes third-party content. Though there are important exceptions for certain criminal and intellectual property-based claims, CDA 230 creates a broad protection that has allowed innovation and free speech online to flourish.

Electric Frontier Foundation, *CDA 230: The Most Important Law Protecting Internet Speech*, EFF, https://www.eff.org/issues/cda230 [http://web.archive.org/web/20200406171104/https://www.eff.org/issues/cda230].

11. Tom Simonite, *Most Deepfakes Are Porn, and They're Multiplying Fast,* WIRED (Oct. 7, 2019), https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/ [https://web.archive.org/web/20200604142322/https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/].

12. *See* RESTATEMENT (SECOND) OF TORTS §§ 558, 581A, 651 (AM. LAW INST. 1977).

Further, while there is currently no enacted legislation to address deepfakes, a statutory cause of action has the potential to be the best way to address the issue of pornographic deepfakes, provided that it is carefully drafted.

Section II of this Note will begin with a more detailed explanation of the deepfake technology and explore some noteworthy cases dealing with First Amendment free speech issues.[13] In addition, it will introduce some of the current causes of action that may be applicable, and what solutions states have put forward in their initial efforts to combat the issue.[14] Section III will explore the potential efficacy of the causes of action introduced in Section II to determine which, if any, would be useful to combat deepfakes.[15]

II. BACKGROUND

Deepfake is an AI neural net technology that automates the process of creating convincing, fake photos in large numbers, which can then be used to generate a video.[16] This automation is a type of deep learning, hence the name deepfake.[17] In late 2017, a user on a social media platform released his image processing code for public use.[18] Through the contributions of multiple people, the code became more user-friendly and easier to use.[19] There are, however, several limitations to the technology as it exists now: while deepfake can run on consumer-level computers, it still requires high-end, enthusiast-level computer components to work effectively.[20] With respect to the source material, the technology requires a significant number of photos of a subject's face. Ideally, these photos need to consist of a variety of different lighting and angles that will also match the video of the face a user is

13. *See infra* Part II.A.

14. *See infra* Parts II.B–E.

15. *See infra* Part III.

16. *Frequently Asked Questions*, *supra* note 1.

17. Jun Wu, *AI, Machine Learning, Deep Learning Explained Simply*, TOWARDS DATA SCI. (July 1, 2019), https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960 [https://web.archive.org/web/20200624084159/https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960]; *Frequently Asked Questions*, *supra* note 1.

18. *Frequently Asked Questions*, *supra* note 1.

19. *Id.*; iperov, *DeepFaceLab*, GITHUB, https://github.com/iperov/DeepFaceLab [http://web.archive.org/web/20200406171732/https://github.com/iperov/DeepFaceLab] (last visited Apr. 6, 2020). This is an example of a deepfake tool on a popular development platform. The code is free for anyone to download and use. *Id*.

20. *Frequently Asked Questions*, *supra* note 1.

trying to replace.[21] In terms of the number of photos, users have claimed that using thousands of photos produce the best results, but some have had success using hundreds and sometimes fewer.[22]

In the current state of deepfake technology, it is easy to discern if a video of someone is a deepfake.[23] Mismatched facial movements, lifeless eyes, and unnatural blinking patterns are all things that point to the technology still being stuck within the "uncanny valley."[24] That is not to say that current deepfake videos would not fool anyone: if one was not paying close attention, it would be possible to be fooled. Unfamiliarity with the person being impersonated could also contribute to this. In determining whether a victim of involuntary pornography has a valid claim for defamation or another tort claim, it must be clear that the video is of, and concerning, the victim. The reasonable person standard would control this requirement.[25] Plaintiffs might have a hard time meeting the standard simply because no reasonable person would believe that the fake portrayal is real.

There have also been fears that as this technology begins to advance, it will be another "weapon" used in future campaigns of misinformation.[26] One product of the 2016 U.S. presidential election is

---

21. *Id.*

22. *Id.*

23. Chaim Gartenberg, *Deepfake Edits Have Put Harrison Ford into Solo: A Star Wars Story, for Better or for Worse*, THE VERGE (Oct. 17, 2018), https://www.theverge.com/2018/10/17/17990162/deepfake-edits-harrison-ford-han-solo-a-star-wars-story-alden-ehrenreich [http://web.archive.org/web/20200406171955/https://www.theverge.com/2018/10/17/17990162/deepfake-edits-harrison-ford-han-solo-a-star-wars-story-alden-ehrenreich]; *see also* Jeremy Hsu, *Why "Uncanny Valley" Human Look-Alikes Put Us on Edge*, SCI. AM. (Apr. 3, 2012), https://www.scientificamerican.com/article/why-uncanny-valley-human-look-alikes-put-us-on-edge/ [http://web.archive.org/web/20200406172156/https://www.scientificamerican.com/article/why-uncanny-valley-human-look-alikes-put-us-on-edge/].

24. Gartenberg, *supra* note 20; *see also* Hsu, *supra* note 20.

25. *See* RESTATEMENT (SECOND) OF TORTS § 283 (AM. LAW INST. 1965)."Unless the actor is a child, the standard of conduct to which he must conform to avoid being negligent is that of a reasonable man under like circumstances." *Id.*

26. J.M. Porup, *What Are Deepfakes? How and Why They Work*, CSO (Nov. 8, 2018, 3:00 AM), https://www.csoonline.com/article/3293002/fraud/what-are-deepfakes-how-and-why-they-work.html [http://web.archive.org/web/20200406172320/https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html]. Senator Marco Rubio, in a speech, stated:

> In the old days, . . . if you wanted to threaten the United States, you needed [ten] aircraft carriers, and nuclear weapons, and long-range missiles. Today, you just need access to our internet system, to our banking system, to our electrical grid and infrastructure, and increasingly, all you need is the ability to

that "fake news" has remained a hot button issue. There have already been efforts to invest research in new technology capable of recognizing if something is a deepfake.[27] The issues of fake news and misinformation with regard to politics, however, are outside the scope of this Note.

When the technology was first released, it was mainly used to replace adult actresses' faces with those of other celebrities.[28] However, there are now concerted efforts by universities to explore the uses of deepfake in other contexts as well.[29] There have also been huge strides in developing the technology to accurately re-create a person's voice and to generate new speech that may never have been spoken by that person.[30]

> produce a very realistic fake video that could undermine our elections, that could throw our country into tremendous crisis internally and weaken us deeply.

Id. (internal quotation marks omitted); *see also* Mikael Thalen, *Jennifer Buscemi Is the Deepfake That Should Seriously Frighten You*, DAILY DOT (Jan. 30, 2018, 10:22 AM), https://www.dailydot.com/debug/jennifer-buscemi-deepfake/ [http://web.archive.org/web/20200406172515/https://www.dailydot.com/%20debug/jennifer-buscemi-deepfake/]. In 2018, the House of Representatives sent a letter to the Director of National Intelligence Dan Coates stating: "As [deepfake] technology becomes more advanced and more accessible, it could pose a threat to United States public discourse and national security, with broad and concerning implications for offensive active measures campaigns targeting the United States . . . ." *Id.*

27. *See* Dan Robitzski, *DARPA Spent $68 Million on Technology to Spot Deepfakes*, FUTURISM (Nov. 19, 2018), https://futurism.com/darpa-68-million-technology-deepfakes [http://web.archive.org/web/20200406172744/https://futurism.com/darpa-68-million-technology-deepfakes]. The Department of Defense has already begun looking into deepfakes. As of 2018, DARPA has poured sixty eight million dollars into technology research to be able to spot deepfakes. DARPA has created an algorithm that can filter through video frame-by-frame to determine whether a video has been doctored. *Id.*

28. Cole, *supra* note 9.

29. Damon Beres & Marcus Gilmer, *A Guide to 'Deepfakes,' the Internet's Latest Moral Crisis*, MASHABLE (Feb. 2, 2018), https://mashable.com/2018/02/02/what-are-deepfakes/#44pbJVTxvqqV [http://web.archive.org/web/20200406172900/https://mashable.com/2018/02/02/what-are-deepfakes/]; *see also* Michael Kan, *Latest Deepfake Tech Will Have You Dancing Like Bruno Mars*, PCMAG (Aug. 24, 2018), https://www.pcmag.com/news/363321/latest-deepfake-tech-will-have-you-dancing-like-bruno-mars [http://web.archive.org/web/20200406173114/https://www.pcmag.com/news/latest-deepfake-tech-will-have-you-dancing-like-bruno-mars].

30. Benjamin Powers, *Adobe Is Developing Photoshop for Your Voice*, MEDIUM (Feb. 27, 2018), https://medium.com/s/story/adobe-is-developing-photoshop-for-your-voice-f39f532bc75f [http://web.archive.org/web/20200406173436/https://medium.com/s/story/adobe-is-developing-photoshop-for-your-voice-f39f532bc75f]. Adobe presented a new program named Voco, which was showcased in 2016. Comedian Jordan Peele recorded himself saying a sentence. This was input into the program which was able to shift the positions

There are definitely upsides to deepfakes, especially in the field of facial recognition.[31] In a more limited context, there have been applications—such as Snapchat—that use filters to augment reality and transform an individual's face into one with animal features and the like.[32] There is also a tool to swap faces with other people and inanimate objects as well.[33]

Deepfake technology could also have big implications for the film industry.[34] A few movies have digitally replicated actors and actresses who have passed away before the completion of the film.[35] Big strides in deepfake technology could aid film studios in creating a much more immersive film experience that does not strain a person's suspension of disbelief.[36] Outside of the entertainment context, other uses for this technology include therapeutic purposes for soldiers suffering from post-traumatic stress disorder and a multitude of other uses in the future.[37] One consideration that must be taken into account is that any legal response to deepfake must not hamper the upsides to this technology in the quest to obstruct its harmful application.

A. First Amendment and Falsity

The 2012 Supreme Court decision in *United States v. Alvarez* was seminal, as it concerned false statements and the First Amendment.[38] In this case, Alvarez lied about his receipt of a Congressional Medal of Honor in violation of the Stolen Valor Act of 2005, a federal criminal statute that made it a misdemeanor to falsely represent oneself as having received any U.S. military decoration or medal.[39] Alvarez's statements

---

of the words in the sentence and was also able to add new words to the recording even though Peele had not spoken them prior. *Id.*

31. Beres & Gilmer, *supra* note 29.

32. *Id.*

33. *Id.*

34. Erin Winick, *Actors Are Digitally Preserving Themselves to Continue Their Careers Beyond the Grave*, MIT TECH. REV. (Oct. 16, 2018), https://www.technologyreview.com/s/612291/actors-are-digitally-preserving-themselves-to-continue-their-careers-beyond-the-grave/.

35. *Id.*

36. Patrick Shanley & Katie Kilkenny, *Deepfake Tech Eyed by Hollywood VFX Studios*, HOLLYWOOD REP. (May 4, 2018, 7:05 AM), https://www.hollywoodreporter.com/news/deepfake-tech-eyed-by-hollywood-vfx-studios-1087075 [http://web.archive.org/web/20200406173734/https://www.hollywoodreporter.com/news/deepfake-tech-eyed-by-hollywood-vfx-studios-1087075].

37. Beres & Gilmer, *supra* note 29.

38. *See generally* United States v. Alvarez, 567 U.S. 709 (2012).

39. *Id.* at 716.

were not made to secure employment, financial benefits, or admission to privileges granted to those who have received the medal.[40] The Court began by stating that under the First Amendment, the government has no power to restrict expression because of its message, ideas, subject matter, or content.[41]

The Court noted that the burden of showing the constitutionality of a content-based restriction is on the government, and such restrictions on speech are presumed to be invalid.[42] The Court further stated that falsity alone was not enough to bring speech outside of the First Amendment.[43] The Supreme Court found that there was a lack of a direct causal link between the restriction imposed and the injury to be prevented.[44] It held that the Stolen Valor Act of 2005 infringed upon speech protected by the First Amendment because the Act constituted a content-based restriction on protected speech.[45] The Government did not satisfy the strict scrutiny test, which required proving that it had a compelling interest in prohibiting this type of speech and that the Act was the least restrictive means of achieving that interest.[46] Accordingly, any statutory efforts by state legislatures would need to ensure that legislation prohibiting the sharing or posting of involuntary pornography is not struck down as contravening the First Amendment merely because of its propagation of factually false information.

B. Intentional Infliction of Emotional Distress and Speech

The doctrine of intentional infliction of emotional distress (IIED) generally requires that the plaintiff show "(1) the defendant acted

---

40. *Id.* at 714. Alvarez's statements were not made for an actionable fraudulent purpose. *Id.*

41. *Id.* at 717. Courts have employed different standards for evaluating the constitutionality of content-based restrictions on speech and content-neutral restrictions. Content-based restrictions on speech are subject to the most scrutiny (strict) as they "pose the inherent risk that the government seeks not to advance a legitimate regulatory goal, but to suppress unpopular ideas or information or manipulate the public debate through coercion rather than persuasion." 16A AM. JUR. 2D *Constitutional Law* § 480 (2017). In contrast, content-neutral restrictions, which concern regulations that are unrelated to content, are subject to a less exacting level of scrutiny due to the lower risk of such regulations suppressing public ideas or opinions. *Id.*

42. *Alvarez*, 567 U.S. at 717.

43. *Id.* at 719.

44. *Id.* at 725.

45. *Id.* at 724.

46. *Id.* at 729; *see also Strict Scrutiny*, LEGAL INFO. INST., https://www.law.cornell.edu/wex/strict_scrutiny [http://web.archive.org/web/20200406173915/https://www.law.cornell.edu/wex/strict_scrutiny].

intentionally or recklessly; (2) the defendant's conduct was extreme and outrageous; (3) the defendant's actions caused the plaintiff emotional distress; and (4) the emotional distress was severe."[47]

Snyder v. Phelps established why the doctrine of IIED may not conform well when applied to certain kinds of speech.[48] In this case, Westboro Baptist Church members picketed near a soldier's funeral service with signs such as "Thank God for Dead Soldiers," "Fags Doom Nations," "America is Doomed," "Priests Rape Boys," and "You're Going to Hell."[49] The soldier's father brought an action of intentional infliction of emotional distress, among other claims, against the church.[50] The Supreme Court found that the content of Westboro Baptist Church's speech related to broad issues of public interest rather than a matter of purely private concern.[51] It held that the speech of church members who picketed near the funeral was protected under the First Amendment.[52] The fact that the church members conducted their picketing peacefully on a matter of public concern at a public place adjacent to a public street entitled their speech to special protection under the First Amendment.[53] The Court found that such speech could not be restricted simply because it was upsetting or aroused contempt.[54] Therefore, what must be determined is whether deepfakes have social value that must not be suppressed under the doctrine of IIED.

---

47. 38 AM. JUR. 2D *Fright, Shock, Etc.* § 5 (2018). Under the Restatement (Second) of Torts, Intentional Infliction of Emotional Distress is defined as follows:

> (1) One who by extreme and outrageous conduct intentionally or recklessly causes severe emotional distress to another is subject to liability for such emotional distress, and if bodily harm to the other results from it, for such bodily harm.
>
> (2) Where such conduct is directed at a third person, the actor is subject to liability if he intentionally or recklessly causes severe emotional distress
>
> > (a) to a member of such person's immediate family who is present at the time, whether or not such distress results in bodily harm, or
> >
> > (b) to any other person who is present at the time, if such distress results in bodily harm.

RESTATEMENT (SECOND) OF TORTS § 46 (AM. LAW INST. 1965).

48. Snyder v. Phelps, 562 U.S. 443 (2011).
49. *Id.* at 448.
50. *Id.* at 449.
51. *Id.* at 454.
52. *Id.* at 460.
53. *Id.*
54. *Id.* at 458.

C. Defamation

The law of defamation is based on the idea that individuals should have the freedom to enjoy their reputations unhindered by false and defamatory attacks.[55] An action for defamation is based upon a violation of this right.[56] Tort law permits a victim of defamation "to recover for an injury to his reputation caused by the false statement of another."[57] The requirements for a claim of defamation are as follows: (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, and; (d) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.[58]

Although there have not been any defamation cases concerning the use of deepfakes in involuntary pornography, other editing tools such as Adobe Photoshop have been involved in such suits.[59] In *Tharpe v. Lawidjaja*, the plaintiff brought a claim of defamation, among other claims, when the defendant distributed photographs in which the plaintiff was identified as a "porn star."[60] The defendant had altered the photos to depict the plaintiff acting in a sexually explicit manner and attached identifiers to these photos to link them to the plaintiff's place of employment.[61] The defendant moved for summary judgment on all claims, and the court dismissed the motion, holding that the defendant's statements were defamatory per se.[62] The court stated that photographs can constitute a defamatory statement.[63] These statements were found to arguably imply an unfitness for the plaintiff to perform his duties as a youth soccer coach and the statements prejudiced the plaintiff in his profession or trade.[64] While we can analogize that defamation would apply to deepfakes where the creator had the intent to harm the victim's reputation, it is less clear where the deepfake has been clearly labeled as false.

---

55. 50 AM. JUR. 2D *Libel and Slander* § 2 (2018).
56. *Id.*
57. *Id.*
58. RESTATEMENT (SECOND) OF TORTS § 559 (AM. LAW INST. 1965).
59. Tharpe v. Lawidjaja, 8 F. Supp. 3d 743 (W.D. Va. 2014).
60. *Id.*
61. *Id.* at 786.
62. *Id.*
63. *Id.*
64. *Id*.

D. False Light

The tort doctrine of invasion of privacy actually comprises "several causes of action that overlap each other and the tort of defamation."[65] The right to privacy may be invaded in four ways: "(1) unreasonable intrusion upon the seclusion of another; (2) appropriation of the other's name or likeness; (3) unreasonable publicity given to the other's private life; and (4) publicity that unreasonably places the other in a false light before the public."[66] Of these causes of actions, only the fourth is potentially relevant to involuntary pornography.

Douglass v. Hustler Magazine, Inc. applied this doctrine in a case where a "provocative" magazine published nude photographs of an actress that she had taken for a different magazine.[67] Although there is some overlap between the torts for false light and defamation, the Seventh Circuit explained that the false-light tort rests on "an awareness that people who are made to seem pathetic or ridiculous may be shunned, and not just people who are thought to be dishonest or incompetent or immoral."[68] The plaintiff argued that defendant *Hustler* insinuated that she was a lesbian and also that she was the kind of person willing to be shown naked in *Hustler*.[69] The court engaged in an analysis of the magazine's character, and it found that it would not be irrational for a jury to find the plaintiff's association with the magazine as degrading.[70] The court held that the plaintiff had a cause of action against *Hustler* for portraying her in a false light.[71] Although the actress had willingly posed for the photographs for the magazine *Playboy*, the court stated that *Hustler* had violated her rights under the commercial-appropriation branch of the right to privacy by publishing those photographs without

65. 62A AM. JUR. 2D *Privacy* § 26 (2018). The right to privacy was heavily influenced by the Restatement Second of Torts and a renowned law review article written in 1890 by Samuel D. Warren and Louis D. Brandeis (before he was appointed to the Supreme Court). They contended that the government could protect individual privacy interests without running up against the First Amendment and argued for the development of an invasion of privacy tort. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890). The invasion of the right to privacy is comprised of four different wrongs. RESTATEMENT (SECOND) OF TORTS § 652A cmt. b (AM. LAW INST. 1977). According to the Restatement (Second), their only relation to each other "is that each involves interference with the interest of the individual in leading, to some reasonable extent, a secluded and private life, free from the prying eyes, ears and publications of others." *Id.*

66. RESTATEMENT (SECOND) OF TORTS § 652A (AM. LAW. INST. 1977).

67. Douglass v. Hustler Magazine, Inc., 769 F.2d 1128, 1131 (7th Cir. 1985).

68. *Id.* at 1134.

69. *Id.* at 1135.

70. *Id.*

71. *Id.* at 1138.

her specific consent.[72] Facially, it appears that false light could be a good fit to address deepfakes, but this cause of action also requires a false statement.[73] Therefore, it will encounter some of the same problems as defamation in applying to labeled deepfakes.

E. State Legislatures' Initial Efforts

New York lawmakers recently proposed a bill to combat involuntary pornography.[74] However, the proposed bill is currently dead.[75] The bill aimed to make it illegal to use digital replicas of individuals without their permission.[76] The effort was intended to amend New York's right of publicity statute and indirectly regulate the deepfake technology.[77] The bill would have prohibited the use of a digital replica in a pornographic work: "[I]f done without the consent of the individual if the use is in an audiovisual pornographic work in a manner that is intended to create and that does create the impression that the individual represented by the digital replica is performing."[78] After this bill was proposed, there was significant push back from leaders in the film industry who felt that the bill was written too hastily and could have harmful, unintended consequences, such as making biopics about people impossible.[79] As it was written, entertainment companies argued that the bill would be too restrictive to creativity and storytelling and that it was likely unconstitutional.[80] The New York legislature adjourned without voting on the bill, and as of April 2019, the bill remains dead and no other bills have been proposed to tackle this issue.[81]

---

72. *Id*.
73. Machleder v. Diaz, 801 F.2d 46, 53–54 (2d Cir. 1986).
74. S. 5857, 240th Leg., Reg. Sess. § 51(4) (N.Y. 2017).
75. *Id*.
76. *Id.*
77. *Id.*
78. *Id.*
79. Katyanna Quach, *New York State Is Trying to Ban 'Deepfakes' and Hollywood Isn't Happy*, THE REGISTER (June 12, 2018, 10:22 PM), https://www.theregister.co.uk/2018/06/12/new_york_state_is_trying_to_ban_deepfakes_and_hollywood_isnt_happy/ [https://web.archive.org/web/20200325222501/https://www.theregister.co.uk/2018/06/12/new_york_state_is_trying_to_ban_deepfakes_and_hollywood_isnt_happy/].
80. *Id.*
81. N.Y. S.B. 5857.

III. ANALYSIS

As discussed, defamation, false light, and IIED are potential legal solutions that victims of deepfakes might be able to rely on to help them recover damages for the harms they have suffered from the technology.[82] However, it is uncertain if the courts will find any of these causes of action are a good fit. This section will analyze whether modern tort doctrine can allow a victim to recover damages as a result of the use of deepfakes and, if not, propose how it should be addressed.

The justice system is slow, and the relative newness of the deepfake technology has meant that no court has yet had the opportunity to address this problem. Despite this, we can examine other cases where claims of defamation, IIED, and false light have been brought to determine whether such frameworks might suit a claim concerning a harmful deepfake. With regards to statutory efforts, absent any currently enacted legislation addressing deepfakes, we can look to the relatively recent trend of states enacting nonconsensual pornography, or "revenge porn," regulations to determine the best way to avoid a successful First Amendment challenge.

A. Defamation

At first glance, defamation seems as though it might be suitable to address the issue of deepfakes. The tort focuses on injury to one's reputation and generally requires a false and defamatory statement.[83] In the scenario where a creator of a pornographic deepfake has disseminated a video falsely alleging that the video is, in fact, real, it is likely that a defamation claim would be appropriate to allow the victim to recover provided that he or she can satisfy all the elements. Less certain, however, is the scenario where a creator generates a pornographic deepfake of the victim but clearly labels it as a fake, making it clear that there is no false statement of fact.

In *Hustler v. Falwell,* the Supreme Court affirmed the interpretation of the Fourth Circuit Court of Appeals that the jury's finding that an advertisement parody published by a magazine was not reasonably believable, and thus, it did not constitute libel.[84] In *Geary v. Goldstein*, the District Court for the Southern District of New York found for the defendant on the claim of defamation because the defendant's creation of a parody of the plaintiff, in the context that it was shown, would not have

82. *See supra* Parts II.B, II.C, III.D.
83. RESTATEMENT (SECOND) OF TORTS § 559 (AM. LAW INST. 1977).
84. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 57 (1988).

been mistaken for an actual advertisement.[85] The court concluded that a reasonable viewer who saw the parody in context would not have believed that the plaintiff consented to her image being used for the parody.[86] It is doubtful whether traditional defamation would be appropriate in a situation like this.

The aspect of "actual malice" is another aspect of defamation that must be taken into account in certain circumstances.[87] This elevated standard applies to three categories of persons: public officials,[88] all-purpose public figures,[89] and limited-purpose public figures.[90] A victim of deepfakes that falls within one of these categories will have a higher burden to overcome than a private individual would to recover for defamation. Due to the abundance of celebrity photos available, most deepfakes that have been made so far are of those in the public eye.[91] As such, those victims will likely have a tougher time overcoming their burden of proving all the elements required for defamation.

If labeled as false, it is likely that because of this disclosure, a deepfake's creator would be able to overcome a claim of defamation. What is less certain is the very real possibility that subsequent disseminators could post or send a deepfake video without the label. The nature of the internet almost ensures that such a video, if it goes "viral," would be widely disseminated with or without the label. In this situation, does the creator of the deepfake bear some responsibility for this? The

---

85. Geary v. Goldstein, No. 91-CIV-6222, 1996 WL 447776, at *1 (S.D.N.Y. Aug. 8, 1996).

86. *Id.* at *3.

87. *Falwell*, 485 U.S. at 46, 56.

88. 53 C.J.S. *Libel and Slander; Injurious Falsehood* § 143 (2020). Public officials include government employees who have or appear to have substantial responsibility for, or control over, governmental affairs, in whose qualifications and performance the public has a special interest. *Id.*; *see also* Rosenblatt v. Baer, 383 U.S. 75, 85 (1966) (stating "[t]here is, first, a strong interest in debate on public issues, and, second, a strong interest in debate about those persons who are in a position significantly to influence the resolution of those issues").

89. 53 C.J.S. *Libel and Slander; Injurious Falsehood* § 141 (2020). An individual may achieve such pervasive fame or notoriety that he or she becomes a public figure for all purposes and in all contexts, including for purposes of a qualified privilege defense to a defamation action applicable to communications concerning public figures. *Id.*; *see also* Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974) (stating public figures can occupy a position a persuasive power and influence. In effect, they attract attention and comment).

90. 53 C.J.S. *Libel and Slander; Injurious Falsehood* § 142 (2020). An individual who is involved in a particular public controversy may become a limited-purpose public figure. *Id.*; *see also Gertz*, 418 U.S. at 345 (stating "an individual voluntarily injects himself or is drawn into a particular public controversy and thereby becomes a public figure for a limited range of issues").

91. *See generally* Cole, *supra* note 9.

innovation of the internet has made the sharing of information trivial. A person disseminating a deepfake for titillation, or otherwise, likely knows that it will be reposted countless of other times to other people and websites.[92] Despite being "out of the creator's control" at the moment the video has been uploaded, it seems foreseeable that the posting of a deepfake video could be widely disseminated by other people. Thus, it seems fair to hold the originator partly responsible for these subsequent disseminations.

While a victim of a deepfake video might be able to recover against its originator, it is generally accepted today that nothing is ever truly deleted from the internet.[93] Future publications by other people are very likely to occur if such a video is popular, and a court would probably determine whether it was foreseeable that this would occur and look at the amount of harm that could be attributed to the originator of the deepfake. However, just because a plaintiff has managed to recover damages from that originator does not mean that the harm has stopped. Future publications may be ongoing and may never stop. Where a plaintiff has recovered against the originator of a deepfake for a defamatory publication, for the purposes of the statute of limitations, the single publication rule will preclude that plaintiff from seeking future damages from him.[94]

In the law of defamation, the single-publication rule lays out that a single communication seen simultaneously by two or more people is a single publication.[95] Further, for any single publication, only one cause of action may be brought for which damages are recoverable.[96] In the 1800s, the common-law view was that every individual copy of a defamatory statement would be treated as a separate publication.[97] Therefore, "a newspaper that sold a million copies of a defamatory statement distributed across [fifty] states could theoretically be liable in a million different lawsuits litigated in [fifty] states."[98] The advent of mass circulation of newspapers and magazines, as well as the use of radios and

---

92. Fady Zaki, *'Innocence of Islam' Incident Proves Governments Can't Control Speech*, AM.'S FUTURE FOUND. (Feb. 22, 2013), https://americasfuture.org/innocence-of-islam-incident-proves-governments-cant-control-speech/ [https://web.archive.org/web/20200325231323/https://americasfuture.org/innocence-of-islam-incident-proves-governments-cant-control-speech/].

93. *Id.*

94. 1 LAW OF DEFAMATION, *Liability for Republication—Single-publication Rule* § 4:93 (2019).

95. *Id.*

96. *Id.*

97. *Id.*

98. *Id.*

televisions soon made such a rule untenable.[99] Today, the majority of states have recognized the single-publication rule.[100] Additionally, courts have long held that the single-publication rule applies to allegedly defamatory statements made on internet websites.[101] The idea is that a publisher should not be perpetually liable for everything that he has communicated in the past.[102] The question then becomes: what constitutes a single publication?

In a D.C. Circuit court case, *Jankovic v. International Crisis Group*, the plaintiff brought claims of defamation against a non-profit organization, which had produced international policy reports that had linked the plaintiff and his companies to a former Serbian president who had been put on trial as a war criminal before his death.[103] Although the plaintiff brought his defamation claim after the one-year statute of limitations had tolled on one of the published reports, he argued that foreseeable republication of the defendant's reports on the internet should reset the one-year clock.[104] The court rejected this contention and reasoned that in applying the single-publication rule, it must keep in mind the purpose of the rule's implementation, which was to avoid multiple suits, the harassment of defendants, and potential hardship on the plaintiff.[105] Although the plaintiff alleged that the defamatory language had been further disseminated through various other websites and publications, the court held that the defendant could only be required to account for a single publication of one report.[106] Analogizing from the rule's application in the world of print media, where the copying of an article by a reader, even for wide distribution, did not constitute a new publication, the court determined that the plaintiff's alleged third-party reproductions only constituted "mere continuing impact from past violations that [was] not actionable as a new cause of action."[107]

Therefore, should a victim successfully recover against the originator of a deepfake but discover that there have been subsequent publications by other persons, she will likely be barred from bringing a claim against

---

99. *Id*.

100. Keeton v. Hustler Magazine, Inc., 465 U.S. 770, 777 n.8 (1984) (citing RESTATEMENT (SECOND) OF TORTS § 577A (AM. LAW INST. 1977) (Reporter's Note)).

101. *See* Pippen v. NBCUniversal Media, LLC, 734 F.3d 610, 615 (7th Cir. 2013) (stating that "excluding the Internet from the single-publication rule would eviscerate the statute of limitations and expose online publishers to potentially limitless liability").

102. *See* Clark v. Viacom Int'l Inc., 617 F. App'x 495, 502 (6th Cir. 2015) (stating "even false historical statements at some point deserve legal repose").

103. Jankovic v. Int'l Crisis Grp., 494 F.3d 1080 (D.C. Cir. 2007).

104. *Id.* at 1086.

105. *Id.*

106. *Id.* at 1088.

107. *Id.* at 1087.

the originator for those third-party publications.[108] Of course, this does not mean that a victim would be without any recourse. Courts have held that each person who takes a responsible part in a defamatory publication may be held liable for the publication.[109] Further, under the single publication rule, multiple causes of action against multiple defendants may be brought in one proceeding for a single defamatory statement.[110] In the last decade, there have been several other internet-related contexts that courts have looked at concerning the single-publication rule.[111]

If a potential victim of deepfakes is limited in bringing claims against the creator for subsequent disseminations of deepfake media, is there an avenue for them to recover against those subsequent distributors? In defamation law, there are different requirements of fault that apply to hold someone liable for different forms of publication.[112] Therefore, there has been an effort to establish three categories of participants: "primary publishers, secondary publishers or disseminators, and those who are suppliers of equipment and facilities and are not publishers at all."[113] "Primary publishers were held to a strict liability standard, whereas secondary publishers were only liable for publishing defamation with actual or constructive knowledge of its defamatory character."[114] Secondary publishers have also been labeled as distributors.[115] Thus, while it is generally understood that every repetition of a defamation is a publication in itself, liability will depend on how that person is classified and their level of fault.[116] Historically, although it has been accepted that there is such a distinction between publisher liability

---

108. *Id*.

109. Traditional Cat Ass'n, Inc. v. Gilbreath, 13 Cal. Rptr. 3d 353, 359 (Ct. App. 2004).

110. *Id.*

111. Pippen v. NBCUniversal Media, LLC, 734 F.3d 610, 614 (7th Cir. 2013) (discussing the passive maintenance of a website to which the defamatory statement is posted); *In re* Philadelphia Newspapers, LLC, 690 F.3d 161 (3d Cir. 2012) (discussing the creation of hypertext links to previously published statements); Roberts v. McAfee, Inc., 660 F.3d 1156, 1167–68 (9th Cir. 2011) (discussing the failure to remove a statement from a website after receiving notice of its falsity); Jankovic v. Int'l Crisis Grp., 494 F.3d 1080, 1087 (D.C. Cir. 2007) (discussing a third party posting the statement elsewhere on the internet); Firth v. State, 775 N.E.2d 463, 466–67 (N.Y. 2002) (discussing the addition of an unrelated story to the web page that hosts the allegedly defamatory statement).

112. 1 LAW OF DEFAMATION *Liability for republication—Distinction between primary and secondary publishers* § 4:92 (2019).

113. *Id.*

114. *Id.*

115. Cubby, Inc. v. CompuServe, Inc., 776 F. Supp. 135, 139 (S.D.N.Y.1991).

116. 1 LAW OF DEFAMATION, *Liability for Republication—Distinction Between Primary and Secondary Publishers* § 4:92 (2019).

and distributor liability in defamation law, courts have generally been reluctant to construe this secondary publisher category too widely.[117]

An especially popular video could be reposted and re-uploaded hundreds if not thousands of times on any number of websites. It is not uncommon for these re-posters to change the name of the original video to gain more views or to pass it off as their own creation. Picture the situation where a deepfake video has been created, posted online, and labeled as deepfake. The first person to repost this video decides to do so with a whole new title. It is likely that this person has just published a defamatory statement. A day later, a new person sees that re-uploaded version of the video without the label and reposts that video on another website with another new name. Assuming that this new re-poster had no knowledge of this video's defamatory character, it is likely that he would not be liable under a secondary publisher theory of liability. But has his conduct of changing the video's title pushed him into the primary publisher category of liability? A secondary publisher has been described as a mere conduit and transmitter. In print media, these have included news dealers, bookstores, and libraries.[118]

Ordinarily, these types of distributors do not alter the media they sell, they merely take them in their original form and pass them along to their customers.[119] It is uncertain whether this alteration is enough to push this person from the secondary to the primary publisher category. Further, for a plaintiff bringing multiple suits against several defendants, it would likely be difficult to determine how much of the world of harm could be attributed to each publisher.

Defamation in the context of deepfakes is inherently problematic due to its requirement that there be a false statement of fact. Pornographic deepfakes present a confusing paradigm because their uses so far have generally not been to smear or defame, but to show off the capability of this technology albeit in a manner that undoubtedly causes harm to their victims.[120] Nevertheless, defamation has its place, and it will apply where someone has created a deepfake to pass it off as real footage. As discussed, however, it cannot be the end-all solution.

---

117. Zeran v. Am. Online, Inc., 129 F.3d 327 (4th Cir. 1997).

118. 1 LAW OF DEFAMATION *Liability for Republication—Distinction Between Primary and Secondary Publishers* § 4:92 (2019).

119. *Id.*

120. *See generally* Cole, *supra* note 9.

B. False Light

For those jurisdictions that recognize the tort of false light, this cause of action focuses on the offensiveness of the conduct in question rather than the effect on a person's reputation.[121] Due to the nature of the claim and its closeness to defamation, both claims are usually brought together, though recovery is always limited to one or the other.[122] False light concerns what is highly offensive to a reasonable person.[123] Similar to defamation, however, a misrepresentation or false statement is required for this cause of action.[124]

Assuming that a plaintiff could get over the first hurdle of a false statement of fact, would an offensiveness standard be easier to satisfy than to prove harm to one's reputation? Comment c to this section of the Restatement provides that it "is only when there is such a major misrepresentation of his character, history, activities or beliefs that serious offense may reasonably be expected to be taken by a reasonable man in his position, that there is a cause of action for invasion of privacy."[125]

Unlike defamation, however, information in a false light claim must be publicized as opposed to published.[126] The information must be disseminated to "the public at large, or to so many persons that the matter must be substantially certain to become public knowledge."[127] Leaving aside the problematic issue of the deepfake's falsity, is there greater potential for recovery now that defamation's publisher liability categories have been left behind? With regard to publicity, the Restatement (Second) further provides that false light concerns a communication "that reaches, or is sure to reach, the public."[128]

In the Second Circuit case of *Lerman v. Flynt*, the plaintiff was misidentified as a nude woman in a photograph contained in a magazine.[129] She brought an action against the magazine publisher for violations of the state privacy statute and common law right to publicity.[130] The lower court found for Lerman and awarded her damages

121. RESTATEMENT (SECOND) OF TORTS § 652E (AM. LAW INST. 1977).
122. *Id.* at cmt. b.
123. *Id.* at cmt. c.
124. *Id.* at § 652E.
125. *Id.* at cmt. c.
126. *Id.* at cmt. a.
127. RESTATEMENT (SECOND) OF TORTS § 652D cmt. a. (AM. LAW INST. 1977).
128. *Id.*
129. Lerman v. Flynt Distributing Co., Inc., 745 F.2d 123, 127 (2d Cir. 1984).
130. *Id.*

of ten million dollars.[131] The Second Circuit reversed, holding that there was no violation of the state privacy statute and that Lerman, as a limited-purpose public figure, had not made a sufficient showing under the heightened actual malice standard.[132] Strangely, although Lerman did not actually allege a false light claim in her complaint, the Second Circuit court underwent a false light analysis upon the facts of this case.[133] The court stated that the publicity given to the captioned photo of a nude actress that was not Lerman would have been highly offensive to a reasonable person.[134]

The *Lerman* case involved a situation that is similar to the harm caused by deepfakes. Both involve media that allegedly depict the victim as doing something that they never actually did. Where it differs, however, is with the potential for harm. People close to Lerman would know that the nude actress in the photo was not her, whereas, with deepfakes, it is not just a misidentified caption, but the actual victim's face that has been put in the video. Despite the court's pseudo- false light analysis being dicta, it suggests that false light could apply to pornographic deepfakes in some circumstances.[135]

Similar to defamation, because a false statement of fact is required for this tort, there will be issues if a creator of a deepfake has labeled it as such. On the other hand, a plaintiff may have more success with satisfying the false light offensiveness requirement as opposed to showing that her reputation was harmed. Public plaintiffs will again have to satisfy the heightened actual malice standard. Unfortunately, false light has only been accepted as a viable cause of action in thirty-one states.[136] Therefore, it may not be available for some plaintiffs, depending on their jurisdiction.

C. Intentional Infliction of Emotional Distress

The Restatement (Second) defines IIED as "[o]ne who by extreme and outrageous conduct intentionally or recklessly causes severe emotional distress to another is subject to liability for such emotional distress, and if bodily harm to the other results from it, for such bodily

131. *Id.*
132. *Id.* at 142.
133. *Id.* at 134–36.
134. *Id*. at 136.
135. *Id.*
136. Sandra F. Chance & Christina M. Locke, *When Even the Truth Isn't Good Enough: Judicial Inconsistency in False Light Cases Threatens Free Speech*, 9 FIRST AMEND. L. REV. 546, 557 (2011).

harm."[137] The comments to Section 46 of the Restatement (Second) of Torts provide that conduct is considered outrageous if the conduct is "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."[138] Assuming that the requisite intent and showing of emotional distress have been established, the question then must be, would deepfakes of someone in involuntary pornography satisfy this standard? Law Professors Bobby Chesney and Danielle Citron believe so.[139] In their research, they claim that pornographic deepfakes would constitute outrageous conduct under the tort of IIED because they would fall "outside the norms of decency."[140]

This favorable view of IIED as an adequate tool to combat deepfakes falls in line with another author's view that the common law of torts is not always exact, which allows it to remain open to new kinds of developments.[141] In Benjamin C. Zipursky's view, the concept of "outrageousness" serves as a rubric for the courts to determine when something is an actionable wrong or not.[142] A victim of a pornographic deepfake is likely to feel extremely humiliated and distressed from such a portrayal of themselves.[143] Such conduct would likely satisfy the outrageous element of IIED. Unfortunately, for those in the public eye, the analysis cannot stop there. Where deepfakes fall into the area of matters of public concern, victims will have to satisfy the standard of actual malice to overcome the First Amendment and to succeed under

---

137. RESTATEMENT (SECOND) OF TORTS § 46 (AM. LAW INST. 1965).

138. *Id.*

139. Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753 (2019) https://poseidon01.ssrn.com/delivery.php?ID=992074118067116085006067072000086102120084005013002000092118077092024112069123103011106062111041056027030024103111094086064064050015071081017112117125110118120126089053011068001095004001096065104126081115096102028018118127083096087109089117071095005&EXT=pdf [https://web.archive.org/web/20200623143734/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954].

140. *Id.* at 1794.

141. Benjamin C. Zipursky, Snyder v. Phelps, *Outrageousness, and the Open Texture of Tort Law*, 60 DEPAUL L. REV. 473 (2011).

142. *Id.* at 503.

143. Hannah-Rose Yee, *The Terrifying Rise of Deepfake Porn*, WHIMN (July 18, 2018), https://www.whimn.com.au/talk/news/the-terrifying-rise-of-deepfake-porn/news-story/7c704c6ae3029ca71c3de2afb2ee9e52 [https://web.archive.org/web/20200326021537/https://www.whimn.com.au/talk/news/the-terrifying-rise-of-deepfake-porn/news-story/7c704c6ae3029ca71c3de2afb2ee9e52].

IIED.[144] Actual malice is not malicious intent, rather it is knowing that a statement is false or acting with reckless disregard as to the statement's falsity.[145]

Snyder v. Phelps re-affirmed the notion that the First Amendment can serve as a valid defense to tort claims, including intentional infliction of emotional distress.[146] In *Snyder*, whether the Westboro Baptist Church could be held liable for its speech turned on the public or private nature of that speech.[147] The current uses of deepfakes have been primarily focused on celebrities and those in the public eye.[148] This is, in part, because there are numerous, easily accessible, high-quality photos of these public figures; such photos are required for a successful output of the deepfake algorithm.[149] Any claim that concerns a matter in the public interest will have a much harder time overcoming the obstacle of the Free Speech Clause of the First Amendment.[150] The Court in *Snyder* re-stated the notion that for matters of only private concern, "First Amendment protections are often less rigorous."[151] The *Snyder* Court also provided some guidance as to what constituted a matter of public concern, stating that:

> Speech deals with matters of public concern when it can be fairly considered as relating to any matter of political, social, or other concern to the community, or when it is a subject of legitimate news interest; that is, a subject of general interest and of value and concern to the public . . . .[152]

This guidance was consolidated from several previous Supreme Court cases.[153] In fact, *Snyder* built upon many of the principles that

144. Hustler Magazine, Inc. v. Falwell, 485 U.S. at 46, 56 (1988).

145. New York Times Co. v. Sullivan, 376 U.S. 254, 280 (1964).

146. Snyder v. Phelps, 562 U.S. 443, 443 (2011) (citing *Falwell*, 485 U.S. at 50–51).

147. *Id*. at 444.

148. Cole, *supra* note 9.

149. Tyler Huckabee, *Here's What You Need to Know About 'Deepfakes,' the Fake Porn That Should Terrify Us All*, RELEVANT (Dec. 31, 2018), https://relevantmagazine.com/culture/tech/heres-what-you-need-to-know-about-deepfakes-the-fake-porn-that-should-terrify-us-all/ [https://web.archive.org/web/20200326022308/https://relevantmagazine.com/culture/tech/heres-what-you-need-to-know-about-deepfakes-the-fake-porn-that-should-terrify-us-all/].

150. *Snyder*, 562 U.S. at 444.

151. *Id.* at 452.

152. *Id.* at 453 (internal quotation marks and citations omitted).

153. *See* City of San Diego v. Roe, 543 U.S. 77 (2004); Connick v. Myers, 461 U.S. 138 (1983); *see also* Rankin v. McPherson, 483 U.S. 378, 387 (1987) (finding that

informed the decision in *Hustler Magazine, Inc. v. Falwell*.[154] In *Hustler*, a nationally known minister and political commentator sued *Hustler Magazine* for IIED and libel after it published an "advertisement parody" that portrayed Falwell as having engaged in drunken sexual relations with his mother.[155]

The Court of Appeals held for Falwell on his IIED claim, but the Supreme Court reversed.[156] It likened the advertisement to political cartoons and caricatures, which have occupied an important role in public and political debate.[157] Neither party disputed that Falwell was a public figure; therefore, the Court concluded that public figures may not recover for IIED without a showing of actual malice and that the publication comprises a false statement of fact.[158] For deepfakes that fall into the categories outlined by the *Snyder* Court, victims will have the added challenge of establishing actual malice.[159] For private individuals where the deepfake concerns only purely private matters, the bar to recovery is much lower. Therefore, it is likely that IIED could be a promising cause of action for private individuals who fall victim to deepfakes. If the plaintiff in *Falwell* had been a purely private individual rather than a renowned minister, it is likely that the result would have been different.

Snyder espouses the notion that even vulgar advocacy is part of the public debate which has social value.[160] This "matters of public concern" test informed the Court's decision in this case.[161] The Supreme Court has stated that speech that constitutes a matter of public concern "can be fairly considered as relating to any matter of political, social, or other concern to the community, or when it is a subject of legitimate news interest; that is, a subject of general interest and of value and concern to the public."[162] Therefore the question that must be asked is whether there is any value in deepfakes that we need to worry about suppressing. It is

---

"inappropriate or controversial character of a statement is irrelevant to the question whether it deals with a matter of public concern").

154. *Snyder*, 562 U.S. at 451 (citing Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 50–51 (1988)).

155. *Falwell,* 485 U.S. at 56.

156. *Id.*

157. *Id.* at 54.

158. *Id.* at 56 (defining "actual malice" as referring to knowledge that the statement was false or with reckless disregard as to whether or not it was true).

159. *Snyder*, 562 U.S. at 444.

160. *Id.*

161. *Id.* at 452 (stating "[t]he First Amendment reflects a profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open" (internal quotation marks and citation omitted)).

162. *Id.* at 453.

very clear that the context of the deepfake will be the deciding factor here. The *Snyder* Court found that the Westboro Baptist Church was advocating on topics that concerned the political and moral conduct of the United States, which is of public import.[163] Where a deepfake falls into this "matter of public concern" category, courts will likely be much less enthusiastic about allowing such a claim to proceed.

As we have seen, IIED is problematic with regard to speech due to the First Amendment defense.[164] Plaintiffs that are victims of a deepfake will have to prove that the deepfake video has not become a matter of public concern, which would entitle it to heightened First Amendment protection. On the other hand, for purely private individuals, IIED could be an adequate tool to recover damages for a pornographic deepfake.

D. Statutory Solutions

If the range of options available to victims in the law of torts is limited, then another path to addressing the harm of deepfakes could lie in state statutes. New York's initial proposal was to amend its "right to publicity" statute to include a provision banning the nonconsensual use of a digital replica of an individual in a pornographic work.[165] This was the first of such state efforts to combat deepfakes.[166]

Federal lawmakers have also begun looking more closely at deepfakes, and one Senator from Nebraska has introduced a bill to combat deepfakes.[167] Senator Ben Sasse put forth a bill at the end of 2018 to criminalize the malicious creation and distribution of deepfakes.[168] The bill states the following:

---

163. *Id.* at 454.

164. *See supra* Part II.B.

165. S. 5857, 240th Leg., Reg. Sess. (N.Y. 2017).

166. Eriq Gardner, *Disney Comes Out Against New York's Proposal to Curb Pornographic "Deepfakes"*, HOLLYWOOD REP. (June 11, 2018, 4:01 PM), https://www.hollywoodreporter.com/thr-esq/disney-new-yorks-proposal-curb-pornographic-deepfakes-1119170 [https://web.archive.org/web/20200326025434/https://www.hollywoodreporter.com/thr-esq/disney-new-yorks-proposal-curb-pornographic-deepfakes-1119170].

167. S. 3805, 115th Cong. § 1041 (2018). The bill was introduced the day before the shutdown and fell under the radar until it expired ten days later. Kaveh Waddell, *Lawmakers Plunge into "Deepfake" War*, AXIOS (Jan. 31, 2019), https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html [https://web.archive.org/web/20200326025938/https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html]. Although it expired at the end of the year, the senator has said that he intends to reintroduce it. *Id.*

168. S. 3805, 115th Cong. § 1041 (2018).

> It shall be unlawful to, using any means or facility of interstate or foreign commerce— (1) create, with the intent to distribute, a deep fake with the intent that the distribution of the deep fake would facilitate criminal or tortious conduct under Federal, State, local, or Tribal law; or (2) distribute an audiovisual record with— (A) actual knowledge that the audiovisual record is a deep fake; and (B) the intent that the distribution of the audiovisual record would facilitate criminal or tortious conduct under Federal, State, local, or Tribal law.[169]

This proposed bill would, in effect, target two groups: (1) individual creators of deepfakes, provided that they are disseminating it with the intent to contravene federal, state, local or tribal law;[170] and (2) distribution platforms but only if they know that what they are publishing or disseminating is, in fact, a deepfake.[171]

While some commentators have stated that this piece of legislation will be a step forward in targeting individual creators of deepfakes, it has also been criticized for being too broad with respect to platforms that are involved in distributing these videos.[172] One law professor has warned that Senator Sasse's bill could have a chilling effect on some speech.[173] In their quest to avoid liability for publishing deepfakes, platforms might be overeager to take down those posts that are reported as deepfakes.[174] Legitimate posts could, thus, be easily swept into this without further checks, which may not be feasible for platforms depending on how many reports are being made.[175]

---

169. *Id.* at § 1041(b).

170. *Id.*

171. *Id.*

172. Waddell*, supra* note 167.

> (d) Limitations.—(1) IN GENERAL.—For purposes of this section, a provider of an interactive computer service shall not be held liable on account of— (A) any action voluntarily taken in good faith to restrict access to or availability of deep fakes; or (B) any action taken to enable or make available to information content providers or other persons the technical means to restrict access to deep fakes.

S. 3805, 115th Cong. § 1041(d) (2018).

173. Waddell*, supra* note 167.

174. *Id.*

175. *Id.* In their paper, Danielle Citron and Robert Chesney argue that some deepfakes could potentially fall under already existing statutes: if perpetrators post deepfakes in connection with the targeting of individuals, for example, they might violate the federal cyberstalking law, 18 U.S.C. § 2261A (2018), as well as analogous state statutes. Among other things, under Section 2261A:

> [I]t is a felony to use any "interactive computer service or electronic communication service" to "intimidate" a person in ways "reasonably expected

As this is a content-based regulation, it must be subject to the test of strict scrutiny to determine if it violates the First Amendment.[176] To satisfy strict scrutiny, the law must have been passed by the legislature to further a compelling government interest, and it must also have been narrowly tailored to achieve that interest.[177] The courts are ever vigilant to ensure that regulations do not sweep in more speech than is necessary to achieve its legislative purpose.[178]

Rebecca Delfino, a professor at Loyola Law School, has also examined this statute in her article *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, which discusses the potential for a federal criminal statute solution to the problem of deepfakes.[179] With regard to Senator Sasse's Malicious Deep Fake Prohibition Act of 2018, Professor Delfino argues that it is an inadequate solution to the issue of deepfakes.[180] She lays out three deficiencies; (1) overbreadth; (2) its condition of facilitating criminal or tortious conduct; and (3) its focus on solely the consequences of political deepfakes.[181]

First, in regard to overbreadth, the statute defines deepfakes as including any "audiovisual record created or altered in a manner that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of an individual[,]" which Professor Delfino argues would sweep in a wide range of media, including "non-offensive content like computer-generated imagery [(CGI)] in films."[182] Second, Professor Delfino claims that, with this condition, the statute is under-inclusive because there is currently an absence of civil or criminal remedies that address deepfakes, and it is over-inclusive because creators could find themselves criminally liable under this statute for a wholly unrelated tort.[183] Finally, she argues that pornographic deepfakes are not the focus of this statute, and thus, for this

---

> to cause substantial emotional distress . . . ." This reflects the fact that, even when cyberstalking victims do not fear bodily harm, "their lives are totally disrupted . . . in the most insidious and frightening ways." . . . Some deep fakes will fit this bill.

Chesney & Citron, *supra* note 139 (internal quotation marks and citations omitted).

176. Brown v. Entm't Merchants Ass'n, 564 U.S. 786, 799 (2011).

177. *Id.*

178. *Id.*

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

180. *Id.* at 923–24.

181. *Id.*

182. *Id.* at 923.

183. *Id.* at 924.

specific issue, it is not adequate to protect victims of these kinds of deepfakes.[184]

Historically, a distinction has been made between high-value speech and low-value speech.[185] Content-based restrictions on what constitutes high-value speech are presumed to be invalid.[186] On the other hand, low-value speech is not entitled to the same, high degree of protection, if any at all.[187] The Supreme Court has justified this distinction between high-value and low-value speech by claiming that the categories of low-value speech are extremely limited.[188] In the 1942 case *Chaplinsky v. New Hampshire*, the Court stated that there were "well-defined and narrowly limited classes of speech" that had never caused a constitutional issue as to its prevention and punishment.[189] The Court noted that the lewd and obscene, the profane, the libelous, and the insulting, or "fighting words," had such little social value that the interest in social morality and order outweighed any benefit that could be found in them.[190] Following this case, the Court has continuously emphasized these historically unprotected categories of low-value speech.[191]

In *United States v. Stevens*, the Court re-emphasized that these categories of low-value speech could be distinguished as a "previously recognized, long-established category of unprotected speech" or categories of speech that have been "historically unprotected but have not yet been specifically identified or discussed as such in our case law."[192] Therefore, analysis of a purported category of low-value speech would have to begin with a look to history to determine if this particular kind of speech had previously been excluded from protection.[193] In *Stevens*, the issue at hand was the constitutionality of a federal statute that criminalized the commercial creation, sale, or possession of certain depictions of animal cruelty.[194] As one of its arguments, the Government put forth the contention that a balancing test should apply to determine whether a certain type of speech should be excluded from First Amendment protection.[195] That is, the balancing of "the value of the

184. *Id.*
185. Genevieve Lakier, *The Invention of Low-Value Speech*, 128 HARV. L. REV. 2166, 2173 (2015).
186. *Id.* at 2171.
187. *Id.*
188. *Id.* at 2173.
189. Chaplinsky v. New Hampshire, 315 U.S. 568, 571 (1942).
190. *Id.* at 572.
191. United States v. Stevens, 559 U.S. 460 (2010).
192. *Id.* at 471–72.
193. *Id.* at 468.
194. *Id.* at 460.
195. *Id.* at 470.

speech against its societal costs."[196] The Court wholly rejected this argument, stating that: "The First Amendment's guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits."[197] The Court found that there was no evidence that animal cruelty had been historically unprotected and that it was overbroad.[198] Therefore, it held that the statute was invalid under the First Amendment.[199]

Of these defined categories, it is not clear that deepfakes fit in any of them. At first glance, one might be tempted to say that deepfakes should fit under the category of obscenity. Historically, however, obscenity has generally been a very high standard to meet.[200] Pornographic deepfakes are pornography, which is protected under the First Amendment unless it falls into the two categories of obscenity and child pornography. Courts have, however, long struggled with what is considered obscene.[201] Nonetheless, in *Miller v. California*, the Supreme Court gave a few threshold examples of what a state statute could regulate and be in compliance with the First Amendment: "(a) Patently offensive representations or descriptions of ultimate sexual acts, normal or perverted, actual or simulated; and (b) Patently offensive representation or descriptions of masturbation, excretory functions, and lewd exhibition of the genitals." If a deepfake were to contain any of these things, a statute that regulated these kinds of obscene deepfakes could potentially be able to overcome a First Amendment challenge.[202]

In crafting a statutory solution for deepfakes, some guidance may be available from looking at revenge porn statutes that have relatively recently been enacted in forty-one states.[203] Also titled as the distribution

---

196. *Id.*

197. *Id*.

198. *Id.* at 482.

199. *Id.*

200. In *Miller v. California*, the Court laid out a three-part test for jurors:

> (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, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and (3) whether the work, taken as a whole, lacks serious literary, artistic, political or scientific value.

Miller v. California, 413 U.S. 15, 24 (1973).

201. Ashcroft v. Am. Civil Liberties Union, 535 U.S. 564, 574 (2002).

202. Delfino, *supra* note 179, at 926.

203. *46 States + DC + One Territory Now Have Revenge Porn Laws*, CYBER C.RS. INITIATIVE, https://www.cybercivilrights.org/revenge-porn-laws/ [https://web.archive.org/web/20200326162504/https://www.cybercivilrights.org/revenge-porn-laws/]. Currently, the states that have enacted non-consensual pornography laws are: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware,

of non-consensual pornography, these laws generally criminalize the knowing disclosure of someone "whose intimate parts are exposed or who is engaged in a sexual act with knowledge of or reckless disregard for the fact that the person depicted did not consent to such disclosure."[204] It is possible that these statutes could be amended to include a provision prohibiting the knowing distribution of deepfakes.

Like revenge porn, a statute addressing deepfakes could constitute a content-based restriction on free speech.[205] Therefore, the clause itself would have to be very narrowly construed to not fall afoul of the First Amendment. On the other hand, revenge porn is unlike deepfakes in that revenge porn statutes will likely be harder to invalidate under the First Amendment. Whereas deepfakes are a false depiction of someone doing something or acting in a certain way, revenge porn is the disclosure of private, intimate pictures or videos of someone without their consent. It is much easier to say that from a First Amendment perspective, revenge porn should not be afforded as much First Amendment protection because of the intrusion into an individual's privacy. Deepfakes are more problematic because there are clearly uses for the technology that do not include causing harm. We want to avoid prohibiting the good uses of it with the bad.

A few cases have explored the constitutionality of these revenge porn statutes, subjecting them to strict scrutiny.[206]

In *State v. VanBuren*, a defendant charged with violating the nonconsensual pornography statute (NCP) in Vermont moved to dismiss the charge in the lower court, which ruled that the NCP statute was unconstitutional and granted the defendant's motion.[207] The Vermont Supreme Court disagreed and held that the statute survived strict scrutiny even though it did not fall into any of the existing exclusions from the First Amendment.[208] According to the court, the NCP statute defined

---

District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Michigan, Minnesota, Missouri, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, North Dakota, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, and Wisconsin. *Id.*

204. Mary Anne Franks, *CCRI Model State Law*, CYBER C.RS. INITIATIVE, https://www.cybercivilrights.org/wp-content/uploads/2019/01/CCRI-Model-State-Law.pdf
[https://web.archive.org/web/20200326162724/https://www.cybercivilrights.org/wp-content/uploads/2019/01/CCRI-Model-State-Law.pdf].

205. State v. VanBuren, 214 A.3d 791 (Vt. 2019).

206. *See id.*; *see also Ex parte* Jones, No. 12-17-00346-CR, 2018 WL 2228888, at *1 (Tex. App. May 16, 2018).

207. *VanBuren*, 214 A.3d at 791.

208. *Id.* at 807.

nonconsensual pornography narrowly and was limited to a specific class of content.[209] Further, the statute made use of a rigorous intent element and consisted of a limitation to only those images that were necessary to achieve the State's compelling interest.[210] Thus, the state supreme court concluded that the NCP statute was constitutional on its face.[211]

Conversely, in Texas, the state appellate court found that Texas's version of an NCP statute was overbroad and did not survive strict scrutiny.[212] The court found that, because of the disjunctive language used in the statute, both a person who knew that the depicted person expected the material to remain private and a person who disclosed the material without any knowledge of the depicted person's expectation of privacy could be punished.[213] Therefore, the Texas court held that the Texas NCP statute constituted an invalid content-based restriction in violation of the First Amendment.[214]

Because revenge porn statutes are not as problematic as deepfakes with regard to the First Amendment, it likely would not be prudent to allow a deepfake statutory provision to piggyback off of a revenge porn statute. It would be more reasonable to draft a separate deepfake statute. Such a statute would need to be drafted sufficiently narrowly with respect to who will be punished and the level of fault required so that it is truly the least restrictive means necessary to achieve the goal of punishing those who seek to do harm using deepfakes.

Currently, while there are no adequate statutory measures available to victims of deepfakes, it is likely that, if properly drafted, a federal criminal statute would be the best option for plaintiffs seeking to redress the harms they have suffered. A custom-crafted statute could address the

---

209. *Id.* at 812.

210. *Id.* The court listed several reasons why the statute was suitably tailored narrowly:

> [The statute] defines unlawful nonconsensual pornography narrowly, including limiting it to a confined class of content, a rigorous intent element that encompasses the nonconsent requirement, an objective requirement that the disclosure would cause a reasonable person harm, an express exclusion of images warranting greater constitutional protection, and a limitation to only those images that support the State's compelling interest because their disclosure would violate a reasonable expectation of privacy. Our conclusion on this point is bolstered by a narrowing interpretation of one provision that we offer to ensure that the statute is duly narrowly tailored. The fact that the statute provides for criminal as well as civil liability does not render it inadequately tailored.

Id.

211. *Id.* at 814.

212. *Ex parte* Jones, No. 12-17-00346-CR, 2018 WL 2228888, at *1 (Tex. App. May 16, 2018).

213. *Id.* at *5–6.

214. *Id.* at *8.

situations that tort doctrines are unable to cover. Further, as a federal statute, it would provide a much more significant disincentive for pornographic deepfake creators.[215] Leaving this to the states would likely result in a very slow implementation, as has been seen with the enacting of revenge porn statutes.[216] As Professor Delfino advocated, implementing a federal criminal statute would demonstrate the seriousness of the harm and give the impression that the government was taking a strong stance against pornographic deepfakes.[217] This, in turn, would send a message to individual creators as well as incentivize big internet publishers to take steps to address this issue.

IV. CONCLUSION

The potential for deepfake technology is almost limitless, with a multitude of different uses across various industries. With it, however, comes the certainty that it will continue to be used in ways that cause people harm. The question must be: what will be the best way to combat the harmful uses of deepfakes while also preserving the value it has to society? Torts such as false light and defamation appear to be a good fit where the deepfakes are not labeled as such, thereby satisfying the requirement of a false statement of fact. For purely private individuals, IIED will likely be a more suitable tool to combat the harmful uses of this media. Finally, while there are no current statutes that can adequately aid deepfake victims, a custom-crafted statutory solution would likely be the most effective measure against pornographic deepfakes, provided that it is not overbroad and complies with the First Amendment.

Thus far, deepfakes have drawn relatively little attention, especially regarding its place in facilitating involuntary pornography. Historically, technology has often outpaced the ability of the law to constrain it, and, despite the technology's release in late 2017, there has yet to be a single case brought dealing with deepfakes. It is likely that as deepfakes become more advanced and widespread, legislators and judges will take notice and have more of an opportunity to tackle this issue. Unfortunately, in the meantime, victims of deepfakes will have to rely on already established torts to limited effect. Time will tell whether our current framework is sufficient to deal with deepfakes or if a more tailored response will be required.

---

215. Delfino, *supra* note 179, at 927.
216. *Id.*
217. *Id.*