"DEEPFAKES": THE NEWEST WAY TO COMMIT ONE OF THE OLDEST CRIMES

Russell Spivak[*]

CITE AS: 3 GEO. L. TECH. REV. 339 (2019)

INTRODUCTION

Last year, a widely read technology blog turned heads with the deeply disturbing headline: "We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now."[1] While deliberately provocative, it was—and remains—unfortunately true. An unnamed individual on the popular discussion board, Reddit, superimposed images of celebrities such as Gal Gadot (*Wonder Woman*), Masie Williams (*Game of Thrones*), and Daisy Ridley (*Star Wars*) onto the bodies of adult video stars in pornographic films.[2] That Reddit poster's handle, or moniker, was "deepfake." Hence, the term deepfake now refer to a video that superimposes hyper-realistic faces onto the bodies of others with the intent of creating a new video with fake representations.[3]

The initial Reddit post containing the altered video led to the proliferation of computer-generated pornographic videos starring anyone and everyone. As *The Atlantic* correctly sums up, "[i]n a dank corner of

[*] J.D., Harvard Law School, 2017; B.S., Massachusetts Institute of Technology, 2013. All errors are my own. My deepest gratitude goes out to the editors of the Georgetown Law Technology Review, without whom this article would not be possible.

[1] Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, MOTHERBOARD (Jan. 24, 2018, 1:13 PM) [hereinafter Cole, *We Are Truly Fucked*], https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley [https://perma.cc/VK95-AVGW].

[2] *Id.*

[3] Note that deepfakes are not synonymous with simple face-swapping. As well, throughout this article, to distinguish between the anonymous individual versus the technology, the former will be indicated with an "@" to denote its status as a username handle.

the internet, it is possible to find actresses from *Game of Thrones* or *Harry Potter* engaged in all manner of sex acts."[4]

Importantly, this image-based technology—which is simply an intelligent algorithm, explained in more depth *infra*—can perform similar mimicry for auditory sounds. In other words, it can match one's vocal tone and pattern with user-generated scripts, *à la* lip-synching.

The purpose of this article is not to debate the morality of this technology and whether it ought to be legal to purchase or download; this article, instead, leaves these decisions to ethicists and, ultimately, policymakers.[5] The article also does not attempt to address national security or political questions that this technology raises.[6] Rather, this article endeavors to discuss the remedies available to private victims of this technology. Put plainly, how can Daisy Ridley et al. pursue legal recourse against their digital manipulators?

To do so, this article begins in Section I with an in-depth explanation of generative adversarial networks, the technology that enables deepfakes. Section II outlines whether lawmakers looking to

[4] Franklin Foer, *The Era of Fake Video Begins*, ATLANTIC (May 2018), https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/ [https://perma.cc/FM8Q-49XL].

[5] Policymakers are already taking this problem seriously and have said "they want to start working on fixes to the problem before it's too late." Ali Breland, *Lawmakers Worry About Rise of Fake Video Technology*, HILL (Feb. 19, 2018, 9:41 AM), http://thehill.com/policy/technology/374320-lawmakers-worry-about-rise-of-fake-video-technology [https://perma.cc/N9ZC-4V9G].

[6] *See, e.g.*, Roberty Chesney & Danielle Citron, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?*, LAWFARE BLOG (Feb. 21, 2018, 10:00 AM), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy [https://perma.cc/53AF-ZVVK] (Chesney and Citron offer the following examples: "[f]ake videos could feature public officials taking bribes, uttering racial epithets, or engaging in adultery"; "[p]oliticians and other government officials could appear in locations where they were not, saying or doing horrific things that they did not"; "[f]ake videos could place them in meetings with spies or criminals, launching public outrage, criminal investigations, or both"; "[s]oldiers could be shown murdering innocent civilians in a war zone, precipitating waves of violence and even strategic harms to a war effort"; "[a] deep fake might falsely depict a white police officer shooting an unarmed black man while shouting racial epithets"; "[a] fake audio clip might "reveal" criminal behavior by a candidate on the eve of an election"; "[a] fake video might portray an Israeli official doing or saying something so inflammatory as to cause riots in neighboring countries, potentially disrupting diplomatic ties or even motivating a wave of violence"; "[f]alse audio might convincingly depict U.S. officials privately 'admitting' a plan to commit this or that outrage overseas, exquisitely timed to disrupt an important diplomatic initiative"; "[a] fake video might depict emergency officials 'announcing' an impending missile strike on Los Angeles or an emergent pandemic in New York, provoking panic and worse.").

regulate this emerging technological field can do so under the First Amendment. Concluding no such regulation is constitutionally permissible, the article then turns to other available remedies notwithstanding any potential legislative solutions. Section III addresses the viability of potential state law causes of action, including defamation, privacy torts, and right of publicity that victims can pursue to redress their violated rights. Section IV looks to federal law and asks whether the Communications Decency Act, namely 47 U.S.C. § 230, would protect website hosts where deepfakes are housed. Section V then reviews the applicability of copyright law to the question of deepfakes. Finally, the article concludes with some final thoughts about deepfakes and how the United States should act quickly to develop and adopt stronger prevention mechanisms.

I. DEEPFAKES FOR DUMMIES

A. The Science

"[T]he earliest known surviving photograph made in a camera, was taken by Joseph Nicéphore Niépce in 1826 or 1827,"[7] but 150 years passed before people began digitally editing photographs. Photoshop was developed in 1987 by Thomas and John Knoll.[8] "Thomas Knoll, a doctoral candidate in computer vision, was trying to write . . . computer code to display grayscale images on a black-white bitmap monitor."[9] After rewriting the code for color mapping and image formats, as well as creating Adobe's hallmark feature Layers, Photoshop was born.[10] Within a few years, the neologism "photoshopping" entered common vocabulary.[11] By 2006, the verb "photoshop" was entered into the Oxford English Dictionary.[12]

[7] *The First Photograph*, HARRY RANSOM CTR: UNIV. OF TEX. AT AUSTIN, http://www.hrc.utexas.edu/exhibitions/permanent/firstphotograph/ [https://perma.cc/DPM5-JH5M].

[8] Jeff Schewe, *Thomas & John Knoll*, PHOTOSHOPNEWS.COM (Feb. 2000), http://www.photoshopnews.com/feature-stories/photoshop-profile-thomas-john-knoll-10/ [https://perma.cc/BZ8Q-ABFA].

[9] *Id.*

[10] *See id.*

[11] Dictionaries list its origin as the 1990's. *See Photoshop*, ENGLISH OXFORD LIVING DICTIONARIES, https://en.oxforddictionaries.com/definition/photoshop [https://perma.cc/G9SZ-6YGF].

[12] The Oxford English Dictionary officially added the term in 2006. *See September 2006 Update*, OXFORD ENGLISH DICTIONARY, https://public.oed.com/the-oed-today/recent-

Photoshop's ubiquity has given way to "a general belief that manipulated photos are prevalent," "mak[ing] people . . . generally skeptical about the veracity of photos . . . ."[13] These "[c]onvincing Photoshop-esque techniques for video have arrived"[14] in the form of deepfakes. A working knowledge of how deepfakes are made will help inform our ability to address its improper uses.

Deepfakes are created using discriminative algorithms and generative algorithms. "Discriminative algorithms try to classify input data; that is, given the features of a data instance, they predict a label or category to which that data belongs."[15]

> For example, given all the words in an email, a discriminative algorithm could predict whether the message is spam or not-spam. [S]pam is one of the labels, and the bag of words gathered from the email are the features that constitute the input data. When this problem is expressed mathematically, the label is called y and the features are called x. The formulation p(y|x) is used to mean "the probability of y given x", which in this case would translate to "the probability that an email is spam given the words it contains."
>
> So discriminative algorithms map features to labels.[16]

updates-to-the-oed/previous-updates/september-2006-update/ [https://perma.cc/K7AZ-B33T].

[13] Sophie J. Nightingale, Kimberley A. Wade & Derrick G. Watson, *Can People Identify Original and Manipulated Photos of Real-World Scenes?*, 2 COGNITIVE RES.: PRINCIPLES & IMPLICATIONS 30, 40 (2017).

[14] Adrienne Lafrance, *The Technology That Will Make It Impossible for You to Believe What You See*, ATLANTIC (Jul. 11, 2017), https://www.theatlantic.com/technology/archive/2017/07/what-do-you-do-when-you-cannot-believe-your-own-eyes/533154/ [https://perma.cc/8H3C-5H8Y].

[15] *A Beginner's Guide to Generative Adversarial Networks (GANs)*, DEEP LEARNING FOR JAVA, https://deeplearning4j.org/generative-adversarial-network [https://perma.cc/UT2M-TVR8].

[16] *Id*; *see also* Andrew Ng, Lecture at Stanford University: Generative Learning Algorithms, http://cs229.stanford.edu/notes/cs229-notes2.pdf [https://perma.cc/CVD9-2XPZ] ("Consider a classification problem in which we want to learn to distinguish between elephants (y = 1) and dogs (y = 0), based on some features of an animal. Given a training set, [a discriminative algorithm] tries to find a straight line—that is, a decision boundary—that separates the elephants and dogs. Then, to classify a new animal as either an elephant or a dog, it checks on which side of the decision boundary it falls, and makes its prediction accordingly.").

Unlike discriminative algorithms, generative algorithms do not just aim to classify the correct label; rather, a generative model provides a way to *generate* data that looks like it came from the dataset. Instead of predicting a label given certain features, it attempts to predict features given a certain label.[17] Harkening back to the spam/not-spam example: "The question a generative algorithm tries to answer is: Assuming this email is spam, how likely are these features? . . . They allow you to capture . . . the probability of x given y, or the probability of features given a class."[18] Said plainly, generative algorithms assume a classification and establish the particular *features* of the classification.

Generative Adversarial Networks, or GANs, pit these two types of algorithms against one another. GANs were introduced by Ian Goodfellow and other researchers at the University of Montreal.[19] Goodfellow et al. describe it in the following way: GANs are a "framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that generates artificial samples, and a discriminative model D that estimates the probability that a sample came from the training data rather than G."[20] In lay terms, researchers create two separate computer models: "One neural network, called the *generator*, generates new data instances, while the other, the *discriminator*, evaluates them for authenticity; i.e. the discriminator decides whether each instance of data it reviews belongs to the actual training dataset or not."[21] This is also the case with deepfake videos.

> The generator is creating new images that it passes to the discriminator. It does so in the hopes that they, too, will be deemed authentic, even though they are fake. The goal of the generator is to generate passable hand-written digits, to lie without being caught. The goal of the discriminator is to identify images coming from the generator as fake.[22]

[17] *A Beginner's Guide to Generative Adversarial Networks (GANs)*, *supra* note 15.
[18] *Id.*; *see also* Ng, *supra* note 16.
[19] *A Beginner's Guide to Generative Adversarial Networks (GANs)*, *supra* note 15 (citing Ian J. Goodfellow et al., Generative Adversarial Networks (June 10, 2014) (unpublished paper) (https://arxiv.org/pdf/1406.2661.pdf [https://perma.cc/LG7F-G5X8])).
[20] *Id.*
[21] *Id.*
[22] *Id.*

"After enough of this 'training,'" the algorithm is refined enough to "convincingly manipulat[e] video on the fly,"[23] meaning it will generate images into each individual video frame such that when played regularly, the video appears seamless. This process produces a deepfake.

A literary comparison may be instructive. Some of the most famous authors often have their own styles, be it Ernest Hemingway's concise sentences,[24] James Joyce's stream of consciousness,[25] David Foster Wallace's "winding sentences and novelistic footnotes,"[26] or Emily Dickinson's non-traditional meter and punctuation.[27] Even famous jurists are known for their distinct writing styles—e.g. Justice Scalia's exceptional metaphors[28] or Justice Kagan's pragmatism.[29] Well-versed readers may only need a few sentences—if not less—to determine the author.

Now imagine 'training' the discriminator to learn a particular author's style *so* well that it can pinpoint the author's style among numerous texts. The discriminator is then handed a new page of prose produced by the fraudster, or generator. He or she must then determine whether or not the new sample is written by the original author. Or

[23] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, MOTHERBOARD (Dec. 11, 2017, 2:18 PM) [hereinafter Cole, *AI-Assisted Fake Porn Is Here*], https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/8QSX-BU76].

[24] *See, e.g.*, ERNEST HEMINGWAY, A FAREWELL TO ARMS 274 (Hemingway Library ed. 1929) ("She's just having a bad time. The initial labor is usually protracted. She's only having a bad time. Afterward we'd say what a bad time and Catherine would say it wasn't really so bad. But what if she should die? She can't die. Yes, but what if she should die? She can't, I tell you. Don't be a fool. It's just a bad time.").

[25] *Stream of Consciousness*, ENCYCLOPAEDIA BRITANNICA, https://www.britannica.com/art/stream-of-consciousness [https://perma.cc/63QD-RFN3] (citing Joyce as an exemplar of stream of consciousness).

[26] *See* Spencer Kornhaber, *Advice: Don't Try to Write Like David Foster Wallace: Marvel at His Style, but Don't Imitate It*, ATLANTIC (Feb. 3, 2015), https://www.theatlantic.com/entertainment/archive/2015/02/advice-dont-try-to-write-like-david-foster-wallace/384753/ [https://perma.cc/G2H3-3RXA].

[27] *See Major Characteristics of Dickinson's Poetry*, EMILY DICKINSON MUSEUM, https://www.emilydickinsonmuseum.org/poetry_characteristics [https://perma.cc/S8D4-YQJJ].

[28] *See, e.g.*, Yury Kapgan, *What Made Antonin Scalia a Great Writer*, SLATE (Feb. 13, 2016, 8:01 PM), http://www.slate.com/articles/news_and_politics/jurisprudence/2016/02/antonin_scalia_s_writing_assessed.html [https://perma.cc/Y534-HBX5].

[29] Ross Guberman, *The Supreme Writer on the Court: The Case for Kagan*, VOLOKH CONSPIRACY (July 9, 2013, 4:26 PM), http://volokh.com/2013/07/09/the-supreme-writer-on-the-court-the-case-for-kagan [https://perma.cc/ZSC4-4YLZ].

consider an art forger trying to establish the legitimacy of a fake Vincent Van Gogh or Georges Seurat. The two, like many renowned painters, had particular styles and visions. If the counterfeiter hopes to make a passable replica, using the wrong color palette or the improper technique is a dead giveaway. A discerning eye could tell the difference.

Deepfakes simply apply this process from text or art to videos. With deepfakes, the generator constructs new video frames, while the discriminator tries to discern whether the frame, with its superimposed subject, is authentic (say, an actual video frame of the original actor) or fake (a doctored video frame of the actor in a compromising position). If the discriminator cannot tell the real images from the false images, a human may not be able to either.

B. Proliferation: Pornography and Otherwise

After Reddit user @deepfakes posted his creation, other users, called "Redditors," caught on quickly. Within a month of Motherboard's initial reporting of these altered pornographic videos, over 15,000 Redditors subscribed to @deepfakes's dedicated discussion board, posting videos of their own.[30] This rapid proliferation of deepfake videos was possible because the technology needed to create them was already widely available to the public. Additionally, another Redditor, wanting to break down barriers to entry for this technology even further, "created an app specifically designed to allow users without a computer science background to create AI-assisted fake porn. All the tools one needs to make these videos are free, readily available, and accompanied with instructions that walk novices through the process."[31] The app—appropriately titled "FakeApp"—opened the door to even more deepfake creators.[32]

Indeed, Gadot, Williams, and Ridley are not the only female celebrities that have had their likeness grafted onto adult films: Jessica Alba,[33] Natalie Dormer,[34] Scarlett Johansson,[35] Chloe Bennet,[36] Taylor

[30] Cole, *We Are Truly Fucked*, *supra* note 1.

[31] *Id.*

[32] Randall Coburn, *Oh No: They've Moved Past Nic Cage and Are Now Digitally Inserting Trump into Videos*, A.V. CLUB (Feb. 1, 2018), https://www.avclub.com/oh-no-theyve-moved-past-nic-cage-and-are-now-digitally-1822627349 [https://perma.cc/SC9X-C832].

[33] Cole, *We Are Truly Fucked*, *supra* note 1.

[34] Dave Lee, *Deepfakes Porn Has Serious Consequences*, BBC (Feb. 3, 2018), https://www.bbc.com/news/technology-42912529 [https://perma.cc/EN2K-XF97].

Swift,[37] Sophie Turner,[38] Katy Perry,[39] Cara Delevingne,[40] and Aubrey Plaza[41] have become victims of this technology as well. In one example, "a deepfake of Emma Watson taking a shower was reuploaded by CelebJihad—a celebrity porn site that regularly posts hacked celebrity nudes—as a 'never-before-seen video . . . from my private collection, [which] appears to feature Emma Watson fully nude and flaunting her naked sex organs while showering with another girl.'"[42]

Once Reddit itself got wind of this development, it shut down the discussion board where this was posted and banned similar content.[43] Reddit additionally updated its site-wide rules regarding its ban on "involuntary pornography" and "sexual or suggestive content involving minors."[44] The ban on involuntary pornography includes revenge porn, the spreading of private nudes, and any sexualized image "apparently created or posted without [the subject's] permission, including depictions that have been faked."[45] Reddit was not the only website to ban this type of content. For example, Pornhub, one of the Internet's largest databases of online adult videos, established similar bans.[46]

Deepfakers have also focused on generating celebrity videos outside the adult film industry. In a slightly more good-natured use of the technology, the deepfake community turned to one particular movie star for comedic relief—Nicolas Cage.[47] Cage's face was superimposed onto

[35] Cole, *AI-Assisted Fake Porn Is Here*, *supra* note 23.

[36] Cole, *We Are Truly Fucked*, *supra* note 1.

[37] Cole, *AI-Assisted Fake Porn Is Here*, *supra* note 23.

[38] Charlie Warzel, *Pornhub Banned Deepfake Celebrity Sex Videos, but the Site Is Still Full of Them*, BUZZFEED NEWS (Apr. 18, 2018), https://www.buzzfeed.com/charliewarzel/pornhub-banned-deepfake-celebrity-sex-videos-but-the-site?utm_term=.qrB7wqoKr#.ktzLar29G [https://perma.cc/BF2U-JFAP].

[39] *Id.*

[40] *Id.*

[41] Cole, *AI-Assisted Fake Porn Is Here*, *supra* note 23.

[42] Cole, *We Are Truly Fucked*, *supra* note 1.

[43] Aja Romano, *Reddit Finally Bans its Forum for Creepy Fake Celebrity Porn*, VOX (Feb. 8, 2018, 1:00 PM), https://www.vox.com/culture/2018/2/8/16987098/reddit-bans-deepfakes-celebrity-face-swapping-porn [https://perma.cc/5S5W-KWJ7].

[44] *Id.*

[45] *Id.*

[46] Warzel, *supra* note 38. In practice, it appears that Pornhub's ban has been largely ineffective: "[w]hile banned material frequently slips through the cracks on large sites that allow users to upload content, the deepfake violations on Pornhub are especially flagrant." *Id.*

[47] Clayton Purdom, *Deep Learning Technology Is Now Being Used to Put Nic Cage in Every Movie*, A.V. CLUB (Jan. 29, 2018, 12:03 PM), https://www.avclub.com/deep-

Harrison Ford's Indiana Jones in *Raiders of the Lost Ark* and onto Amy Adams' Lois Lane in *Man of Steel*.[48] In one particularly humorous and meta deepfake, Cage's face was superimposed onto Andy Samberg's face in a *Saturday Night Live* sketch in which Samberg was impersonating Cage.[49]

Unsurprisingly, deepfakes have already made their way into the political arena. Admittedly, the groundwork was there; disinformation campaigns have been effectively deployed as political weapons for decades.[50] In a recent iteration of this disgraceful tactic, U.S. Ambassador to Russia Michael McFaul was targeted by a Russian disinformation campaign. McFaul recalled the episode in the *Washington Post*: "State propagandists and their surrogates crudely photoshopped me into pictures, spliced my speeches to make me say things I never uttered and even accused me of pedophilia."[51]

learning-technology-is-now-being-used-to-put-nic-c-1822514573 [https://perma.cc/XD8X-EEZQ].

[48] *Id.*

[49] *Id.* For a compilation of Nicolas Cage deepfakes, see Usersub, *Nick Cage DeepFakes Movie Compilation*, YOUTUBE (Jan. 31, 2018), https://www.youtube.com/watch?v=BU9YAHigNx8 [https://perma.cc/3M7X-XF74]. Cage is not the only actor to be spoofed. *See* Louise McCreesh, *New Tool Swaps Nicolas Cage with Every Actor in Every Film Ever*, DIG. SPY (Jan. 31, 2018), http://www.digitalspy.com/movies/news/a848882/put-nicolas-cage-in-any-film-fakeapp-deepfakes-subreddit-reddit [https://perma.cc/P6QU-8FCK].

[50] *See, e.g.*, Adam Taylor, *Before 'Fake News,' There Was Soviet 'Disinformation'*, WASH. POST (Nov. 11, 2016), https://www.washingtonpost.com/news/worldviews/wp/2016/11/26/before-fake-news-there-was-soviet-disinformation [https://perma.cc/J7K8-TML3]; Philip Girardi, *The CIA's 70-Year History of Disinformation: How the CIA Funded the Opinion Magazines in Europe*, AM. HERALD TRIB. (Jan. 25, 2018), https://freepress.org/article/cia%E2%80%99s-70-year-history-disinformation-how-cia-funded-opinion-magazines-europe [https://perma.cc/YX3D-24QH]. This article avoids any and all discussion of the 2016 Presidential election's alleged disinformation campaign by other sovereign actors.

[51] Notably, the foundation for this sort of political disinformation has already been laid. *See* Michael McFaul, *The Smear That Killed the 'Reset'*, WASH. POST (May 11, 2018), https://www.washingtonpost.com/news/posteverything/wp/2018/05/11/feature/putin-needed-an-american-enemy-he-picked-me [https://perma.cc/6FYQ-DUC5]; *see also* Isaac Chotiner, *"I'm Scared of That World": A Former U.S. Ambassador to Russia on the Disinformation Campaign Against Him—and Russia's Increasingly Sophisticated Attacks on Reality*, SLATE (May 10, 2018), https://slate.com/news-and-politics/2018/05/michael-mcfaul-on-russian-disinformation-and-the-lost-promise-of-medvedev.html [https://perma.cc/2AVH-8L8L] (quoting Ambassador McFaul saying that Russia "would Photoshop my image on posters, and made it sound like I was trying to overthrow the regime.").

Deepfakes with specific face-swaps of U.S. elected leaders in compromising position have not yet been produced. They have, however, been deployed mockingly. For example, two random Internet posters turned President Trump into the main character from the television show *The Office*.[52] Additionally, a quick Google search reveals President Trump's likeness superimposed onto German Chancellor Angela Merkel,[53] onto *Back to the Future* villain Biff Tannen,[54] and onto *Austin Powers* villain Dr. Evil.[55]

Deepfakes are also used to superimpose an average member of the public onto a celebrity's body. As one blogger wrote, "we can leverage these celebrities for other things, such as inserting your friends and family into blockbuster movies and shows!"[56] That blogger then turned his wife's likeness—on the body of Anne Hathaway—into an interviewee opposite David Letterman and a film star opposite Steve Carrell.[57] In his words:

> I personally think it's fun, can be innocent, and even makes for a nice surprise/gift. . . . [N]ow you can put your best friend into his favourite movie: have her dance with Patrick Swayze and have the time of her life, or have an alien burst out of his stomach.[58]

[52] Coburn, *supra* note 32.

[53] PotatoKaboom, *Merkel Trump Deepfake*, YOUTUBE (Jan. 28, 2018), https://www.youtube.com/watch?v=5hZOcmqWKzY [https://perma.cc/7CQG-YEKV].

[54] ZeroCool22, *Donald Trump as Biff Tannen, Back to the Future. (Deepfakes Method)*, YOUTUBE (Feb. 2, 2018), https://www.youtube.com/watch?v=QgXp_trk9DA [https://perma.cc/58GJ-TQYG].

[55] Deep Fried Country, *Dr Evil Trump Deep Fake*, YOUTUBE (Feb. 1, 2018), https://www.youtube.com/watch?v=UKizH9aBifs [https://perma.cc/8KVN-HBYQ].

[56] Sven Charleer, *Family Fun with Deepfakes. Or How I Got My Wife onto the Tonight Show*, SVEN CHARLEER: BLOG (Feb. 2, 2018), http://svencharleer.com/blog/2018/02/02/family-fun-with-deepfakes-or-how-i-got-my-wife-onto-the-tonight-show/ [https://perma.cc/WT5H-KESP].

[57] *Id.*

[58] *Id.*

C. The Current Status and Imminent Growth of Deepfakes: Better Technology, Better Source Material, and an Incentivized Private Sector

In some cases, current deepfakes are very "believable,"[59] yet "[d]eepfake technology remains brittle and prone to failure in many scenarios," according to Tim Hwang, the Director of the Ethics and Governance of AI Initiative at the Harvard Berkman-Klein Center and the MIT Media Lab.[60] "The computing power required to generate a believable fake remains a barrier for casual computer users."[61] Thus, a deepfake video often comes out as "a blurry, semi-believable version" of the targeted victim.[62] Nevertheless, as technology improves, these limiting factors will slowly fall by the wayside.

For example, computing power limitations are already dissipating. "According to [@]deepfakes—who declined to give his identity . . . to avoid public scrutiny—the software is based on multiple open-source libraries."[63] Moreover, "a decent, consumer-grade graphics card could process this effect in hours, but a CPU[64] would work just as well, only more slowly, over days."[65] Thus, Dr. Hwang's technological limitations may soon expire. As average graphics cards and CPUs continue to improve in performance,[66] the time needed to turn individuals into un-

[59] Samantha Cole, *Deepfakes Were Created As a Way to Own Women's Bodies—We Can't Forget That*, VICE: BROADLY (June 18, 2018, 10:10 AM), https://broadly.vice.com/en_us/article/nekqmd/deepfake-porn-origins-sexism-reddit-v25n2 [https://perma.cc/92DE-3JH2] (stating, in response to viewing the deepfake pornography, "holy shit did they look believable").

[60] Jeremy Hsu, *Experts Bet on First Deepfakes Political Scandal*, IEEE SPECTRUM (June 22, 2018, 6:00 PM), https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/experts-bet-on-first-deepfakes-political-scandal [https://perma.cc/W9AY-W3YY].

[61] *Id.*

[62] Kristen Dold, *Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy*, ROLLING STONE (Apr. 17, 2018), https://www.rollingstone.com/culture/culture-features/face-swapping-porn-how-a-creepy-internet-trend-could-threaten-democracy-629275/ [https://perma.cc/M2FW-7ATW].

[63] Cole, *We Are Truly Fucked*, *supra* note 1.

[64] A computer processor unit, the main microchip powering a computer.

[65] *Id.*

[66] *See, e.g.*, Joel Hruska, *Graphene-Coated Copper Could Dramatically Boost Future CPU Performance*, EXTREME TECH (Feb. 21, 2017), https://www.extremetech.com/computing/244693-graphene-coated-copper-dramatically-boost-future-cpu-performance [https://perma.cc/N3V2-XWVJ?type=image]; *After Moore's Law*, ECONOMIST (Mar. 12, 2016), https://www.economist.com/technology-quarterly/2016-03-12/after-moores-law [https://perma.cc/9FSL-BJNJ].

consenting subjects of videos, adult or otherwise, will shrink exponentially.

Another potential limiting factor is sufficient source material, but the digital era in which we live has rendered this factor a non-issue. With only one image of the victim, for example, a person—much less an artificially intelligent algorithm—cannot learn much about him or her. But the more one can augment the dataset, the more realistic the fake will be. For celebrities, a simple Google search provides enough source material. But for the general public, a Google search may be insufficient—for now. As more and more pictures populate the Internet, laypersons are no longer immune to the potentially harmful effects of deepfakes.

Aside from better technology and better source material, an additional propellant in the proliferation of deepfakes is the private sector. As told by Dr. Neil DeGrasse Tyson, one of the foremost advocates for scientific exploration—in truth, deepfakes are simply discoveries and explorations in computer science—"[t]he history of exploration has never been driven by exploration. But Columbus himself was a discoverer. So was Magellan. But the people who wrote checks were not. They had other motivations."[67] For example, in defending his work, @deepfakes pointed out "that he is using an algorithm similar to one developed by Nvidia,"[68] a software capable of turning "snowy roads into summer, and day into night" instantaneously.[69] That the private sector fuels the flame should come as no shock. Indeed, John Knoll, the aforementioned co-creator of Photoshop later worked at Industrial Light and Magic, the visual effects department of Lucasfilm.[70]

The private sector likely continues to see utility in these sorts of technological advances due to their commercial applications—imagine, for instance, paying to correct the fumbling of a best man's toast or a commencement address. Given this success, companies will likely continue to devote substantial resources to advancing and finding new uses for them. How such advances may be misused is yet to be seen. Tellingly, when *Mashable* reached out to Nvidia after @deepfakes marshaled support for his work by citing the company's work, "[t]he Nvidia researchers who developed the algorithm declined to comment on this possible application."[71]

[67] JAMES EGAN, 1000 HISTORIC QUOTES 144 (2015).
[68] Cole, *We Are Truly Fucked*, *supra* note 1.
[69] *Id.*
[70] *See* Schewe, *supra* note 8.
[71] Cole, *We Are Truly Fucked*, *supra* note 1.

D. Related, Truth-Defying Technologies

At least three other technological developments run parallel to deepfakes, arguably as destabilizing, albeit less popular: manipulating video with voice-overs, face-to-face capture and reenactment, and audio-to-video conversion.

1. Audio/Visual Manipulation

Audio/Visual manipulation is the ability to "manipulate and digitally alter the footage of [one speaker] to a script written and performed by [another]."[72] Thus far, this technology has only made its appearance in popular culture once: comedian-filmmaker Jordan Peele produced a video with *Buzzfeed* CEO Jonah Peretti of Peele's voice (impersonating Obama's) onto a video of President Obama. The frightening part about it, however, is that President "Obama's lips move in sync with Peele's voice."[73] Though the distinct audio modification is enough to alert any casual observer, a more pitch-perfect match could fool even a keen ear.[74]

2. Face-to-Face Capture and Reenactment

Face-to-face capture and reenactment technologies internalize every movement of a speaker's face. Then, the software recreates a digital face with the same movements. The researchers' "goal is to animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion."[75] Not only can the technology capture—and replicate in a video—the speaker's motions, it

[72] David Mack, *This PSA About Fake News from Barack Obama Is Not What It Appears*, BUZZFEED NEWS (Apr. 17, 2018, 11:26 AM), https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed [https://perma.cc/J9VJ-AJPF].

[73] Morgan Gstalter, *'Obama' Voiced by Jordan Peele in PSA Video Warning About Fake Videos*, HILL (Apr. 17, 2018, 12:08 PM), http://thehill.com/blogs/in-the-know/in-the-know/383525-obama-voiced-by-jordan-peele-in-psa-video-warning-about-fake [https://perma.cc/J3C6-Q4F4].

[74] *Id.*

[75] Thies et al., Face2Face: Real-time Face Capture and Reenactment of RBG videos (2016) (unpublished paper in connection with the Conference on Computer Vision and Pattern Recognition), https://web.stanford.edu/~zollhoef/papers/CVPR2016_Face2Face/paper.pdf [https://perma.cc/CBV8-Q5WT].

can even capture “real-time facial expressions, including distinct movements such as eyebrow raises.”[76] In practice, “[a]n actor speaks to the webcam and his facial expressions and speech are copied by George Bush, Vladimir Putin, Donald Trump, and Barack Obama.”[77]

3. *Audio-to-Video Conversion*

The final technological development is the ability to “take audio of someone talking and turn that into a realistic video of someone speaking those words.”[78]

> University of Washington researchers have developed new algorithms that solve a thorny challenge in the field of computer vision: turning audio clips into a realistic, lip-synced video of the person speaking those words . . . . In a visual form of lip-syncing, the system converts audio files of an individual’s speech into realistic mouth shapes, which are then grafted onto and blended with the head of that person from another existing video.[79]

These videos are not stilted or robot-like: these fakes include swaying, pacing, facial cues, and other distinctly human ticks. As one researcher stated, “We’re learning how to capture human personas.”[80] One could take snippets of speeches and splice them together in a hyper-realistic way to create a video that has the look and feel of the speaker’s mannerisms.

[76] Adario Strange, *Face-Tracking Software Lets You Make Anyone Say Anything in Real Time*, MASHABLE (Mar. 20, 2016), https://mashable.com/2016/03/20/face-tracking-software/#KP7NWfgiH8qV [https://perma.cc/VGD6-AK6Y].

[77] Mark Burgess, *Make Putin Pout With This Creepy Face-Tracking Tech*, WIRED (Mar. 21, 2016), http://www.wired.co.uk/article/face2face-face-recognition-copy-putin-bush-trump [https://perma.cc/68JC-D4HH].

[78] Lafrance, *supra* note 14.

[79] Jennifer Langston, *Lip-Syncing Obama: New Tools Turn Audio Clips into Realistic Video* (July 11, 2017), UW NEWS, https://www.washington.edu/news/2017/07/11/lip-syncing-obama-new-tools-turn-audio-clips-into-realistic-video/ [https://perma.cc/58S4-472V].

[80] Lafrance, *supra* note 14.

E. Prevention Mechanisms

Due to the realistic nature of the films, scientists are researching new methods to identify when an image has been faked. After the Lawfare blog published multiple essays promoting a grim and foreboding view of deepfakes, Dr. Herb Lin, a senior research scholar for cyber policy and security at the Center for International Security and Cooperation, and Hank J. Holland, Fellow in Cyber Policy and Security at the Hoover Institution, offered a "ray of hope":

> Consider the technology of digital signatures, which enable a party to sign a digital object in such a way that proves he or she was the one who signed it. Now imagine that a vendor produces cameras and sound recorders (i.e., devices) that digitally sign every video or audio file the user creates. Further, the vendor keeps records so that the purchaser of any given device is known in the future—that is, the device and its public signature key is registered in a database accessible to anyone. Any video or audio file released in the future, accompanied by a digital signature, could then be associated with a specific purchaser.
>
> This scheme does not produce 100-percent confidence . . . . But this scheme would certainly provide more confidence in the authenticity of the video or audio than for a video that was *not* accompanied by a signature that could be traced to a registered device.[81]

The private sector is already deploying the first generation of this technology. For example, Canon's Original Data Security Kit "enhances security by providing image data encryption and decryption features in addition to a verification function that authenticates image originality."[82] Nikon also offers a similar software package.[83] Yet, it is worth noting that both such technologies have been hacked and, thus, "rendered useless."[84]

[81] Herb Lin, *The Danger of Deep Fakes: Responding to Bobby Chesney and Danielle Citron*, LAWFARE BLOG (Feb. 27, 2018, 7:00 AM), https://www.lawfareblog.com/danger-deep-fakes-responding-bobby-chesney-and-danielle-citron [https://perma.cc/84SL-3LWC].

[82] *OSK-E3*, CANON, http://www.canon.co.jp/imaging/osk/osk-e3/index.html [https://perma.cc/VAB3-MLAS].

[83] *See Image Authentication Software*, NIKON, http://imaging.nikon.com/lineup/software/img_auth/index.htm [https://perma.cc/7MGM-D2WY] (stating that Nikon's Image Authentication software "enables the authentication

Digital-watermarks are not the only possible solution. There are multiple large-scale projects in academia, industry, and government aimed at ferreting out manipulated and falsified images grounded in genuine ones.[85] A Columbia University project, for example, is taking pedestrian reverse image search technology[86] "to the next level, and starting to find parts of images that have been repurposed from other images."[87] Thus, if even *part* of an image is identical to another publicly available image, the technology can flag it as a potentially manipulated image. This technology may well be expanded to videos.

Additionally, digital color analysis may provide a possible solution. Dr. Hany Farid, a computer science professor at Dartmouth College, said:

> Almost every image is stored in a JPEG file, which throws away some information to save on storage. There is a huge amount of variation in how each camera does that. If a JPEG is unpacked—opened in Photoshop—and then put back together, it is always repackaged slightly differently, and we can detect that.[88]

of an image captured by the camera and can determine whether or not it has been altered since capture, verifying the image as well as information attached to it").

[84] Eric Reagan, *Canon "Original Data Security" Cracked, Rendered Useless*, PHOTOGRAPHY BAY (Nov. 30, 2010), http://www.photographybay.com/2010/11/30/canon-original-data-security-cracked-rendered-useless/ [https://perma.cc/MCF2-6SGH]; *see also* John E. Dunn, *Nikon's Image Authentication Algorithm Cracked*, COMPUTERWORLD (Apr. 28, 2011), https://www.computerworld.com/article/2508282/cybercrime-hacking/nikon-s-image-authentication-algorithm-cracked.html [https://perma.cc/C5CC-9JUJ]. "[I]ronically," Lin notes, both were cracked "by [the same] well-known Russian company." Lin, *supra* note 81.

[85] *See* Karen Hao, *Deepfake-Busting Apps Can Spot Even a Single Pixel Out of Place*, MIT TECH. REV. (Nov. 1, 2018), https://www.technologyreview.com/s/612357/deepfake-busting-apps-can-spot-even-a-single-pixel-out-of-place [https://perma.cc/4KBZ-XEYF].

[86] *See, e.g.*, *Frequently Asked Questions,* TINEYE, https://www.tineye.com/faq#what [https://perma.cc/VBN2-4ANT] ("TinEye is a reverse image search engine. You can submit an image to TinEye to find out where it came from, how it is being used, if modified versions of the image exist, or to find higher resolution versions. TinEye uses image recognition technology rather than keywords, metadata or watermarks.").

[87] Elizabeth Gibney, *The Scientist Who Spots Fake Videos*, NATURE (Oct. 6, 2017), https://www.nature.com/news/the-scientist-who-spots-fake-videos-1.22784 [https://perma.cc/GS4P-G4CC] (citing *DARPA MEMEX Project*, COLUM. UNIV., http://www.ee.columbia.edu/ln/dvmm/memex/index.html#About [https://perma.cc/C46K-AQX2]).

[88] *Id.*

Farid notes that that technology has an equivalent for videos "based on the observation that computer-generated content lacks the imperfections that are present in a recorded video. It's created in almost too perfect a world. So one of the things we look at is, are we not seeing the statistical and geometric patterns we'd expect to see in the physical world?"[89]

A final technological breakthrough is based on humans' natural blood flow. An algorithm perceives "periodic pulsatile motion within a narrow temporal passband centered around the heart rate" in tissue.[90] In lay terms, it can calculate one's pulse by measuring the frequency of subtle color changes to tissue. This technology allows us to measure a person's pulse in a video of him or her speaking into a camera. It can, therefore, flag computer-altered or computer-generated videos because a computer-generated video of humans would not exhibit these subtle changes.[91]

F. In Defense of Technological Exploration

Despite the potential invidious uses of this technology, there are many possible benefits and important uses for this technology as well. For example:

> Dr. [Louis-Philippe] Morency [the director of Carnegie Mellon University's MultiComp Lab] said soldiers suffering from post-traumatic stress disorder could eventually video-conference with doctors using similar technology. An individual could face-swap with a generic model without sacrificing the ability to convey his or her

[89] *Id.*

[90] Alborz Amir-Khalili et al., *Auto Localization and Segmentation of Occluded Vessels in Robot-Assisted Partial Nephrectomy*, 17 MED. IMAGE COMPUTING & COMPU.-ASSISTED INTERVENTION 407, 409 (2014), https://www.cs.sfu.ca/~hamarneh/ecopy/miccai2014d.pdf [https://perma.cc/64AQ-FBL4].

[91] The Defense Advanced Research Projects Agency (DARPA) has also dedicated resources to the Medical Forensics, or MediFor, Project, which aims to "automatically detect manipulations, provide detailed information about how these manipulations were performed, and reason about the overall integrity of visual media to facilitate decisions regarding the use of any questionable image or video." Dr. Matt Turek, *Media Forensics, (MediFor)*, DEF. ADV. RES. PROJECTS AGENCY, https://www.darpa.mil/program/media-forensics [https://perma.cc/2CQT-GH5C].

> emotions. In theory, this would encourage people to get treatment who might otherwise be deterred by a perceived stigma, and the quality of their treatment wouldn't suffer due to a doctor being unable to read their facial cues.
>
> Another one of Dr. Morency's possibilities—and its own can of worms—would be to use models in video interviews to remove gender or racial bias when hiring.[92]

Additionally, deepfakes are created using advanced machine learning technologies, which have a wide array of applications, from driverless cars to facial recognition software. Derailing research that improves deepfake technology, thus, may inadvertently impact these other industries as well. Further, research into deepfake technology may have unexpected positive impacts. It is often the case that "great achievement has no road map. The X-Ray [and] penicillin[,] neither were discovered with a practical objective in mind. [W]hen the electron was discovered in 1897, it was useless. And now we have an entire world run by electronics."[93]

Deepfakes have quickly permeated multiple facets of society, from parody to pornography, using this fascinating new technology. What's more, deepfakes are on the rise: in part due to the private sector's economic incentive, in part from academia's push for exploration, and in concert with other interesting yet precarious technologies. With this deeper understanding of deepfakes—their history, proliferation, and related technologies—we can begin to apply it to the law.

II. DEEPFAKES AND THE FIRST AMENDMENT

Before questioning how victims of deepfakes can seek legal redress, a threshold question must be addressed: does the First Amendment protect deepfakes and deepfakers? More directly, can the government regulate, if not altogether ban, the production and dissemination of deepfakes?[94] To address this question, we begin with first principles.

[92] 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/#dnpjFgfXHqqb [https://perma.cc/47GH-DCQA].

[93] *The West Wing: Dead Irish Writers* (NBC television broadcast Mar. 6, 2002).

[94] As stated previously, lawmakers are indeed considering legislative action. *See* Breland, *supra* note 5.

A. First Amendment, First Principles

The First Amendment provides that: "Congress shall make no law . . . abridging the freedom of speech."[95] "Under that Clause, a government, including a municipal government vested with state authority, 'has no power to restrict expression because of its message, its ideas, its subject matter, or its content.'"[96] For that reason, "content-based restrictions on speech [are] presumed invalid[, and] the Government bear[s] the burden of showing their constitutionality.[97] In lay terms, this rule means that no matter how abhorrent one may find a message—from promoting Nazism to segregation—and no matter the size of the group that agrees with a position, disagreement with the message alone is insufficient to merit inhibiting the speaker's permission to enter the "marketplace of ideas"[98] without a compelling government interest.

Per the Supreme Court's 2015 holding in *Reed v. Town of Gilbert, Arizona*, "[g]overnment regulation of speech is content-based if a law applies to particular speech because of the topic discussed or the idea or message expressed," thereby distinguishing between the speaker and the message.[99] While previous constitutional doctrine may have deployed case-specific rules depending on the subject matter of said speech, "[t]he majority opinion in *Reed* effectively abolishes any distinction between content regulation and subject-matter regulation."[100]

"If a law is unconstitutional [because] its restrictions 'depend entirely on the communicative content' of what is regulated, then any restriction of revenge pornography is in deep trouble."[101] This reasoning applies with equal force to deepfakes. Therefore, despite both federal[102] and state[103] officials' recent interest in enacting legislation to curb

[95] U.S. CONST. amend. I.

[96] Reed v. Town of Gilbert, Ariz., 135 S. Ct. 2218, 2226 (2015) (quoting Police Dept. of Chi. v. Mosley, 408 U.S. 92, 95 (1972)).

[97] United States v. Alvarez, 567 U.S. 709, 716–17 (2012) (quotation and citation omitted).

[98] Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 52 (1988).

[99] 135 S. Ct. at 2227.

[100] Norton v. City of Springfield, 806 F.3d 411, 412 (7th Cir. 2015).

[101] Andrew Koppelman, *Revenge Pornography and First Amendment Exceptions*, 65 EMORY L.J. 661, 665 (2016) (footnote omitted) (quoting *Reed*, 135 S. Ct. at 2227).

[102] *See* Breland, *supra* note 5.

[103] *See, e.g.*, Katyanna Quach, *New York State is Trying to Ban 'Deepfakes' and Hollywood isn't Happy*, 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://perma.cc/MP57-YLPJ].

deepfakes, any law banning, or even regulating, deepfakes would be presumptively invalid, given that such a law would fall squarely into content-based or message-based regulation.

Thankfully, that's not the end of the story. Despite its sweeping language, "it is well understood that the right of free speech is not absolute at all times and under all circumstances."[104] As the Supreme Court has affirmed:

> There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or 'fighting' words—those which by their very utterance inflict injury or tend to incite an immediate breach of the peace.[105]

And the government has indeed availed itself of this semi-permeable bar.[106] To pass constitutional muster, deepfake regulations must fall into one of these exceptional categories; whether it in fact does so, however, is dubious at best.

B. Exceptions

Certain obscenities, including child pornography, are exempt from First Amendment's protections.[107] And indeed, deepfakes may be, and likely are, used to create obscene and child-pornographic videos. Thus, these issues are examined with regard to this new technology. Ultimately, a regulation or an outright ban on deepfakes is unlikely to fit neatly within the obscenity or child pornography exceptions set out in our nation's First Amendment jurisprudential framework because not all uses of deepfakes

[104] Chaplinsky v. New Hampshire, 315 U.S. 568, 571 (1942); *see also* Whitney v. California, 274 U.S. 357, 371 (1927) ("That the freedom of speech which is secured by the Constitution does not confer an absolute right . . . is not open to question.").

[105] *Chaplinsky*, 315 U.S. at 571–72 (footnotes omitted).

[106] *See* Brandenburg v. Ohio, 395 U.S. 444, 447 (1969); *see also* Virginia v. Black, 538 U.S. 343, 359 (2003) (holding that the First Amendment does not protect "true threats" of violence); In re R.M.J., 455 U.S. 191, 203 (1982) ("Truthful advertising related to lawful activities is entitled to the protections of the First Amendment"); Cohen v. California, 403 U.S. 15, 18 (1971) (holding that the First Amendment does not protect the "intent to incite disobedience").

[107] *See Chaplinsky*, 315 U.S. at 571–72

are obscene. Therefore, such a law is unlikely to withstand judicial scrutiny.

1. Obscenity

That “obscenity is not within the area of constitutionally protected speech or press”[108] is beyond contestation.[109] The Supreme Court first recognized this in 1942.

> It has been well observed that [lewd and obscene] utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.[110]

The Court re-affirmed this notion in the following decade, explaining that the history of regulating obscene speech further supported the rejection of First Amendment protections for obscenities.

> [I]mplicit in the history of the First Amendment is the rejection of obscenity as utterly without redeeming social importance. This rejection for that reason is mirrored in the universal judgment that obscenity should be restrained, reflected in the international agreement of over 50 nations, in the obscenity laws of all of the 48 States, and in the 20 obscenity laws enacted by the Congress from 1842 to 1956.[111]

The more complex question is what constitutes obscenity. This question remains particularly germane with respect to deepfakes. Justice Potter Stewart famously—or infamously—failed to delineate a bright line rule as to what constitutes obscenity: “I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it[.]”[112] Yet, for lower courts and the rule of law, the Supreme Court later pronounced a three-pronged inquiry

[108] Roth v. United States, 354 U.S. 476, 485 (1957).
[109] *See, e.g.*, Miller v. California, 413 U.S. 15, 24 (1973).
[110] *Chaplinsky*, 315 U.S. at 571–72.
[111] *Roth*, 354 U.S. at 484–85.
[112] Jacobellis v. State of Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring).

in *Miller v. California*[113] to determine when speech crosses from cringe-inducing, yet protected, speech into unprotected obscenities.

> The basic guidelines for the trier of fact must be: (a) whether 'the average person, applying contemporary community standards' would find that the work, taken as a whole, appeals to the prurient interest; (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.[114]

To this point, "[t]he *Miller* standard . . . was an accommodation between the State's interests in protecting the 'sensibilities of unwilling recipients' from exposure to pornographic material and the dangers of censorship inherent in unabashedly content-based laws."[115] Given this multi-pronged standard, whether deepfakes, or computer-generated pornography, are obscene is not easily answered.

That being said, what constitutes obscenity is a decision left up to each individual state to decide for itself.[116] For example, the District of Columbia has determined that under the District's statute barring obscenity, materials depicting or live performances of oral sex are per se obscene, meaning "the Government need not proffer any evidence of national community standards."[117] Similarly, the Court of Appeals of South Carolina has determined "[n]ude dancing per se is not illegal."[118]

[113] Justice Stewart's *Jacobellis* concurrence was far from the only time the Justices were unable to determine a standard. "Apart from the initial formulation in *Roth*, no majority of the Court has at any given time been able to agree on a standard to determine what constitutes obscene, pornographic material subject to regulation under the States' police power." *Miller*, 413 U.S. at 22 (citing Redrup v. New York, 386 U.S. 767, 770–71 (1967)).

[114] *Id.* at 24.

[115] New York v. Ferber, 458 U.S. 747, 756 (1982); *cf.* Reno v. American Civil Liberties Union, 521 U.S. 844, 874 (1997) ("In evaluating the free speech rights of adults, we have made it perfectly clear that '[s]exual expression which is indecent but not obscene is protected by the First Amendment.'").

[116] *Miller*, 413 U.S. at 32–33 ("It is neither realistic nor constitutionally sound to read the First Amendment as requiring that the people of Maine or Mississippi accept public depiction of conduct found tolerable in Las Vegas, or New York City. . . . People in different States vary in their tastes and attitudes, and this diversity is not to be strangled by the absolutism of imposed uniformity.").

[117] Morris v. U.S., 259 A.2d 337, 340 (D.C. 1969).

[118] State v. Bean, 490 S.E.2d 16, 17 (S.C. 1997) (citations omitted).

Because deepfakes are simply images dynamically superimposed onto preexisting videos, whether they are obscene depends solely on whether the underlying video is deemed obscene.[119] As applied, deepfakes' obscenity is therefore coterminous with the obscenity of the underlying video. For example, if a state were to ban a particular type of pornographic video (*e.g.*, a video depicting rape) as obscene, this ban would extend to deepfakes superimposing one's face onto an unedited video that violated this provision. However, if the state did not ban the original pornographic video, its deepfake counterpart would similarly be permitted. Therefore, deepfakes are not on their face obscene speech. As such they require some protection, meaning any legislation regulating deepfakes would not pass muster under this obscenity exception.

2. *Child Pornography*

"The Supreme Court has repeatedly recognized that children are different than adults, and that . . . justice systems must reflect that."[120] This difference is reflected in the treatment of child pornography.[121] Despite the protections afforded to pornography, the Supreme Court has held that "States are entitled to greater leeway in the regulation of pornographic depictions of children."[122] In arriving at this conclusion, the Supreme Court in *New York v. Ferber* considered four

[119] *See infra* notes 130–133 and accompanying text (explaining that, under Supreme Court precedent, a video that purported to depict child pornography was not *per se* obscene because the actual underlying video was of consenting adults modified via computer-generated images).

[120] B.R. v. McGivern, 714 F. App'x 528, 538 (6th Cir. 2017) (Stranch, J., concurring) (citing Miller v. Alabama, 567 U.S. 460, 471, 473, 477–78 (2012)) (affirming that children are "constitutionally different" from adults and that the "characteristics" and "incompetencies" of youth, including their lack of sophistication in dealing with the criminal justice system, must be taken into account); *see also* J. D. B. v. North Carolina, 564 U.S. 261, 264–65 (2011) (holding that "a child's age properly informs the *Miranda* custody analysis" because it is "beyond dispute that children will often feel bound to submit to police questioning when an adult in the same circumstances would feel free to leave"); Graham v. Florida, 560 U.S. 48, 68 (2010) (acknowledging "fundamental differences" between adults and youth); Roper v. Simmons, 543 U.S. 551, 569–70 (2005) (consulting scientific studies, among other sources, in recognizing that developmental and environmental differences, such as immaturity and lesser control over their environments, can result in young people being "more vulnerable or susceptible to negative influence").

[121] *Miller*, 567 U.S. at 480–81.

[122] New York v. Ferber, 458 U.S. 747, 756 (1982).

independent substantive reasons[123] why child pornography is not protected by the First Amendment: (1) "a State's interest in 'safeguarding the physical and psychological well-being of a minor' is 'compelling'"[124]; (2) "[t]he distribution of photographs and films depicting sexual activity by juveniles is intrinsically related to the sexual abuse of children"[125]; (3) "[t]he advertising and selling of child pornography provide an economic motive for and are thus an integral part of the production of such materials, an activity illegal throughout the Nation"[126]; and (4) "[t]he value of permitting live performances and photographic reproductions of children engaged in lewd sexual conduct is exceedingly modest, if not *de minimis*."[127]

Thus, while adult pornography—save for truly obscene images therein—may not be obscene, the same images of children are deemed obscene. It would be reasonable to assume, therefore, that deepfakes involving children are necessarily not subject to the First Amendment's strong shield, but the inquiry does not end there.

In 2001, the Supreme Court heard oral argument for *Ashcroft v. Free Speech Coalition*.[128] The case concerned the expansion of the Child Pornography Prevention Act of 1996 to include not only pornographic images made using actual children but also "'any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture,' that 'is, or appears to be, of a minor engaging in sexually explicit conduct.'"[129]

The Court held, however, that banning virtual child pornography, or pornography depicting children created entirely through computer-generated graphics, went too far:

> Where the images are themselves the product of child sexual abuse, *Ferber* recognized that the State had an interest in stamping it out without regard to any judgment

[123] The Court's fifth and final justification was that this determination was "incompatible with [its] earlier decisions." *Ferber*, 458 U.S. at 763.

[124] *Id.* at 756–57 (1982) (quoting Globe Newspaper Co. v. Superior Court, 457 U.S. 596, 607 (1982)); *cf.* Prince v. Massachusetts, 321 U.S. 158, 168 (1944) ("A democratic society rests, for its continuance, upon the healthy, well-rounded growth of young people into full maturity as citizens.").

[125] *Ferber*, 458 U.S. at 759.

[126] *Id.* at 761.

[127] *Id.* at 762.

[128] 535 U.S. 234 (2002) (quoting 18 U.S.C. § 2256(8)(B)).

[129] *Id.* at 241.

> about its content. The production of the work, not its content, was the target of the statute. . . .
>
> *Ferber* upheld a prohibition on the distribution and sale of child pornography, as well as its production, because these acts were "intrinsically related" to the sexual abuse of children in two ways. First, as a permanent record of a child's abuse, the continued circulation itself would harm the child who had participated. . . . Second, because the traffic in child pornography was an economic motive for its production, the State had an interest in closing the distribution network. . . . Under either rationale, the speech had what the Court in effect held was a proximate link to the crime from which it came. . . .
>
> In contrast to the speech in *Ferber,* speech that itself is the record of sexual abuse, the CPPA prohibits speech that records no crime and creates no victims by its production. Virtual child pornography is not 'intrinsically related' to the sexual abuse of children, as were the materials in *Ferber.* While the Government asserts that the images can lead to actual instances of child abuse, the causal link is contingent and indirect. The harm does not necessarily follow from the speech, but depends upon some unquantified potential for subsequent criminal acts.[130]

Deepfakes fall into this second *Ashcroft* category. Even if deepfakes were to involve children, they are not necessarily created with the sexual abuse and exploitation of children. As explained *infra*, an intelligent algorithm merely needs perfectly appropriate and normal pictures of minors—e.g. a child playing at the beach—to twist them into a child pornographic deepfake. Thus, as abhorrent as we may consider superimposed underage children in illicit videos, the First Amendment likely protects these deepfakes, notwithstanding the child exploitation exception.[131]

[130] *Id.* at 249–50.

[131] It is worth noting that Justice Thomas concurred in the judgment but was taken by the government's "prosecution rationale—that persons who possess and disseminate pornographic images of real children may escape conviction by claiming that the images are computer generated, thereby raising a reasonable doubt as to their guilt." *Id.* at 259.

> While this speculative interest cannot support the broad reach of the CPPA [in 2002,] technology may evolve to the point where it becomes impossible to enforce actual child pornography laws because the

Due to deepfakes' inputs and outputs, they are unlikely candidates to fall within the First Amendment's obscenity or child pornography exceptions. Thus, laws barring or even regulating their creation are unlikely to survive First Amendment litigation. Victims must therefore utilize other means of protection—namely, deterrence via civil litigation.

III. STATE REMEDIES: DEFAMATION, PRIVACY TORTS, AND THE RIGHT OF PUBLICITY

In light of the potential havoc deepfakes and related technologies can wreak, scholars and legislators alike ought to consider how to structure relevant legal regimes. The constitutionality of proactive legislation is dubious (discussed *supra*). Thus, the focus on rectifying harms to victims should explore other methods to obtaining just ends.

Currently, scholarly literature is limited in its exploration of this subject.[132] Discussions of deepfakes reside almost exclusively in newspapers, magazines, and online articles, many of which are cited throughout this article. In such periodicals, one conclusion is clear, although it lacks expounding analysis: the obvious remedy is state tort law. Indeed, Rebecca Crootof, executive director of the Information Society Project and a research scholar and lecturer in law at Yale Law School, "suggested that tort law may be the better mechanism for dealing

> Government cannot prove that certain pornographic images are of real children. In the event this occurs, the Government should not be foreclosed from enacting a regulation of virtual child pornography that contains an appropriate affirmative defense or some other narrowly drawn restriction.

Id. It may well be the case that Justice Thomas—and potentially his robed brethren—consider deepfakes and the algorithmic technology powering them to have arrived at this point. However, Thomas noted that "the Government asserts only that defendants *raise* such defenses, not that they have done so successfully. In fact, the Government points to no case in which a defendant has been acquitted based on a 'computer-generated images' defense." *Id.* Therefore, while deepfakes may bring us closer to Thomas' perceived inflection point, there remains no such case in which deepfake technology served as the foundation of a successful defense. Until such a case occurs, Thomas and others are unlikely to view this as persuasive justification to overturn *Ashcroft*.

[132] At the time of this article's publication, only one substantive law review-type paper has substantively evaluated deeepfakes, though it approaches the subject far differently than this one. *See* Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. (forthcoming 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954 [https://perma.cc/2YLZ-LXVZ].

with Deepfakes technology on a 'tailored, case-by-case basis' in courtrooms."[133]

This section surveys multiple state tort causes of action. Because common law tort actions are filed pursuant to state law, such actions must take into account two nuances of common law doctrines. First, different standards apply in each jurisdiction. While these causes of action are similar across all fifty states, they are not identical. However, for simplicity, this article uses a commonly accepted standard for the causes of action as captured in texts and treatises.

Second, in nearly all jurisdictions, liability standards are victim-dependent. When victims are private citizens, they are afforded an increased measure of protection.[134] On the other hand, courts employ a heightened standard—"actual malice"—in privacy tort actions when brought by public officials and public figures.[135]

[133] Hsu*, supra* note 60.

[134] "[T]he 'actual malice' standard does not apply to the tort of appropriation of a right of publicity." Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 52 (1988) (citing and describing the holding of Zacchini v. Scripps-Howard Broad. Co.*,* 433 U.S. 562 (1977)).

[135] What constitutes a public official or public figure is not clearly defined. In the Supreme Court's own words:

> In some instances an individual may achieve such pervasive fame or notoriety that he becomes a public figure for all purposes and in all contexts. More commonly, 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. In either case such persons assume special prominence in the resolution of public questions.

Gertz v. Robert Welch, Inc., 418 U.S. 323, 351 (1974). The designation has also been further bifurcated and delineated:

> Extending the *Sullivan* line of cases, the United States Supreme Court has identified two types of public figures for purposes of defamation: all purpose public figures, such as politicians, who are widely recognized, and limited purpose public figures, who may not be well known on every issue but who are sufficiently involved in a particular area to be considered as public figures for that purpose.

Medure v. New York Times Co., 60 F. Supp. 2d 477, 484 (W.D. Pa. 1999). Despite these uncertain labels, some categories of persons are undisputed. For example, "[a] politician is the archetypal public figure." Jones v. New Haven Register, Inc., No. 393657, 2000 WL 157704, at *2 (Conn. Super. Ct. Jan. 31, 2000) (citing *Partington v. Bugliosi,* 825 F.Sup. 906, 917 (D. Hawaii 1993)). Courts have also recognized "actors and actresses, professional athletes, public officers, noted inventors, explorers, [and] war heroes," among others, as public figures. Carlisle v. Fawcett Publications, Inc., 201 Cal. App. 2d 733, 746 (Ct. App. 1962).

Use of the actual malice standard emanates from the Supreme Court's "landmark"[136] opinion in *New York Times Co. v. Sullivan*.[137] In *Sullivan*, the *New York Times* published a full-page advertisement titled "Heed Their Rising Voices."[138] In the Supreme Court's own words, "[i]t is uncontroverted that some of the statements contained in the [advertisement] were not accurate descriptions of events which occurred in Montgomery."[139]As a result, Sullivan, the subject of the inaccurate statement, filed a libel claim against both Alabama signatories to the ad[140] and the *Times* itself.[141]

Quoting Justice Brandeis' famed concurrence[142] in *Whitney v. California*,[143] the Court held: "Believing in the power of reason as applied through public discussion, [the Founding Fathers] eschewed silence coerced by law—the argument of force in its worst form. Recognizing the

[136] *E.g.*, Obsidian Fin. Group, LLC v. Cox, 740 F.3d 1284, 1289 (9th Cir. 2014); Railey v. Webb, 540 F.3d 393, 404 n.4 (6th Cir. 2008); Moore v. Vislosky, 240 F. App'x 457, 464 (3rd Cir. 2007); Wells v. Liddy, 186 F.3d 505, 520 (4th Cir. 1999); Jefferson Cty. Sch. Dist. No. R-1 v. Moody's Inv'r.'s Servs., Inc., 175 F.3d 848, 852 (10th Cir. 1999); Woods v. Evansville Press Co., Inc., 791 F.2d 480, 483 (7th Cir. 1986); Walker v. Pulitzer Pub. Co., 394 F.2d 800, 803 (8th Cir. 1968).

[137] 376 U.S. 254 (1964).

[138] *Id.* at 256. The advertisement, which was paid for and "signed at the bottom of the page by the 'Committee to Defend Martin Luther King and the Struggle for Freedom in the South,'" *id.* at 257, purported to solicit funds to defend Dr. King against two charges of perjury, for which he was indicted in Alabama. *Id.* at 257, 259–60. In addition to this solicitation, the advertisement discussed other important incidents related to the Montgomery, Alabama Police department: arrests of Dr. King "for 'speeding,' 'loitering' and similar 'offenses,'" *id.* at 258, as well as allegations that the police department attempted to "starve [protestors] into submission," *id.* at 257. L.B. Sullivan, on the other hand, was one of the three elected Commissioners of the City of Montgomery, Alabama and was in charge of supervising, among other things, the Police and Fire Departments. *See id.* at 256. Moreover, Sullivan was not involved in Dr. King's arrests, he was not elected when Dr. King's home had been bombed, and his police department had not been subsequently implicated in the attack, despite allegations to the contrary in the advert. *Id.* at 259.

[139] *Id.* at 258.

[140] The signatories included celebrities such as Harry Belafonte, Marlon Brando, Nat King Cole, Sammy Davis Jr., Mahalia Jackson, Langston Hughes, Sidney Poitier, and Jackie Robinson as well as civil rights leaders and politicians like Rev. Ralph Abernathy, John Lewis, and Eleanor Roosevelt. *See Heed Their Rising Voices*, N.Y. TIMES (Mar. 29, 1960), http://recordsofrights.org/records/177/heed-their-rising-voices [https://perma.cc/43MA-URSV].

[141] *Sullivan*, 376 U.S. at 256.

[142] Brandeis was joined by another "Supreme Court Superstar," Oliver Wendell Holmes. Bernard Schwartz, *Supreme Court Superstars: The Ten Greatest Justices*, 31 TULSA L. J. 93 (2013).

[143] 274 U.S. 357, 372–80 (1927).

occasional tyrannies of governing majorities, they amended the Constitution so that free speech and assembly should be guaranteed."[144] Therefore,

> all officials are protected unless actual malice can be proved. The reason for the official privilege is said to be that the threat of damage suits would otherwise "inhibit the fearless, vigorous, and effective administration of policies of government" and "dampen the ardor of all but the most resolute, or the most irresponsible, in the unflinching discharge of their duties."[145]

The Supreme Court has since applied its "actual malice" standard to other torts involving public figures[146] beyond defamation, including the intentional infliction of emotional distress,[147] and state courts have also extended this standard to causes of action alleging a false light invasion of privacy.[148]

All deepfakes, by definition, rise to the level of actual malice, should that standard apply.[149] Per the Supreme Court's *New York Times Co. v. Sullivan* ruling, "actual malice" equates to "knowledge that it was false or with reckless disregard of whether it was false or not."[150]

> Reckless disregard, it is true, cannot be fully encompassed in one definition . . . . [R]eckless conduct is not measured by whether a reasonably prudent man would have published, or would have investigated before publishing. There must be sufficient evidence to permit the conclusion

[144] *Sullivan*, 376 U.S. at 259 (quoting *Whitney*, 274 U.S. at 375–76).

[145] *Id.* at 282 (quoting Barr v. Matteo, 360 U.S. 564, 571 (1959)).

[146] While *Sullivan* involved a public "official," the Supreme Court extended this protection to all public figures just three years later. *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 335–36 (1974) (citing Associated Press v. Walker, 388 U.S. 130, 162 (1967)).

[147] *See* Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 56 (1988).

[148] *E.g.*, West v. Media Gen. Convergence, Inc., 53 S.W.3d 640, 647 (Tenn. 2001) (citing Goodrich v. Waterbury Republican–American, Inc., 448 A.2d 1317, 1330 (Conn. 1982); Lovgren v. Citizens First Nat'l Bank of Princeton, 534 N.E.2d 987, 991 (Ill. 1989); McCall v. Courier–Journal & Louisville Times Co., 623 S.W.2d 882, 888 (Ky. 1981); Dean v. Guard Pub. Publ'g Co., Inc., 699 P.2d 1158, 1160 (Or. 1985)). Notably, the "actual malice" standard does not apply to the tort of appropriation of a right of publicity. *See Hustler*, 485 U.S. at 52 (citing Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562 (1977)).

[149] *See, e.g.*, Ashby v. Hustler Mag., Inc., 802 F.2d 856, 860 (6th Cir. 1986).

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

> that the defendant in fact entertained serious doubts as to the truth of his publication. Publishing with such doubts shows reckless disregard for truth or falsity and demonstrates actual malice.[151]

Because deepfakers create a given deepfake video by combining two distinct sources into one, its creator must *know* the final result is fraudulent, thereby satisfying the standard.

A. Defamation

"In today's world, one's good name can too easily be harmed through publication of false and defaming statements on the Internet."[152] Deepfakes are an archetypal example of that.

Defamation is one means of civil recourse for pursuing deepfakers. "A defamatory statement is defined as a communication that tends to harm the reputation of another as to lower him in the estimation of the community or to deter third persons from associating or dealing with him."[153] "In American law, defamation is not about compensating for damage done to a false reputation by the publication of hidden facts. Instead, it protects a good reputation honestly earned."[154] Victims of deepfakes may be entitled to recovery under a defamation action. For example, an individual who spends a lifetime cultivating a given reputation only to have it obliterated by a fraudulent video depicting false actions, which he or she appears to partake in, when those actions run contrary to said reputation, likely has a cause of action that satisfies the standard for defamation.

"Defamation is the generic term for the twin torts of libel and slander."[155] But the line of demarcation between the two is neither clear nor settled. The Restatement reads:

> (1) Libel consists of the publication of defamatory matter by written or printed words, by its embodiment in physical

[151] St. Amant v. Thompson*,* 390 U.S. 727, 730–31 (1968).

[152] W.J.A. v. D.A., 210 N.J. 229, 248 (2012).

[153] Gleason v. Smolinski, 125 A.3d 920, 947 (Conn. 2015) (citations omitted).

[154] Bustos v. A & E Television Networks, 646 F.3d 762, 764 (10th Cir. 2011); *see also* Hancock v. Variyam, 400 S.W.3d 59 (Tex. 2013) ("Defamation is generally defined as the invasion of a person's interest in her reputation and good name.").

[155] RODNEY SMOLLA, 1 LAW OF DEFAMATION § 1:10 (2d ed. Nov. 2018).

> form or by any other form of communication that has the potentially harmful qualities characteristic of written or printed words.
> (2) Slander consists of the publication of defamatory matter by spoken words, transitory gestures or by any form of communication other than those stated in Subsection (1).
> (3) The area of dissemination, the deliberate and premeditated character of its publication and the persistence of the defamation are factors to be considered in determining whether a publication is a libel rather than a slander.[156]

A video, obviously, is neither written nor printed. As a result, whether a video should fall under libel or slander is not unanimously settled law.

As recently as 1998, the Tennessee Court of Appeals observed that "[t]here is no clear consensus among" courts as to "whether a television broadcast should be designated as libel or slander."[157] Indeed, some courts have adopted a more discerning test, mandating some proof of a prepared script or historical record.[158]

The Restatement, however, states definitively that "[b]roadcasting of defamatory matter by means of radio or television is libel, whether or not it is read from a manuscript."[159] And indeed, many courts have adopted this reasoning,[160] including more recent reviews of the

[156] RESTATEMENT (SECOND) OF TORTS, § 568 (1977).

[157] Ali v. Moore, 984 S.W.2d 224, 227 (Tenn. Ct. App. 1998).

[158] *See* Charles Parker Co. v. Silver City Crystal Co., 116 A.2d 440, 443 (Conn. 1955) ("The basis of the distinction between libel and slander is the written or printed word or passage. Having been reduced to permanent form and published, the written or printed word has greater capabilities of harm. We can see no difference between the reading of defamatory words from a prepared manuscript to a group of people within the presence of the reader, which constitutes libel, and reading defamatory words from a prepared manuscript to be broadcast by the facilities of a radio station. The latter simply carries the defamatory words farther because the defamer has used a medium for dissemination which reaches listeners far beyond the ordinary limits of the human voice. The law of libel is applicable to the case at bar.").

[159] RESTATEMENT (SECOND) OF TORTS, § 568A.

[160] *See* Brown v. Hearst Corp., 862 F. Supp. 622, 627 (D. Mass. 1994), *aff'd*, 54 F.3d 21 (1st Cir. 1995) ("[B]ecause the allegedly offensive statements were fixed, recorded, and widely distributed in a television program, if defamation does exist in this case, it is libel and not slander.").

question.[161] Two separate policy justifications reinforce this reasoning. First, a broadcast's wide dissemination puts 'the broadcaster upon the same footing as the publisher of a newspaper.'" [162] Second, with technological progress, few things are unrecorded and kept private in the way a conversation does. Therefore, the general rule applies that "[a] defamatory statement addressed to the eye, such as a writing or a photograph, is libel. One addressed to the ear, such as a spoken word, is slander."[163] Because a deepfake—as opposed to any captions or descriptions thereof—are permanent, defamation actions arising from deepfakes would likely be exclusively the purview of libel actions.[164]

"Libel is governed predominantly by state law, and the elements of libel vary by jurisdiction."[165] But many, if not most, elements of libel remain common across state lines. Consider New York:

> In order to state a cause of action for libel under New York law, a plaintiff must plead: (1) a written false and defamatory statement of fact concerning the plaintiff; (2) that was published by the defendant to a third party; (3) due to the defendant's negligence or actual malice, depending on the status of the person libeled; and (4) special damages or *per se* actionability.[166]

[161] *See* Sabino v. WOIO, L.L.C., 2016-Ohio-491, ¶ 41, 56 N.E.3d 368, 376 (Ohio Ct. App. 2016) ("defamatory matter broadcast by means of radio or television is classified as libel").

[162] *See* Eisenstein v. WTVF-TV, News Channel 5 Network, LLC, 389 S.W.3d 313, 317 n.4 (Tenn. Ct. App. 2012) (quoting RESTATEMENT (SECOND) OF TORTS § 568A cmt. 1 (1977))).

[163] Franco v. Diaz, 51 F. Supp. 3d 235, 244 (E.D.N.Y. 2014) (citation omitted); *see also* Hardesty v. Waterworks Dist. No. 4 of Ward Four, 954 F. Supp. 2d 461, 475 (W.D. La. 2013) ("Libel is defamation which is 'expressed by print, writing, pictures, or signs', while slander is communicated by 'oral expressions or transitory gestures.'" (citation omitted)); Doe v. Mobile Video Tapes, Inc., 43 S.W.3d 40, 48 (Tex. App. 2001) ("Libel is defamation in written or other graphic form that tends to injure a person's reputation, exposing the person to public hatred, contempt, or ridicule. Slander is orally communicated defamation." (internal citations omitted)); Hardesty v. Waterworks Dist. No. 4 of Ward Four, 954 F. Supp. 2d 461, 475 (W.D. La. 2013) ("Libel is defamation which is 'expressed by print, writing, pictures, or signs', while slander is communicated by 'oral expressions or transitory gestures.'" (citation omitted)).

[164] And, relatedly, any discussions of defamation pertain exclusively to libel.

[165] Kevin L. Kite, Note, *Incremental Identities: Libel-Proof Plaintiffs, Substantial Truth, and the Future of the Incremental Harm Doctrine*, 73 N.Y.U. L. REV. 529, 532 (1998) (internal footnote omitted).

[166] Daytree at Cortland Sq., Inc. v. Walsh, No. 15CV2298JFBAYS, 2018 WL 3869247, at *9 (E.D.N.Y. Aug. 15, 2018) (citing Celle v. Filipino Reporter Enters., 209 F.3d 163, 176

New York libel law, like other states, considers photographs as "statements" sufficient to allege libel.[167] Elements two and three are also satisfied because any deepfaker clearly publishes the content to a third party, i.e. not to the subject of the fake, and exhibits actual malice, as he or she knowingly creates falsified content. Indeed, in some states, malice may entitle deepfake victims to additional punitive damages.[168] However, evaluating deepfakes' applicability to libel actions, then, requires further consideration of elements one and four: Is a deepfake defamatory? And what damages do deepfakes impose?

1. Is a Deepfake Defamatory?

Defamation is broadly defined as a false "statement that tends to expose the [individual] to public contempt, ridicule, aversion, or disgrace, or induce an evil opinion of him in the minds of right-thinking persons, and to deprive him of their friendly intercourse in society."[169] Outside of clear parodies, which do not give rise to defamation action,[170] determining whether something is defamatory can be particularly difficult when edited video is involved. As Judge Raymond Dearie of the Eastern District of

(2d Cir. 2000)). These are fairly representative nationwide. *Compare id. with, e.g.*, Janiszewski v. Belmont Career Ctr., 86 N.E.3d 613, 630 (Oh. Ct. App. 2017) (Under Ohio law, "[t]he elements of a defamation action, whether slander or libel, are: (1) the defendant made a false and defamatory statement concerning another; (2) the false statement was published without privilege to a third person; (3) the plaintiff was injured; and (4) the defendant acted with the required degree of fault that was defamatory per se or caused special harm to the plaintiff."), *and* McGettigan v. Di Mare, 173 F. Supp. 3d 1114, 1125-26 (D. Colo. 2016) (Under Colorado law, "[t]he elements of a libel claim are: (1) a written defamatory statement of and concerning the plaintiff; (2) published to a third party; (3) with the publisher's fault amounting to at least negligence; and (4) either actionability of the statement irrespective of special damages or the existence of special damages caused by the publication." (quotation omitted)), *and* Wal-Mart Stores, Inc. v. Smitherman, 872 So. 2d 833, 840 (Ala. 2003) (Under Alabama law, "[t]he elements of a cause of action for defamation are: 1) a false *and* defamatory statement concerning the plaintiff; 2) an unprivileged communication of that statement to a third party; 3) fault amounting at least to negligence on the part of the defendant; and 4) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication of the statement.") (emphasis in original).

[167] *See, e.g.*, Palmisano v. Modernismo Publications, Ltd., 98 A.D.2d 953, 954 (Sup. Ct. N.Y. 1983)

[168] *See, e.g.*, Gleason v. Smolinski, 125 A.3d 920, 948 (Conn. 2015).

[169] Rinaldi v. Holt, Rinehart & Winston, Inc., 366 N.E.2d 1299, 1305 (N.Y. 1977) (quotation omitted).

[170] *See supra* notes 162–164 and accompanying text.

New York, now the namesake to the courthouse's atrium,[171] presciently observed in 1994:

> [T]elevision broadcasts add new and potentially significant variables to the defamation analysis. Courts must scrutinize the juxtaposition of the audio and video portions of a television program. In subtle ways, a television director can alter the tone of an otherwise innocuous broadcast. With the emerging popularity of self-styled "magazine" news programs, courts should be sensitive to the possibility that a transcript which appears relatively mild on its face may actually be, when the total mix of creative ingredients are considered, highly toxic. Indeed, a clever amalgamation of half-truths and opinion-like statements, adorned with orchestrated images and dramatic audio accompaniment, can be devastating when packaged in the powerful television medium.[172]

Like such television broadcasts, deepfakes are primed to add further wrinkles to defamation analysis.

However, some deepfakes are quite simple to navigate with respect to defamation. Some, for example, are decidedly not defamatory. Consider the example referenced above in which a man, in homage to his wife's favorite show, placed her in that show.[173] There is nothing disgraceful about such a video so as to invoke defamation.

Deepfakes made with an obvious humorous intent may also be easily disposed of should they prompt a victim to sue. A deepfake turning President Trump into Biff Tannen, the villainous character in *Back to the Future*, was obviously made as a parody.[174] The character is fictional, and the connection is meant to be a caricature of the President. To be sure, "[a] defendant cannot escape liability for defamatory factual assertions simply by claiming that the statements were a form of ridicule, humor, or

[171] *See Courthouse Atrium Dedicated to Judge Dearie*, N.Y. L. J. (May 1, 2018, 10:00 AM), https://www.law.com/newyorklawjournal/2018/05/01/courthouse-atrium-dedicated-to-judge-dearie/ [https://perma.cc/8XL5-3FSQ].

[172] Corp. Training Unlimited, Inc. v. Nat'l Broadcasting Co., Inc., 868 F. Supp. 501, 507 (E.D.N.Y. 1994).

[173] *See supra* notes 56–58 and accompanying text.

[174] *See supra* note 53.

sarcasm."[175] But "if the allegedly defamatory statement could not be reasonably understood as describing actual facts about the plaintiff or actual events in which he participated, the publication will not be libelous."[176] Deepfakes like Trump-as-Biff fall squarely into that category, which defamation law has long dealt with smoothly.

Deepfake pornography is different. Without question, falsely placing someone in an adult video without his or her consent could seriously '"harm the reputation of [the victim] as to lower him [or her] in the estimation of the community or to deter third persons from associating or dealing with him [or her].'"[177] This conclusion is also analogically supported by case law in which courts have upheld defamatory causes of action where edited videos appear to show the subject acting in a way other than what in fact occurred.[178] To immunize deepfakers from defamation action would undermine society's "pervasive and strong interest in preventing and redressing attacks upon reputation," the public policy driving defamation.[179]

Content is not the only factor considered, however; the video's context (e.g. its caption) play a role in defamation analysis. "[T]he rule of innocent construction" states "'[a] written or oral statement is to be considered in context, with the words and the implications therefrom, given their natural and obvious meaning; if so construed the statement may reasonably be innocently interpreted . . . it cannot be actionable *per se*.'"[180] Applying this thinking, if the deepfaker is quite clear about the fact that the video is fabricated or fantastical, he or she has a stronger defense that the video does not inflict the same harm on the video's subject. On the

[175] Solaia Tech., LLC v. Specialty Pub. Co., 852 N.E.2d 825, 840 (Ill. 2006) (modification in original) (quotation omitted).

[176] Bollea v. World Championship Wrestling, Inc., 610 S.E.2d 92, 96 (Ga. Ct. App. 2005) (citing Pring v. Penthouse Intl., 695 F.2d 438, 442 (10th Cir. 1982)).

[177] Crump v. Beckley Newspapers, Inc., 320 S.E.2d 70, 77 (W. Va. 1983) (quoting RESTATEMENT (SECOND) OF TORTS § 559 (1977)).

[178] *See, e.g.*, Cummins v. Bat World Sanctuary, No. 02-12-00285-CV, 2015 WL 1641144, at *22 (Tex. Ct. App. Apr. 9, 2015); *cf.* N.B.C. v. Gonzalez, No.04-95-00219-CV, 1995 WL 624549, at *5 (Tex. Ct. App. Oct. 25, 1995) ("The portion that was broadcast was a truthful depiction of appellee's conduct, and as such, is not defamatory. While inclusion of the entire video and audio may have been more flattering to appellee, this Court will not sit as a senior editor to television stations.") (internal citations omitted); Newton v. Nat'l Broad. Co., 930 F.2d 662, 686 (9th Cir. 1990), *cert. denied,* 502 U.S. 866 (1991) (failure to broadcast complete statement of plaintiff is not indicative of actual malice).

[179] Rosenblatt v. Baer, 383 U.S. 75, 86 (1966).

[180] Paul v. Premier Elec. Const. Co., 581 F. Supp. 721, 723 (N.D. Ill. 1984) (quoting Chapski v. Copley Press, 92 Ill.2d 344, 352 (Ill. 1982)).

other hand, if the individual creating and posting the video makes no effort to dispel the mistruth of the video's subject, the video is more likely to be considered legitimate and thus harm one's reputation. Ultimately, it is likely the case that a deepfake satisfies the first element of a libel action and clears the first hurdle, though not absolutely certain.

2. What Damages Regime Do Deepfakes Fall Under?

Assuming a deepfake is defamatory, a court must then determine appropriate damages. Defamatory statements can either be actionable *per se* or *per quod*. *Per se* actionable statements mean "its harm is obvious and apparent on its face."[181] "Statements falling outside of these categories may only be actionable as libel *per quod* which requires that special damages be alleged."[182] In the latter situation, the words' "injurious nature appears only in consequence of extrinsic facts."[183]

Typically, "[w]ords tending to impute criminal offense, loathsome disease, business misconduct, or serious sexual misconduct constitute defamation *per se*."[184] This final sub-category of *per se* defamatory is quite broad: for example, "[t]he traditional common law position is that the imputation of unchastity" meets this standard.[185] To this end, courts have repeatedly affirmed that statements pertaining to one's sexual life, including reports of one's alleged extramarital affairs or sexual habits, are *per se* actionable.[186] Thus, a falsified video that purports to demonstrate one's sexual actions, the filming thereof, and its subsequent publication, could indeed be considered defamatory *per se*.

Even if a pornographic deepfake is not *per se* defamatory, the special damages requirement in a defamatory *per quod* cause of action is likely met. "In a defamation *per quod* action, damage to the plaintiff's reputation is not presumed and the plaintiff must plead and prove special

[181] Solaia Tech., LLC v. Specialty Pub. Co., 852 N.E.2d 825, 839 (Ill. 2006).

[182] Paul v. Premier Elec. Const. Co., 581 F. Supp. 721, 723 (N.D. Ill. 1984) (citation omitted).

[183] Joseph v. Scranton Times L.P., 959 A.2d 322, 344 n.23 (Pa. Super. Ct. 2008).

[184] Sottosanti-Mack v. Reinhart, 173 F. Supp. 3d 94, 104 (E.D. Pa. 2016).

[185] 2 LAW OF DEFAMATION § 7:18 (2d ed.).

[186] *See, e.g.*, Hoskins v. Fuchs, 517 S.W.3d 834, 843 (Tex. Ct. App. 2016), *review denied* (Feb. 16, 2018) (concluding that statements made in law student's complaint to university's office of equal opportunity services, alleging that student's girlfriend was having sexual relationship with professor, qualified as defamation per se);
Moreau v. Brenan, 466 So. 2d 572, 574 (La. Ct. App. 1985) (holding that wife and husband were defamed by allegations that wife was having extramarital sexual relations, and those allegations constituted defamation *per se*).

damages."[187] Special damages are "actual damages of a pecuniary nature."[188]Thus, a plaintiff must plead that pecuniary damages are appropriate as a remedy to his or her cause of action.

In many instances, a deepfakes create substantial risk of financial harm for its victims because of the inherent value of one's reputation, particularly when that person is a public official.

> [A] public image is a valuable asset. A favorable public image enables a public figure to earn large fees for lecturing or for endorsing products. It is a source of influence in politics, entertainment, sports, religion, education, or other fields. It may be an important source of self-esteem and personal satisfaction. A person who enjoys a positive public image thus may be injured by defamation, even if there is no harm to his existing or future personal relations.[189]

This problem is only amplified by the blindingly fast pace at which news, particularly harmful news, spreads on the Internet. One case is particularly instructive: In *Stephen G. Perlman, Rearden LLC v. Vox Media, Inc.*,[190] *The Verge*, a website that "examine[s] how technology will change life in the future for a massive mainstream audience," published an article that Perlman claimed defamed both himself and his company, OnLive, following its bankruptcy.[191] The statistics regarding the speed with which the article was shared are staggering:

> In the first fifteen minutes after The Verge published the August 28 Article, various journalists and editors associated with The Verge, Polygon, and Vox

[187] Doctor's Data, Inc. v. Barrett, 170 F. Supp. 3d 1087, 1103 (N.D. Ill. 2016).

[188] Imperial Apparel, Ltd. v. Cosmo's Designer Direct, Inc., 882 N.E.2d 1011, 1018 (Ill. 2008).

[189] David A. Anderson, *Reputation, Compensation and Proof*, 25 WM. & MARY L. REV. 747, 766 (1984); *see also* Denny v. Mertz, 318 N.W.2d 141, 151 (Wisc. 1982) ("A person's reputation and good name is of inestimable value to him and once it has been besmirched by another through carelessness or malice restoration is virtually impossible.") (internal footnote omitted).

[190] *The Verge* is owned by Vox, the defendant in this case. Stephen G. Perlman, Rearden LLC v. Vox Media, Inc., No. CV 10046-VCP, 2015 WL 5724838, at *2 (Del. Ch. Sept. 30, 2015).

[191] *Id.* at *1.

> promoted the article as the "definitive account" based on "exhaustive proof," despite the fact that they had not fact-checked the article . . . using social media platforms such as Facebook, LinkedIn, Twitter, Tumblr, and Google+ to reach hundreds of thousands, if not millions, of readers. Readers quickly posted 300 comments (288 in the first two days) responding to the August 28 Article, and the article spread rapidly through social media networks. Soon the August 28 Article became a top Google search result for "OnLive," behind only OnLive's own corporate and service web pages and the OnLive Wikipedia page.[192]

Additionally, the court took note of the permanence of the allegedly defamatory article:

> In fact, two years later the August 28 Article was still the fourth Google result for "OnLive." Also, when Internet users use Google to search for "Steve Perlman," Google provides three "In-depth articles," which it identifies as "high-quality content to help [users] learn about or explore a subject;" the August 28 Article appears alongside two articles from www.businessweek.com and www.smithsonianmag.com, respectively, both highly credible publications.[193]

In no small part due to the article's widespread dissemination and permanence, the court held that the damage done to Perlman was genuine, compelling the judge to deny Vox's motion to dismiss.[194]

So, too, could be the fate of a deepfake victim. If the video is to be believed—whether sexual or otherwise—the reputational damage could be swift and lasting. Therefore, pending specifics, a cause of action for defamation may be the ideal avenue for any deepfake victim.

[192] *Id.* at *5.

[193] *Id.*

[194] *See id.* at *21.

B. Privacy Torts

The right to privacy "dates back to a law review article published in December of 1890 by two young Boston lawyers, Samuel Warren and Louis Brandeis."[195] "A specific suggestion of [Warren's], as well as [Warren's] deep-seated abhorrence of the invasions of social privacy . . . led to [their] taking up the inquiry."[196] The first cases recognizing an actionable invasion of the right to privacy were heard and decided a decade later.[197]

Despite the Supreme Court's statement that its right to privacy jurisprudence "def[ied] categorical description,"[198] Dean William Prosser, the father of modern American tort law, described the state of privacy law as follows:

> The law of privacy comprises four distinct kinds of invasion of four different interests of the plaintiff, which are tied together by the common name, but otherwise have almost nothing in common except that each represents an interference with the right of the plaintiff, in the phrase coined by Judge Cooley "to be let alone."[199]

The four privacy torts are intrusion upon seclusion, publicity given to private life, publicity in false light, and wrongful appropriation. Each tort, and its application to deepfakes, is examined herein.

[195] Dorothy J. Glancy, *The Invention of the Right to Privacy*, 21 ARIZ. L. REV. 1, 1 (1979).

[196] *Id.* at 6 n.29 (1979) (first and second alterations in original) (quoting Letter from Brandeis to Warren (April 8, 1905)). Glancy offers an excellent glimpse into the publication's critical and popular acclaim. *See id.* at 6–7 (citing *The Right to Be Let Alone*, 67 ATLANTIC MONTHLY 428–29 (1891)); *The Defense of Privacy*, 66 SPECTATOR 200 (Feb. 7, 1891); *Comment*, 3 GREEN BAG 524, 525 (1891)).

[197] *See* 57 A.L.R.4th 22 (originally published in 1987) (citing Pavesich v. New England Life Ins. Co., 50 S.E. 68 (1905)); Roberson v. Rochester Folding Box Co., 64 N.E. 442 (N.Y. 1902).

[198] Paul v. Davis, 424 U.S. 693, 713 (1976). In a whimsical metaphor, the Third Circuit called it a "haystack in a hurricane." Ettore v. Philco Television Broad. Corp., 229 F.2d 481, 485 (3d Cir. 1956).

[199] William Prosser, *Privacy*, 48 CAL. L. REV. 383, 389 (1960) (footnote omitted).

1. Intrusion Upon Seclusion

> *One who intentionally intrudes, physically or otherwise, upon the solitude or seclusion of another or his private affairs or concerns, is subject to liability to the other for invasion of his privacy, if the intrusion would be highly offensive to a reasonable person.*[200]

Intrusion upon seclusion is not dependent on the "the truth or falsehood of the information itself"; instead, it "deals with the manner in which Defendant obtained the information."[201] For example, in *Peterson v. Moldofsky*[202] the plaintiff "claim[ed] that Defendant intruded on her privacy by emailing photographs of her engaged in group sex to several people."[203] However, the court held that "no intrusion occurred, as [the Plaintiff] knew of and consented to [the Defendant's] presence and his taking of pictures during the sex acts [meaning] there 'is no evidence of an intrusion as based on the manner in which the information is obtained[.]'"[204]

The import of *Peterson*'s reasoning is made clear when compared with *DePiano v. Atlantic County*.[205] In that case, the plaintiff, Gregory DePiano, was a corrections officer and Sergeant at the Atlantic County Justice Facility (ACJF).[206] While DePiano served in that capacity, an ACJF warden and internal affairs officer, Gary Merline, disseminated photographs from DePiano's personnel files in which he was dressed in women's clothing, which he admitted "is, or at some point was, part of his sexual life."[207] Merline, by abusing the access afforded by his position, therefore intruded upon DePiano's seclusion.[208]

Juxtaposing *Peterson* and *DePiano* demonstrates that it is the manner of the intrusion that makes all the difference. In *Peterson*, the purported "intrusion" occurred with one's consent to obtain information—

[200] RESTATEMENT (SECOND) OF TORTS § 652B (1977).
[201] Trundle v. Homeside Lending, Inc., 162 F. Supp. 2d 396, 401 (D. Md. 2001).
[202] Peterson v. Moldofsky, No. 07-2603-EFM, 2009 WL 3126229 (D. Kan. Sept. 29, 2009).
[203] *Id.* at *3.
[204] *Id.* (quoting Haehn v. City of Hoisington, 702 F. Supp. 1526, 1531 (D. Kan. 1988)).
[205] DePiano v. Atlantic Cty., No. CIV.02-5441 RBK, 2005 WL 2143972 (D.N.J. Sept. 2, 2005).
[206] *Id.* at *1.
[207] *Id.* at *3.
[208] *See id.* at *11.

photographs—of the plaintiff in compromising situations, even if the plaintiff failed to restrict how such information was disseminated. On the other hand, in *DePiano*, the victim in no way permitted the intruder access to such information. In sum, "the tort of intrusion upon seclusion is based upon the manner in which an individual obtains information,"[209] not whether the private information was thereafter disseminated.

Deepfake creators are more likely in the *Peterson* camp rather than the *DePiano* camp. True, the victim's did not consent to the way in which the photos were used. But as in *Peterson*, the victims knowingly consented to their creation in the first place, not to mention their dissemination into and throughout the public domain. In many (though not all) cases, the deepfake subject has either put the photos into the public by posting them online or consented to their collection by posing for paparazzi. Deepfakers have not violated anyone's personal space to obtain the necessary information to create and publish their work. Thus, as was the case in *Peterson*, a deepfake victim is unlikely to prevail on an intrusion of seclusion claim.

2. Publicity Given to Private Life

> *One who gives publicity to a matter concerning the private life of another is subject to liability to the other for invasion of his privacy, if the matter publicized is of a kind that*
>
> - *(a) would be highly offensive to a reasonable person, and*
> - *(b) is not of legitimate concern to the public.*[210]

This second tort is not a candidate for a deepfake victim for one reason: "an essential element of the tort of public disclosure of private facts is that the facts at issue be true."[211]

In the "earliest non-consensual pornography lawsuit,"[212] the infamous publication Hustler was adjudged to have invaded LaJuan

[209] *Haehn*, 702 F. Supp. at 1531.

[210] RESTATEMENT (SECOND) OF TORTS § 652D (1977).

[211] Tyne ex rel. Tyne v. Time Warner Entm't Co., L.P., 204 F. Supp. 2d 1338, 1344 (M.D. Fla. 2002), *aff'd sub nom.* Tyne v. Time Warner Entm't Co., L.P., 425 F.3d 1363 (11th Cir. 2005).

[212] Amanda Levendowski, *Using Copyright to Combat Revenge Porn*, 3 N.Y.U. J. INTELL. PROP. & ENT. L. 422, 434 (2014) (citing Alexa Tsoulis-Reay, *A Brief History of Revenge Porn*, N.Y. MAG. (Jul. 21, 2013), http://nymag.com/news/features/sex/revenge-porn-2013-7 [https://perma.cc/9RQR-QCW6]).

Wood's privacy by publishing a stolen photograph of her in the nude with a "falsely attributed lewd fantasy."[213] The court held Hustler liable for publicity in a false light—discussed substantively *infra*—rather than a private facts theory because the fantasy did not truthfully reflect Wood's private life.[214] The same reasoning holds true for all deepfakes, in which "none of the facts disclosed by the picture are alleged to be true."[215] This theory of liability is thus altogether foreclosed.

3. Publicity in False Light

> *One who gives publicity to a matter concerning another that places the other before the public in a false light is subject to liability to the other for invasion of his privacy, if*
>
> - *(a) the false light in which the other was placed would be highly offensive to a reasonable person, and*
> - *(b) the actor had knowledge of or acted in reckless disregard as to the falsity of the publicized matter and the false light in which the other would be placed.*[216]

Dean Prosser identified the seminal false light invasion of privacy case: the successful 1816 suit by the famous English poet Lord Byron to enjoin circulation of a volume of bad poetry falsely attributed to him.[217] Two centuries later, false light has a distinct application to deepfakes.

Deepfakes, by definition, place an individual before the public in a false light. Deepfakes, "[n]onconsensual [videos] created through digitally manipulated images of victims[, are] entirely false because the victim never posed for the image."[218] This is most certainly the case with non-pornographic videos.[219]

[213] Wood v. Hustler Mag., Inc., 736 F.2d 1084, 1085 (5th Cir. 1984).

[214] *Id.* at 1090 (citing RESTATEMENT (SECOND) OF TORTS §§ 652D cmt. a, b, 625E cmt. b (1977)).

[215] Tyne ex rel. Tyne v. Time Warner Entm't Co., L.P., 204 F. Supp. 2d 1338, 1344 (M.D. Fla. 2002), *aff'd sub nom.* Tyne v. Time Warner Entm't Co., L.P., 425 F.3d 1363 (11th Cir. 2005).

[216] RESTATEMENT (SECOND) OF TORTS § 652E (1977).

[217] W. PAGE KEETON ET AL., PROSSER AND KEETON ON TORTS § 117, 863 (5th ed. 1984) (citing Lord Byron v. Johnston (1816) 2 Mer 29, 35 Eng. Rep. 851)).

[218] Levendowski, *supra* note 212, at 434 (citing Tsoulis-Reay, *supra* note 212).

[219] While the idea of acting in a traditional film may seem innocuous enough, one need only look to the actors who have sworn off working with Woody Allen in light of

Consider again the aforementioned *Wood v. Hustler* case. There, the court was persuaded that Hustler was liable for falsely representing that Wood consented to the submission and publication of a photograph depicting her in the nude in the coarse and sex-centered magazine. Moreover, the publication falsely attributed a lewd fantasy to Wood.[220] To be sure, the same could be said for any actor that "appears" in a sexually explicit video via a deepfake, or any deepfake for that matter. Without question, the fabricated video would ascribe conduct to an actor, which he or she did not participate in, nor would such an actor likely consent to the dissemination of video suggesting they did participate in these illicit acts.

4. Wrongful Appropriation

> *One who appropriates to his own use or benefit the name or likeness of another is subject to liability to the other for invasion of his privacy.*[221]

A victim of a deepfake may have a cognizable claim for wrongful appropriation, otherwise called misappropriation.[222] "The tort of wrongful appropriation requires that the defendant appropriate the plaintiff's likeness to his own *use* or benefit."[223] Usually, such use or benefit is attributed to a commercial or financial benefit.

Though opponents may rebut that they are not benefitting commercially,[224] victims have two substantial arguments. First,

allegations of his misconduct. *See, e.g.*, Lisa Respers France, *List of Actors Refusing to Work with Woody Allen Grows*, CNN (Jan. 19, 2018), https://www.cnn.com/2018/01/19/entertainment/woody-allen-actors/index.html [https://perma.cc/M732-8NH4]. The same phenomenon occurred previously with Mel Gibson, when a cameo of his was canceled because his inclusion "ultimately did not have the full support of [the] entire cast and crew." Sam Jones, *Mel Gibson Film Cameo Cancelled After Protests From Cast and Crew*, GUARDIAN (Oct. 22, 2010), https://www.theguardian.com/film/2010/oct/22/mel-gibson-the-hangover-2 [https://perma.cc/2MSX-YV3M]. Admittedly, however, this may be a tougher sell vis-à-vis suggesting that such placement would be highly offensive to a reasonable person.

[220] Wood v. Hustler Mag., Inc., 736 F.2d 1084, 1089 (5th Cir. 1984).

[221] RESTATEMENT (SECOND) OF TORTS § 652C (1977).

[222] Often titled "Appropriation of Name or Likeness." *Id.*

[223] Ault v. Hustler Mag., Inc., 860 F.2d 877, 883 (9th Cir. 1988) (citation omitted) (emphasis added).

[224] Anderson v. Fisher Broad. Co., 712 P.2d 803, 813 (Or. 1986) ("Publication of . . . [a] photograph is not appropriation for commercial use simply because the medium itself is operated for profit.").

individuals who drive consumers to the website hosting the videos, particularly those that advertise the purported video alongside "promotional images," are acting with a commercial purpose.[225] Consider, for example, an individual who publishes his or her own blog that hosts the deepfake and offers pedestrian digital ads on the very same page. By driving traffic to the page via the deepfake, he or she stands to earn additional revenue because advertisers pay more money to advertise on pages visited more frequently.

Moreover, only four states—New York, Oklahoma, Utah, and Virginia—specifically articulate that the appropriation *must* "be for advertising, or for purposes of trade."[226] Thus, deepfakers are unlikely to successfully defend themselves on the argument that because they acted without a commercial purpose, they are not liable for a wrongful appropriation cause of action.

Second, deepfakers are nevertheless *using* the individual's likeness without consent and "injure[] the economic interests of the plaintiff due to commercial exploitation[.]"[227] Whether the individual in question is a celebrity or layperson is irrelevant; it is not a requirement that one be a public official to have his or her likeness appropriated without her consent for economic reasons.[228]

Moreover, a deepfaker cannot hide behind "the general rule . . . that incidental use of a name or likeness does not give rise to liability for invasion of privacy by appropriation."[229] Because a deepfake tries to attract attention based on the false premise of its purported subject and because the victim is the chief—and ever-present—subject in the deepfake, its use cannot be considered incidental.

[225] Bosley v. Wildwett.com, 310 F. Supp. 2d 914, 922 (N.D. Ohio 2004).

[226] RESTATEMENT (SECOND) OF TORTS § 652C (1977).

[227] *Ault*, 860 F.2d at 883.

[228] *See generally* Manger v. Kree Inst. of Electrolysis, 233 F.2d 5 (2d Cir. 1956) (affirming the general 'right of privacy' violation for manipulating a contest winner's submission for an advertisement and running the altered material without her consent); Colgate-Palmolive Co. v. Tullos, 219 F.2d 617, 619 (5th Cir. 1955) (affirming a right of privacy violation for misappropriating one's likeness in an advertisement when the individual in question was not a celebrity).

[229] Aligo v. Time-Life Books, Inc., No. C 94-20707 JW, 1994 WL 715605, at *2 (N.D. Cal. Dec. 19, 1994) (collecting cases). To determine if a use is incidental, consider (1) whether the use has a unique quality or value that would result in commercial profit to the defendant, (2) whether the use contributes something of significance, (3) the relationship between the reference to the plaintiff and the purpose and subject of the work, and (4) the duration, prominence or repetition of the name or likeness relative to the rest of the publication. See *id.* at *3 (citations omitted).

Ultimately, appropriation may yet be a lost cause (of action). Wrongful appropriation cases, particularly those involving digital images of one's likeness, are almost always using the victim's likeness to endorse or advertise a particular product.[230] A deepfake, thus, presents an atypical fact pattern because deepfakers may not be attempting to create their own commercial benefit like the typical defendant in a wrongful appropriation case. For example, a deepfaker that creates an explicit video of a celebrity and posts it online to a site from which they derive no revenue does not serve an economic purpose.

So, courts may be reluctant to recognize that a deepfaker's personal use and enjoyment of a fabricated video, even if it is disseminated on the Internet for others' personal, analogous use and enjoyment. Without any promise of monetary value, personal deepfakes are likely insufficient to satisfy the elements of appropriation.

C. Right of Publicity

What may instead prove to be the most direct source of redress is a cause of action alleging a violation of the victim's right of publicity, an interrelated[231] but distinct right. Plainly, "the right of publicity is an economic right to use the value of one's own celebrity."[232]

The right of publicity exists to "prevent[] unjust enrichment by the theft of good will. No social purpose is served by having the defendant get free some aspect of the plaintiff that would have market value and for

[230] *See, e.g.*, Kyser-Smith v. Upscale Commc'ns, Inc., 873 F. Supp. 1519, 1526 (M.D. Ala. 1995).

[231] *See* Toffoloni v. LFP Publ'g Grp., LLC, 572 F.3d 1201, 1205 (11th Cir. 2009) (defining the "right of publicity [as] protect[ing] against 'the appropriation of another's name and likeness'") (quoting Martin Luther King, Jr., Ctr. for Social Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697, 703 (Ga. 1982)).

[232] In re NCAA Student-Athlete Name & Likeness Licensing Litig., 724 F.3d 1268, 1284 n.1 (9th Cir. 2013) (Thomas, J., dissenting); *see also* ETW Corp. v. Jireh Publ'g, Inc., 332 F.3d 915, 928 (6th Cir. 2003) ("The right of publicity is an intellectual property right of recent origin which has been defined as the inherent right of every human being to control the commercial use of his or her identity.") (citing MCCARTHY ON PUBLICITY AND PRIVACY, § 1:3); Comedy III Prods., Inc. v. Gary Saderup, Inc., 21 P.3d 797, 807 (Cal. 2001) ("[t]he right of publicity holder possesses is not a right of censorship, but a right to prevent others from misappropriating the economic value generated by the celebrity's fame").

which he would normally pay."[233] In light of this reasoning, "[a]ll that a plaintiff must prove in a right of publicity action is that she has a pecuniary interest in her identity, and that her identity has been commercially exploited by a defendant."[234]

But "[t]he distinctive aspect of the common law right of publicity is that it recognizes the commercial value of the picture or representation of a prominent person or performer, and protects his proprietary interest in the profitability of his public reputation or 'persona.'"[235] Thus, unsurprisingly, the quintessential right of publicity cases involve cases in which celebrities' distinct yet replicable traits are used without their permission—"so-called 'impersonator' cases"[236]—including *Midler v. Ford Motor Co.*[237] and *Waits v. Frito-Lay, Inc.*[238] In both, famous actors with distinct voices refused to partake in advertisements, and companies responded by circumventing their refusal and recreating celebrities' trademark voices with sound-alike voice actors after the stars declined to participate in the advertisement themselves.[239]

The same is true with deepfakes: celebrities are deprived of their ability to control their likeness or image. And while [d]amages from such evident abuse of a plaintiff's property right in his public reputation are plainly difficult to measure by monetary standards,[240] courts, depending on relevant state law, are open to awarding both the "market value"[241] of the celebrity's persona used and damages to compensate for any "induce[d] humiliation, embarrassment, and mental distress."[242] Therefore, celebrity deepfake victims may succeed on a right to publicity claim against the deepfaker. However, a right to publicity cause of action is far

[233] Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 576 (1977) (quoting Harry Kalven, Jr., *Privacy in Tort Law Were Warren and Brandeis Wrong?*, 31 L. & CONTEMP. PROB. 326, 331 (1966)).

[234] Parks v. LaFace Records, 329 F.3d 437, 460 (6th Cir. 2003) (citations omitted).

[235] Ali v. Playgirl, Inc., 447 F. Supp. 723, 728 (S.D.N.Y. 1978).

[236] Jonathan Faber, *A Brief History of the Right of Publicity*, RIGHT OF PUBLICITY (July 31, 2015), http://rightofpublicity.com/brief-history-of-rop [https://perma.cc/87T5-QU7L].

[237] Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1989).

[238] Waits v. Frito-Lay, 978 F.2d 1093 (9th Cir. 1992), *abrogated on other grounds by* Lexmark Int'l, Inc. v. Static Control Components, Inc., 572 U.S. 118 (2014).

[239] *See* Faber, *supra* note 236.

[240] *Ali*, 447 F. Supp. at 729 (citing Myers v. U.S. Camera Publ'g Corp., 167 N.Y.S.2d 771, 774 (City Ct. N.Y. 1957)).

[241] *Waits*, 978 F.2d at 1103.

[242] *Id.* (quoting Motschenbacher v. R.J. Reynolds Tobacco Co., 498 F.2d 821, 824 (9th Cir. 1974)).

from a certain victory for the victim; deepfakers have substantial counterarguments in the form of satire.

In some instances, defendant deepfakers have a substantial defense against right of publicity claims: parody. "The right of publicity derived from public prominence does not confer a shield to ward off caricature, parody and satire. Rather, prominence invites creative comment."[243] Political cartoonists, for example cannot be held civilly liable for depicting a celebrity or politician in what any observer reasonably recognizes is a lampoon.[244] This rule should similarly hold true for other obvious forms of parodies, including deepfakes, such as the aforementioned parody deepfake depicting President Trump as television or movie characters.[245]

But even satire has its limits; the Supreme Court, in its *Bresler–Letter Carriers–Falwell* line of cases, provides protection for statements that cannot "reasonably [be] interpreted as stating actual facts" about an individual.[246] The very point of though deepfakes is to create video so seamlessly superimposed that the reasonable person *cannot* discern fact from fiction. Common sense tells us that videos falsely portraying individuals in compromising and intimate affairs cannot be considered a parody.

"The right of publicity a holder possesses is not a right of censorship, but a right to prevent others from misappropriating the economic value generated by the celebrity's fame."[247] But the right is not conditioned on celebrity; in the last nearly forty years, the right of publicity doctrine has dramatically expanded to include laypersons.

In 1982, the Supreme Court of Georgia compared one's "right not to have another appropriate one's photograph" in two cases—one involving a private person, one involving a public figure. [248] The court concluded that "private citizens have the right of privacy, public figures have a similar right of publicity, and that the measure of damages to a

[243] Guglielmi v. Spelling-Goldberg Prods., 603 P.2d 454, 460 (Cal. 1979) (en banc) (Bird, C.J., concurring).

[244] *See id.* at 460 n.12 ("For example, Garry Trudeau, creator of the satiric cartoon strip 'Doonesbury,' regularly fictionalizes events and dialogue involving prominent political figures. It cannot be seriously maintained that one such satirized notable could successfully pursue an action for an infringement on his right of publicity based on such use.").

[245] *See supra* notes 53–55 and accompanying text.

[246] Milkovich v. Lorain Journal Co., 497 U.S. 1, 20 (1990).

[247] Comedy III Prods., Inc. v. Gary Saderup, Inc., 21 P.3d 797, 807 (Cal. 2001).

[248] Martin Luther King, Jr., Ctr. for Social Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697, 703 (Ga. 1982).

public figure for violation of his or her right of publicity is the value of the appropriation to the user."[249]

To be sure, given the very nature of celebrity, in right of publicity causes of action involving high-profile plaintiffs, "the mere allegation that the plaintiff was not compensated has been deemed sufficient to satisfy the injury prong."[250] But courts now recognize that this right belongs to the entire population.

A California statute codified this right.[251] "The [relevant] statutory text makes no mention of preexisting value, and in fact can be read to presume that a person whose name, photograph, or likeness is used by another for commercial purposes without their consent is 'injured as a result thereof.'"[252] Consequently, "California courts have clearly held that 'the statutory right of publicity exists for celebrity and non-celebrity plaintiffs alike.'"[253]

Logic similarly dictates this result. In *KNB Enterprises v. Matthews*, the copyright owner of erotic photographs of non-celebrity models brought a cause of action when said photos were displayed without authorization, and for profit, on the Internet.[254] The court specifically noted that in terms of damages, the models' "anonymity . . . is allegedly a valuable asset in the marketing of erotic photographs."[255] Further, "[i]n a society dominated by reality television shows, YouTube, Twitter, and online social networking sites, the distinction between a 'celebrity' and a 'non-celebrity' seems to be an increasingly arbitrary one."[256] Therefore, the deepfake "need not be a national celebrity to prevail" in a right to publicity action.[257]

[249] *Id.*

[250] Fraley v. Facebook, Inc., 830 F. Supp. 2d 785, 807 (N.D. Cal. 2011) (citing Solano v. Playgirl, Inc., 292 F.3d 1078, 1090 (9th Cir. 2002); Newcombe v. Adolf Coors Co*.*, 157 F.3d 686, 693 (9th Cir. 1998)).

[251] *See* Cal. Civ. Code § 3344. So, too, did Nevada. *See* Hetter v. Eighth Jud. Dist. Ct. of State In and For County of Clark, 874 P.2d 762, 763 (Nev. 1994) (citing NRS 598.980–988).

[252] *Id.* at 806 (N.D. Cal. 2011) (citing Cal. Civ. Code § 3344).

[253] *Id.* at 807 (quoting KNB Enterprises v. Matthews, 92 Cal. Rptr. 2d 713, 722 n.12 (Cal. Ct. App. 2000)).

[254] *See generally* KNB Enterprises v. Matthews, 92 Cal. Rptr. 2d 713 (Cal. Ct. App. 2000).

[255] *Id.* at 718.

[256] Fraley v. Facebook, Inc., 830 F. Supp. 2d 785, 808 (N.D. Cal. 2011).

[257] Landham v. Lewis Galoob Toys, Inc., 227 F.3d 619, 624 (6th Cir. 2000).

IV. THE VIABILITY OF A § 230 DEFENSE

Even when deepfake victims are able to successfully sue the deepfaker, the deepfaker, likely an individual, may simply not have sufficient monetary funds to compensate the victim. Facing the potential to recoup only paltry sums, the victim may also choose to go after the publisher, namely the website, that hosts the video, a (likely) wealthier entity. However, the publisher would likely assert a defense from liability under § 230 of the Communications Decency Act (CDA).[258] Yet, this famous—or infamous, depending on one's perspective—shield may in fact be penetrable by deepfake victims.

A. The History, Text, and Exceptions of § 230 of the Communications Decency Act

Understanding § 230's protections and exceptions requires a thorough review of § 230's verbiage, Congress's intent in enacting it, and the interpretation of the act since the 1990s.

In October 1994, an anonymous Internet user wanted to alert the public to what he felt was fraudulent and illegal securities trading activity by Stratton Oakmont.[259] To do so, the user posted his suspicions on a message board entitled *Money Talk*, which was run by Prodigy Communications Corporation (Prodigy),[260] a leading Internet Services Provider at the time.[261] Stratton, none too pleased at the accusation, sued Prodigy as well as the particular administrator of the *Money Talk* message board for libel in New York Supreme Court.[262] Because Prodigy "held itself out to the public and its members as controlling the content of its computer bulletin boards" and "implemented this control through its automatic software screening program," the court ruled that Prodigy was

[258] *See* Communications Decency Act, 47 U.S.C. § 230 (2018).

[259] This is the same Stratton Oakmont whose founders and executives would be jailed for perpetrating myriad frauds that were given notoriety by co-founder Jordan Belfort's novel *The Wolf of Wall Street* and thereafter by Martin Scorsese's eponymous film.

[260] *See* Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995).

[261] *See* Eben Shapiro, *The Media Business; New Features Are Planned by Prodigy*, N.Y. TIMES (Sept. 6, 1990), https://www.nytimes.com/1990/09/06/business/the-media-business-new-features-are-planned-by-prodigy.html?sq=prodigy+second-largest&scp=1&st=nyt [https://perma.cc/BAC6-5BPV] ("Prodigy has become the second-largest and fastest-growing computer-information company since it was introduced in 1988.").

[262] *See Stratton Oakmont*, 1995 WL 323710 at *1.

indeed a publisher that could be found liable.[263] The court also found that the administrator acted as Prodigy's agent and thus could similarly be found liable.[264]

After the enactment of Section 230, the court noted that Congress was aware of the *Stratton* decision and:

> remove[d] the disincentives to selfregulation [sic] created by the *Stratton Oakmont* decision. . . . Fearing that the specter of liability would . . . deter service providers from blocking and screening offensive material, Congress enacted § 230's broad immunity "to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material."[265]

But the *Stratton* decision was not Congress's sole impetus for § 230. At that time, the public was just starting to understand the vast potential of the Internet and was, thus, just beginning to comprehend the sheer quantity of data and information that it could transmit. As Judge Wilkerson writes in *Zeran v. America Online, Inc.*, the "seminal case"[266] explicating the statute:

> The amount of information communicated via interactive computer services is . . . staggering. The specter of tort liability in an area of such prolific speech would have an obvious chilling effect. It would be impossible for service providers to screen each of their millions of postings for possible problems. Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted. Congress considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect.[267]

[263] *Id.* at *4.

[264] *Id.* at *6.

[265] Zeran v. Am. Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997) (quoting Communications Decency Act, 47 U.S.C. § 230(b)(4) (2018)).

[266] Bennett v. Google, LLC, 882 F.3d 1163, 1166 (D.C. Cir. 2018).

[267] *Zeran,* 129 F.3d at 331.

With dual purpose, Congress enacted § 230. The operative part reads:

> (1) . . . 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.
> (2) . . . No provider or user of an interactive computer service shall be held liable on account of—
> (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
> (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).[268]

So, to summarize, the provision protects:

> [W]ebsites against suits based on torts committed by users. For instance, Wikipedia cannot be held liable for defamation posted by a user. This intermediary liability protection encourages websites to engage in content moderation without fear that their efforts to screen content will expose them to liability for defamatory material that slips through.[269]

As the D.C. Circuit described, "the intent of the CDA is thus to promote rather than chill internet speech."[270] And in light of such protections, it has understandably been "lauded as 'the most important law protecting internet speech' and called 'perhaps the most influential law to protect the kind of innovation that has allowed the Internet to thrive.'"[271]

[268] Communications Decency Act, 47 U.S.C. § 230 (2018).

[269] Note, *Section 230 As First Amendment Rule*, 131 HARV. L. REV. 2027, 2027 (May 10, 2018) [hereinafter *First Amendment Rule*], https://harvardlawreview.org/2018/05/section-230-as-first-amendment-rule/ [https://perma.cc/TL4R-28JE].

[270] *Bennett*, 882 F.3d at 1166 (citing *Zeran*, 129 F.3d at 331).

[271] *First Amendment Rule*, *supra* note 269 (footnotes omitted).

On the other hand, the statute can, and does, protect online platforms and publishers from defamation suits brought by those who claim to have been defamed. Thus, if a victim wishes to pursue the publisher for monetary or equitable reasons, defamation causes of action, among others, may be foreclosed. Other causes of action, however, may be available through § 230's exceptions.

Despite its broad protections, § 230 is not without limits. Congress carved out four important exceptions in which an ISP *is* liable for what resides on its (digital) pages:

> (1) No effect on criminal law
>
> Nothing in this section shall be construed to impair the enforcement of . . . any . . . Federal criminal statute.
>
> (2) No effect on intellectual property law
>
> Nothing in this section shall be construed to limit or expand any law pertaining to intellectual property.
>
> (3) State law
>
> Nothing in this section shall be construed to prevent any State from enforcing any State law that is consistent with this section. No cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.
>
> (4) No effect on communications privacy law
>
> Nothing in this section shall be construed to limit the application of the Electronic Communications Privacy Act of 1986 or any of the amendments made by such Act, or any similar State law.[272]

Via these carve-outs, the CDA incentivizes service providers to actively curate their platforms and excise impermissible content.[273]

Of late, the outer bounds of these exceptions have been tested. Most notably, the first exception has come under fire, as litigants have questioned whether social media giants like Facebook and Twitter should be held liable for the actions terrorists have taken by using their platforms

[272] Communications Decency Act, 47 U.S.C § 230(e) (2018).

[273] *See Bennett*, 882 F.3d at 1166 (quoting *Zeran*, 129 F.3d at 331).

to help carry out attacks.[274] But the second exception, which withholds the statute's protections for violations of intellectual property, may create sufficient judicial daylight for deepfake victims seeking federal redress.

B. Clarifying the Intellectual Property Exception

The intellectual property exception under § 230 states, in lay terms, that individuals may still sue digital content platforms if the platform publishes copyrighted material. Digital platforms are thus heavily incentivized to remove all such content. Because deepfakes manipulate likely copyrighted videos, it is worth examining whether copyright law can provide a basis for victims' legal redress.

But a deepfake victim's ability to assert a cause of action under this exception is not a guarantee. Indeed, copyright protections may be inapposite for deepfake victims for two reasons. First, the victim likely does not own the copyright interest in the manipulated video and thus cannot claim a cause of action pursuant to property he or she does not own.[275] Second, the manipulation may be so egregious as to render the video transformative.[276]

However, intellectual property law is not *solely* constrained to copyright protections. Instead, victims may still pursue a cause of action against platforms that publish harmful and destructive deepfakes by asserting a right of publicity, a different intellectual property right. But legal hurdles and defenses, as well as strategic considerations, may foreclose this avenue of remediation. This theory of liability and two of its potential hurdles are discussed below.

[274] *See, e.g.*, Benjamin Wittes & Zoe Bedell, *Did Congress Immunize Twitter Against Lawsuits for Supporting ISIS?*, LAWFARE BLOG (Jan. 22, 2016), https://www.lawfareblog.com/did-congress-immunize-twitter-against-lawsuits-supporting-isis [https://perma.cc/9NBA-D2KW].

[275] *See* Megan Farokhmanesh, *Is It Legal to Swap Someone's Face into Porn Without Consent? Yes, No, Maybe*, VERGE (Jan. 30, 2018), https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal [https://perma.cc/B6JV-JN76].

[276] *See infra* Section V.A on Fair Use.

1. Privacy or Intellectual Property?

"[O]ne might argue that the right of publicity is a privacy issue, not an intellectual property right at all."[277] To be sure, "[t]he right of publicity is, somewhat paradoxically, an outgrowth of the right of privacy,[278] but that does not mean that the right is solely a question of a right to privacy, which would not be actionable under § 230. To the contrary, the Eleventh Circuit wrote that "there appears to be no dispute that the right of publicity is a type of intellectual property right."[279] Legal scholars agree.[280] Therefore, the right of publicity should clear this first legal hurdle, fitting within the intellectual property exception, without tremendous obstacle.

2. What Laws Define Intellectual Property under Section 230?

Were a victim to pursue a right of publicity claim under the intellectual property exemption to § 230, other issues would arise. One of these issues is whether § 230's intellectual property exception includes both state intellectual property law and federal intellectual property law. Courts are divided on this issue. The Southern District of New York, the First Circuit, and the Middle District of Florida have said "[c]laims based on intellectual property laws are not subject to § 230 immunity,"[281] while the Ninth Circuit Court of Appeals has said otherwise.[282]

In *Atlantic Recording Corp. v Project Playlist, Inc.*, a corporation named Project Playlist (Playlist) "operate[d] a website . . . that provides an index of links to songs available on third-party websites . . . [that allowed users to] download the songs from the third-party websites."[283]

[277] Jesse Lempel, *Combatting Deep Fakes Through the Right of Publicity*, LAWFARE BLOG (Mar. 30, 2018 8:00 AM), https://www.lawfareblog.com/combatting-deep-fakes-through-right-publicity [https://perma.cc/A2UU-FL9R].

[278] ETW Corp. v. Jireh Pub., Inc., 332 F.3d 915, 928 (6th Cir. 2003).

[279] Almeida v. Amazon.com, Inc., 456 F.3d 1316, 1323 (11th Cir. 2006).

[280] *See* J. THOMAS MCCARTHY, THE RIGHTS OF PUBLICITY AND PRIVACY § 1:7 (2d ed. 2018).

[281] Universal Commc'n Sys., Inc. v. Lycos, Inc., 478 F.3d 413, 422–23 (1st Cir. 2007); *see also* Malibu Media, LLC v. Weaver, No. 8:14-CV-1580-T-33TBM, 2016 WL 1394331, at *8 (M.D. Fla. Apr. 8, 2016); Atl. Recording Corp. v. Project Playlist, Inc., 603 F. Supp. 2d 690 (S.D.N.Y. 2009).

[282] Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1108 (9th Cir. 2007).

[283] Atlantic Recording Corp. v. Project Playlist, Inc., 603 F. Supp. 2d 690, 692–93 (S.D.N.Y. 2009).

> [Plaintiffs,] six of the world's largest record companies[,] sue[d] Playlist for copyright infringement and unfair competition. Plaintiffs own[ed] copyrights to the majority of sound recordings in the United States, and claim[ed] that most of the songs on the third-party websites to which Playlist provides links are posted without plaintiffs' permission, and therefore infringe[d] plaintiffs' copyrights.[284]

Facing this state law copyright claim, Playlist tried to limit the CDA to apply only to federal claims, filing a motion to dismiss under § 230 on the basis that the exception in question "means that nothing in the CDA should be construed to limit any *federal* intellectual property law."[285] The court disagreed and found that § 230's plain text did not support that contention.[286]

> The fact that Section 230(e)(3) addresses state law does not mean that a reference in another subsection to "any law" is meant to only encompass federal law. Indeed, Section 230(e)(1) refers specifically to federal criminal law, *see* 230(e)(1) (referring to "any other Federal criminal statute"), and the specific reference would be unnecessary if Playlist were correct that subsections (1), (2), and (4) covered only federal law. Playlist's contention is also contradicted by subsection (4), which refers to, *inter alia,* "any similar State law."[287]

The First Circuit reached the same conclusion, albeit without similarly rigorous analysis.[288] And, relying on these two decisions, district

[284] *Id.* at 693.

[285] *Id.* at 702.

[286] *See id.* at 702–04.

[287] *Id.* Playlist also argued that the CDA "preempts all state laws relating to intellectual property, because those laws are inconsistent with the CDA." *Id.* Judge Chin similarly discarded this claim. *See id.*

[288] *See* Universal Commc'n. Sys., Inc. v. Lycos, Inc., 478 F.3d 413, 422–23 (1st Cir. 2007) (stating "[c]laims based on intellectual property laws are not subject to Section 230 immunity").

courts in the Middle District of Florida[289] and the District of New Hampshire[290] have parroted this conclusion.

Setting up a potential Supreme Court battle due to circuit split,[291] the Ninth Circuit takes the opposing stance. In *Perfect 10, Inc. v. CCBill LLC*, Perfect 10, the publisher of an adult entertainment magazine and the owner of the subscription website perfect10.com, alleged that CCBill and CWIE violated copyright, trademark, and state unfair competition, false advertising, and right of publicity laws by providing services to websites that posted images stolen from Perfect 10's magazine and website. [292]

As the *Perfect 10* court notes, "[t]he CDA does not contain an express definition of 'intellectual property,' and there are many types of claims in both state and federal law which may—or may not—be characterized as 'intellectual property' claims."[293] The panel held that the lack of uniformity among state intellectual property laws foreclosed the possibility that the patchwork state law could form the basis of the intellectual property exception:

> Such laws may bear various names, provide for varying causes of action and remedies, and have varying purposes and policy goals. Because material on a website may be viewed across the Internet, and thus in more than one state at a time, permitting the reach of any particular state's definition of intellectual property to dictate the contours of this federal immunity would be contrary to Congress's expressed goal of insulating the development of the Internet from the various state-law regimes. In the absence of a definition from Congress, we construe the term "intellectual property" to mean "federal intellectual property."[294]

That is the extent of Ninth Circuit's analysis.

[289] *See* Malibu Media, LLC v. Weaver, No. 8:14-CV-1580-T-33TBM, 2016 WL 1394331, at *8 (M.D. Fla. Apr. 8, 2016).

[290] Doe v. Friendfinder Network, Inc., 540 F. Supp. 2d 288, 298–302 (D.N.H. 2008).

[291] *Cf.* Braxton v. United States, 500 U.S. 344, 347 (1991) ("A principal purpose for which we use our certiorari jurisdiction, and the reason we granted certiorari in the present case, is to resolve conflicts among the United States courts of appeals and state courts concerning the meaning of provisions of federal law.").

[292] 488 F.3d 1102, 1108 (9th Cir. 2007).

[293] *Id.* at 1118.

[294] *Id.* at 1118–19.

The Southern District has the better of the argument.[295] Indeed, Judge Chin[296] convincingly grappled with—and summarily disposed of—the Ninth Circuit's reasoning to show that its reasoning "lacks any support in the plain language of the CDA."[297]

In four different points in § 230(e), Congress specified whether it intended a subsection to apply to local, state, or federal law. It is therefore clear from the statute that if Congress wanted the phrase "any law pertaining to intellectual property" to actually mean "any *federal* law pertaining to intellectual property," it knew how to make that clear, but chose not to.

> Moreover, the modifier "any" in Section 230(e)(2), employed without any limiting language, "amounts to 'expansive language [that] offers no indication whatever that Congress intended [a] limiting construction.'" This conclusion is bolstered by the fact that, as discussed above, the "surrounding statutory language" supports the conclusion that Congress intended the word "any" to mean any state or federal law pertaining to intellectual property.[298]

Therefore, state law intellectual property claims—such as the right of publicity—ought to fall under the intellectual property exception, thereby clearing another hurdle. But even if the right of publicity is considered a state intellectual property claim, and further, if state intellectual property claims are considered within the § 230 intellectual property exception, liability of the deepfaker is not guaranteed. That is, while the deepfaker may not shield itself from liability under § 230 because of the intellectual property exception, the elements under a right of publicity claim, discussed *supra* Section III.B, must still be met and its relevant defenses considered.

[295] Unsurprisingly, more courts not bound by either court's opinion as precedent favored the Southern District. *See, e.g,* Malibu Media, LLC v. Weaver, No. 8:14-CV-1580-T-33TBM, 2016 WL 1394331, at *8 (M.D. Fla. Apr. 8, 2016); Parisi v. Sinclair, 774 F. Supp. 2d 310, 318 (D.D.C. 2011) ("I am not inclined to extend the scope of the CDA immunity as far as the Ninth Circuit.").

[296] This was prior to his elevation to the Second Circuit.

[297] Atl. Recording Corp. v. Project Playlist, Inc., 603 F. Supp. 2d 690, 703 (S.D.N.Y. 2009).

[298] *Id.* at 703–04 (citations omitted).

V. COPYRIGHT LAW IS NO WHITE KNIGHT, BUT THE DMCA MAY BE

Unlike in the case of revenge pornography,[299] copyright law is unlikely to provide victims an avenue of redress against perpetrators of deepfakes, despite the assertions of some commentators.[300] The Patent and Copyright Clause of the Constitution affords Congress the power "[t]o promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."[301] But because the victims of deepfakes do not own the underlying copyright of the source material, victims have no copyright claim. Additionally, only the owner of the copyrighted source material from which the deepfake was created could file a copyright infringement suit, and given the expense of litigation and the limited returns that a copyright holder may receive from a deepfake creator, it is unlikely a copyright holder would pursue a copyright infringement suit to vindicate deepfake victims. Nevertheless, the Digital Millennium Copyright Act (DMCA) may still provide victims with some reprieve.

A. Copyright Infringement

The Copyright Act effectuates that power bestowed on copyright owners as envisioned in the Patent and Copyright Clause of the U.S. Constitution. Section 102 of the Copyright Act reads:

> Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise

[299] *See generally* Levendowski, *supra* note 212 (arguing that copyright law provides a persuasive vehicle for revenge porn victims).

[300] *See, e.g.*, Cale Guthrie Weissman, *Are Deepfakes Legal? Here's What the Law Says About the Creepy Video Mashups*, FAST CO. (Feb. 13, 2018), https://www.fastcompany.com/40530634/are-deepfakes-legal-heres-what-the-law-says-about-the-creepy-video-mashups [https://perma.cc/B6HL-LPAY]; David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELEC. FRONTIER FOUND. (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them [https://perma.cc/74UN-GF6S]; Farokhmanesh, *supra* note 275.

[301] U.S. CONST. art. I, § 8, cl. 8.

> communicated, either directly or with the aid of a machine or device.[302]

"To establish infringement, two elements must be proven: (1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original."[303] First, ownership can be established by demonstrating that the claimant is the initial author or that the work was a work for hire.[304] Second, copying may occur in two different ways: through exact copying or by making a substantially similar copy. Obviously, deepfakes, because elements of the original video have been changed, are not exact copies, and thus, a court would only need to consider, under this second element, whether the deepfake was substantially similar to the original video.

A deepfake victim does not have a copyright infringement claim because he or she is not the original author and would not have created the deepfake as a work for hire. Thus, the victim is not the copyright owner, removing any viable argument for a copyright infringement claim. However, a production company—presumably the creator of the original work—is the likely owner of the copyright. Yet, even though such a production company, as the copyright owner, likely satisfies this first element, they have little economic incentive to invest in and pursue a lawsuit on behalf a deepfake victim for two reasons: (1) the uncertainty of litigation in its pursuit of an infringement claim is likely not worth the financial risk, and (2) even if the copyright owner succeeds, the deepfaker (whether attempting to create explicit videos or setting out to make a non-pornographic parody) will likely not have the funds to pay any damages. So, under element two, copying, regardless of whether the deepfake is substantially similar or not to the original work, the owner of the original work is unlikely to bring suit.[305]

Additionally, while a producer might pursue a deepfaker for their own economic purposes, it is unlikely they will pursue a cause of action for a victim's economic benefit, again given the substantial cost of

[302] 17 U.S.C. § 102 (2018) (emphasis added) (demonstrating that ownership.

[303] Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991).

[304] 17 U.S.C. § 201 (2018).

[305] In the event a copyright owner pursued an infringement suit on moral or principled grounds, the parties would need to consider whether the deepfake and the original video were substantially similar and whether the deepfaker had a fair use exception to infringement. For fair use factors, see 17 U.S.C. § 107.

litigation.[306] In other words, even if the copyright owner succeeded in the litigation and even if a deepfaker paid a damages award, the monetary damages paid to the copyright owner would not help the deepfake victim.

B. An Alternative to Litigation

While the copyright owner of the underlying original video may lack a financial interest to pursue a cause of action against deepfakers in litigation, the copyright owners may be willing to pursue a substantially less costly alternative to litigation—a DMCA takedown notice.

The Digital Millennium Copyright Act (DMCA)[307] provides an alternative to litigation that may help reduce the harm to victims. The DMCA provides a safe harbor to online services from copyright infringement for hosting copyrighted material on their platform if, and only if, the platform makes a good faith effort to take down the material upon being notified of its existence.[308] "Merely by sending a proper takedown notice, a copyright owner can prompt an Internet Service Provider ("ISP") to swiftly remove an allegedly infringing item from its servers; ISPs earn immunity from infringement liability if they provide that swift removal and thus are incentivized to comply." [309] Therefore, if the copyright owner of a video is informed by a deepfake victim of the deepfake and if the copyright owner informs the video hosting platform

[306] Albeit, there may be scenarios where a wealthy individual would be willing to pay a copyright owner to bring suit against the deepfake creator and pay the litigation costs, but this in turn raises a question of whether a case or controversy actually exists in such a matter, a discussion which is beyond the scope of this paper.

[307] The law, signed by President Clinton on October 28, 1998, "implements two 1996 World Intellectual Property Organization (WIPO) treaties: the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty." U.S. COPYRIGHT OFFICE, THE DIGITAL MILLIENNIUM COPYRIGHT ACT OF 1888: U.S. COPYRIGHT OFFICE SUMMARY 1 (1998), https://www.copyright.gov/legislation/dmca.pdf [https://perma.cc/YV7Y-B5N3].

[308] *See* 17 U.S.C. § 512(b)(2)(e). For more information on the takedown process, see *What is the DMCA notice and takedown process?*, COPYRIGHT ALLIANCE, https://copyrightalliance.org/ca_faq_post/dmca-notice-and-takedown-process/ [https://perma.cc/69UU-VU4B].

[309] It should be noted that copyright holders would likely be able to pursue notice and take down actions. Under the 1998 Digital Millennium Copyright Act, Lydia Pallas Loren, *Deterring Abuse of the Copyright Takedown Regime by Taking Misrepresentation Claims Seriously*, 46 WAKE FOREST L. REV. 745, 747 (2011). Such actions would help victims minimize the harm from a deepfake but would provide no deterrent to deepfake creators. Moreover, it improperly puts the burden on victims to supervise the Internet for deepfakes.

that the deepfake has been posted, the online service provider will likely take down the video to protect itself from litigation.

Given the choice between a resource-intensive and time-consuming lawsuit and a swift, viable alternative, copyright owners are likely to prefer enforcing their rights with a DCMA takedown. Additionally, the relatively low cost and minimal effort of the DMCA takedown process may even incentivize copyright owners to file takedown notices on behalf of deepfake victims. So even though a victim may not be able to procure monetary damages from a copyright lawsuit, because they do not own the underlying copyright of the original work, the DMCA takedown provision, with the cooperation of the copyright owners, provides a promising alternative to litigation that might achieve what the deepfake victim really desires: removal of the victimizing video from the Internet.

CONCLUSION

Deepfakes are a problem. That much is certain. The technology is easily deployable, growing in prevalence, and seeing its technological underpinnings improve. Additionally, deepfakes have the potential to be weaponized in serious and global ways.[310] Simultaneously, efforts to combat deepfakes, though growing, do not appear to keep pace with the technological prominence.

As this article demonstrates, this problem lacks a clear-cut solution. Neither an outright ban of deepfakes nor a bill seeking only to regulate their production is unlikely to survive a court challenge. The First Amendment likely provides sufficient refuge for deepfakers to guard against such measures.

As discussed, four possible state tort remedies may intuitively come to mind—intrusion upon seclusion, publicity given to private life, publicity in false light, and wrongful appropriation. But wrongful appropriation is likely the only tort action in which a deepfake victim may successfully seek refuge.

Irrespective of the theory pursued, § 230 of the Communications Decency Act creates a shield for content providers, preventing victims from naming the host platform as a defendant, thereby limiting recovery to

[310] *See* Chesney & Citron, *supra* note 6; Kaveh Waddell, *The Impending War Over Deepfakes*, AXIOS (July 22, 2018), https://www.axios.com/the-impending-war-over-deepfakes-b3427757-2ed7-4fbc-9edb-45e461eb87ba.html [https://perma.cc/KUJ9-7BU7].

the deepfake creator. Moreover, the costs of litigation—temporally and monetarily—when counterbalanced against potentially limited damages, will likely dissuade a victim from pursuing legal action.

The lack of available remedies should compel reflection on both the legal frameworks at play, as well as the technical precipice on which we sit. Reconsideration of what speech is and is not protected by the First Amendment may be warranted, and courts may want to consider whether non-consensual pornography of *any* kind, revenge porn or deepfakes, should be the foundation of a new exception to the First Amendment's broad protections.[311] By the same token, Congress may want to reflect on to whether or not § 230 needs revisiting to accommodate an Internet that has changed considerably since 1996.

[311] *See* Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1312 (2017).