

50 Seton Hall L. Rev. 177

Seton Hall Law Review

2019

Comment

Elizabeth Caldera ^{a1}

Copyright © 2019 by Seton Hall University School of Law; Elizabeth Caldera

“REJECT THE EVIDENCE OF YOUR EYES AND EARS”:¹ DEEPFAKES AND THE LAW OF VIRTUAL REPLICANTS

I. INTRODUCTION

Former President Barack Obama sits in front of the American flag as he delivers an address.² “[We are] entering an era in which our enemies can make it look like anyone is saying anything at any point in time,” he warns.³ Although he is using familiar inflections and hand gestures, there is something just slightly off about the video. Obama's face appears slightly contorted, and his voice, even with the inflections, sounds flat and forced.⁴ It is difficult to place exactly what is wrong with the video, and it only gets stranger from there. Obama references *Black Panther* and *Get Out*, and, in an out-of-character move, calls President Trump a “total and complete dipshit.”⁵ The video is unbelievable, and it is supposed to be. At the thirty-six second mark, the screen splits, and it becomes evident that Oscar-winning filmmaker and comedian Jordan Peele is behind the stunt.⁶ Despite appearances, Obama is not speaking.⁷ Instead, Peele used artificial intelligence to manipulate previous videos of Obama, along with technology to manipulate audio, to create an incredibly realistic video of Obama saying and doing things he has never said or done.⁸

*178 Fake videos of this type are known as “deepfakes.” In the span of about a year, deepfakes have advanced to the point where they are nearly indistinguishable from authentic videos. Using a mix of artificial intelligence and machine learning, the technology behind them will only continue to advance. As more Internet users learn how to harness deepfake technology, these videos will become more widespread and begin to creep into the public consciousness. As deepfakes become more popular, the ability to distinguish between which videos are authentic and which are doctored will begin to diminish, causing the potential for social, legal, and political harms in a variety of areas in our daily lives. But as of 2019, deepfakes are unregulated, and no clear area of law governs them. This Comment will argue that placing regulatory authority in the hands of federal agencies, and specifically the Federal Trade Commission (the “FTC”), is the best method of regulating this technology. It will, accordingly, propose potential regulations for implementation.

Part II of this Comment will discuss exactly what deepfakes are, describe the technology behind them, explain their rise, denote some popular examples, and analyze the types of harm that this technology can cause. This section will demonstrate the need for some form of regulation to address this technology. Part III will explain what measures are currently in place to address the rise of deepfakes, and it will then compare different methods of regulating deepfakes. Part IV will analyze different administrative agencies that could potentially regulate deepfakes, and it will then focus on why the FTC is the best choice currently available. Part V will outline what potential FTC regulations could include. Part VI will address some limitations and challenges the FTC regulation of deepfakes would face. Part VII will conclude.

II. DEEPAKES: WHAT THEY ARE AND WHY THEY ARE DANGEROUS

This section will provide a definition for “deepfakes,” explain the advance of the technology that created them, trace a broader history of photo and video manipulation, and describe the harms this technology can bring.

A. *What Is a Deepfake?*

Combining the words “deep learning”⁹ and “fake,” a deepfake is a “hyper-realistic digital falsification of images, video, and audio.”¹⁰ Put *179 simply, a deepfake is a forged video; it depicts something that has never happened by manipulating previously existing video footage or pictures.¹¹ Jordan Peek's deepfake of Obama utilized real videos of past addresses and used those clips to create an entirely new video, with the ability to depict Obama saying essentially anything Peele wished.¹² The implications of this technology are far-reaching¹³ and will be explored in detail throughout this Comment.

Examples of deepfakes range from the silly to the sinister. Some of the lighter applications of deepfakes include videos putting Nicholas Cage into famous scenes from movies such as *Raiders of the Lost Ark* or videos of a Wall Street Journal reporter performing Bruno Mars's dance moves.¹⁴ But because deepfakes' origins are closely tied to pornography, a darker point of focus for many deepfakes involves creating pornographic videos of famous celebrities.¹⁵ Another disturbing use of deepfakes involved a fake video of gun control activist Emma Gonzalez tearing up a copy of the Constitution.¹⁶ While the original video featured Gonzalez tearing up a target to advocate for gun control, someone manipulated the image for incendiary purposes.¹⁷ This wide range of potential uses for deepfakes encapsulates their potential to harm.¹⁸ While benign utilizations can and will exist, the early prevalence of pornographic applications likely indicates an ongoing problem for deepfakes.¹⁹ And, in a similar vein, the doctored video of Emma Gonzalez *180 demonstrates deepfakes' potential for deepening America's fake news crisis.²⁰ These different uses highlight the impact deepfakes could have upon our society and demonstrate the need to focus on this issue now.

To frame deepfakes in a relevant pop culture context, an elucidating analogy comes from the Ridley Scott science fiction movie *Blade Runner*.²¹ In the film, technology has evolved to create human-like androids called “replicants” that are virtually identical to human beings, aside from their synthetic creation.²² It requires an extensive “Voight-Kampff” test to determine whether a being is a human or a replicant.²³ The film has become part of the science fiction canon, and its cult legacy became cemented thanks in part to the ambiguity surrounding whether even its main character, Rick Deckard, is a human or a replicant.²⁴ One of the film's central tensions revolves around the diminishing boundary between man and machine,²⁵ and this tension highlights the anxieties that surround deepfakes. Like replicants, deepfakes are advancing to a point where it will be impossible to determine whether a video is authentic.²⁶ Currently, tech companies and the US government are developing de facto “Voight-Kampff” tests to accurately determine when a video is a deepfake, but as technology advances, the effectiveness of any test becomes questionable.²⁷ Like the debate surrounding whether Deckard is a replicant, the debate over which videos are fake and which are real could wage for a long time.

*181 B. *The Technology Behind Deepfakes*

The advancement of various forms of technology precipitated the rise of deepfakes. Artificial intelligence,²⁸ machine learning,²⁹ and generative adversarial networks (“GANs”)³⁰ are the tools that allow users to create deepfakes.³¹ Basically, the technology that creates these videos works by having “a computer program find[] common ground between two faces and stitch[] one over the other.”³² By utilizing previously existing images and videos, the technology creates a generated video that

nevertheless looks authentic.³³ One of the technological components behind deepfakes--deep learning--“consists of networks of interconnected nodes that autonomously run computations on input data.”³⁴ Deep learning only allows software to go so far, though, and its main strength is its ability to discriminate between data.³⁵ GANs, however, have helped technology make large strides toward creating, rather than merely manipulating, realistic fake images.³⁶ GANs give software competition as a motivator to create more realistic-looking images.³⁷ Generative software under the GAN model “learn[s] to create images that look real, but are not” by having the software attempt to fool an adversary.³⁸ For audio, GANs use neural networks to learn and then reproduce the properties of a source, modeling speech on a millisecond-by- *182 millisecond basis.³⁹ In short, algorithms are reaching a point where a user need only input a recording of a speech from a public figure into a GAN to create realistic audio of that same public figure saying whatever the user wants him or her to say.⁴⁰ When that manipulated audio combines with a GAN-created video, the result is a video that both looks and sounds like the figure in the video but that in actuality is a fabrication.⁴¹ Some of the more popular deepfakes have been “created with a machine learning algorithm, using easily accessible materials and open-source code that anyone with a working knowledge of deep learning algorithms could put together.”⁴² As one artificial intelligence expert states, “[t]his is no longer rocket science.”⁴³ This is one of the reasons why deepfakes are so dangerous: the materials are open to the public, and anyone with a working knowledge of the technology can use them to create virtually whatever he or she wants.⁴⁴

C. The History of Photo and Video Manipulation

For nearly as long as photography has existed, humans have found ways to manipulate the medium.⁴⁵ One early example is an iconic portrait of Abraham Lincoln dating back to 1860.⁴⁶ Although the image appears authentic, the picture is a combination of photographs of Lincoln's head and John Calhoun's body.⁴⁷ The entire field of spirit photography depended on using techniques such as multiple exposure and combination printing to generate fake images of loved ones with passed-on family members.⁴⁸ *183 Manipulated photos have also had political consequences. Millard Tydings may have lost his 1950 re-election bid to the United States Senate in part due to a manipulated photo depicting him conversing with a leader of the Communist Party.⁴⁹ But the popular photo-editing software Photoshop is currently the most well-known example of photo manipulation technology.⁵⁰ Photoshop was first invented in 1987 and was widely disseminated by 1990.⁵¹ Today, Photoshop is a well-known tool in a photographer's arsenal, used to manipulate everything from magazine covers to Instagram posts.⁵²

Although there is a longer history of photo manipulation, video manipulation also has a long and storied history.⁵³ The first multi-scene motion pictures involved literally cutting and taping pieces of film on an editing table.⁵⁴ More pertinently, though, film can be used to generate images.⁵⁵ The 1970s marked the beginning of computer animation, using layered 2D images to create visual effects.⁵⁶ The first feature-length film wholly created using computer-generated imagery (CGI)⁵⁷ was Pixar's *Toy Story*, which premiered in 1995.⁵⁸ The technology has advanced since then, and has since been used to capture the movements of actors to render CGI-created characters by using motion capture technology.⁵⁹ There are more controversial applications of this technology as well, including discussions over whether or not filmmakers should use CGI to create performances from deceased actors.⁶⁰ These applications, however, do not compare to the reality of deepfakes and the technology behind them. Deepfakes essentially combine the cutting and pasting technique with image-generation *184 technology, editing together a video from previously existing footage to create something that is as fake as a CGI creation.⁶¹ Additionally, one of the hallmarks of CGI is its connection to animation studios and film.⁶² What was previously the domain of a visual effects department or a special effects company can now be created by virtually anyone, at low cost, with the same effect.⁶³

D. *The Rise of Deepfakes*

Tracing the rise of deepfake videos gives a sense of both the technology's rapid development and how the technology may produce harms. What may be considered the spiritual ancestor of deepfakes is the Internet phenomenon known as ElfYourself, where users insert photographs of faces into a preset video of Christmas elves dancing to Christmas songs.⁶⁴ Despite the parallels between how these videos and deepfakes are made, the obvious superimposition of the heads on the fake elf bodies make it sufficiently clear that the ElfYourself videos are fake.⁶⁵ There is also a trend of editing speeches of well-known politicians to make it appear as though they are singing well-known pop songs. For example, the popular YouTube account “baracksdubs”⁶⁶ takes snippets of phrases from former President Barack Obama's speeches to correspond to the lyrics of songs such as “Call Me Maybe.”⁶⁷ The resulting videos are choppy, with virtually no transition between the words of the songs.⁶⁸ With these Internet trends, there is an obvious fakeness to the videos that adds to their humor. The sophisticated deepfakes produced today, though, are not necessarily created for humor; rather, some of them are created for incendiary purposes or for humiliation.⁶⁹ Even though there are similarities between the Internet trends and deepfakes, the differences between them are extreme enough to demonstrate how the swift rise of deepfakes presents a host of problems that these Internet trends *185 do not.

The current iteration of hyper-realistic, simulated deepfakes began on the social media website Reddit.⁷⁰ The first true deepfake⁷¹ appeared on the subreddit r/CelebFakes, which is “mainly devoted to photoshopping celebrities to appear nude.”⁷² But on September 30, 2017, Reddit user ‘deepfakes’ posted a virtual recreation of actress Maisie Williams's face.⁷³ ‘Deepfakes’ then started his own subreddit r/deepfakes, where he publicly released the script he used to create the face-swaps.⁷⁴ Users within the subreddit then began to “[build] on each other's data sets to create even more convincing facial swapping models.”⁷⁵ Today, the technology is more widely distributed than ever, in part due to the release of an app called “FakeApp,” which helps users create deepfakes.⁷⁶ FakeApp allows anyone to make these videos so long as they have “one or two high-quality videos of the faces they want to fake.”⁷⁷ These advancements have allowed deepfakes and the technology that creates them to become both more widespread and more advanced than originally predicted.⁷⁸ The chief computer scientist of the Electronic Frontier Foundation estimated that it would take a year or two for the technology behind deepfakes to advance far enough to make it incredibly difficult to distinguish between an authentic video and a deepfaked video.⁷⁹ Instead, it only took about two months for deepfakes to become “incredibly convincing” as more and more people began to experiment with the AI-assisted model.⁸⁰ Although currently the most prevalent use of the technology is pornographic videos of celebrities, it is easy to foresee how this technology can create future social, legal, and political harm.

186 E. *Analysis of Potential Harms

As the technology behind deepfakes becomes more and more sophisticated at a rapid pace, it has the potential to create serious harms in a variety of different areas, including revenge pornography, fake news, and the reliance of video as a medium. While deepfakes may also have some beneficial uses, as discussed later in the Comment, the detriments may outweigh the benefits.⁸¹

Because deepfakes began to create celebrity pornography, it is easy to imagine that bad actors will use the technology to create revenge porn for non-famous individuals as well.⁸² Revenge porn, also known as “involuntary porn”⁸³ or “nonconsensual pornography,” involves “the distribution of sexually explicit photos or videos of another individual without that individual's consent or knowledge.”⁸⁴ Revenge porn may involve the distribution of explicit photos or videos taken without consent,⁸⁵ taken consensually but with an understanding of privacy,⁸⁶ or created via “sexualised photoshopping.”⁸⁷ With the rise of deepfakes, the possibility of “sexualised photoshopping”⁸⁸ now exists for both images and videos. It is likely that “the majority of victims of fake sex videos will be female,” in part due to revenge porn's popularity.⁸⁹ Indeed, some social media users have

already indicated interest in creating deepfakes with various women in their lives.⁹⁰ There has already been at least one private figure who has been a victim of revenge porn in the form of a deepfake.⁹¹ Noelle Martin of Perth, Australia, had already been a victim of revenge porn for years before anonymous predators photoshopped images of her face onto pornographic pictures of someone else's body.⁹² But recently, the attacks have escalated, *187 “doctoring [her] into pornographic videos which appear to show [her] performing numerous sexual acts.”⁹³ Unfortunately, stories like this are becoming more common as deepfakes become even more widespread and advanced.⁹⁴ While some deepfakes of this kind exist solely for sexual gratification, it is highly probable that others will intend to humiliate the person whose likeness is featured in the video.⁹⁵

America is already a country flooded with fake news.⁹⁶ As Jordan Peek's video of Obama shows, technology has advanced to allow fake videos of prominent political figures to appear alarmingly realistic.⁹⁷ If bad actors use deepfakes to proliferate fake news, the harm to America's media system will only worsen.⁹⁸ Because of the ability to both rapidly create and distribute fake content, a computer science professor from Dartmouth fears a “perfect storm” of disinformation.⁹⁹ Part of what makes deepfakes so dangerous is how they exploit the natural human tendency to rely on observation through senses such as sight and sound.¹⁰⁰ The prevalence of fake videos, however, will disrupt that reliance.¹⁰¹ Conversely, the inability to distinguish between authentic and doctored videos will lead to the possibility that any form of video would be distrusted as “fake news.”¹⁰² This “liar's dividend” will only grow as the public becomes more informed about what deepfakes are and the dangers they pose.¹⁰³

Because of the technology's versatility, there is a high probability that deepfakes could be used in any context that uses regular video. The possibilities of blackmail, extortion, “reputational sabotage,” problems finding employment, and more all point to the ways that individuals will face legal and social problems if they cannot prove that a video appearing to feature them in an unsavory position is actually doctored.¹⁰⁴ But if deepfakes *188 become popular in the mainstream, the value of real videos will diminish.¹⁰⁵ This devaluation of video will have the long-term effect of increasing the effectiveness of deepfakes.¹⁰⁶ If video cannot be trusted, having a corroborating video to debunk a deepfake would no longer be sufficient; the risk of the supposedly corroborating evidence also being a deepfake may be too high if there is no ability to determine if a video has been doctored.¹⁰⁷ One reason videos are so powerful is that we tend to believe the things that we can see and hear.¹⁰⁸ Until now, video has been a relatively reliable source of information.¹⁰⁹ But once deepfakes become more popular, the value of any video, real or fake, will necessarily diminish without a reliable way to determine whether a video has been manipulated or not.

III. CURRENT RESPONSES AND POTENTIAL PATHS FORWARD

Some groups are currently attempting to limit the reach of deepfakes, while others are actively countering their rise.¹¹⁰ This section begins by discussing current technological efforts to detect deepfake technology. It then surveys potential areas of law that could apply to deepfakes, discussing the effectiveness of different fields.

A. *What is Being Done About This Issue?*

Researchers have been attempting to develop algorithms and other AI-assisted tools to determine whether a video is a deepfake or not.¹¹¹ Researchers at Carnegie Mellon University have utilized a tool to determine whether a video is a deepfake by analyzing the pulse of the subject.¹¹² An individual's pulse tends to stay constant, even at different pulse points; however, if a video was created by layering images and videos on top of each other, then what seems to be one individual in a video may have different *189 pulses at various pulse points.¹¹³ The tool picks up those differences as evidence that a video is actually a deepfake.¹¹⁴ Another technological response has been to rely on the “lack of physiological signals intrinsic to human beings”

that often results when creating a “synthesized video[.]” of an individual.¹¹⁵ One such example is analyzing whether and how often the subject of a video blinks in order to determine whether a video is a deepfake.¹¹⁶ Because “most training datasets do not contain faces with eyes closed,” a video created using AI likely will not include blinking or will include blinking at a slower rate than a real subject.¹¹⁷ Therefore, blinking and the lack thereof may be a “telltale sign” of when a video is a deepfake.¹¹⁸ One flaw with technological approaches, though, is that even if a specific algorithm or tool can accurately spot manipulated videos, creators can merely find new ways to produce deepfakes that circumvent these algorithms and tools.¹¹⁹ Consequently, even if researchers or tech companies can develop a reliable method to determine a deepfake, there is always a risk of developers advancing the technology beyond those detection methods.

Another response to limit the spread of deepfakes has been for some websites to ban the use of these videos on their platforms.¹²⁰ Pornhub has begun removing deepfakes from its site, although it appears that the process relies on user reports rather than administrative monitoring or the use of an algorithm.¹²¹ Reddit also has taken action, deleting the subreddit *r/deepfakes* where these videos first began to arise.¹²² Other platforms, such as Discord, Gyfcat, and Twitter, have clarified that face-swap porn is prohibited on their sites, although this does not appear to be a universal ban on deepfake videos.¹²³

The United States government is also aware of the issues that deepfakes *190 raise, and the Department of Defense is developing technology that could help spot deepfakes.¹²⁴ The U.S. Defense Advanced Research Projects Agency (DARPA) changed the mission of its Media Forensics program in order to focus on developing technology to stop deepfakes.¹²⁵ DARPA is also currently “funding a project that will try to determine whether the increasingly real-looking fake video and audio generated by artificial intelligence might soon be impossible to distinguish from the real thing.”¹²⁶ But because the technology has advanced so rapidly, these early efforts at handling the problem may not be sufficient. More urgent action is necessary to effectively address the harms that deepfakes can create.

Congress has also taken notice of this issue, with Senators on both sides of the aisle expressing concerns about the political threat deepfakes could pose.¹²⁷ Democrat Mark Warner, the Vice Chairman of the Senate Select Committee on Intelligence, “absolutely” believes that deepfake videos will be the “next phase of disinformation campaigns.”¹²⁸ Republican Marco Rubio also warned about the power of manipulated videos to “sow discontent and divide [Americans].”¹²⁹ Additionally, a bipartisan group in the House of Representatives penned a letter to the Director of National Intelligence expressing concerns that deepfakes may pose a threat to national security.¹³⁰ The letter asks the Intelligence Community for a “report to Congress and the public about the implications” deepfakes may have when individuals use them in bad faith.¹³¹ The letter's main concern is with malicious foreign actors using deepfakes to spread misinformation throughout America or to blackmail the subjects for political purposes.¹³² The letter ends by requesting the identification of deepfakes created by foreign actors, identification of potential countermeasures that can be adopted, and recommendations about the next steps Congress and the intelligence community can take to stem the *191 rise of deepfakes.¹³³ While the letter indicates more of a concern with national security than the personal harms that can arise from deepfakes, Congress's decision to get involved in this issue may be a positive sign that systems can be put in place to redress at least some of the harms from deepfakes.¹³⁴

B. What Areas of Law Govern Deepfakes?

To complicate the problems deepfakes cause, it is currently unclear what area of law would provide legal recourse for victims.¹³⁵ At least one law professor believes that victims of deepfakes would have little to no legal recourse.¹³⁶ As a threshold issue, victims would have limits in who they would be able to sue.¹³⁷ Because of the prevalence of anonymity on the Internet, if an individual harmed by a deepfake cannot find the creator of the video, that individual may not have an identifiable

party to sue.¹³⁸ Additionally, the Communications Decency Act grants websites immunity for claims about content from third parties.¹³⁹ Therefore, suing a social media website for hosting a deepfake is an unlikely path of success.¹⁴⁰

Beyond this initial limitation, though, is the deeper problem of what area of law governs the use and applications of deepfakes. There is a possibility that defamation claims may be effective “because the person depicted in the video [is not] actually in it.”¹⁴¹ But if a creator makes clear that a video is a deepfake and does not actually feature the person whose likeness appears, the success of a defamation claim may be unlikely.¹⁴² Additionally, victims may face problems in “proving that the creators intended to cause them emotional distress,” adding further difficulties to winning on a defamation claim.¹⁴³

***192** A right of publicity claim could also be an avenue to address harm from deepfakes.¹⁴⁴ Although typically associated with celebrities, “the right of publicity protects the commercial value of any person's identity.”¹⁴⁵ If a creator profits from using another person's image in a deepfake without that person's consent, the person whose likeness appears may be able to bring a right of publicity claim.¹⁴⁶ One benefit for victims bringing this claim is that the right to bring the claim does not depend upon legal ownership of the image.¹⁴⁷ But one of the claim's limitations is that it depends upon the “deepfakes [being] sold or the creator receiv[ing] some other benefit from them;” therefore, this may not be a route all victims of deepfakes could utilize.¹⁴⁸

Copyright infringement would also be an effective area of the law in which to address deepfakes.¹⁴⁹ That route would, however, require the person affected by the deepfake to have taken the video in the first place.¹⁵⁰ Additionally, the person who owns the original video may or may not be the same person the deepfake actually harmed.¹⁵¹ Furthermore, even if the person harmed has a copyright, a deepfake creator may be able to claim that courts would consider deepfakes to be fair use.¹⁵² Although a full discussion of copyright infringement and fair use is beyond the scope of this Comment, the transformative purpose of copyrighted material is a key distinction in qualifying its usage by others as fair use or not.¹⁵³ The essence of deepfakes is taking previously existing images and manipulating them to create a new video.¹⁵⁴ It is certainly possible that a court would consider this type of use to be transformative: the user is transforming those previous images into a new medium, often depicting scenarios that have not actually happened or placing those images into a new context. Therefore, while copyright law ***193** would provide some protections, those protections are limited.¹⁵⁵

Revenge pornography presents similar harms as deepfakes, but current criminal laws addressing revenge porn would not be sufficient to address this problem.¹⁵⁶ Statutes addressing revenge porn often are premised upon violations of privacy, and deepfakes--at least in the pornography context--would likely not be considered a privacy issue in the eyes of the law.¹⁵⁷ The problem with predicating deepfaked revenge porn videos on existing revenge porn statutes is that the underlying video would likely not include the body of the victim.¹⁵⁸ This amalgamation would complicate issues of privacy because “you [cannot] sue someone for exposing the intimate details of your life when [it is] not your life [they are] exposing.”¹⁵⁹ If courts do not “agree that the victim *becomes* the nude person in the deepfake for purposes of nonconsensual pornography statutes,” then the current statutory scheme for revenge pornography would likely be insufficient to provide redress for victims of revenge porn created via deepfake.¹⁶⁰

IV. ANALYSIS OF ADMINISTRATIVE AGENCIES

Because of the potential harms deepfakes present, the lack of clarity surrounding which area of law would govern, and the rapid rise and advancement of the technology that creates these videos, federal administrative agencies would provide the fastest, most effective method of providing a form of regulation for deepfakes.¹⁶¹ Current administrative agencies that may be viable options for creating regulations for deepfakes include the FTC and the Federal Communications Commission (the FCC). But

a new agency may be necessary to more effectively address the problems that deepfakes create and the broader issues the advancement of technology such as artificial intelligence and advanced algorithms pose for our modern society. This section will analyze each option in turn.

A. The Federal Trade Commission

The FTC's mission is to “[work] to protect consumers by preventing anticompetitive, deceptive, and unfair business practices, enhancing *194 informed consumer choice and public understanding of the competitive process, and accomplishing this without unduly burdening legitimate business activity.”¹⁶² The FTC accomplishes these goals through both regulation and litigation.¹⁶³ Its “unique dual mission” of consumer protection and competition protection makes it a potential option for regulating deepfakes.¹⁶⁴

Due to the FTC's ability to “develop rules to establish a vibrant marketplace,”¹⁶⁵ along with its oversight over data security issues,¹⁶⁶ it would likely be able to create effective regulations addressing the use of deepfakes.¹⁶⁷ One of the benefits of having the FTC handle deepfakes is that it may be within the FTC's jurisdiction to hold liable the creator of a deepfake app, such as FakeApp.¹⁶⁸ Because the technology that creates deepfakes “is using someone's data and morphing it onto someone else's,”¹⁶⁹ there is a possibility that the nonconsensual use of data would bring the technology within the range of “unfair or deceptive acts or practices in or affecting commerce.”¹⁷⁰ Using its rulemaking and enforcement abilities, the FTC may be able to create regulations delineating permissible and impermissible uses of deepfakes.¹⁷¹

The FTC could also be a good candidate to regulate deepfakes because of deepfakes' similarities with fake news; if the FTC views fake news more narrowly as false advertisement or spam, then fake news could potentially be seen as an “unfair or deceptive act[] or practice[] in or affecting commerce,” bringing it within the jurisdiction of the FTC.¹⁷² The FTC has *195 already acted against fake news in certain scenarios, shutting down “fake news” sites if they are in a “commercial context[].”¹⁷³ If the FTC were to view fake news as “a kind of commercial offering in which ‘the political misinformation *is* the product,’” then the FTC may be able to prevent its spread.¹⁷⁴ By extending this reasoning to deepfakes, then the FTC may be able to effectively regulate at least some forms of deepfakes.

But the agency's emphasis on commercial practices may present problems for most forms of deepfakes.¹⁷⁵ A commercial component may be key in order for the FTC's jurisdiction to extend.¹⁷⁶ If a deepfake is made for noncommercial reasons, such as sexual gratification, humiliation of the subject, or as a parody for entertainment purposes, then the deepfake may not fall under the FTC's jurisdiction. Therefore, any rulemaking ability the FTC may have in regard to addressing deepfakes would likely be limited, and regulations would need to be narrowly tailored in order to ensure that the FTC does not go beyond the bounds of its jurisdiction.

B. The Federal Communications Commission

The FCC is in charge of “regulat[ing] interstate and international communications by radio, television, wire, satellite and cable,” and it is the “primary authority” on issues including “communications law, regulation[,] and technological innovation.”¹⁷⁷ The FCC would also be a potentially viable candidate for producing regulations surrounding the use of deepfakes because of its involvement with media.¹⁷⁸ One benefit of choosing the FCC to create regulations for deepfakes is that the agency has already created rules regarding false information for broadcasters on television and the radio.¹⁷⁹ If the FCC could provide a similar regulatory role for Internet “broadcasters,” then the FCC would be a viable choice as a regulator of deepfakes.

But there are many questions regarding the FCC's ability to regulate *196 the Internet.¹⁸⁰ With the Restoring Internet Freedom Order in effect, the FCC removed net neutrality protections and took a less active role in regulating the Internet.¹⁸¹ The Order “replaces unnecessary, heavy-handed regulations that were developed way back in 1934 with strong consumer protections, increased transparency, and common-sense rules that will promote investment and broadband deployment.”¹⁸² The passage of this order indicates the FCC's desire to step away from regulating the technology behind the Internet, as well as from regulating the Internet itself.¹⁸³

C. A New Agency

There is also a possibility that deepfake technology is so new and so specialized that any current agency would be unable to properly regulate the use and spread of the technology. Instead, it may be time to implement a new agency to handle more general aspects of Internet law, such as artificial intelligence or robotics. “[B]ig events or changes in behavior” tend to bring about new, complex, specific problems that current regulatory structures may not be fully equipped to handle.¹⁸⁴ New agencies then develop as a means of addressing those new problems more effectively.¹⁸⁵

One potential new agency could be an Agency of Artificial Intelligence. As the rapid development of deepfakes proves, artificial intelligence is an increasingly powerful technology with the potential to create intense changes in our society, both positive and negative. As the premise of *Blade Runner* demonstrates, humanity has long wrestled with questions about the freedoms and limits we should place on artificial intelligence.¹⁸⁶ An agency for artificial intelligence would be better able to address the technology that creates deepfakes, including not only artificial intelligence but also advanced algorithms, deep learning, and machine learning. An example of a regulation from this hypothetical agency may include sourcing images and ensuring consent from parties before using their likenesses to create a deepfake. The ability to regulate the technology that creates deepfakes would allow more *197 effective implementation of these kinds of regulations.

Creating a new agency to handle general issues arising from the increasing development of the Internet might not be feasible if there is a lack of momentum to create a new regulatory body.¹⁸⁷ For example, there is not a “strong push” to create a similar regulatory body for the Internet of Things.¹⁸⁸ Additionally, President Trump has issued an executive order meant to cut back on federal regulations; now, “[F]or every one new regulation issued, at least two prior regulations [must] be identified for elimination.”¹⁸⁹ Although the order does not disallow the issuance of new regulations, it does indicate the lack of a “strong push” for new regulations, and the development of a new regulatory agency would provide logistical difficulties in light of this executive order.¹⁹⁰

Overall, of these three options for agency regulation, the FTC currently provides the strongest avenue for developing effective regulations for deepfakes. Because of its dual capacity to create and enforce regulations, its precedent with handling at least certain types of fake news, and its mission to protect consumers from deceptive practices, the FTC is the most likely agency to have jurisdiction over deepfakes. While the FCC may be able to similarly create guidelines, its move away from regulating the Internet diminishes the likelihood of the FCC taking a more active role in stemming harmful deepfakes. Additionally, while a new agency would likely be the most effective option, it would take time to establish, and current circumstances indicate that there is no strong desire or plan to create a new agency. Therefore, the FTC would be able to quickly produce and implement regulations to minimize and control the harms that deepfakes can produce.

V. POSSIBLE GUIDELINES FOR REGULATION

As discussed throughout this Comment, deepfakes have the capacity to produce a wide variety of harms; however, because deepfakes are so new, it may be unclear to Internet users exactly what deepfakes are and which uses of them are likely to

create harm. Any FTC regulations would have to clearly define deepfakes and delineate what uses the agency would consider permissible and impermissible.

Any guidelines created by an agency would need to officially define what a deepfake is in the eyes of that agency. Just as “fake news” is *198 becoming a catch-all term to the point that the phrase “fake news” is beginning to lose its meaning, the term “deepfakes” may reach a similar point.¹⁹¹ Critics argue that the term “deepfake” has “become a stand-in for ... AI-assisted face swaps.”¹⁹² Including artificial intelligence in a regulatory definition for deepfakes may make any resulting definition under-inclusive. Although AI has made the process quicker and more sophisticated, there may be other methods of creating deepfakes that do not require the use of AI.¹⁹³ Therefore, a definition at this point can be simple: a deepfake is a video appearing to be authentic but that is created from other images, videos, or audio.¹⁹⁴

Providing a taxonomy of deepfakes may be useful for regulators to have a clearer understanding of what types of videos would classify as deepfakes, and understanding the differences between them can help regulators draw clearer lines to address the specific types of deepfakes they encounter.¹⁹⁵ One category of deepfakes would be when an original video is manipulated or altered in a way that distorts the reality of the original video. An example of this would be the “shallowfake”¹⁹⁶ video of Jim Acosta released by the White House,¹⁹⁷ where a single video of Acosta's interaction with an intern had a segment sped up to make it appear as though he “karate chop[ped]” her when she attempted to take his microphone.¹⁹⁸ But in the original video, the contact between the two seemed incidental and not as aggressive as it *199 appeared in the altered video.¹⁹⁹ This kind of editing to mislead²⁰⁰ would be labeled a “shallowfake” or “cheap fake”²⁰¹ because it only manipulates one part of an existing video, rather than splicing together a variety of sources of images. Deepfakes in the second category would combine videos, images, or audio of the same individual to create a new video with the intention of impersonating the individual depicted. An example of this would be Jordan Peele's deepfake of Barack Obama referenced in the introduction of this Comment.²⁰² Deepfakes in the third category would combine videos, images, or audio of a variety of people, even though they appear to impersonate one individual. Deepfaked revenge or celebrity porn, such as the deepfakes grafting actress Scarlett Johansson's face onto different women's bodies in graphic sex scenes, would most likely use this method.²⁰³ The differences in sources, editing techniques, and verifiability of these different categories of deepfakes demonstrate the need for regulators to have a clear conception of the type of video they are examining.

The regulations would also need to address the different uses of deepfakes. Although deepfakes can cause harm to others, they also have many beneficial applications.²⁰⁴ Therefore, an element gauging intent would be useful in order to differentiate between beneficial and harmful deepfakes. A standard to determine “malicious intent” could include information such as whether the creator profited from the video, on what platforms and how often the creator posted the video, and more context-specific clues about why the video was created. This regulation would require a heavy emphasis on the facts of the situation in order to determine the motivation behind the creation of the video.

In addition to adding an intent element to regulation, it would similarly be beneficial to add a section explicitly allowing deepfakes to create spoofs, *200 parodies, or satires.²⁰⁵ Not only would the addition of this explicit section help avoid running afoul of the First Amendment, it also would ensure that regulations would not quell beneficial uses of deepfakes. Another important element for the regulations would be to add a clarification that the exception for spoofs is not intended to be a pretext to allow other types of harm that may result from more personal deepfakes, such as those used in revenge porn settings. This exception also would require closely examining the context and facts of the deepfake in order to determine the motivation behind its creation and the harms that may result.

Another potential regulation could include requiring a disclosure of some kind that a video is indeed a deepfake rather than a real video. While this regulation may not eliminate some of the more personal harms that deepfakes can cause, like emotional distress, it would help mitigate the potential of deepfakes to disrupt a viewer's ability to distinguish between an authentic and

doctored video. One method of achieving this could be a digital signature, either through the use of a watermark to indicate that a video has been faked or through the availability of metadata. Additionally, if technology advances to the point to allow an agency to detect that a video is, indeed, a deepfake, this regulation could parallel the Endorsement Guides the FTC currently utilizes to monitor disclosures about marketing on social media.²⁰⁶ The Endorsement Guides rely on voluntary compliance but reserve the right for the FTC to take corrective action if certain groups of people do not follow the designated practices and the practices used are deemed unlawful.²⁰⁷ The Endorsement Guides require full, clear, and conspicuous disclosures of connections between endorsers and sellers when that connection would otherwise affect the credibility of an endorsement.²⁰⁸ This model could be utilized for deepfakes as well: when a video is created from previously existing images and videos, the FTC could require full, clear, and conspicuous disclosure of how the video was made in order to circumvent the harms that could otherwise result.²⁰⁹

Overall, regulations should not be overly restrictive because of “hypothetical worst-case scenarios, or else best-case scenarios will never come about.”²¹⁰ There are many potentially useful applications of deepfakes, *201 and while it is important to mitigate the harms that deepfakes can create, regulators should keep those beneficial uses in mind while drafting the regulations.²¹¹ Putting regulations in place that clearly delineate what forms of deepfakes can cause harm, while allowing for certain uses, ensures that there will not be a ban on deepfakes as a whole. Rather, by having clarity on which uses are permissible, creators of deepfakes can experiment and innovate in legal, beneficial ways, while understanding that misusing deepfakes comes with legal ramifications.

VI. CHALLENGES

Although this Comment has demonstrated the harms that deepfakes pose, explained the need to regulate them, and delineated a regulatory scheme, there are still many complexities that deepfakes pose that have not been fully addressed. This section acknowledges counterarguments to the idea of FTC regulation of deepfakes and addresses some central concerns.

A. Does This Technology Need to Be Regulated?

This Comment's unstated premise is that deepfake technology needs to be regulated; however, that premise should not go unchallenged. Reddit user “deepfakes,” the man who started this phenomenon, points out that “every technology can be used with bad motivations, and [it is] impossible to stop that.”²¹² Although malicious uses of deepfakes are inevitable, that inevitability should not preclude attempts to prevent or mitigate the potential harms that the technology is likely to cause. “Deepfakes” further points out that it is not necessarily a problem for “more average people [to] engage in machine learning research.”²¹³ While true, it may be a problem if the average person does not fully understand the consequences and ramifications that can arise by wielding powerful technology. Deepfakes may not be authentic videos, but the harms they can produce are real. The average person making a deepfake for their own personal gain or gratification may not foresee the harms to others that their videos can produce. Therefore, having a clear regulatory scheme in place can allow the average person to be aware of what types of deepfakes are and are not permissible. With this knowledge, creators can experiment with machine learning research via creating deepfakes without fear of running afoul of the law.

Additionally, “there is nothing inherently illegal about the technology” used to create deepfakes.²¹⁴ After all, “deepfakes [do not] hurt people, *202 people using deepfakes hurt people.”²¹⁵ But even if the technology itself is neutral, its potential to do damage indicates a need for some form of proscriptive action to be taken. The regulations this Comment proposes do not advocate for limiting the technology itself, apart from potentially requiring a digital signature on deepfakes. This Comment instead advocates for action when a deepfake results in harm to others.

Furthermore, any discussion about regulation regarding the Internet must acknowledge the tension inherent in the idea. When the Internet was first developing, early Internet users believed “not just that the government would not regulate cyberspace--[but] that government *could not* regulate cyberspace.”²¹⁶ A manifesto from 1996 explicitly rejected the idea of external governance of the Internet, declaring:

Governments derive their just powers from the consent of the governed. You have neither solicited nor received ours. We did not invite you. You do not know us, nor do you know our world. Cyberspace does not lie within your borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions.²¹⁷

Lawrence Lessig's conception that “code is law” underscores this idea.²¹⁸ But Lessig argues that code serves as a form of regulation within cyberspace.²¹⁹ In his view, “[w]e can build, or architect, or *code* cyberspace to protect values that we believe are fundamental. Or we can build, or architect, or code cyberspace to allow those values to disappear.”²²⁰ In this way, the regulations this Comment proposes utilize this concept of “code is law.” By requiring the disclosure of alterations to a video, these FTC regulations would require the use of code to protect the fundamental values of understanding reality and authenticity.²²¹

203 B. *Would Regulation of Deepfakes Run Afoul of the First Amendment?

Although there are exceptions, “all content is presumptively protected by the First Amendment.”²²² Even if a deepfake is causing harm, having the FTC attempt to remove the content could easily violate the First Amendment unless the content falls within one of the exceptions to free speech.²²³ The regulations this Comment proposes do not require removing all deepfakes, but rather require transparency regarding the fact that videos are doctored; thus, these recommendations would not conflict with the First Amendment. Furthermore, spoofs, caricatures, parodies, and satires are all typically protected under the First Amendment; therefore, if a user makes it clear that the posted video is doctored, the content may likely fall under First Amendment protection.²²⁴ Accordingly, this Comment proposes regulations that take the poster's intention into account. If someone who creates a deepfake does not have malicious intent or intent to cause harm, that person would likely not fall under the restrictions of the regulations. If a deepfake is a true spoof or parody, even if it may be harsh or mocking, it is unlikely that the proposed regulations would treat it as malicious.

C. *How Effective Would Regulations Be?*

Any regulation's effectiveness may be limited because the “[t]echnologies that can be used to enhance and distort what is real are evolving faster than our ability to understand and control or mitigate it.”²²⁵ It may be too late for any regulations to make an effective difference due to the increasing sophistication of the technology behind deepfakes. While it is likely too late to control the actual technology behind deepfakes, it is not too late to regulate the videos actually produced. It is also currently unclear how to address already-posted deepfakes that would run afoul of the proposed regulations. America does not have “right to be forgotten” laws regarding information posted on the Internet, so there may be additional difficulties in removing an already-existing deepfake.²²⁶ Another limitation on the potential effectiveness of the proposed regulation is technology's current inability to reliably decipher what is and is not a deepfake. If the video is convincing enough and there is no true way for a victim to establish that a video is indeed forged, then the regulations may not be able to provide *204 a remedy. If the FTC had the technology to detect which videos would fall outside of their guidelines, this would amplify the effectiveness of the proposed regulations.

D. Executive Order

Another specific challenge for implementing regulations is President Trump's executive order limiting the creation of new regulations.²²⁷ This executive order may lower the probability that any proposed regulations would go into effect due to the difficulty of eliminating so many others.²²⁸ Regardless, this executive order does not diminish the importance of regulating deepfakes. While it would be more difficult to fully implement these regulations, that difficulty should not preclude putting these regulations in place at all.

E. Would All Deepfakes Require Regulation?

Although deepfakes can cause a wide variety of harms, it is important to remember that they also have many potentially beneficial applications. Some positive uses could include therapeutic applications, education, and art.²²⁹ One powerful example of a beneficial use of deepfakes would be allowing patients who would otherwise worry about stigma to receive treatment for mental health via video conference with a therapist.²³⁰ Another potential use would be creating deepfakes of famous historical figures to make an educational video more exciting and engaging for children.²³¹ With all new technology, there will always be the potential for bad actors. While it is important to be aware of the harm technology can perpetuate, it is equally important to realize the potential for innovation.

VII. CONCLUSION

For the foregoing reasons, deepfakes present the possibility of serious harms to individuals, companies, governments, and society overall. Although some efforts are underway to attempt to address this issue, if a more unified response does not come together soon, the technology may advance beyond limitation. While there are still serious questions of law and policy to address regarding this issue, the implementation of regulations by the FTC would be a way to start the process and mitigate potential harms. *205 By issuing clear guidelines, the FTC can help prevent the harmful uses of deepfakes without stymieing their beneficial uses. In a society filled with fake news and alternative facts, it is more difficult than ever to know the truth. If allowed to proceed unchecked, deepfakes will only exacerbate this issue in our society. There are currently two paths deepfakes may take: they may--like “any other machine[--be] ... either a benefit or a hazard.”²³² To fully enjoy the benefits deepfakes can provide, we must first take action via regulation to mitigate their hazards.

Footnotes

¹ GEORGE ORWELL, 1984 81 (Signet Classics 1977) (1949).

^{a1} J.D. Candidate, 2020, Seton Hall University School of Law; B.A. in English Language and Literature, B.S. in Business Marketing, University of Maryland, College Park, May 2017. I would like to thank my faculty advisor, Professor Najarian Peters, for her invaluable insight and guidance throughout the development of this Comment. I also would like to thank Professor Michael Coenen for his advice on issues regarding administrative law and Professor David Operderbeck for his assistance regarding the technology behind deepfakes. I would also like to thank my family for their endless love and support in all of my endeavors.

² BuzzFeedVideo, *You Won't Believe What Obama Says in This Video!*, YOUTUBE (Apr. 17, 2018), <https://www.youtube.com/watch?v=cQ54GDm1eL0> [hereinafter BuzzFeed Video].

³ *Id.*

⁴ *Id.*

⁵ *Id.*

6 *Id.*

7 *Id.*

8 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-peepe-psa-video-buzzfeed>.

9 “Deep learning” refers to a branch of artificial intelligence where software learns how to recognize patterns out of data. The software learns “in a very real sense” by mimicking how the brain utilizes neurons to think. Robert D. Hof, *Deep Learning*, MIT TECH. REV. (Apr. 23, 2013), <https://www.technologyreview.com/s/513696/deep-learning>.

10 John Brandon, *Terrifying High-Tech Porn: Creepy ‘Deepfake’ Videos Are on the Rise*, FOX NEWS (Feb. 16, 2018), <http://www.foxnews.com/tech/2018/02/16/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-rise.html>; *see also* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. (forthcoming 2019) (manuscript at 4) (on file with author).

11 *See* Chesney & Citron, *supra* note 10, at 4-5.

12 *See id.*

13 *See* Samantha Cole, *AI-Assisted Fake Porn Is Here and We’re All Fucked*, MOTHERBOARD (Dec. 11, 2017, 2:18 PM), https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn; *see also* Chesney & Citron, *supra* note 10, at 16-29 (listing manipulation of elections, jeopardizing national security, and undermining journalism as some of the potential harmful applications of deepfake technology).

14 Hilke Schellmann, *Deepfake Videos Are Getting Real and That’s a Problem*, WALL ST. J. (Oct. 15, 2018, 5:29 AM), <https://www.wsj.com/articles/deepfake-videos-are-ruining-lives-is-democracy-next-1539595787>.

15 *See infra* Part II.D.

16 *See* Chesney & Citron, *supra* note 10, at 2; Gianluca Mezzofiore, *No, Emma Gonzalez Did Not Tear Up a Photo of the Constitution*, CNN (Mar. 26, 2018, 3:30 PM), <https://www.cnn.com/2018/03/26/us/emma-gonzalez-photo-doctored-trnd/index.html>.

17 Chesney & Citron, *supra* note 10, at 2.

18 Chesney & Citron, *supra* note 10, at 14 (describing how the variety of purposes for deepfakes “can inflict a remarkable array of harms”).

19 Deepfakes’ origins in pornography will likely have long-lasting implications for women, especially in terms of revenge porn. *See* Rebecca Ruiz, *Deepfakes Are About to Make Revenge Porn So Much Worse*, MASHABLE (June 24, 2018), https://mashable.com/article/deepfakes-revenge-porn-domestic-violence/#IA8ClkF_tQqF. While this Comment touches on these issues, there is still much room for further exploration of how deepfakes configure into existing revenge porn laws.

20 “The 2016 election season saw the viral distribution of numerous factually inaccurate claims regarding political figures or events,” leading to concerns that this intentional spread of misinformation skewed the electoral results. Lili Levi, *Real “Fake News” and Fake “Fake News”*, 16 FIRST AMEND. L. REV. 232, 233 n.3 (2017).

21 BLADE RUNNER (The Ladd Company 1982).

22 *Id.*

23 *Id.*

24 *See* Michael Schulman, *The Battle for Blade Runner*, VANITY FAIR (Sept. 14, 2017, 8:00 AM), <https://www.vanityfair.com/hollywood/2017/09/the-battle-for-blade-runner-harrison-ford-ridley-scott>.

25 *See id.*


- 26 One fellow of the New America think tank has jokingly created a “‘Blade Runner’ Rule,” wherein the public has a “right to know whether you are interacting ... with a robot or not, or with something that is fake or not.” Olivia Beavers, *Washington Fears New Threat from ‘Deepfake’ Videos*, HILL (Jan. 20, 2019, 10:30 AM), <https://thehill.com/policy/national-security/426148-washington-fears-new-threat-from-deepfake-videos/>.
- 27 *See infra* Part III.
- 28 While the original definition was “thinking machines,” today artificial intelligence definitions “focus on ... how machines can imitate human intelligence.” Bernard Marr, *The Key Definitions of Artificial Intelligence (AI) that Explain Its Importance*, FORBES (Feb. 14, 2018, 1:27 AM), <https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#3b4fe44a4f5d>.
- 29 Like deep learning, machine learning is “a specific subset of AI that trains a machine how to learn.” *Machine Learning: What It Is and Why It Matters*, SAS, https://www.sas.com/en_us/insights/analytics/machine-learning.html (last visited Nov. 1, 2018).
- 30 GANs “are deep neural net architectures comprised of two nets, pitting one against the other.” *A Beginner’s Guide to Generative Adversarial Networks (GANs)*, SKYMIND, <https://skymind.ai/wiki/generative-adversarial-network-gan> (last visited Nov. 1, 2018). GANs operate by “[learning] to mimic any distribution of data.” *Id.*
- 31 Chesney & Citron, *supra* note 10, at 4-6; *Fake News: You Ain’t Seen Nothing Yet*, ECONOMIST (July 1, 2017), <https://www.economist.com/science-and-technology/2017/07/01/fake-news-you-aint-seen-nothing-yet>; John Donavan, *Deepfake Videos Are Getting Scary Good*, HOWSTUFFWORKS (Sept. 5, 2018), <https://electronics.howstuffworks.com/future-tech/deepfake-videos-scary-good.htm>.
- 32 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/#pNi2cZMBtqqM>.
- 33 *See* Chesney & Citron, *supra* note 10, at 4-5.
- 34 Cole, *supra* note 13.
- 35 *Fake News: You Ain’t Seen Nothing Yet*, *supra* note 31.
- 36 *Id.*
- 37 *Id.*
- 38 *Id.*
- 39 *Id.*
- 40 *Id.*
- 41 *See Fake News: You Ain’t Seen Nothing Yet*, *supra* note 31.
- 42 Cole, *supra* note 13.
- 43 *Id.*
- 44 One reason for deepfakes’ rapid ascent is that the technology to create them is easily accessible online. *See* Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, MOTHERBOARD (Jan. 24, 2018, 1:13 PM), https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley. Instead of requiring the expensive equipment necessary for movie studios to create similar videos, any user with an understanding of the code required can inexpensively create a realistic fake video. *Id.* (comparing CGI footage of Carrie Fisher from *Rogue One* on a budget of \$200 million with a deepfake of the same scene created by Reddit user “deepfakes” for free).


- 45 See Megan Garber, *Oprah's Head, Ann-Margaret's Body: A Brief History of Pre-Photoshop Fakery*, ATLANTIC (June 11, 2012), <https://www.theatlantic.com/technology/archive/2012/06/oprahs-head-ann-margarets-body-a-brief-history-of-pre-photoshop-fakery/258369/>.
- 46 *Photo Tampering Throughout History*, GA. TECH. C. COMPUTING, <https://www.cc.gatech.edu/~beki/cs4001/history.pdf> (last visited Feb. 15, 2019).
- 47 See *id.*
- 48 Megan Garber, *When Cameras Took Pictures of Ghosts*, ATLANTIC (Oct. 30, 2013), <https://www.theatlantic.com/technology/archive/2013/10/when-cameras-took-pictures-of-ghosts/281010/>.
- 49 *Photo Tampering Throughout History*, *supra* note 46.
- 50 See Garber, *supra* note 45.
- 51 *Id.*
- 52 See *id.*
- 53 Bill Roberts, *The Evolution of Film Editing*, ADOBE BLOG (Feb. 20, 2015), <https://theblog.adobe.com/the-evolution-of-film-editing/>.
- 54 Roberts, *supra* note 53.
- 55 HuffPost Australia, *How CGI Changed Movies Forever*, HUFFINGTON POST (May 13, 2016, 12:00 PM), https://www.huffingtonpost.com.au/2016/05/12/how-cgi-changed-movies-forever_a_21358758/.
- 56 *Id.*
- 57 CGI works by using a multi-step process to animate all of the frames of a scene requiring CGI, and then using “high-powered graphics computers” to render those images into what looks like a “fluid camera shot[.]” Kyle Neubeck, *This Is How CGI Actually Works*, COMPLEX (May 29, 2015), <https://www.complex.com/pop-culture/2015/05/this-is-how-cgi-actually-works/>.
- 58 HuffPost Australia, *supra* note 55.
- 59 *Id.*
- 60 Kevin Goering et al., *New York Right of Publicity Law: Reimagining Privacy and the First Amendment in the Digital Age*, 36 CARDOZO ARTS & ENT L.J. 601, 603 (2018). In fact, one of the main creators of deepfake content compared the ethics behind deepfake videos with the ethics of the digital recreation of the late Paul Walker in the film *Furious 7*. Cole, *supra* note 14.
- 61 See Chesney & Citron, *supra* note 10, at 4-6.
- 62 See HuffPost Australia, *supra* note 55.
- 63 See Cole, *supra* note 13.
- 64 ELFYOURSELF, <https://www.elfyourself.com/> (last visited Sept. 14, 2018).
- 65 *Id.*
- 66 See Baracksdubs, <https://www.youtube.com/user/baracksdubs> (last visited Sept. 14, 2018).
- 67 Baracksdubs, *Barack Obama Singing Call Me Maybe by Carly Rae Jepsen*, YOUTUBE (June 4, 2012), <https://www.youtube.com/watch?v=hX1YVzdnpe>.
- 68 See *id.*

- 69 See Chesney & Citron, *supra* note 10, at 2 (describing the deepfake of Emma Gonzalez tearing up the Constitution); see also Ally Foster, *Teen's Google Search Reveals Sickening Online Secret About Herself*, NEWS.COM.AU (June 30, 2018), <https://www.news.com.au/technology/online/security/teens-google-search-reveals-sickening-online-secret-about-herself/news-story/ee9d26010989c4b9a5c6333013ebbf2> (describing Noelle Martin's experience with deepfaked revenge porn).
- 70 Aja Romano, *Why Reddit's Face-Swapping Celebrity Porn Craze is a Harbinger of Dystopia*, VOX (Feb. 7, 2018, 5:55 PM), <https://www.vox.com/2018/1/31/16932264/reddit-celebrity-porn-face-swapping-dystopia>.
- 71 There was one precursor to deepfakes from 2016 on the thread, a video that “spliced an interview with Emma Watson over footage of an adult film actress removing her top.” Romano, *supra* note 70.
- 72 *Id.*
- 73 *Id.*
- 74 *Id.*
- 75 *Id.*
- 76 Matt Binder, *The U.S. Defense Department is Ready for the Battle Against Deepfakes*, MASHABLE (Aug. 7, 2018), <https://mashable.com/article/defense-department-fighting-deepfakes/#W6PJhu3Q0aqE>.
- 77 Cole, *supra* note 44.
- 78 *Id.*
- 79 *Id.*
- 80 *Id.*
- 81 See *infra* Part VI.E.
- 82 Romano, *supra* note 70.
- 83 Clare McGlynn & Erika Rackley, *Image-Based Sexual Abuse*, 37 OXFORD J. LEGAL STUD. 534, 535 (2017).
- 84 Caroline Drinnon, *When Fame Takes Away the Right to Privacy in One's Body: Revenge Porn and Tort Remedies for Public Figures*, 25 WM. & MARY J. WOMEN & L. 209, 211 (2017).
- 85 McGlynn, *supra* note 83.
- 86 Amanda L. Cecil, *Taking Back the Internet: Imposing Civil Liability on Interactive Computer Services in an Attempt to Provide an Adequate Remedy to Victims of Nonconsensual Pornography*, 71 WASH. & LEE L. REV. 2513, 2520 (2014).
- 87 See McGlynn & Rackley, *supra* note 83.
- 88 *Id.*
- 89 Chesney & Citron, *supra* note 10, at 17; see also Cecil, *supra* note 86, at 2524 (stating that revenge porn “disproportionately upsets the lives of heterosexual young women”).
- 90 Chesney & Citron, *supra* note 10, at 17. One Reddit user expressed a desire to create a deepfake porn video with his ex-girlfriend, while a Discord user claimed to have already created a deepfake using Facebook photos from a girl he attended high school with. *Id.*
- 91 See Foster, *supra* note 69.
- 92 *Id.*
- 93 *Id.*

- 94 See Jeff John Roberts, *Fake Porn Videos Are Terrorizing Women. Do We Need a Law to Stop Them?*, FORTUNE (Jan. 15, 2019), <http://fortune.com/2019/01/15/deepfakes-law/> (describing survey of 500 victims of revenge porn, wherein 12% had been victims of deepfakes).
- 95 Chesney & Citron, *supra* note 10, at 18.
- 96 See Levi, *supra* note 20, at 233 (“‘Fake news’ has become the central inflammatory charge in media discourse in the United States since the 2016 presidential contest.”).
- 97 BuzzFeed Video, *supra* note 2.
- 98 Chesney & Citron, *supra* note 10, at 27-28.
- 99 Beavers, *supra* note 26.
- 100 Franklin Foer, *The Era of Fake Video Begins*, ATLANTIC (May 2018), <https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/>.
- 101 *Id.* “[It is] natural to trust one’s own senses, to believe what one sees—a hardwired tendency that the coming age of manipulated video will exploit.” *Id.*
- 102 Chesney & Citron, *supra* note 10, at 28.
- 103 *Id.*
- 104 Chesney & Citron, *supra* note 10, at 17-19.
- 105 Foer, *supra* note 100 (“Unedited video has acquired an outsize authority in our culture.”).
- 106 Chesney & Citron, *supra* note 10, at 28-29.
- 107 This has serious implications for a society that becomes more and more dependent on video surveillance. See Milton Heumann et al., *Privacy and Surveillance: Public Attitudes on Cameras on the Street, in the Home, and in the Workplace*, 14 RUTGERS J.L. & PUB. POL’Y 37, 42 (2016) (describing young people’s comfort with video surveillance).
- 108 Chesney & Citron, *supra* note 10, at 28-29.
- 109 See Foer, *supra* note 100.
- 110 See Donie O’Sullivan, *When Seeing Is No Longer Believing*, CNN BUS., <https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/> (last accessed Feb. 15, 2019).
- 111 See O’Sullivan, *supra* note 110.
- 112 Sara Ashley O’Brien, *Deepfakes Are Coming. Is Big Tech Ready?*, CNN BUS. (Aug. 8, 2018, 11:16 AM), <https://money.cnn.com/2018/08/08/technology/deepfakes-countermeasures-facebook-twitter-youtube/index.html>.
- 113 O’Brien, *supra* note 112.
- 114 *Id.*
- 115 Yuezun Li, Ming-Ching Chang & Siwei Lyu, *In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking*, CORNELL U. LIBR., <https://arxiv.org/pdf/1806.02877.pdf> (last visited Feb. 15, 2019).
- 116 *Id.*
- 117 *Id.*
- 118 *Id.*

- 119 See O'Brien, *supra* note 112 (stating that after releasing results of study surrounding blinking, deepfake developers began working on improving videos to avoid detection).
- 120 See Beres & Gilmer, *supra* note 32.
- 121 *Id.*
- 122 Adi Robertson, *Reddit Bans 'Deepfakes' AI Porn Communities*, VERGE (Feb. 7, 2018), <https://www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban>.
- 123 Megan Farokhmanesh, *Deepfakes Are Disappearing from Parts of the Web, But They're Not Going Away*, VERGE (Feb. 9, 2018, 9:00 AM), <https://www.theverge.com/2018/2/9/16986602/deepfakes-banned-reddit-ai-faceswap-porn>.
- 124 Binder, *supra* note 76.
- 125 *Id.*
- 126 Will Knight, *The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery*, MIT TECH. REV. (May 23, 2018), <https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/>.
- 127 See Beavers, *supra* note 26.
- 128 *Id.*
- 129 *Id.*
- 130 James Vincent, *US Lawmakers Say AI Deepfakes 'Have the Potential to Disrupt Every Facet of Our Society'*, VERGE (Sept. 14, 2018, 1:17 PM), <https://www.theverge.com/2018/9/14/17859188/ai-deepfakes-national-security-threat-lawmakers-letter-intelligence-community>.
- 131 Letter from Adam B. Schiff, Member of Congress, Stephanie Murphy, Member of Congress, Carlos Curbelo, Member of Congress, to The Honorable Daniel R. Coats, Director of National Intelligence (Sept. 13, 2018), <https://schiff.house.gov/imo/media/doc/2018-09%20ODNI%C20Deep%C20Fakes%20letter.pdf>.
- 132 *Id.*
- 133 *Id.*
- 134 See Vincent, *supra* note 130.
- 135 See Emma Grey Ellis, *People Can Put Your Face on Porn--And the Law Can't Help You*, WIRED (Jan. 26, 2018, 7:00 AM), <https://www.wired.com/story/face-swap-porn-legal-limbo/>.
- 136 Beres & Gilmer, *supra* note 32 (describing University of Chicago Law School professor Jonathan Masur's belief that existing law will not “cover the vast majority of situations”).
- 137 See Chesney & Citron, *supra* note 10, at 34.
- 138 See *id.*
- 139 Megan Farokhmanesh, *Is It Legal to Swap Someone's Face into Porn Without Consent?*, VERGE (Jan. 30, 2018, 2:39 PM), <https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal>.
- 140 *Id.*
- 141 *Id.*
- 142 Beres & Gilmer, *supra* note 32.

- 143 Ellis, *supra* note 135.
- 144 See Roberts, *supra* note 94; David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELECTRONIC FRONTIER FOUND. (Feb. 13, 2018), <https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them>.
- 145 Tara E. Langvardt, *Reinforcing the Commercial-Noncommercial Distinction: A Framework for Accommodating First Amendment Interests in the Right of Publicity*, 13 VA. SPORTS & ENT. L.J. 167, 172 (2014).
- 146 See Roberts, *supra* note 94.
- 147 Jesse Lempel, *Combating Deepfakes Through the Right of Publicity*, LAWFARE (Mar. 30, 2018, 8:00 AM), <https://www.lawfareblog.com/combating-deep-fakes-through-right-publicity>.
- 148 Greene, *supra* note 144.
- 149 See Farokhmanesh, *supra* note 139.
- 150 Beres & Gilmer, *supra* note 32.
- 151 Farokhmanesh, *supra* note 139.
- 152 See Greene, *supra* note 144.
- 153  [Authors Guild v. Google, Inc., 804 F.3d 202, 214 \(2d Cir. 2015\)](#).
- 154 See Chesney & Citron, *supra* note 10, at 4-5.
- 155 Beres & Gilmer, *supra* note 32.
- 156 See Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 7 DUKE L. & TECH. REV. 99, 120-23 (2019) (providing a detailed example of how current nonconsensual pornography statutes may be insufficient to provide redress for deepfaked pornography).
- 157 Ellis, *supra* note 135.
- 158 *Id.*
- 159 *Id.*
- 160 Harris, *supra* note 156, at 123.
- 161 A full analysis of different methods of controlling deepfakes, such as judicial decisions or legislation, goes beyond the scope of this Comment.
- 162 *About the FTC*, FED. TRADE COMM'N, <https://www.ftc.gov/about-ftc> (last visited Aug. 18, 2019).
- 163 Chesney & Citron, *supra* note 10, at 45.
- 164 *What We Do*, FED. TRADE COMM'N, <https://www.ftc.gov/about-ftc/what-we-do> (last visited Aug. 18, 2019).
- 165 *Id.*
- 166 The FTC is “the country's de facto privacy regulator.” Nicholas Confessore and Cecilia Kang, *Facebook Data Scandals Stoke Criticism That a Privacy Watchdog Too Rarely Bites*, N.Y. TIMES (Dec. 30, 2018), <https://www.nytimes.com/2018/12/30/technology/facebook-data-privacy-ftc.html>.
- 167 *What We Do*, *supra* note 164.

- 168 See Ellis, *supra* note 135.
- 169 *Id.*
- 170 Further strengthening this argument is the doctrine of Chevron deference, wherein “courts accept agency interpretations regardless of whether there are other plausible interpretations.” Gerard M. Stegmaier & Wendell Bartnick, *Psychics, Russian Roulette, and Data Security: The FTC's Hidden Data-Security Requirements*, 20 GEO. MASON L. REV. 673, 679 (2013).
- 171 See John Allen Riggins, *Law Student Unleashes Bombshell Allegation You Won't Believe!: “Fake News” as Commercial Speech*, 52 WAKE FOREST L. REV. 1313, 1334-35 (delineating three-factor definition for “fake news” to determine when fake news falls within FTC jurisdiction).
- 172 Riggins, *supra* note 171, at 1325 (quoting  15 U.S.C. §45(a)(1) (2018)); see Callum Borchers, *How the Federal Trade Commission Could (Maybe) Crack Down on Fake News*, WASH. POST (Jan. 30, 2017), https://www.washingtonpost.com/news/the-fix/wp/2017/01/30/how-the-federal-trade-commission-could-maybe-crack-down-on-fake-news/?utm_term=.3ee33b8216d4.
- 173 Levi, *supra* note 20, at 302-03.
- 174 Borchers, *supra* note 172.
- 175 Chesney & Citron, *supra* note 10, at 46-47.
- 176 Borchers, *supra* note 172.
- 177 *What We Do*, FED. COMMC'N COMM'N, <https://www.fcc.gov/about-fcc/what-we-do> (last visited Aug. 31, 2019).
- 178 *See id.*
- 179 *Broadcasting False Information*, FED. COMMC'N COMM'N, <https://www.fcc.gov/consumers/guides/broadcasting-false-information> (last visited Aug. 31, 2019).
- 180 *See Restoring Internet Freedom*, FED. COMMC'N COMM'N, <https://www.fcc.gov/restoring-internet-freedom> (last visited Aug. 31, 2019).
- 181 *Id.*
- 182 *Id.*
- 183 *Id.*
- 184 Ryan Calo, *The Case for a Federal Robotics Commission*, CTR. FOR TECH. INNOVATION BROOKINGS (Sept. 2014).
- 185 *Id.*
- 186 The title of the novel that *Blade Runner* is based on, *Do Androids Dream of Electric Sheep?*, makes explicit the anxieties surrounding the possible sentience of artificially intelligent beings. Although that discussion is beyond the scope of this Comment, it is important to recognize that advances in artificial intelligence may very well merit the need for a separate regulatory agency in charge of monitoring the sophistication of artificial intelligence.
- 187 Mohana Ravindranath, *Who's in Charge of Regulating the Internet of Things?*, NEXTGOV (Sept. 1, 2016), <https://www.nextgov.com/emerging-tech/2016/09/internet-things-regulating-charge/131208/>.
- 188 *Id.*
- 189 Exec. Order No. 13,771, 82 Fed. Reg. 9339 (Feb. 3, 2017).
- 190 *Id.*; Ravindranath, *supra* note 187.

- 191 Hossein Derakhshan & Claire Wardle, *Ban the Term ‘Fake News’*, CNN (Nov. 27, 2017, 3:12 PM), <https://www.cnn.com/2017/11/26/opinions/fake-news-and-disinformation-opinion-wardle-derakhshan/index.html>.
- 192 Nick Statt, *Fake Celebrity Porn is Blowing Up on Reddit, Thanks to Artificial Intelligence*, VERGE (Jan. 24, 2018, 3:53 PM), <https://www.theverge.com/2018/1/24/16929148/fake-celebrity-porn-ai-deepfake-face-swapping-artificial-intelligence-reddit>.
- 193 *Id.*
- 194 *See* Chesney & Citron, *supra* note 10, at 3-4.
- 195 This idea borrows heavily from Hossein Derakhshan and Claire Wardle's argument for a new vocabulary regarding fake news. *See* Derakhshan & Wardle, *supra* note 191.
- 196 Dawn Stover, *The White House Shallowfake: Press Secretary Uses Manipulated Video in War Against Press*, BULL. ATOMIC SCIENTISTS (Nov. 12, 2018), <https://thebulletin.org/2018/11/the-white-house-shallowfake-press-secretary-uses-manipulated-video-in-war-against-press/>.
- 197 At a press conference on November 7, 2018, Jim Acosta and a White House intern had a brief interaction that resulted in physical contact when she tried to take away Acosta's microphone. Casey Newton, *The Fake Video Era of US Politics Has Arrived on Twitter*, VERGE (Nov. 9, 2018, 9:30 AM), <https://www.theverge.com/2018/11/9/18076418/acosta-cnn-fake-video-deepfakes-dystopia>.
- 198 *See* Didi Martinez, *Kellyanne Conway Says Jim Acosta Video Was ‘Sped Up,’ but Not ‘Doctored’*, NBC NEWS (Nov. 12, 2018, 12:22 PM), <https://www.nbcnews.com/news/all/kellyanne-conway-says-jim-acosta-video-was-sped-not-doctored-n935196>.
- 199 Drew Harwell, *White House Shares Doctored Video to Support Punishment of Journalist Jim Acosta*, WASH. POST (Nov. 8, 2018, 3:23PM), https://www.washingtonpost.com/technology/2018/11/08/white-house-shares-doctored-video-support-punishment-journalist-jim-acosta/?utm_term=.ce7f2eea5cc2.
- 200 To be clear, it is not my intention to suggest that the White House was the party that edited the video. Whether the video was intentionally altered or differences occurred in the conversion to a GIF from a video is still unclear. *See* Lauren Aratani, *Altered Video of CNN Reporter Heralds a Future Filled with ‘Deep Fakes’*, FORBES (Nov. 8, 2018, 8:12 PM), <https://www.forbes.com/sites/laurenaratani/2018/11/08/altered-video-of-cnn-reporter-jim-acosta-heralds-a-future-filled-with-deep-fakes/#5c80cb823f6c>.
- 201 *Id.*
- 202 BuzzFeed Video, *supra* note 2.
- 203 Drew Harwell, *Scarlett Johansson on Fake AI-Generated Sex Videos: ‘Nothing Can Stop Someone from Cutting and Pasting My Image’*, WASH. POST (Dec. 31, 2018, 4:14 PM), https://www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/?utm_term=.e44e22e8b1e5.
- 204 *See infra* Part VI.E.
- 205 Farokhmanesh, *supra* note 139.
- 206 *The FTC's Endorsement Guides: What People Are Asking*, FED. TRADE COMM'N, <https://www.ftc.gov/tips-advice/business-center/guidance/ftcs-endorsement-guides-what-people-are-asking>.
- 207 16 C.F.R. § 255(a) (2018).
- 208 *Id.* § 255.5.
- 209 Another potential regulatory model to follow could be the standards that require “disclosures on campaign ads detailing who funded the advertisement.” Beavers, *supra* note 26.
- 210 Ravindranath, *supra* note 187.
- 211 *Id.*

- 212 Cole, *supra* note 13.
- 213 *Id.* (alteration in original).
- 214 Greene, *supra* note 144.
- 215 James Vincent, *A Porn Company Promises to Insert Customers into Scenes Using Deepfakes*, VERGE (Aug. 21, 2018, 11:26 AM), <https://www.theverge.com/2018/8/21/17763278/deepfake-porn-custom-clips-naughty-america>.
- 216 LAWRENCE LESSIG, CODE 3 (2d ed. 2006).
- 217 John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), <https://www.eff.org/cyberspace-independence>.
- 218 LESSIG, *supra* note 216, at 5.
- 219 *Id.* at 5-6.
- 220 *Id.* at 6.
- 221 *See id.*
- 222 Farokhmanesh, *supra* note 139.
- 223 *Id.*
- 224 Ellis, *supra* note 135.
- 225 Charlie Werzel, *He Predicted the 2016 Fake News Crisis. Now He's Worried About an Information Apocalypse*, BUZZFEED NEWS (Feb. 11, 2018, 8:45 PM), <https://www.buzzfeednews.com/article/charliewerzel/the-terrifying-future-of-fake-news>.
- 226 Ellis, *supra* note 135.
- 227 Exec. Order No. 13,771, 82 Fed. Reg. 9339 (Feb. 3, 2017).
- 228 *See id.*
- 229 *See* Beres & Gilmer, *supra* note 32; *see also* Chesney & Citron, *supra* note 10, at 14-15.
- 230 Beres & Gilmer, *supra* note 32 (describing potential for soldiers suffering from posttraumatic stress disorder to use this technology for this purpose).
- 231 Chesney & Citron, *supra* note 10, at 14.
- 232 BLADE RUNNER (The Ladd Company 1982).

50 SHLR 177

End of Document

© 2019 Thomson Reuters. No claim to original U.S. Government Works.