# LYING BENEATH THE SURFACE: THE IMPACTS OF DEEPFAKE TECHNOLOGY ON THE PRIVACY AND SAFETY OF THE LGBTQ+ COMMUNITY

SERGIO E. MOLINA*



## I. INTRODUCTION

The year is 1938 and New Jersey residents have just turned on their radios to hear Herbert George ("H.G.") Wells broadcast fake news bulletins that warn of an alien invasion.[1] What the listeners do not seem to realize is that Wells is performing a radio adaptation of his science-fiction novel, *The War of the Worlds*.[2] What results is nationwide hysteria that causes a flurry of phone calls from anxious listeners to police stations, newspaper offices, and other radio stations with fears of an imminent Martian maraud—a predictable result of a population believing without seeing.[3] Fast forward to the 21st century, and now, even seeing is no longer believing; citizens can no longer

* Sergio E. Molina is a commercial litigator in Miami, Florida. Sergio obtained his Juris Doctor from Nova Southeastern University's Shepard Broad College of Law with the College's concentration in Intellectual Property, Technology, and Cybersecurity Law, and his bachelor's degree in finance with minors in economics and psychology from Florida International University.

1. A. Brad Schwartz, *The Infamous "War of the Worlds" Radio Broadcast Was a Magnificent Fluke*, SMITHSONIAN MAG. (May 6, 2015), http://www.smithsonianmag.com/history/infamous-war-worlds-radio-broadcast-was-magnificent-fluke-180955180/.

2. *Id.*

3. *Id.*

trust their own eyes or ears.[4] Claims such as these have moved out of the realm of fake radio bulletins, hyperboles, or even hypotheticals and into what is now our new, technologically-advanced reality.[5] To what do citizens of today's society owe this belief in absolute disbelief?[6] Enter "deepfakes," a term that combines the phrases "deep learning" and "fake," that refers to a wide variety of hyper-realistic images, videos, and audio recordings that are fabricated through the use of machine learning.[7]

Deepfakes are synthetic audiovisual ("AV") media with seemingly limitless applications—a type of media that can do everything from the recreation of voices to the swapping of faces from one person onto another.[8] Below is a compilation of images that depict the manner in which deepfake technology employs face-swapping methodology to create synthesized media of Donald Trump and Elizabeth Warren.[9]


*Deepfake Media of Donald Trump and Elizabeth Warren*[10]

4. *See* Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, CATH. U. J.L. & TECH., Fall 2018, at 51, 51.

5. See Nicholas Diakopoulos & Deborah Johnson, Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections, 23 NEW MEDIA & SOC'Y 2072, 2073 (2021).

6. See id.

7. *Id.*; Elizabeth Caldera, Comment, *"Reject the Evidence of Your Eyes and Ears": Deepfakes and the Law of Virtual Replicants*, 50 SETON HALL L. REV. 177, 178 (2019); BRITT PARIS & JOAN DONOVAN, DATA & SOC'Y, DEEPFAKES AND CHEAP FAKES: THE MANIPULATION OF AUDIO AND VISUAL EVIDENCE 2 (2019), http://datasociety.net/library/deepfakes-and-cheap-fakes/.

8. PARIS & DONOVAN, *supra* note 7, at 2; Diakopoulos & Johnson, *supra* note 5, at 2073.

9. *See* Will Knight, *Facebook, Google, Twitter Aren't Prepared for Presidential Deepfakes*, MIT TECH. REV. (Aug 6, 2019), http://www.technologyreview.com/2019/08/06/639/facebook-google-twitter-arent-prepared-for-presidential-deepfakes/. Visual aids are used throughout this Article to assist the reader in seeing the effectiveness of some of the deepfake media currently available to the public. Elizabeth G. Porter, *Taking Images Seriously*, 114 COLUM. L. REV. 1687, 1709 (2014) ("On the rare occasions where journals did include images, they were startlingly effective.").

10. Knight, *supra* note 9.

With deepfakes already spreading throughout various facets of society, the strongest embrace is most notable from both the arts and entertainment fields.[11] Given deepfakes' ability to superimpose the faces of actors onto the bodies of stunt doubles or even to simulate actors' scenes altogether, it is no surprise that Hollywood has taken notice of the vast opportunities that the new technology presents.[12] But, it does not stop there.[13] The realms of art and entertainment have seen the uses of deepfakes taken as far as to bring long-deceased actors or public figures "back to life."[14]

Although deepfakes present numerous benefits to the creative arts, they also introduce a concerning reality.[15] As the number of online deepfakes grows by the day, many have questioned the harmful implications of the technology and the effects that it may have when compounded by current social and political climates.[16] However, the potential for harm is not exclusively reserved for public figures, nor is it reduced only to simplified forms of AV manipulation.[17] As society continues to see the democratization of more advanced technologies, deepfakes have begun to present individuals with novel methods of "exploitation, intimidation, and sabotage."[18] The most concerning example of this has perhaps been the widespread use of deepfake technology to fabricate pornography with the images of both public figures and private individuals without their consent.[19] This is just the tip of the iceberg.[20] Data suggests that minority communities, particularly women, are

11. Diakopoulos & Johnson, *supra* note 5, at 2073.

12. *See id.* at 2074; Hall, *supra* note 4, at 57.

13. *See* Diakopoulos & Johnson, *supra* note 5, at 2073.

14. *Id.*

15. *See* Marcus Baram, *How Deepfakes Evolved So Rapidly in Just a Few Years*, FAST CO. (Oct. 8, 2019), http://www.fastcompany.com/90414479/how-deepfakes-evolved-so-rapidly-in-just-a-few-years; Diakopoulos & Johnson, *supra* note 5, at 2072; Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1754 (2019).

16. Diakopoulos & Johnson, *supra* note 5, at 2072; PARIS & DONOVAN, *supra* note 7, at 3.

17. Diakopoulos & Johnson, *supra* note 5, at 2080; PARIS & DONOVAN, *supra* note 7, at 5–6.

18. Chesney & Citron, *supra* note 15, at 1754.

19. U.S. GOV'T ACCOUNTABILITY OFF., GAO-20-379SP, SCIENCE & TECH SPOTLIGHT: DEEPFAKES 1 (2020), http://www.gao.gov/assets/gao-20-379sp.pdf; Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV., Nov. 2019, at 39, 43.

20. *See* Robert Size, *Publishing Fake News for Profit Should Be Prosecuted as Wire Fraud*, 60 SANTA CLARA L. REV. 29, 30–31 (2020).

more greatly affected by the harms that deepfake technologies present.[21] It is likely that minority communities with a greater stake in information and personal privacy, like the LGBTQ+ community, could stand to lose more in the wake of misused deepfake technology.[22]

While some believe that the discussions surrounding deepfakes' potential threats are overstated, it can hardly be denied that technology is better today than it was yesterday, and yet still not as good as it will be tomorrow.[23] This is a dynamic that necessitates a discussion on the law's evolution in order to effectively address technological advancements and the harms that they may impose.[24] In that regard, deepfakes do not provide an exception to this claim, but instead serve to reinforce its validity.[25]

This Article serves to address the current landscape of deepfake technology in modern culture and its impacts on marginalized communities, particularly the LGBTQ+ community, in four subsequent parts.[26] Part II offers a technical glimpse into the creation of deepfakes; how deepfakes came into being and why deepfakes circulate society with great frequency.[27] Part III looks at the threats that deepfake technology can pose when put in the hands of individuals seeking to harm or extort members of marginalized communities, such as the LGBTQ+ community, by providing a historical overview of similar forms of exploitation that the LGBTQ+ community has faced in the past.[28] Part IV explores the existing regulatory frameworks that serve to address the harms of deepfake technology along with the suggested evolutions and amendments of those frameworks.[29] This Article concludes by

---

21. Robert Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 886 (2019); Baram, *supra* note 15; U.S. GOV'T ACCOUNTABILITY OFF., *supra* note 19, at 1.

22. *See* Chesney & Citron, *supra* note 21, at 886; Baram, *supra* note 15; Sergio E. Molina, *DL and Looking? So Are the Data Miners, and They Already Know What You're Into*, OUTSIDE INFLUENCE, Fall/Winter 2019, at 4–5.

23. Russell Brandom, *Deepfake Propaganda Is Not a Real Problem*, VERGE (Mar. 5, 2019, 12:25 PM), http://www.theverge.com/2019/3/5/18251736/deepfake-propaganda-misinformation-troll-video-hoax; Hayley Duquette, Note, *Digital Fame: Amending the Right of Publicity to Combat Advances in Face-Swapping Technology*, 20 J. HIGH TECH. L. 82, 103 (2020).

24. Duquette, *supra* note 24 at 103; *see also* David Dorfman, *Decoding Deepfakes: How Do Lawyers Adapt When Seeing Isn't Always Believing?*, OR. ST. B. BULL., Apr. 2020, at 18, 20.

25. Duquette, *supra* note 24, at 103; Dorfman, *supra* note 24, at 20.

26. *See* discussion *infra* Parts I–III.

27. *See* discussion *infra* Part II.

28. *See* discussion *infra* Part III.

29. *See* discussion *infra* Part IV.

advocating for the adoption of an amended regulatory scheme via Section 230 of the Communications Decency Act[30] and for the re-appropriation of deepfake technology until such time that federal legislation better promotes and protects the online privacy and personal safety of members of the LGBTQ+ community.[31]

## II. DEEP DIVING INTO DEEPFAKES

Audiovisual manipulation is by no means a novel concept; however, the newest stage in its evolutionary journey incorporates an added layer of technological advancements that makes its existence not only more widespread, but also more intricate.[32] In order to develop a better sense of the threats that deepfakes pose and the manners in which deepfakes may be mitigated, it is important to understand exactly where deepfakes came from, how they are made, and why their availability is growing.[33]

### A. *Remember to Rewind*

Society's understanding of deepfakes has become popularized at a time when "fake news"—or, as some define it, false, inaccurate, or misleading information designed, presented, and promoted to further interests—is front and center.[34] It is important to note that fake news serves as an umbrella term under which misinformation and disinformation exist[35]—misinformation being the unintentional furtherance of misleading or inaccurate information and disinformation being its intentional equivalent.[36] While the root of these concepts are ancient, social media structures and the rise of deepfakes have helped these concepts branch out into a post-truth society where "objective

30. Communications Decency Act of 1996, Pub. L. No. 104–104, 110 Stat. 133 (codified as amended at 47 U.S.C. § 223 (Supp. II 1997)).

31. *See* discussion *infra* Part V.

32. *See* Chesney & Citron, *supra* note 21, at 884–85.

33. Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 342 (2019); Westerlund, *supra* note 19, at 40; Diakopoulos & Johnson, *supra* note 5, at 2074.

34. *See* Cristian Vaccari & Andrew Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, SOC. MEDIA & SOC'Y, Jan.–Mar. 2020, at 1, 2.

35. Fernando Nuñez, Note, *Disinformation Legislation and Freedom of Expression*, 10 U.C. IRVINE L. REV. 783, 785–86 (2020).

36. Kyle Anderson, Note, *Truth, Lies, and Likes: Why Human Nature Makes Online Misinformation a Serious Threat (and What We Can Do About It)*, 44 LAW & PSYCH. REV. 209, 211 (2019–2020).

facts are less influential in shaping public opinion than appeals to emotion and personal belief."[37]

Recent events have highlighted the manners in which all of these concepts intersect.[38] However, these concepts are the latest iteration of a longstanding practice.[39] For example, the earliest known surviving photograph was taken in the mid-to-late 1820s.[40] Since then, the practice of photo editing has been almost as long-standing as the history of the photograph itself.[41] Although photo editing became a developed practice long before the creation of the first computer, the emergence of photoshop in the 1980s allowed for the practice to popularize among both professionals and amateurs alike.[42] The twentieth century saw a similar pattern occur in film—like the development of the world's first editing machine in the 1920s—the development of the videotape recorder in the 1950s, and the introduction of non-linear editing with the help of modern computers.[43]

Advancements in the ability to manipulate all of these forms of media, in one way or another, could produce hundreds of takes and seamlessly string them together into one desired output that, as far as the consumer of the media knows, occurred in one attempt.[44] These edits of audiovisual footage without the use of machine learning are known as "cheapfakes," or "shallowfakes," the most common of which include photoshopped images, recontextualized media, and sped up or slowed down video.[45] As technology advanced and computers began running more intricate programs, the practice of physically splicing reels of film fell out of practice, and the adoption of more cutting-edge techniques like computer-generated imagery ("CGI") became the norm.[46]

Today, deepfake technology has brought society face-to-face with the latest version of the tried-and-true practices of its predecessors.[47] One of the

---

37. Hall, *supra* note 4, at 54.

38. *See* Diakopoulos & Johnson, *supra* note 5, at 2073.

39. Chesney & Citron, *supra* note 21, at 884–85.

40. Spivak, *supra* note 33, at 341.

41. *See* Michael Scott Henderson, Note, *Applying Tort Law to Fabricated Digital Content*, 2018 UTAH L. REV. 1145, 1147 (2018).

42. Spivak, *supra* note 33, at 341.

43. Henderson, *supra* note 41, at 1149.

44. *See* PARIS & DONOVAN, *supra* note 7, at 14–15 (explaining that consumer software and free mobile apps allow for this manipulation).

45. *Id.* at 5–6.

46. *See* Diakopoulos & Johnson, *supra* note 5, at 2074; Marie-Helen Maras & Alex Alexandrou, *Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos*, 23 INT'L J. EVIDENCE & PROOF 255, 256 (2019); David Song, *A Short History of Deepfakes*, MEDIUM (Sept. 23, 2019), http://www.medium.com/@songda/a-short-history-of-deepfakes-604ac7be6016.

47. *See* Diakopoulos & Johnson, *supra* note 5, at 2074.

more notable uses of deepfake technology and synthetic media is its use by the Dalí Museum to "resurrect," or rather, "reincarnate" Salvador Dalí for a more immersive, interactive guest experience that the museum's website describes as allowing "visitors an opportunity to learn more about Salvador Dalí's life from the person who knew him best: the artist himself."[48] But the technology is not limited only to living things.[49] For example, whereas the Dalí Museum in St. Petersburg, Florida, uses deepfakes to reproduce Salvador Dalí himself, Russian researchers have used similar software to animate intimate subjects that include the works of other great artists, such as Johannes Vermeer's *Girl with a Pearl Earring*, and Leonardo DaVinci's *Mona Lisa*, as depicted below.[50]

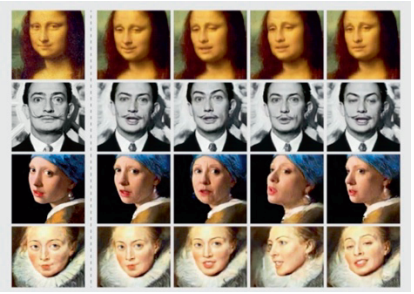

*Image of Works of Art Animated with Deepfake Technology*[51]

Much like copies of the works produced by some of art's great masters, deepfakes have been described as forgeries of photos, videos, and audios made with the assistance of artificial intelligence.[52] In many ways, referring to deepfakes as forgeries is a misnomer of sorts in that, at least in the colloquial sense, forgeries are almost exact copies of works already in existence.[53] Deepfakes, on the other hand, operate more as a hyper-realistic collage in that they synthesize a wide number of already existing works to

48. *Dalí Lives (Via Artificial Intelligence*), SALVADOR DALÍ MUSEUM, http://www.thedali.org/exhibit/dali-lives/ (last visited Apr. 12, 2022).

49. *See* Herbert B. Dixon Jr., *Deepfakes: More Frightening Than Photoshop on Steroids*, JUDGES' J., Summer 2019, at 35, 36.

50. *Id.*

51. Gregory Barber, *Deepfakes Are Getting Better, but They're Still Easy to Spot*, WIRED (May 26, 2019, 7:00 AM), http://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/.

52. U.S. Gov't Accountability Off., *supra* note 19, at 1.

53. *Forgery*, MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/forgery (last visited Apr. 12, 2022).

create an entirely new product that looks seamless and real.[54] Deepfakes fall under the larger umbrella of audiovisual manipulation, which is generally identified as the means for influencing the interpretation of media.[55] Audiovisual manipulation splits into two branches: either deepfakes, which incorporate artificial intelligence, or *cheapfakes*, which, as mentioned before, employ less technologically-advanced techniques.[56] Cheapfakes require that individuals upload media onto a computer and manually make adjustments—a process that, although still yielding a realistic product, can be incredibly labor-intensive and time-consuming, given its less technical nature.[57] However, the introduction of artificial intelligence, described in more detail below, provides a solution that cuts down on the time, as well as the amount of manual work needed to create a convincing product.[58]

It is important to note that "artificial intelligence" is often synonymized with "machine learning," however, the two terms are distinct.[59] Artificial intelligence is modeled after the human brain and reacts to incoming data, rather than relying on programmed rules, in order to operate rationally and intelligently.[60] To do this, artificial intelligence incorporates both algorithms—instructions or sets of instructions—and machine learning.[61] Machine learning is a branch of artificial intelligence that resembles the human trial and error process by allowing computer systems to learn directly from observing examples, data, and experiences.[62]

Deepfakes are created with a similar process that incorporates "deep learning," a deep neural network that takes in a multitude of data from an input layer and autonomously runs it through various nodes until it produces an output layer.[63] Oftentimes, this is done either with an autoencoder, which is an artificial neural network trained to reconstruct inputs from a simpler representation, or with a Generative Adversarial Network ("GAN").[64] GANs

---

54. KELLEY M. SAYLER & LAURIE A. HARRIS, CONG. RSCH. SERV., IF11333, DEEP FAKES AND NATIONAL SECURITY (2021), http://crsreports.congress.gov/product/pdf/IF/IF11333.

55. PARIS & DONOVAN, *supra* note 7, at 5–6.

56. *Id.*

57. Jessica Ice, Note, *Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RSRV. L. REV. 417, 420 (2019).

58. *Id.* at 421.

59. Herbert B. Dixon Jr., What Judges and Lawyers Should Understand About Artificial Intelligence Technology, JUDGES J., Winter 2020, at 36, 36 (2020).

60. Maras & Alexandrou, *supra* note 46, at 256.

61. Dixon, *supra* note 59, at 36.

62. Maras & Alexandrou, *supra* note 46, at 256.

63. Ice, *supra* note 57, at 421.

64. *Id.* at 421–22.

use two programs: on one end, there is a generator, and on the other end, a discriminator.[65] The creator first obtains a wide variety of images or sounds of a single individual, the "data set," and inputs the data set into the generator.[66] The generator then processes the data and uses it to synthesize a new image that is the conglomeration of all distinguishable features of the subject.[67] After creating a new image or sound, the generator passes the media along to the discriminator, which assesses the image for realism and determines whether it is sufficiently convincing to pass off as real.[68] If the discriminator believes that the discriminator's synthesized media is sufficiently convincing, it produces the deepfake; however, if the discriminator believes that the synthesized footage is not sufficiently convincing, it sends it back to the generator to try again.[69] The two programs go back and forth, competing with each other until the two successfully chisel out a realistic product.[70]

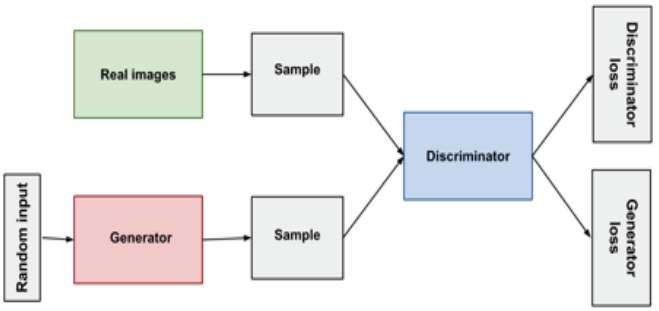


*Image of GAN Process*[71]

This process can be utilized to create either a visually-synthesized product or a vocally-synthesized product.[72] Video synthesis techniques are split into two forms: face-swapping or facial re-enactment.[73] The former places one person's face onto the face of another; the latter leaves the face of

65. Chesney & Citron, *supra* note 15, at 1760.
66. SAYLER & HARRIS, *supra* note 54.
67. *Id.*
68. Spivak, *supra* note 33, at 343–45.
69. Ice, *supra* note 57, at 422.
70. Lauren Renaud, Comment, Will You Believe It When You See It? How and Why the Press Should Prepare for Deepfakes, 4 GEO. L. TECH. REV. 241, 244 (2019).
71. *Overview of GAN Structure*, GOOGLE DEVELOPERS, http://developers.google.com/machine-learning/gan/gan_structure (last updated May 24, 2019).
72. *See* Diakopoulos & Johnson, *supra* note 5, at 2073.
73. *Id.* at 2074.

the subject as it, but moves it in accordance with the movements of another individual.[74] Voice synthesis techniques follow a similar pattern where the product will either mimic an audio recording and use it to create a video that matches up perfectly to the sound, or use a small clip of audio to then dictate any form of speech that is read in the voice of the subject.[75]

## B. *Spreading like Wildfire*

In today's techno-feudalistic society—the technology creators are the sovereign, its regulators are the nobility, its owners are the vassals, and its users are the peasants.[76] While the internet has provided history with a new dimension, it has also amplified previously restrictive notions of accessibility.[77] This has not only led to the democratization of technology, but also to the potential for harm that it brings.[78] In this techno-feudalistic world, although the simplicity with which technology has allowed deepfakes to be made is a concerning thought, one of the more troubling traits of deepfake technology is its recent and continued attainability.[79] After all,

> Modern technology has not only provided new, convincing, false content, it has also facilitated its dissemination. Social media platforms have made sharing content faster than ever by the retweeting, sharing, or reposting mechanisms they have implemented. This may not be a problem on its own, but recent research suggests that not all content spreads at the same rate. Research from Massachusetts Institute of Technology (MIT) suggests that false content spreads up to six times faster than factual content on social media sites and false news stories are seventy percent more likely to be shared.[80]

---

74. *Id.*; *see* U.S. GOV'T ACCOUNTABILITY OFF., *supra* note 19.

75. Spivak, *supra* note 33, at 352.

76. Alex Hagan, Comment to *Future of Work: What Is "Techno-Feudalism"?*, QUORA (July 6, 2015, 1:37 AM), http://www.quora.com/Future-of-Work-What-is-techno-feudalism.

77. Erwin Chemerinsky, Dean of L., Univ. Cal. Berkley Sch. L., Fake News, Weaponized Defamation and the First Amendment, Keynote Address at Southwestern Law School (Jan. 26, 2012), *in* 47 SW. L. REV. 291, 291.

78. *Id.*; Nuñez, *supra* note 35, at 786, 788.

79. Katarina Kertysova, Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered, 29 SEC. & HUM. RTS. 55, 63–64, 67 (2018).

80. Nuñez, *supra* note 35, at 786.

Importantly, the hurdles on the path to mastering the creation and dissemination of deepfake media are not the technical skills required to create deepfakes, per se, but rather the attainability of processors with sufficient capacity to run large programs.[81] Because GANs make outputs a product of inputs, the greater its data training set, the easier it is for the program to develop a credible piece of deepfake media.[82] This requires that a creator obtain a graphic processing unit ("GPU") sizeable enough, and with a vast amount of memory, to work through the large quantities of photos, videos, or audios of the target.[83] More specifically, to perform deep learning, train a neural network to reconstruct patterns effectively, and ultimately create deepfake media, one would need a GPU greater than those found in commercially available laptops (at least as of 2019), and would require an understanding of "torrenting, path configuration, file structures, and application versioning."[84]

In reality, to create an effective deepfake, a user need only a computer comparable to a high-quality gaming laptop that retails for well under $3,000—a far smaller technological obstacle for gamers and avid computer hobbyists.[85] In fact, even that may not be entirely necessary, as anyone with basic computer skills has the means by which to create deepfakes.[86] What makes this heightened accessibility of creative processes possible is the rise of more readily available software in the open market and internet tutorials on the deepfake-media-making process that, together, work to lift technological constraints.[87] For example, FakeApp is a relatively accessible program that does not require complex equipment and creates deepfake media in as little as eight to twelve hours.[88]

Today, more programs are being cheaply sold, with some of the GPUs needed to make deepfake media selling for as low as $160 USD.[89] For those that do not have the financial means or interests to purchase, these types of

81. *See* Dorfman, *supra* note 24, at 20; SAYLER & HARRIS, *supra* note 54.

82. *See* J.M. Porup, *How and Why Deepfake Videos Work — and What Is at Risk*, CSO (Mar. 18, 2021, 2:00 AM), http://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html.

83. *Rise of the Deepfakes*, WEEK (June 9, 2018), http://www.theweek.com/articles/777592/rise-deepfakes.

84. Ice, *supra* note 57, at 425–26.

85. *See id.* at 426; Maras & Alexandrou, *supra* note 46, at 256.

86. U.S. GOV'T ACCOUNTABILITY OFF., *supra* note 19.

87. Kertysova, *supra* note 79, at 63–64; *Science & Tech Spotlight: Deepfakes*, *supra* note 19.

88. Hall, *supra* note 4, at 57.

89. SAYLER & HARRIS, *supra* note 54; Ice, *supra* note 57, at 426.

GPUs are also available to rent.[90] But even absent the necessary GPUs altogether, most people already have access to programs that can develop deepfake media.[91] Mobile applications like Snapchat, Doublicat, and Reface are allowing users to make deepfakes right from the palms of their hands.[92] This may help explain the growing interest that social media users have in the use of deepfake technology.[93] TikTok—one of the newer social media platforms circulating in popular culture—has seen its own buzz around deepfake media with the videos posted by a user very credibly impersonating actor Tom Cruise with deepfake technology.[94] As of the writing of this Article, that TikTok account, @deeptomcruise, now has over 943,600 followers and an approximated forty million views across only six videos.[95]

TikTok does not stand alone, as there are other social media sites with deepfake capabilities.[96] There is a wide field of social media platforms, all of which have seen a fair share of deepfake media uploads, along with, a vast body of literature addressing the issue and the factors that aggravate it.[97] Deep Trace, self-described as the world's first visual threat intelligence company, identified the existence of at least 14,678 deepfakes circulating online at the time of its report—a statistic that shows not only the ease with which deepfakes can be created, but also the simplicity with which social media platforms disseminate them, or at least play a substantive role in doing so.[98]

---

90. SAYLER & HARRIS, *supra* note 54.

91. *See* Rick Andreoli, *Face Swapping App Doublicat/Reface is Hot! — But Is It Safe?*, PARENTOLOGY (July 30, 2020), http://www.parentology.com/the-hottest-new-app-is-doublicat-reface-but-is-it-safe/.

92. *Id.*

93. *See id.*

94. Mitchell Clark, *This TikTok Tom Cruise Impersonator is Using Deepfake Tech to Impressive Ends*, VERGE (Feb. 26, 2021, 5:54 PM), http://www.theverge.com/22303756/tiktok-tom-cruise-impersonator-deepfake.

95. *See* Tom (@deeptomcruise), TIKTOK, http://vm.tiktok.com/ZMeCkpGpt/ (last visited Apr. 15, 2022).

96. Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)*, 118 MICH. L. REV. 1073, 1081 (2020); *see* Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, VA. J.L. & TECH., Spring 2020, at 1, 22.

97. *See* Citron, *supra* note 96, at 1081; Diakopoulos & Johnson, *supra* note 5, at 2073; Duquette, *supra* note 24, at 85; Brown, *supra* note 96, at 7; Chesney & Citron, *supra* note 21, at 883–84; Mbilike M. Mwafulirwa, *Smoke and Mirrors: Constitutional Ideals When Fact and Fiction Can't Be Separated*, OKLA. BAR J., Mar. 2020, at 12, 13; Bruce Bimber & Homero Gil de Zúñiga, *The Unedited Public Sphere*, 22 NEW MEDIA & SOC'Y 700, 703 (2020); Cathay Y. N. Smith, *Truth, Lies, and Copyright*, 20 NEV. L.J. 201, 203 (2019); Nuñez, *supra* note 35, at 784; Caldera, *supra* note 7, at 178; Anderson, *supra* note 36, at 212.

98. Baram, *supra* note 15.

## III. THE FOE WITH A THOUSAND FACES

There are two sides to every coin, and with the good, there comes bad.[99] Deepfakes first hit the scene on Reddit for a troubling purpose: Creating synthetic pornography that featured the faces of well-known celebrities superimposed on existing pornography videos, all without their knowledge or consent.[100] As time went on, this function of deepfake technology became more and more prevalent, and targeted even those not operating as public figures in society.[101] In fact, research suggests that deepfake technology seems to be disproportionately impacting women, whether public figures or private individuals.[102] As one Article author put it, "[t]he harm wrought by [deepfakes] is not simply that a viewer might be deceived into believing that they are watching a video that actually portrays the subject (although, that harm may also exist). Rather, it is the dignitary harm inflicted on the subject herself."[103] Greater still is the disproportionate impact that deepfakes can have when used to blackmail people within vulnerable populations, like the LGBTQ+ community, that oftentimes find themselves hiding in the shadows.[104] The concept of data exploitation for the purposes of blackmailing or harming the LGBTQ+ community is not a new one, and historical data, in addition to modern concerns over dating apps, seems to suggest that the threat is magnified for such communities.[105]

This kind of data exploitation is very much in line with the more archaic forms of data exploitation that have threatened the LGBTQ+ community throughout various points in history.[106] During the height of the Nazi regime, the Gestapo raided sex research institutions and confiscated extensive lists containing the names and addresses of local homosexuals.[107] Those listed became the targets of the Reich Central Office for the Combatting of Homosexuality and Abortion.[108] The Nazis arrested over 100,000 men as homosexuals and took some of these men to concentration camps where they

---

99. *See* Hall, *supra* note 4, at 57–58, 61.
100. *Id.* at 57.
101. PARIS & DONOVAN, *supra* note 7, at 40.
102. Chesney & Citron, *supra* note 21, at 886; Baram, *supra* note 15.
103. Thomas E. Kadri, *Drawing Trump Naked: Curbing the Right of Publicity to Protect Public Discourse*, 78 MD. L. REV. 899, 953 (2019).
104. *See* Kertysova, *supra* note 79, at 67.
105. *See* Molina, *supra* note 22, at 4–5.
106. *See id.*
107. FRANK RECTOR, THE NAZI EXTERMINATION OF HOMOSEXUALS (Stein & Day, Inc., 1981).
108. *Id.*

were denied support groups, experimented on, and murdered.[109] Similarly, in the United States, during the McCarthyist anti-communist campaign of the mid-1900s, the federal government gathered data on homosexuals through community member interrogations and raided community safe spaces to investigate "the alleged employment of homosexuals in the government service" through a congressional subcommittee created for that particular purpose.[110] Any federal employee suspected of being homosexual was terminated and outed publicly—exposing hundreds to lost livelihoods, financial instability, and reduced esteem among their fellow peers and community members, among many other concerns.[111]

As technology advanced, more and more LGBTQ+ individuals have been antagonized by breaches of online privacy and information exploitation.[112] Tragedies like those of Tyler Clementi,[113] Channing Smith,[114] and many others share that common factor.[115] In 2017, LGBTQ+ Chechens saw the latest iteration of this problem.[116] During the last week of February 2017, Chechen officials detained a young man who was suspected of being under the influence of a controlled substance.[117] At the time, Chechen officials searched the man's phone without permission and discovered intimate photographs and messages exchanged with other men which led to the investigation of his social media platforms.[118] The Chechen Officials raided the man's private electronic communications and tortured him to compile a list of other suspected Chechen homosexuals who were then tortured for the same purpose.[119] This sparked the Chechen anti-gay purges, which included the unofficial detention, humiliation, starvation, and torture of Chechen men

---

109. *Id.*

110. Molina, *supra* note 22, at 4, 5.

111. *See id.*

112. *See* AJ Abell, *Coffee Co. Family Says Cyber Bullying Caused High School Student to Take his Own Life*, FOX 17 (Sept. 25, 2019), http://fox17.com/news/local/coffee-co-family-says-cyber-bullying-caused-high-school-student-to-take-his-own-life.

113. Kelly Ebbels, *Tragic end for a true talent*, NORTHJERSEY.COM (Oct. 1, 2010), http://web.archive.org/web/20121017154404/http://www.northjersey.com/news/104132029_Tragic_end_for_a_true_talent.html?page=all.

114. Abell, *supra* note 112.

115. *See id.*; Ebbels, *supra* note 113.

116. TANYA LOKSHINA, "THEY HAVE LONG ARMS AND THEY CAN FIND ME" 1 (Rachel Denber ed., 2017), http://www.hrw.org/sites/default/files/report_pdf/chechnya0517_web.pdf.

117. *Id.* at 16.

118. *Id.* at 16–17.

119. *Id.* at 17.

suspected of being gay.[120] While many of these men were returned to their families, release was often coupled with suggestions that forced disappearances and *honor killings* be carried out.[121]

Although technology has provided, for many in the LGBTQ+ community, access to resources and communities that they once worked secretly to identify, it has also been shown to serve as the target that many in society place on the backs of LGBTQ+ members, as well as the key with which others gain access to them.[122] The abuse of the LGBTQ+ community that we have seen throughout history makes this clear and offers a warning as deepfake technology begins to circulate more prevalently.[123] However, recent events indicate that we need not engage in extensive thought experiments to identify the harms that deepfakes can pose to the LGBTQ+ community.[124]

In 2019, a sex tape was made public of Azmin Ali, the Malaysian Minister of Economic Affairs, engaging in an intimate relationship with the male aid of a rival minister.[125] Aware that homosexuality is illegal in Malaysia, Ali and his allies downplayed the tape and its insinuation by claiming that it was fabricated with deepfake technology and not real.[126] Some digital forensic professionals have yet to find any evidence to suggest that the footage is a deepfake.[127] This circumstance exposes what experts call the "liar's dividend," or when a skeptical public aware of deepfake technology becomes primed to doubt the authenticity of real audio and video evidence.[128] Putting aside questions as to the veracity of the footage, the Ali controversy serves as a reminder that many members of the LGBTQ+ community still live in locations where the exposure of their sexual and gender identity can deny

120. *See id.*

121. LOKSHINA, *supra* note 116, at 1.

122. *See* Molina, *supra* note 22, at 4.

123. *See id.*

124. *See* Veronica Cordoba, *Malaysians Gets First Hand Experience of Deepfake Tech in Scandal Rich Country*, INDEP. NEWS & MEDIA (June 16, 2019), http://theindependent.sg/malaysians-gets-first-hand-experience-of-deepfake-tech-in-scandal-rich-country/.

125. *Id.*; Jarni Blakkarly, *A Gay Sex Tape is Threatening to End the Political Careers of Two Men in Malaysia*, SBS NEWS (June 17, 2019, 3:50 PM), http://www.sbs.com.au/news/a-gay-sex-tape-is-threatening-to-end-the-political-careers-of-two-men-in-malaysia.

126. Cordoba, *supra* note 124; Blakkarly, *supra* note 124.

127. *Digital Forensics Experts Not Convinced that Gay Sex Videos are Fake*, FREE MALAY. TODAY (June 17, 2019, 4:10 PM), http://www.freemalaysiatoday.com/category/nation/2019/06/17/digital-forensics-experts-not-convinced-that-gay-sex-videos-are-fake.

128. SAYLER & HARRIS, *supra* note 54.

them opportunities, employment, liberty, and even life.[129] However, whether the Ali video is ultimately identified as real or a deepfake is irrelevant in this discussion as the possibility alone highlights the reality that many in the LGBTQ+ community might soon face.[130] Even if no scandalous footage exists, any maliciously-intentioned individual could create a deepfake of a target that depicts them engaging in homosexual acts, and use it either to have them denied opportunities altogether or have them extorted for their own purposes.[131]

## IV. STOP, DROP, OR ROLL WITH IT

While numerous scholars, technologists, and government representatives have advocated for the development of new policy initiatives to specifically address the growing concern surrounding deepfake media and its implications, this Article focuses more squarely on the adaptation of current laws in order to provide recourse to those harmed by deepfake media, and on the legislative efforts that can be taken to secure the privacy and safety of members of the LGBTQ+ community in light of the potential for that content's misuse.[132]

### A. *The Regulation Race*

With the public's understanding of deepfake technology continuing to grow, and its concerns for its misuse growing with it, one question has become more prevalent among many others—if technology created the problem, shouldn't technology be the thing to offer the solution?[133] In 2020, Congress passed the first deepfake-specific statute—not one addressing any regulatory structure, but rather one incentivizing research into the development of deepfake detection software similar to the one depicted below that uses a blue box to track head rotation, red dots to map facial expressions, and green beams to detect the direction of eye movement.[134]

129. *See* Blakkarly, *supra* note 125; LOKSHINA, *supra* note 116, at 1.
130. *See* Blakkarly, *supra* note 125.
131. *See id.*; SAYLER & HARRIS, *supra* note 54.
132. *See* discussion *infra* Part III.
133. *See* Dorfman, *supra* note 24, at 21, 23–24.
134. *Id.* at 21.

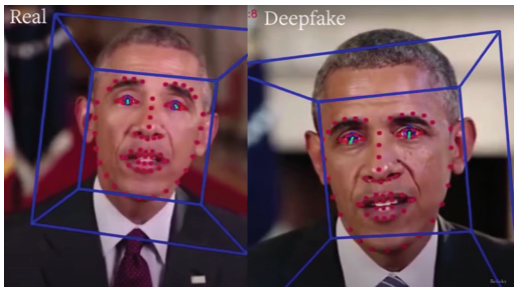


*Image Demonstrating Deepfake Detection Program* [135]

Although the development and implementation of deepfake detection software presents a compelling solution to the potential misuse of deepfake media, if detection technology lacks the capacity to keep pace with the continued advancement of development technology, it would fail to provide an effective solution.[136] After all, the very nature of the GANs used to create deepfake media is to identify the manners in which credibility is detected, and to work over and over to create an output that can successfully deceive its observer.[137] If programmed to understand the methods employed by detection technology, the GANs themselves could probably be trained to create outputs that overcome the detectors from the onset.[138] The most effective way to exercise a useful degree of control over deepfake technology is to take a multifaceted approach.[139]

Although some states have passed legislation regulating the use of deepfakes in particular circumstances, deepfakes still remain unregulated by federal law and not one area of jurisprudence fully governs their use and misuse.[140] Even regulatory agencies like the Federal Trade Commission ("FTC") do not have on-point policies to address this concern.[141] Although the FTC does note that, "[i]f a company's use of . . . deepfakes . . . misleads consumers, that company *could* face an FTC enforcement action."[142] Undoubtedly, crafting an effective legal framework to ensure judicial

135. UC Berkeley, *New Technique for Detecting Deepfake Videos*, YOUTUBE (June 18, 2019), http://www.youtube.com/watch?v=51uHNgmnLWI.

136. Chesney & Citron, *supra* note 15, at 1787.

137. Ice, *supra* note 57, at 422.

138. *See* Chesney & Citron, *supra* note 15, at 1787.

139. Sharon D. Nelson et al., *Detecting Deepfakes*, LAW PRAC. MAG., Jan/Feb. 2020, at 42, 45.

140. Caldera, *supra* note 7, at 178.

141. *See* Chesney & Citron, *supra* note 15, at 1807

142. Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N: BUSINESS BLOG (Apr. 8, 2020), http://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms (emphasis added).

accountability is not without its difficulties.[143] For example, it is likely that an outright ban of deepfake media would survive constitutional muster under the First Amendment.[144]

Additionally, one of the more challenging obstacles that deepfakes present to those that wish to challenge them in the civil arena is the fact that finding their creators is very difficult given that the metadata needed to determine a deepfake's provenance may be inadequate for proper identification of its creator.[145] Moreover, although deepfake media necessitates the exploitation of copyrighted material for its development, a would-be challenger would still have to take on the herculean task of locating the large quantities of input media, sifting through all of the inputs to determine if they have rights in any one, or a few, of the materials used and, if so, still overcome arguments of fair use and transformative use.[146] By that same token, even if the creator of a deepfake is identified, given the geographically diverse nature of the internet, it may be likely that a deepfake creator is domiciled outside of the United States making the exercise of jurisdiction over the creator yet another difficult hurdle to overcome.[147] And of course, civil suits often come at a high cost to both the plaintiff's financial interests as well as to their regard in the public eye, ultimately introducing the possibility of exacerbating their harms.[148] While overcoming these initial road bumps is not impossible, the current legal landscape seems to be a difficult one under which a would-be plaintiff could find the solution they seek.[149] However, that is not to say that there may not be an effective path forward.[150]

### B. *Hope on the Horizon: Amending and Modifying Section 230*

Much of what makes deepfakes so harmful is not just the content itself, but also its ability to spread so rapidly on social media platforms.[151] However, what makes the latter of those two issues possible is not so much the product of the deepfakes themselves as it is the platforms that house them and the laws used to regulate them.[152] The most relevant among them is the

143. *See* Chesney & Citron, *supra* note 15, at 1789.
144. *See id.* at 1790–1791.
145. *See* Chesney & Citron, *supra* note 21, at 889.
146. *See* Chesney & Citron, *supra* note 15, at 1793.
147. *Id.* at 1792.
148. *Id.*
149. *See* Diakopoulos & Johnson, *supra* note 5, at 2086.
150. *See id.*
151. *See* Chesney & Citron, *supra* note 15, at 1768.
152. *Id.* at 1795.

Communications Decency Act,[153] a federal law passed by Congress in 1996—particularly, Section 230 of the Act.[154] Section 230 of the Communications Decency Act states in pertinent part:

> (c) Protection for "Good Samaritan" blocking and screening of offensive material
>
> (1) Treatment of publisher or speaker
>
> *No provider or user of an interactive computer service shall be treated as the publisher or speaker* of any information provided by another information content provider.
>
> (2) Civil liability
>
> 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).[155]

Prior to the enactment of Section 230, websites that sought to moderate harmful or offensive material posted by third parties were treated as publishers and would be held liable if they were unsuccessful in removing the harmful material.[156] However, this presented an interesting loophole as it allowed for websites to stick their heads in the sand, so to speak, and evade the imposition of liability by ignoring any harmful content that they knew

153. Communications Decency Act of 1996, Pub. L. No. 104–104, § 110 Stat. 133 (codified as amended at 47 U.S.C. § 223 (Supp. II 1997)).

154. Citron, *supra* note 96, at 1088.

155. 47 U.S.C. § 230(c)(1) (emphasis added).

156. Brown, *supra* note 96, at 43.

existed on their platform.[157] This was seen in action in 1994 when an individual accused Stratton Oakmont of fraudulent and illegal securities trading practices on Money Talk, a message board run by Prodigy Communication Corporation—a leading Internet Service Provider at the time.[158] Stratton responded by suing both Prodigy and Money Talk's administrator for libel.[159] The court concluded that Prodigy was a publisher because it held itself out to the public as controlling the content of its computer bulletin boards and practiced such control through its use of an automatic software screening program.[160] Following the ruling in the *Stratton Oakmont, Inc. v. Prodigy Services Co.*,[161] Congress members grew concerned that the court's ruling would disincentivize the moderation of offensive content posted to internet service providers by third parties.[162] As a result, Congress passed the Communications Decency Act and included in it, Section 230, which offers internet service providers a broad shield of immunity from liability for moderating too much or too little of their third party users' speech.[163]

However, the absence of narrow language in Section 230's immunity clauses has come to be interpreted in a manner that is so broad that it has created yet another problem—immunity remains available even if an internet service provider intentionally encourages the posting of harmful content.[164] In fact, when it has been applied, this expansive immunity has allowed internet service providers to republish content with the knowledge that it violates the law, alter their platform to prevent the capture of criminals, and allow the sale of illegal or dangerous products.[165] For example, Grindr, a dating app marketed primarily to the gay community, often sees fake profiles on its platform wherein a user appropriates the images, whether more commonplace or intimate, of another ("catfishing").[166] Recently, one of Grindr's users, Matthew Herrick, sued the company for the negligent design of its application after Herrick's ex-boyfriend began impersonating him on the app by creating a fake profile in his name, spreading his nude photographs, and sharing rape

157. *See id.*
158. Spivak, *supra* note 33, at 387.
159. *Id.*; Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *4 (N.Y. Sup. Ct. May 24, 1995).
160. Spivak, *supra* note 33, at 387–88; *Stratton Oakmont, Inc.*, 1995 WL 323710, at *4.
161. No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).
162. Citron, *supra* note 96, at 1088–89.
163. *Id.* at 1089.
164. *See* Chesney & Citron, *supra* note 15, at 1797.
165. *Id.* at 1798; *see also* Citron, *supra* note 96, at 1089.
166. *See* Chris Fox, *Why Do Gay Apps Struggle to Stop Catfish?*, BBC (Oct. 28, 2019), http://www.bbc.com/news/technology-50138390; Citron, *supra* note 96, at 1089.

fantasies with other users, among many other things.[167] Herrick notified Grindr of his ex-boyfriend's behaviors over one hundred times to no avail.[168] At one point, Herrick's ex-boyfriend shared Herrick's address publicly on the app, which led to over twenty strangers coming to his apartment on any given day, totaling more than a thousand trespassers.[169] The court dismissed Herrick's suit against Grindr on the grounds of Section 230 immunity, noting that his claims were all based on content provided by another Grindr user, not by Grindr itself, and that to the extent that Grindr had contributed to the profiles impersonating Herrick, it was only through "neutral assistance," for which Section 230 has been interpreted to provide immunity.[170] The ruling was later affirmed on appeal.[171]

Looking more directly at the role of Section 230 in the perpetuation of the threats that deepfake technologies present particularly to the LGBTQ+ community, the facts of *Herrick v. Grindr, LLC*[172] can be adapted to illustrate the point in action.[173] User A of a social media platform either creates or obtains a deepfake that impersonates User B, and depicts B in some intimate act without their knowledge or consent.[174] User A uploads the harmful or offensive deepfake to a social media platform, where the content spreads and gets shared.[175] Upon discovery of the deepfake content, B requests that the platform remove it on account of the professional, reputational and psychological harm that the synthetic media creates.[176] The platform never removes the content, it continues to spread and harm B, and B sues the platform.[177] Under the current framework of Section 230, it would be unreasonable to expect that the adapted facts would yield a conclusion different than the one in *Herrick*, despite the platform's actual knowledge of the harmful or offensive nature of the content in question and of the damage that it causes.[178] Herein lies the heart of the problem, worse even, when the

167. Citron, *supra* note 96, at 1089.
168. *Id.*
169. *Id.* at 1089–90.
170. Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 584, 589 (S.D.N.Y. 2018), *aff'd*, 765 F. App'x 586 (2d Cir. 2019).
171. Herrick v. Grindr, LLC, 765 F. App'x 586, 593 (2d Cir. 2019).
172. 306 F. Supp. 3d 579 (S.D.N.Y. 2018), *aff'd*, 765 F. App'x 586 (2d Cir. 2019).
173. *See id.* at 585; Citron, *supra* note 96, at 1089, 1091; Chesney & Citron, *supra* note 21, at 884–85.
174. *See Herrick*, 306 F. Supp. 3d at 584–85.
175. *See id.* at 585.
176. See id.
177. See id.
178. *See id.* at 589.

facts are adapted once more so that B is a closeted LGBTQ+ community member whom A exploits and extorts financially or professionally on social media platforms through the use of the deepfake media.[179] The risks seem endless and could even magnify to implicate national security.[180]

As it stands now, Section 230 provides nearly no incentive for social media platforms to monitor and moderate the content on their interfaces, even in the face of actual knowledge of their harmful and offensive nature.[181] However, growing calls to amend Section 230 could provide a solution that not only opens the door to the imposition of liability on account of deepfake media perpetuation, but also more broadly offers recourse for those harmed by it.[182] After all, amendments to Section 230's immunity clauses are not a foreign or far away idea.[183] Only three years prior to the time of this writing, Section 230's text was amended by the passage of the Allow States and Victims to Fight Online Sex Trafficking Act of 2017 ("FOSTA").[184] More recently, legislators on both sides of the aisle—like current Speaker of the House Nancy Pelosi, Senators Ted Cruz and Josh Hawley, and even a former White House technology advisor—have all suggested repealing or amending Section 230.[185] There is merit in Section 230's interest in fostering free speech, and making a blanket repeal could be costly with regard to social discourse; however, modification presents an appealing solution that allows for the maintenance of both the free speech and privacy protection interests.[186]

As Jon M. Garon, Professor and former Dean at Nova Southeastern University's Shepard Broad College of Law, has noted, there are two ways in which Section 230 can be modified to strike a more equitable balance between the two interests.[187]

> First, once content has been determined by a court to be libelous or harassing, the [Internet Service Provider] should have an obligation to remove that content immediately upon notification. Second, if an [Internet Service Provider] refuses to remove content someone

179. *See Herrick*, 306 F. Supp. 3d at 585, 589; Molina, *supra* note 22, at 4.

180. *See Herrick*, 306 F. Supp. 3d at 584; Chesney & Citron, *supra* note 15, at 1783; Molina, *supra* note 22, at 4.

181. *See* Citron, *supra* note 96, at 1088–89.

182. Chesney & Citron, *supra* note 15, at 1799.

183. *See id.* at 1798–99.

184. Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115–164, 132 Stat. 1253 (2018) (codified as amended at 18 U.S.C. § 2421A); Brown, *supra* note 96, at 45.

185. Brown, *supra* note 96, at 44–45.

186. *See* Jon M. Garon, *How to Fix the Internet*, LAW360 (Feb. 17, 2017, 4:41 PM), http://www.law360.com/articles/889115/how-to-fix-the-internet.

187. *Id.*

> believes to be defamatory or an invasion of privacy, that person should be permitted to go to court to have the content removed. If a court determines the speech is harmful, then the [Internet Service Provider] should be obligated to take down or block the speech. That order would then apply to other [Internet Service Providers] as well.[188]

A similar proposed modification has been offered by Danielle Citron and Benjamin Wittes—both leading voices in the area of deepfake technology and its proposed regulation—that conditions immunity on reasonable content moderation practices [189] As suggested by Citron and Wittes, the proposed amendment to Section 230(c)(1) would read:

> No provider or user of an interactive computer service that *takes reasonable steps to prevent or address unlawful uses of its services* shall be treated as the publisher or speaker of any information provided by another information content provider *in any action arising out of the publication of content provided by that information content provider*.[190]

Although there is no simple legislative solution to specifically address the complications that deepfakes present, there is no shortage of reasonable amendments to Section 230's existing language that could remedy the present situation.[191] Amending Section 230 to at least allow individuals threatened by the posting of deepfake media to more fairly challenge its presence, not only establishes recourse where there currently seems to be none, but also shelters individuals, like those in the LGBTQ+ community, whose privacy and safety are at a heightened risk of exploitation and extortion.[192]

### C. *Embracing our New Reality: The LGBTQ+ Community's Information Anonymization Efforts with Deepfake Technology*

The LGBTQ+ community has long been a champion of reappropriating the tools of its oppressors for purposes of finding

---

188. *Id.*

189. *See* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 FORDHAM L. REV. 401, 419 (2017).

190. *Id.*

191. *See* Garon, *supra* note 186; Citron & Wittes, *supra* note 189, at 419.

192. *See* Spivak, *supra* note 33, at 399; Garon, *supra* note 186; Citron & Wittes, *supra* note 189, at 419.

empowerment and freedom.[193] For a community in which the pendulum between its two driving values—visibility and privacy—swings quickly from one side to the other, it comes as no surprise that some of its members have now embraced deepfake technology.[194] Despite deepfakes posing potential privacy and safety threats, members of the LGBTQ+ community have been quick to get in front of deepfake technology and adopt it for anonymization and privacy protection purposes.[195] As of the time of this writing, the most notable use of such a tactic was by the creators of *Welcome to Chechnya*, a documentary exposing the realities of the victims of the Chechen anti-gay purges that uses deepfake technology to mask and protect the identities of the victims and informants featured in the film.[196]

Visual effects expert Ryan Laney described *Welcome to Chechnya*'s technological endeavor as "a digital prosthetic where 100[%] of the motion, the emotion, and the essence of what the subject is doing is there."[197] In order for the documentary's visual effects team to develop this advanced form of anonymization, individuals volunteered a personal image and consented to its application on the content of the film's subjects, ultimately resulting in a sort of digital marionette puppet.[198] Two things make this process different than the ones in which other deepfakes are usually seen.[199] First, the deepfakes were created with the consent of both the subject—the person on which the altered image is placed, or, in other words, the person anonymized—and the target—the person whose image is being transposed on the subject, or, in other words, the anonymizer.[200] Second, the purpose of such a deepfake is to safeguard the interest of the subject rather than to exploit it.[201]

193. *See* Juliette Rocheleau, *A Former Slur is Reclaimed, and Listeners Have Mixed Feelings*, NPR (Aug. 21, 2019, 10:33 AM), http://www.npr.org/sections/publiceditor/2019/08/21/752330316/a-former-slur-is-reclaimed-and-listeners-have-mixed-feelings.

194. Rebecca Heilweil, *How Deepfakes Could Actually do Some Good*, VOX (June 29, 2020, 11:10 AM), http://www.vox.com/platform/amp/recode/2020/6/29/21303588/deepfakes-anonymous-artificial-intelligence-welcome-to-chechnya.

195. *Id.*

196. *Id.*

197. *Id.*

198. *Id.*

199. *See* Heilweil, *supra* note 194.

200. *See id.*

201. See *id.*

*Image of Deepfake Technology used to Disguise Source Featured in Welcome to Chechnya*[202]

*Welcome to Chechnya* has reappropriated technology that is potentially dangerous to members of the LGBTQ+ community around the world and turned that sword, and its threat, into the shield behind which they can find privacy and protection.[203] In fact, this may already be a growing trend that members of the LGBTQ+ community can jump on board with.[204] Laney and other startups such as D-ID and Alethea AI have begun to take an interest in developing entities that facilitate and democratize the creation of "digital veils" to cloak individuals in danger.[205] Until legislation is created or amended to better protect against the threats that deepfake technology could pose to individuals like those in the LGBTQ+ community, the adoption of these "digital veil" programs could not only help to bring about peace of mind, security of liberty, and protection of life to members of the LGBTQ+ community.[206] It could also introduce them to the next evolutionary chapter in its long history of adaptation and self-preservation.[207]

## V. CONCLUSION

As society ventures into a new world where technology develops at an evolutionary rate faster than usual, society must remain mindful, in addition to being cautious, not only of the many implications that advancements have on the technical aspect of our society but also on the social implications that may arise as the natural byproduct.[208] While deepfakes present the newest, in

202. *Id.*; Joshua Rothkopf, *Deepfake Technology Enters the Documentary World*, NY TIMES, http://www.nytimes.com/2020/07/01/movies/deepfakes-documentary-welcome-to-chechnya.html (July 29, 2020).
203. *See* Heilweil, *supra* note 194.
204. See id.
205. *Id.*
206. See id.
207. See id.
208. Chesney & Citron, *supra* note 21, at 889–90.

a long history, of audiovisual manipulations, a closer look at society's response to deepfake technology through an equally evolutionary perspective is necessary.[209] The most effective manner of doing this is in the same way that we have for all other challenges that we face in our day-to-day lives: with the guidance of the law, whose essential function is to provide recourse where individuals are harmed, to carve out a path to such an outcome where there does not exist such a form of redress, and to disincentivize malicious wrongdoers from their misdeeds.[210]

Society must remain hopeful that experts, scholars, technologists, and legislators will move to introduce specific policies that help society counteract the potentially negative implications that deepfake technology may present.[211] However, until a one-size-fits-all policy is adopted, it is necessary to adapt existing regulatory frameworks like Section 230 so as to effectively face deepfakes' problems as they come and protect the interests of those communities like the LGBTQ+ community that remain vulnerable to them.[212]

209. *Id.* at 889.
210. Citron, *supra* note 96, at 1074.
211. *See id.* at 1087.
212. *See id.* at 1090.