UIC John Marshall Law Review

Volume 54 | Issue 1 Article 4

2021

# Is Truth Hanging on by a Thread?, 54 UIC J. Marshall L. Rev. 315 (2021)

Thomas Ryan

# IS TRUTH HANGING ON BY A THREAD?

THOMAS RYAN*



## I. INTRODUCTION

On February 14, 2018, a gunman armed with a semiautomatic AR-15 rifle shot and killed 14 students and three educators at Marjory Stoneman Douglas High School in Parkland,Florida.[1] The story quickly hit the news cycle and spread

*Juris Doctor Candidate, UIC John Marshall Law School, 2021. I want to thank John Conklin for his unwavering guidance and encouraging me to write this Comment. Many thanks to the editors of the UIC John Marshall Law Review. I also wish to thank my wife for her unparalleled support and a special thanks to my mom for everything.

1. Audra D. S. Burch & Patricia Mazzei, *Death Toll is at 17 and Could Rise in Florida School Shooting*, N.Y. TIMES (Feb. 14, 2018), www.nytimes.com/2018/02/14/us/parkland-school-shooting.html [perma.cc/S5EH-5Q2G]. *See also* Amy Held, '*We Live with It Every Day:' Parkland Community Marks 1 Year Since Massacre*, NAT'L PUB. RADIO (Feb. 14, 2019), www.npr.org/2019/02/14/694688365/we-live-with-it-every-day-parkland-community-marks-one-year-since-massacre [perma.cc/2TT7-DH9W] (providing a retrospective on the

online.[2] While journalists and law enforcement sought accurate information, an anonymous user of the notorious forum 4chan[3] posted a screenshot of what appeared to be a Buzzfeed News article on 4chan's message board.[4] The article was titled "Why We Need to Take Away White People's Guns Now More Than Ever" and was supposedly written by Richie Horowitz.[5] The screenshot contained a fabricated quote from the Broward County Sheriff, Scott Israel, the sheriff who was working the Marjory high school shooting.[6] The article was completely fake.[7] There was no Richie Horowitz working at Buzzfeed; nor was there an article published at Buzzfeed or any media outlet with similar content.[8] The headline was intended to pull on the reader's emotional strings.[9] But it did not matter whether the article was real or fake because the fabricated screenshot "pulsed through right-wing outrage channels[10] and was boosted by activists on Twitter."[11]

The fake article was created by an individual using Photoshop,[12] intended to mislead the public and stir up emotions

---

community and students who were affected by the Marjory Stoneman Douglas High School shooting).

2. *See generally* Bethania Palma, *Did BuzzFeed Advocate for Taking Away White People's Guns*, SNOPES (Feb. 14, 2018), www.snopes.com/fact-check/buzzfeed-white-people-guns [perma.cc/KUJ7-GRBN] (dissecting the origins and spread of the fake article).

3. Caitlin Dewey, *Absolutely Everything You Need to Know to Understand 4chan, The Internet's Own Bogeyman*, WASH. POST (Sept. 25, 2014), www.washingtonpost.com/news/the-intersect/wp/2014/09/25/absolutely-everything-you-need-to-know-to-understand-4chan-the-internets-own-bogeyman [perma.cc/9399-5GBA] (explaining that 4chan is a website forum with virtually no rules or consequences for their anonymous users where "nearly every evil of the internet begins, or picks up steam"). *See also* Emma Grey Ellis, *4Chan Is Turning 15—And Remains the Internet's Teenager,* WIRED (June 1, 2018), www.wired.com/story/4chan-soul-of-the-internet [perma.cc/DU7Q-Q2MK] (exploring 4chan's reach to the general public despite its low user count and highlighting the good and bad communities that began there).

4. EJ Gibney (@EJGibney), Twitter (Feb. 14, 2018), twitter.com/EJGibney/status/963951037416726528 (identifying the origin of the fabricated article).

5. Palma, *supra* note 2.

6. *Id.*

7. *Id.*

8. Kevin Roose, Here Come the Fake Videos, Too, N.Y. Times (Mar. 4, 2018), www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html [perma.cc/H8YG-YX2N].

9. *See generally* Palma, supra note 2 (documenting the spread of the fabricated article starting with 4chan all the way to a White House correspondent).

10. Palma, *supra* note 2 (explaining that after being posted on 4chan, a white supremacist Twitter account MAGA Pill tweeted about it). "MAGA Pill gained fame when U.S. President Donald Trump retweeted the account in November 2017." *Id.*

11. Roose, *supra* note 8.

12. Photoshop is an image-editing software. *All About Photoshop*, GCFGLOBAL (2019), edu.gcfglobal.org/en/photoshopbasics/what-is-photoshop/1

among right-leaning news outlets.[13] The doctored image of the fake article was effective in presenting a false narrative.[14] The viral spread of disinformation can go viral in seconds,[15] especially with the help of bots.[16] This kind of false information gets 50,000 shares before any kind of debunking happens, and then information about the debunking only receives 200 shares.[17] Companies like Facebook, YouTube, Twitter, and Google are all hubs for disinformation.[18] As long as the individual has access to the internet, they can spread disinformation by utilizing something with a strong emotional appeal like the fake Buzzfeed article.

Manipulating images and words to tell a story with the intent to deceive or mislead is not new.[19] Now, manipulating videos through the application of machine learning has entered the fray.[20] Society is entering the age of manipulated videos and deepfakes and must now be skeptical of what we see in video form.[21] Instead of a doctored image of a Buzzfeed article, society is presented with a realistic video or audio of someone saying something or doing something that they did not actually say or do.[22] Hypothetically, a

---

[perma.cc/WT99-LLUF].

13. *See* Roose, *supra* note 8. (explaining that the spread of misinformation is not new to the internet).

14. *Id.*

15. Shelly Banjo, *Facebook, Twitter and the Digital Disinformation Mess*, WASH. POST (Oct. 2, 2019), www.createai.io/blog/post/page/facebook-twitter-and-the-digital-disinformation-mess---washington-post [perma.cc/3TQV-266E] (defining disinformation as "false content spread with the specific intent to deceive, mislead or manipulate.").

16. Jennifer Ouellette, *Study: It only takes a few seconds for bots to spread misinformation*, ARS TECHNICA (Nov. 21, 2018), www.arstechnica.com/science/2018/11/study-it-only-takes-a-few-seconds-for-bots-to-spread-misinformation [perma.cc/PHJ4-96TA?type=image].

17. Roose, *supra* note 8.

18 *See* Banjo, *supra* note 15 (urging companies to take action against disinformation).

19. *See* Meg Neal, *The First Photo was Faked 150 Years Before Photoshop Existed*, GIZMODO (Apr. 10, 2015), www.gizmodo.com/the-first-photo-was-faked-150-years-before-photoshop-ex-1697072182 [perma.cc/MAG9-FMU3] (discussing the first recorded faked photo from the mid-1800s by a photographer faking his own death).

20. Karen Hao, *What is Machine Learning?*, MIT TECH. REV. (Nov. 17, 2018), www.technologyreview.com/s/612437/what-is-machine-learning-we-drew-you-another-flowchart [perma.cc/2HRP-ZGEA] (explaining that machine learning is data collection software that essentially "finds the pattern, [and then] appl[ies] the pattern"). "Machine-learning algorithms use statistics to find patterns in massive amounts of data. And data…encompasses a lot of things—numbers, words, images, clicks . . . If it can be digitally stored, it can be fed into a machine-learning algorithm." *Id.*

21. Tom Simonite, *Prepare for the Deepfake Era of Web Video*, WIRED (Oct. 6, 2019), www.wired.com/story/prepare-deepfake-era-web-video [perma.cc/YL4P-HJU8] (explaining that manipulated videos are video "clips altered or fabricated with an artificial intelligence technique called machine learning").

22. *See* Robert Chesney & Danielle K. Citron, *Deep Fakes: A Looming*

fake video of a left leaning politician could show the individual arguing about why white people's guns should be taken away from them.[23] In reality, the politician may have never said what they are saying in the fake video, but the video looks real so the uninformed audience believes it is real. A fake video, which is more engaging than a still image, will spread faster than the truth, especially when it is related to a politically charged topic.[24]

No longer is society too concerned about someone falsely shouting fire in a crowded theater as Justice Oliver Wendell Holmes warned of 100 years ago.[25] Today, a social media user is the person shouting "fire," except now this person has a megaphone that can reach anyone with internet access. "[F]alse cries in the form of deep fakes go viral, fueled by the persuasive power of hyper-realistic evidence in conjunction with the distribution powers of social media."[26] A well-timed deep fake will not be a slight annoyance but, rather, it will cause a public panic that might involve property destruction, personal injuries, and/or death.[27]

The goal of this Comment is to encourage lawmakers to pressure social media companies to fight disinformation by applying 47 U.S.C. § 230 of the United States Communication Decency Act ("Section 230" or "CDA"). This Comment will look at the emergence of deepfakes, how they are made, and the potential benefits and harms of them. This Comment will also look at the history of Section 230 and its application. Additionally, this Comment analyzes the legal effects of Section 230 and the FOSTA-SESTA exception. Lastly, this Comment proposes three potential solutions to prevent deepfakes and a test that the courts could apply to social media

---

*Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 4 (2019) (analyzing deepfakes' potential effects on society).

23. *Cf.* Palma, *supra* note 2. (discussing fake images in the Introduction).

24. *See* Brian Resnick, *False News Stories Travel Faster and Farther on Twitter than the Truth*, VOX (Mar. 19, 2018), www.vox.com/science-and-health/2018/3/8/17085928/fake-news-study-mit-science [perma.cc/N67U-7LEF] (analyzing a published study showing "false news stories and rumors spread on social media at a frightening speed — often outpacing the truth"). *See also* Soroush Vosoughi et al., *The Spread of True and False News Online*, SCIENCE (Mar. 9, 2018), science.sciencemag.org/content/359/6380/1146/tab-pdf [perma.cc/PYE9-N3VX] (investigating how quickly true and false stories spread on Twitter). "Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information." *Id*. The investigation also found that "the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information." *Id.*

25. Schenck v. United States, 249 U.S. 47, 52 (1919) ("The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic").

26. Chesney & Citron, *supra* note 22.

27. *See e.g., Id.* (explaining how in early 2018, Hawaii was put in a state of panic after an employee of Hawaii's Emergency Management Agency accidentally issued a text warning about an incoming ballistic missile).

companies.

## II. BACKGROUND

This section will give an overview of deepfakes and its relation to Section 230 of the CDA. First, this section discusses how deepfakes are made and the benefits or problems they can cause. Then, this section discusses the history of Section 230 and its exceptions.

### *A. What are Deepfakes?*

The search results of Googling "best deepfake" will result in links to YouTube videos with compilations of video clips of celebrities and politicians whose faces have been seamlessly swapped. The result is somewhat humorous but also unsettling due to how the face swapping is almost unnoticeable. These deepfake videos are enjoyable to watch, but there is a more sinister use of them: to harass or discredit people, mostly women journalists and activists.[28]

The term "deepfake" originated from a Reddit[29] user whose username was "deepfakes."[30] The term "deepfake" was a portmanteau[31] of "deep learning"[32] and "fakes."[33] The term has since

28. *See* Tom Simonite, *Prepare for the Deepfake Era of Web Video*, WIRED (Oct. 6, 2019 ), www.wired.com/story/prepare-deepfake-\era-web-video [perma.cc/9SGH-938V] (discussing deepfakes with program director of Witness, a human rights nonprofit).

29. Jake Widman & Will Nicol, *What is Reddit? A Beginner's Guide to the Front Page of the Internet*, DIGITAL TRENDS (May 22, 2019), www.digitaltrends.com/web/what-is-reddit [perma.cc/Z78A-U6ZJ] (explaining that Reddit bills itself as the "front page of the internet."). It is the town hall of the internet because it is "a massive collection of forums, where people can share news and content or comment on other people's posts." *Id.*

30. *See* James Vincent, *Why We Need a Better Definition of 'Deepfake'*, VERGE (May 22, 2018), www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news [perma.cc/7N2A-563L] (examining deepfakes and distinguishing between them from other fake forms of media).

31. Portmanteau is defined as a word or morpheme whose form and meaning are derived from a blending of two or more distinct forms (such as smog from smoke and fog). *Portmanteau*, Merriam-Webster, www.merriam-webster.com/dictionary/portmanteau [perma.cc/B85U-G6LU].

32. *See* Bernard Marr, *What Is Deep Learning AI? A Simple Guide With 8 Practical Examples*, FORBES (Oct. 1, 2018), www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/?sh=656d868a8d4b [perma.cc/44AJ-UJHH]; (providing examples of deep learning, such as virtual assistants, machine translators, chatbots, etc). *See generally*, Brian S. Haney, *The Perils and Promises of Artificial General Intelligence,* 45 J. LEGIS. 151, 157-58 (2018) (providing background information on deep learning).

33. Vincent, *supra* note 30.

been used to describe edited videos and imagery that uses machine learning.[34] In December 2017, the anonymous Reddit user, "deepfakes," used artificial intelligent (AI) tools to paste celebrity faces onto performers in pornographic video clips.[35]

Pornographic deepfakes brought deepfakes into the mainstream.[36] By January 2018, a free app called FakeApp was created and was downloaded over 120,000 times.[37] The app utilized Generative Adversarial Network (GAN) technology[38] and allowed anyone who had access to the internet to create fake videos freely and with relative ease.[39] Fake pornography exploded online as well.[40] Soon afterwards, Twitter[41] banned deepfake nonconsensual pornographic videos and Reddit closed several deepfake communities or Subreddits,[42] "including one with nearly 100,000 members."[43]

Deepfakes are produced by Generative Adversarial Networks

---

34. *Id.*

35. *Id. See also Rise of the Deepfakes*, THE WEEK (June 9, 2018), www.theweek.com/articles/777592/rise-deepfakes [perma.cc/MF7L-357R] (explaining the technology and origin of deepfakes) .

36. *Id.*

37. *Id.*

38. Kashyap Vyas, *Generative Adversarial Networks: The Tech Behind DeepFake and FaceApp*, INTERESTING ENGINEERING (Aug.12, 2019), interestingengineering.com/generative-adversarial-networks-the-tech-behind-deepfake-and-faceapp [perma.cc/3YWP-NY25] (explaining that the "best analogy that we can use for GAN is that it is a two-player game where each player is trying their hardest to beat one another").

39. *Deepfakes: What are they and why would I make one?* BBC, www.bbc.co.uk/bitesize/articles/zfkwcqt [perma.cc/Y5MG-G3CJ] (last visited Feb. 14, 2020). *See also* Roose, *supra* note 8 (explaining a step-by-step process on creating a deepfake). First, the creator should "[f]ind, or rent, a moderately powerful computer." *Id.* The next step would be to download FakeApp or similar deepfake program and begin to collect the "right source data." *Id.* Source data can be found in the images and videos widely available on the internet. *Id.* Similar looking faces, short video clips, and single angle shots provide better source data than long video clips and multiangle shots. *Id.* After all the source data is gathered, the creator can enter the data into FakeApp. *Id.* FakeApp then identifies patterns and similarities between the two faces. *Id.* The more powerful the computer's processor the less time it will take to produce the deepfake. *Id.*

40. Tom Simonite, *Most Deepfakes Are Porn, and They're Multiplying Fast* WIRED (July 10, 2019), www.wired.com/story/most-deepfakes-porn-multiplying-fast [perma.cc/7ZUK-T3SS] (exploring a startup company's findings that "96 percent of the deepfakes circulating in the wild were pornographic").

41. *See* Marvin Ammori, *The New York Times: Free Speech Lawyering in the Age of Google and Twitter*, 127 HARV. L. REV. 2259, 2260 (2014) (describing how Twitter considers itself a "medium for free speech").

42. *See* Widman & Nicol, *supra* note 29 (defining a subreddit as a likeminded community interested in a particular topic, "[f]or example, /r/nba is a subreddit where people talk about the National Basketball Association, while /r/boardgames is a subreddit for people to discuss board games").

43. Roose, *supra* note 8.

or GANs.[44] A GAN is a type of generative model.[45] Generative modeling allows for the generation "of new photographs that are generally similar but specifically different from a dataset of existing photographs."[46] GANs, then, are two neural networks, or a set of algorithms, that learn from each other.[47] GAN technology can be thought of as a game of cat and mouse.[48] For example, in the movie Catch Me if You Can, a conman-counterfeiter was being chased by an FBI agent, a classic cat and mouse story.[49] The counterfeiter made and cashed fake checks while the agent chased the counterfeiter.[50] The agent would try to detect what was fake and then would guess the counterfeiter's next move.[51] The agent was always one step behind, but getting better at understanding the counterfeiter.[52] However, as the agent learned about the counterfeiter's methods and moves, the counterfeiter improved his game of trickery, each side learning from the other.[53] Essentially,

---

44. Ian Goodfellow et al., *Generative Adversarial Nets*, INT'L CONF. ON NEURAL INFO. PROCESSING SYS. (June 10, 2014) (introducing GANs for the first time). The generative model is pitted against an adversary: a discriminative model that learns to determine whether a sample is from the model distribution or the data distribution. *Id.* The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. *Id.* Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles. *Id.*

45. *See* Jason Brownlee, *Best Resources for Getting Started With GANs*, MACHINE LEARNING MASTERY (June 12, 2019), www.machinelearningmastery.com/resources-for-getting-started-with-generative-adversarial-networks [perma.cc/53ZV-GFPG] (explaining Generative Adversarial Networks as "techniques behind the startlingly photorealistic generation of human faces, as well as impressive image translation tasks such as photo colorization, face de-aging, super-resolution.").

46. *Id.*

47. Chris Nicholson, *A Beginner's Guide to Neural Networks and Deep Learning*, PATHMIND, wiki.pathmind.com/neural-network [perma.cc/XWH7-ZAVK] (last visited Feb. 4, 2021) (providing counterfeiter-cop analogy).

48. *Cat and Mouse*, Merriam-Webster, www.merriam-webster.com/dictionary/cat%20and%20mouse [perma.cc/4DRW-SBFK] (describing the cat's constant pursuit of the mouse, but unable to capture it).

49. CATCH ME IF YOU CAN (Amblin Entertainment 2002) (starring Leonardo DiCaprio as the conman counterfeiter and Tom Hanks as the FBI agent).

50. *Id.*

51. *Id.*

52. *Id.*

53. Chris Nicholson, *A Beginner's Guide to Generative Adversarial Networks (GANs)*, SKYMIND, wiki.pathmind.com/generative-adversarial-network-gan [perma.cc/T38N-L6HU] (last visited Oct. 6, 2019). GANs can be seen as the opposition of a counterfeiter and a cop in a game of cat and mouse, where the counterfeiter is learning to pass false notes, and the cop is learning to detect them. *Id.* Both are dynamic; i.e. the cop is in training, too (to extend the analogy, maybe the central bank is flagging bills that slipped through), and each side comes to learn the other's methods in a constant escalation. *Id.*

system A is learning from system B until system B can reproduce an identical copy of system A.

The generative model (generator), or the counterfeiter in the analogy, tries to produce the fakes or images. Concurrently, its adversary, the discriminative model (discriminator), or the agent in the analogy, tries to detect the fakes or images.[54] During this back and forth, both the generator and discriminator improve.[55] The generator creates fakes closer to the true image while the discriminator refines its detection skills.[56] The battle of creating better fakes to outwit the agent eventually leads to nearly perfect fakes "indistinguishable from the genuine articles."[57]

This technique utilized by GANs allows for the indiscernible generation of human faces or the technology known as deepfakes.[58] Deepfakes collect common characteristics from existing images and learn how to stitch those characteristics onto another image.[59] For example, in a video interview with Bill Hader, his face is morphed into Tom Cruise every time Hader impersonates Cruise.[60] This is possible because there are millions of images of Tom Cruise that can be collected and then superimposed onto Bill Hader's face.[61]

### *1. Deepfake Problems*

The harms of deepfakes affect society on a global scale and will not only affect the present and future but also the past.[62] Deepfakes could alter the past by manipulating what is remembered by planting the population with false memories.[63] For example, in 2016, video footage from a 2011 attack on an airport in Moscow and

54. *Id.*
55. *Id.*
56. *Id.*
57. *Id.*
58. *See* Brownlee, *supra* note 45 (providing an overview of GAN technology).
59. *Id.*
60. Ctrl Shift Face, *Bill Hader channels Tom Cruise [DeepFake]*, YOUTUBE (Aug. 6, 2019), www.youtube.com/watch?v=VWrhRBb-1Ig [perma.cc/M6X2-Q29S].
61. *See* Elle Hunt, *Deepfake Danger: What A Viral Clip of Bill Hader Morphing into Tom Cruise Tells Us,* THE GUARDIAN (Aug. 13, 2019), www.theguardian.com/news/shortcuts/2019/aug/13/danger-deepfakes-viral-video-bill-hader-tom-cruise [perma.cc/5XZA-3WE2] (questioning the dangers of deepfakes and briefly explaining that deepfakes are more successful with higher resolution images).
62. *See* Franklin Foer, *The Era of Fake Video Begins*, ATLANTIC (May 2018), www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877 [perma.cc/X59R-D5BH]. (hypothesizing that "manipulated video will ultimately destroy faith in our strongest remaining tether to the idea of common reality").
63. Brian Resnick, *We're Underestimating the Mind-Warping Potential of Fake Video*, VOX (July 24, 2018), www.vox.com/science-and-health/2018/4/20/17109764/deepfake-ai-false-memory-psychology-mandela-effect [perma.cc/4SN3-S85G].

a metro station in Minsk were posted on YouTube claiming to be current attacks on an airport in Brussels.[64] The video clips were not deepfakes, but were slightly doctored so that they were in black and white and flipped horizontally.[65] The videos quickly spread online and were picked up by reputable news organizations resulting in public confusion as to whether the airport in Brussels was actually attacked.[66]

Artificially constructed deepfakes could potentially create more confusion if used in a similar way as the fake Brussels airport video.[67] The deepfake could not be easily traced back to an original source because it is not one image, but millions of images used to create realistic impersonations or alterations.[68] Fake police cams, public surveillance videos, or mobile recordings could end up being used in court.[69] Over time, as the technology becomes more sophisticated, it will become more difficult for an internet user to determine what is real and what is fake.[70]

### *2. Possible Benefits in Using Deepfake Technology*

Deepfakes could benefit society as a tool in education, art, and autonomy.[71] Deepfakes could also provide a therapeutic use. [72] For example, soldiers suffering from post-traumatic stress disorder could video-conference with doctors using deepfake technology.[73] An

64. *See* Jasper Jackson, *Fake Brussels YouTube Videos Prove Ease of Digital Disinformation*, GUARDIAN (Mar. 23, 2016), www.theguardian.com/media/2016/mar/23/fake-youtube-videos-brussels-attacks-facebook-twitter [perma.cc/92YF-CA7U] (reporting on the manipulated footage that was intended to mislead the public about attack in Brussels in 2016); *see also* Brayne et al., *Visual Data and the Law*, 43 LAW & SOC. INQUIRY 1149, 1156 (2018) (noting the compelling power of visual evidence when compared with verbal accounts).

65. Jackson, *supra* note 64.

66. *Id*.

67. Chesney & Citron, *supra* note 22 (emerging deepfake technology will be more difficult to debunk than doctored images).

68. *Id.* (presenting a deepfake example of former President Barack Obama, "for whom plentiful video footage was available to train the network[]," making it show him saying things that he did not say).

69. *See* Catherine F. Brooks, *Faked Video Will Complicate Justice by Twitter Mob*, WIRED (June 18, 2018), www.wired.com/story/faked-video-could-end-justice-by-twitter-mob/ [perma.cc/4QT7-36Y2] (reporting on the accessibility of creating fake videos and their influence on the public).

70. *Id.*

71. Chesney & Citron, *supra* note 22 (providing a thorough analysis of how deepfakes can be beneficial in the arts, education, and individual autonomy).

72. Damon Beres & Marcus Gilmer, *A Guide To 'Deepfakes,' The Internet's Latest Moral Crisis*, MASHABLE (Feb. 2, 2018), www.mashable.com/2018/02/02/what-are-deepfakes [perma.cc/5MUC-GSR4] (illustrating an example of therapeutic use by Dr. Louis-Philippe Morency, director of the MultiComp Lab at Carnegie Mellon University).

73. *Id.*

individual could face-swap with a generic model without sacrificing the ability to convey his or her emotions.[74] In theory, this would encourage people to get treatment who might otherwise be deterred by a perceived stigma, and the quality of their treatment would not suffer due to a doctor being unable to read their facial cues.[75]

Deepfakes could also benefit prospective employees by removing gender and racial bias when being interviewed via video.[76] Patients that struggle with their body image could use deepfake technology to help them see a version of their body that makes them more comfortable.[77] Deepfake-related technology could also be used to dub celebrities in other languages to help bring awareness to issues.[78]

However, for society to benefit from deepfake technology, people must allow researchers to gather data about them.[79]

## *B. Section 230 of the Communication Decency Act 47 USCS § 230*

Twenty-six words created the internet in 1996.[80] These words read, "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."[81] These twenty-six words under subsection c of Section 230 of the Communication Decency Act were set up to address the internet and the arising issues. Their result was broad immunity for websites and internet service providers.[82]

### *1. Application of Section 230*

Section 230 was a response to a 1995 New York State court case known as *Stratton Oakmont v. Prodigy*, where the court ruled

74. *Id.*
75. *Id.*
76. *Id.*
77. *See Deepfakes: What are they and why would I make one, supra* note 39. (explaining how researchers are advancing artificial intelligence and how deepfakes are evolving.)
78. *David Beckham's 'Deep Fake' Malaria Awareness Video*, REUTERS (Apr. 10, 2019), www.reuters.com/video/watch/idOVA9QVUY3 [perma.cc/XCK2-2ZEA] (presenting video of David Beckham delivering his malaria awareness message in nine languages helping launch a global appeal to end malaria).
79. *See generally* Beres & Gilmer, *supra* note 72 (highlighting Dr. Morency's need for more data to use deepfake technology in a positive way).
80. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 7 (2019) (examining Section 230 of the Communication Decency Act).
81. 47 U.S.C.§ 230(c)(1) (2021).
82. KOSSEFF, *supra* note 80 at 8.

against defendant Prodigy.[83] Prodigy was an online service that "exercised editorial control over the content of messages posted on its computer bulletin boards."[84] Prodigy was moderating some user content on its site, but did not delete content that defamed the plaintiff.[85] Because they were moderating some content, the court ruled in favor of the plaintiff stating that "Prodigy's conscious choice, to gain the benefits of editorial control, has opened it up to a greater liability."[86] If Prodigy did not moderate any content, the court would have considered them the electronic equivalent of a newsstand or book store and ruled in their favor.[87] This ruling alerted online platforms that they could reduce their liability by not moderating content and allowing users to post whatever content they wished.[88] However, Congress wanted the online platforms to be able to moderate content, and thus, Section 230 was formed.[89]

Section 230 allows users to upload videos to YouTube, post reviews on Amazon or Yelp, sell an item on Craigslist, or simply communicate with people all over the world via Facebook and Twitter.[90] CDA's goal is to promote the continued development of the internet and technologies that utilize an internet user's access to information.[91] CDA's goal is accomplished by shielding online intermediaries from potential liability for their user's actions.[92] As internet access continues to grow, it becomes more difficult for a company to stop harmful content from showing up on their website.[93] The European nations, Canada, Japan, and many other countries do not have a similar statute that protects online

---

83. *Id.*

84. Stratton Oakmont Inc. v. Prodigy Servs. Co., INDEX No. 31063/94, 1995 N.Y. Misc. LEXIS 229, at *3 (Sup. Ct. May 24, 1995).

85. KOSSEFF, *supra* note 80.

86. *Stratton Oakmont*, 1995 N.Y. Misc. LEXIS 229, at *13.

87. *See* Adi Robertson, *Why the Internet's Most Important Law Exists and How People Are Still Getting It Wrong,* VERGE (Jun. 21, 2019), www.theverge.com/2019/6/21/18700605/section-230-internet-law-twenty-six-words-that-created-the-internet-jeff-kosseff-interview [perma.cc/M5ZD-D4X5] (interviewing Jeff Kosseff about Section 230).

88. *Id.*

89. *Id. See also* H.R. REP. NO. 104-223 at 3 (providing Representative Cox's plan for Section 230 to protect content providers from liability).

90. *See generally Section 230 of the Communications Decency Act*, ELEC. FRONTIER FOUNDATION, www.eff.org/issues/cda230 [perma.cc/4E6H-TVB5] (last visited Feb.7, 2021) [hereinafter EFF *Section 230*] (providing an overview of Section 230 of the CDA and its effect on the internet)

91. *See* 42 U.S.C.§ 230(b) (2021) (providing policy of the law).

92. *See* 42 U.S.C.§ 230(c)(1) (2021) (stating that "[n]o provider or user of an interactive computer service shall be treated as the publisher . . ."). *See also Section 230, supra* note 90 (providing an example of a liability shield for blogger who host content on their blogs).

93. *See* EFF *Section* 230, *supra* note 90 (illustrating that Facebook "has more than one billion users, and YouTube users upload 100 hours of video every minute").

intermediaries hosting controversial or political speech.[94] "No matter how vile or damaging" a user's comments, pictures, or videos are, the website hosting the vile content would not be liable with few exceptions.[95] However, federal criminal law and intellectual property law still apply.[96] Only recently has Section 230 begun to face backlash as well as restrictions, which have had consequences of their own.

### *2. Exceptions to Section 230*

Section 230 does not apply to federal criminal law, intellectual property law, electronic communications privacy law, or sex trafficking.[97] Online service providers enjoy a broad sweeping immunity; however, this immunity has been chipped away with respect to sex workers.[98] Since Section 230 has been enacted, there has only been one amendment.[99] This amendment was The Fight Online Sex Trafficking Act and the Stop Enabling Sex Traffickers Act, or FOSTA-SESTA, which was signed into law in 2018. FOSTA-SESTA removed Section 230's immunity for services that "promote and facilitate prostitution."[100] The amendment led to many websites self-censoring.[101] Further, the amendment has made the lives of sex workers more dangerous and more difficult for police investigators to find sex traffickers.[102] Before FOSTA-SESTA, police investigating sex trafficking were able to set up stings by using Backpage.com, an advertising website similar to craigslist, but which became notorious for buying and selling sex.[103] After FOSTA-SESTA, the police have been unable to set up successful stings to

94. *Id.*

95. KOSSEFF, *supra* note 80 at 8 (describing the reach of Section 230's liability shield).

96. *See* 42 U.S.C.§ 230(e)(3) (2021). *See also* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans 230 Immunity*, 86 FORDHAM L. REV. 401, 404 (2017) (explaining that "Section 230(e)(3) preempts contrary state laws but does not 'prevent any State from enforcing any State law that is consistent with this section.' Federal criminal law, intellectual property law, and the Electronic Communications Privacy Act are not covered by the immunity provision" (citing § 230(e)(3)).

97. 47 U.S.C.§ 230(e) (2021).

98. Matt Laslo, *The Fight Over Section 230—and the Internet as We Know It,* WIRED (Aug. 13, 2019), www.wired.com/story/fight-over-section-230-internet-as-we-know-it [perma.cc/35U6-45YH].

99. *Id.*

100. *Id.*

101. *Id.*

102. *See* Karol Markowicz, *Congress' Awful Anti-Sex-Trafficking Law Has Only Put Sex Workers In Danger And Wasted Taxpayer Money,* BUS. INSIDER (Jul. 14, 2019), www.businessinsider.com/fosta-sesta-anti-sex-trafficking-law-has-been-failure-opinion-2019-7 [perma.cc/6H72-KV5S] (examining the effects of FOSTA-SESTA 15 months after the bill was signed into law).

103. *Id.*

catch sex traffickers, so instead they arrest sex workers.[104] Furthermore, the amendment has resulted in more sex workers on the street because they can no longer directly connect to their clients through the internet.[105] Online sex work made it safer for sex workers because they could vet their clients by communicating with other sex workers online.[106] Now, without the internet as a buffer, they are being put back on the streets.[107]

Section 230 has created the internet as we know it today. Section 230 has protected the websites and online service providers from being sued for acting as a message board or town hall.[108] The immunity shield has allowed social media and video sharing sites to flourish, encouraging people to speak to each other and make new things.[109] However, some websites have taken advantage of Section 230's immunity power and have prospered.[110]

## III. ANALYSIS

Part A of this section will discuss the broad immunity of Section 230 of the CDA and its current effect on the internet, including deepfakes. Part B will analyze how Section 230 affects individuals and two ways how false information quickly circulates and becomes viral on the internet. Part C will explore content virality. Finally, Part D will discuss the crossroad between Section 230 and deepfakes.

### *A. Legal Effects of Section 230*

Section 230 of the CDA provides a liability shield for online content platforms, such as Google, YouTube, and Facebook.[111] Essentially, the liability shield makes it difficult for someone to sue an online platform for hosting harmful content created or developed

104. *Id.*

105. *See* Lux Alptraum, *The Internet Made Sex Work Safer. Now Congress Has Forced It Back into the Shadows*, VERGE (May 1, 2018), www.theverge.com/2018/5/1/17306486/sex-work-online-fosta-backpage-communications-decency-act [perma.cc/A8YA-3KZH] (noting the impact on sex workers as a result of FOSTA).

106. *Id.*

107. *Id.*

108. *Id.*

109. *Id.*

110. *Id.*

111. 47 U.S.C.§ 230 (2021). *See* Daisuke Wakabayashi, *Legal Shield for Social Media Is Targeted by Lawmakers* N.Y. TIMES (May 28, 2020), www.nytimes.com/2020/05/28/business/section-230-internet-speech.html [perma.cc/9QRF-6RQL] (examining Section 230's liability protection and noting that on December 9, 2020, "chief executives of Google, Facebook and Twitter testified before a Senate committee and delivered a full-throated defense of speech on their platforms and supporting Section 230").

by users.[112] There are two ways Section 230 protects online content platforms.[113] The first is through Section 230(c)(1) which provides that "[n]o 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."[114] The second way Section 230 protects online content providers is through Section 230 (c)(2) which protects the content provider from civil liability as long as their action is in good faith.[115]

### *1. Protection of the Content Provider*

This first protection provides that the content provider cannot be sued for harmful or problematic content that was published by another user or platform.[116] For example, Adam, a user, publicly posts a defamatory or otherwise actionable comment about his employer on Facebook. The employer would be unable to sue Facebook for allowing Adam's comment to be posted.[117] "As courts have interpreted Section 230, online platforms enjoy immunity from liability for user-generated content even if they deliberately encourage the posting of that content."[118]

### *2. Protection from Civil Liability*

The second way Section 230 protects online content providers is through Section 230(c)(2).[119] The content provider is protected from civil liability if it, in good faith, voluntarily restricts[120] or enables[121] access to offensive content that would be considered

112. 47 U.S.C.§ 230 (2021).
113. 47 U.S.C.§ 230(c)(1) (2021).
114. *Id.*
115. 47 U.S.C.§ 230(c)(2) (2021).
116. *Id.*
117. *See* Force v. Facebook, Inc., 934 F.3d 53 (2d Cir. 2019) (holding that Facebook was not liable for hosting terrorist content). "Facebook falls within the heartland of what it means to be the "publisher" of information under Section 230(c)(1)." *Id.* at 65.
118. Chesney & Citron, *supra* note 22.
119. 47 U.S.C.§ 230(c)(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). *Id.*

120. 47 U.S.C.§ 230(c)(2)(A).
121. 47 U.S.C.§ 230(c)(2)(B).

"obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."[122] For example, in *E360insight, LLC v. Comcast Corp.*, an internet service provider, Comcast, filtered out unsolicited and bulk e-mail messages sent by E360insight's, a marketer, to Comcast subscribers.[123] E360insight sued and Comcast filed a motion for judgment on the pleadings arguing that the CDA protects Comcast from all of E360insight's claims.[124] The court ruled in favor of Comcast because Comcast's voluntary actions were in good faith and therefore had immunity under Section 230(c)(2).[125] The duality of allowing a content provider to both restrict and enable offensive content was intended to "remove the disincentive to self-regulation that liability otherwise might produce."[126]

### *3. Purpose of Section 230*

To understand why Section 230 provides such broad immunity to online content providers, one must look back to 1996 when the internet was in its early stages.[127] Section 230 was passed by Congress because the "First Amendment did not adequately protect large online platforms that processed vast amounts of third-party content."[128] In the internet's early days, content providers were deterred from moderating user posts because First Amendment immunity did not apply to all distributors.[129] If the content provider moderated any content on their website, such as deleting offensive content, then the provider risked being liable for any content posted on their website.[130] Republican Congressman Chris Cox, one of the two authors of CDA, first noticed this First Amendment problem after the 1995 New York state ruling against Prodigy, the largest online service provider at that time.[131] Cox saw this ruling as way to stifle the internet and "punish[] [companies] for trying to keep things clean."[132] The ruling in Prodigy was a threat to the growth of

122. 47 U.S.C.§ 230(c)(2).

123. E360insight, LLC v. Comcast Corp., 546 F. Supp. 2d 605, 606 (N.D. Ill. 2008).

124. *Id.* at 607.

125. *Id.* at 609.

126. Chesney & Citron, *supra* note 22 (citing 47 U.S.C. § 230(c)(2)) (suggesting that an online platform would refrain from content filtering if it were liable for the content posted by users).

127. *Id.*

128. KOSSEFF, *supra* note 80 at16.

129. *Id.* at 16-17.

130. *Id.*

131. *Id.* at 8.

132. Alina Selyukh, *Section 230: A Key Legal Shield for Facebook, Google Is About to Change,* NPR MORNING EDITION (Mar. 21, 2018), www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-

the internet and could potentially harm consumers.[133] Computer companies would be confronted with legal liability if they made any effort to monitor anything posted on their site, which would result in the company discontinuing all monitoring or providing any communication services.[134] Section 230 was enacted in order to protect the content providers so that they could moderate offensive content but not be held liable if they did not moderate all offensive content.[135]

## *B. Rational of Section 230's Reach*

Since Section 230's enactment in 1996, its immunity provision has continued to broaden, and courts have adhered to Congress's "policy choice . . . not to deter harmful online speech through the separate route of imposing tort liability on companies that serve as intermediaries for other parties' potentially injurious messages."[136]

### *1. Who Section 230 Affects*

This immunity from liability has resulted in the "purveying [of] systematic disinformation and falsehoods (state-sponsored or otherwise)."[137] The U.S. Intelligence Community[138] confirmed that the 2016 U.S. presidential election faced disinformation threats online from Russian state actors in order to "undermine the U.S.-led liberal democratic order."[139] The Russian campaign helped

facebook-google-is-about-to-change [perma.cc/5TDU-JZRW] (providing background on Section 230 and highlighting representative Chris Cox's vision of Section 230).

133. KOSSEFF, *supra* note 80 at 70-71. (quoting R. Hayes Johnson Jr., *Defamation in Cyberspace: A Court Takes a Wrong Turn on the Information Superhighway in Stratton Oakmont, Inc. v. Prodigy Services Co.,* 49 ARK. L. REV. 589 (1996)).

134. *Id.*

135. *See* KOSSEFF, *supra* note 80 at 8 (noting that Representatives Cox and Wyden hoped the ability to moderate would "encourage the companies to… adopt basic conduct codes and delete material that the companies believe [were] inappropriate").

136. Zeran v. Am. Online, Inc., 129 F.3d 327, 330–31 (4th Cir. 1997).

137. Chesney & Citron, *supra* note 22.

138. *About the Committee*, U.S. SENATE SELECT COMMITTEE ON INTELLIGENCE, www.intelligence.senate.gov/about [perma.cc/5GMK-4GLP] (last visited Mar. 14, 2021) (explaining that the Committee is made up of 15 U.S. Senators: eight from the majority party and seven from the minority and, among other things, is tasked to "oversee and make continuing studies of the intelligence activities and programs of the United States Government").

139. NAT'L INTELLIGENCE COUNCIL, OFFICE OF THE DIR. OF NAT'L INTELLIGENCE, ICA 2017-01D, ASSESSING RUSSIAN ACTIVITIES AND INTENTIONS IN RECENT U.S. ELECTIONS (2017). ("This report includes an analytic assessment drafted and coordinated among The Central Intelligence Agency (CIA), The Federal Bureau of Investigation (FBI), and The National

successfully spread a conspiracy theory known as "Pizzagate" against the Democratic nominee Hilary Clinton.[140] Rumors of Clinton being involved in a child sex ring and murdering children spread on Twitter and Youtube, with the help of Russian-linked Twitter accounts.[141] The disinformation resulted in an armed man searching for what he believed to be Clinton's underground vaults containing a child sex ring at a neighborhood pizza joint.[142]

The courts have also extended Section 230's immunity so that it applies when the content provider: (1) knowingly violates the law by republishing content;[143] (2) encourages submissions of illegal or personally harmful content while protecting the identities of the people who post the harmful content;[144] (3) deliberately restructure a website interface to ensure that illegal content could not be traced back to the user;[145] and (4) sells dangerous products.[146]

In *Jones v. Dirty World Enter. Recordings, LLC*, Plaintiff Sarah Jones, a teacher and cheerleader for the National Football League's Cincinnati Bengals, sued Dirty World, which hosts a website that solicits "dirt" about local nonpublic figures by anonymous users.[147] The users posted offensive content towards Jones which humiliated her, allegedly undermining her position as an educator, her

---

Security Agency (NSA), which draws on intelligence information collected and disseminated by those three agencies.").

140. Marc Fisher et al., *Pizzagate: From Rumor, To Hashtag, To Gunfire in D.C.*, WASH. POST (Dec. 6, 2016), www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html [perma.cc/K6LP-3THT] (detailing the timeline of a fake story that resulted in a shooting at local D.C. pizza shop).

141. *See* Amanda Robb, *Pizzagate: Anatomy of a Fake News Scandal,* ROLLING STONE (Nov. 16, 2017), www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877 [perma.cc/VEG8-GLYB] (suggesting that Russian operatives played a part in the fake story and fanned the flames of the conspiracy).

142. *Id.* Edgar Madison Welch was armed with an AR-15 semiautomatic rifle, .38 handgun, and a folding knife when he entered the pizza restaurant. *Id.* He had told his friends that the "raid" on a "pedo ring" might require them to "sacrifice the lives of a few for the lives of many." *Id.*

143. Shiamili v. Real Est. Group of N.Y., 952 N.E.2d 1011 (N.Y. 2011) (holding that defendant content providers were shielded from liability after they republished allegedly defamatory comments on their website about their real estate competitor).

144. *See* Jones v. Dirty World Enter. Recordings, LLC, 755 F.3d 398 (6th Cir. 2014) (holding that defendants who ran a message board type website were not liable for the user uploaded content that was harmful to the plaintiff).

145. Doe v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016) (arguing that defendants tailored their advertising website "to make sex trafficking easier").

146. *See* Hinton v. Amazon, 72 F. Supp. 3d 685, 687 (S.D. Miss. 2014) (holding that eBay, an interactive computer service under the CDA, was immune from claims alleging it improperly permitted a retailer to advertise and sell recalled hunting equipment because the advertisements and product information published on its site originate from third-party retailers).

147. *Jones*, 755 F.3d at 403.

membership as a cheerleader, and her personal life.[148] The Sixth Circuit ruled in favor of Dirty World finding that the CDA bars Jones's claims "[b]ecause (1) the defendants are interactive service providers, (2) the statements at issue were provided by another information content provider, and (3) Jones's claim seeks to treat the defendants as a publisher or speaker of those statements."[149]

### *2. FOSTA Exception*

In *Doe v. Backpage.com*, the First Circuit permitted the extension of Section 230 by allowing a website to deliberately restructure its website interface to ensure that illegal content could not be tracked to the user who posted it.[150] Backpage.com provides online classified advertising similar to the website Craigslist.[151] Users of the site can post in different categories, such as "ads for clothing, event tickets, furniture, and other products."[152] The category at issue here was a subcategory titled "Escorts."[153] The three women who sued were all minors at the time of their victimization.[154] They allege that Backpage.com, in an effort to maximize their profits, deliberately structured its website to facilitate sex trafficking.[155] The three women also allege that "Backpage.com selectively removed certain postings made in the 'Escorts' section (such as postings made by victim support organizations and law enforcement 'sting' advertisements) and tailored its posting requirements to make sex trafficking easier"[156]

Backpage.com continued to avoid liability with Section 230's protection.[157] Public attention then grew after a documentary was released about Backpage.com's connection to sex trafficking.[158] In 2018, Congress enacted the Allow States and Victims to Fight Online Sex Trafficking Act (known as "FOSTA") to prevent websites from facilitating sex trafficking.[159] FOSTA chips away at Section 230's immunity defense when the website "knowingly assist[s],

148. *Id.* at 405.
149. *Id.* at 417.
150. *Doe,* 17 F.3d at 16.
151. *Id.*
152. KOSSEFF, *supra* note 80 at 300.
153. *Doe,* 17 F.3d at 16. (noting that the sex trafficking posts occurred in the "Escorts" section).
154. *Id.*
155. *Id.*
156. *Id.*
157. *Id.*
158. KOSSEFF, *supra* note 80 at 314 (referring to the documentary titled I Am Jane Doe). Oscar-nominee Jessica Chastain narrated the documentary. Id. It "told the tragic stories of sex trafficking victims who had sued Backpage but had their cases dismissed because of Section 230." Id.
159. Allow States and Victims to Fight Online Sex Trafficking Act of 2017, 115 P.L. 164, 132 Stat. 1253, H.R. Con. Res. 1865, 115th Cong. (2018) (enacted).

support[s], or facilitate[s] sex trafficking.[160]

The consequences of FOSTA are mostly felt by marginalized communities and groups, "especially organizations that provide support and services to victims of trafficking and child abuse, sex workers, and groups and individuals promoting sexual freedom."[161] The law's goal was to protect and fight sex trafficking, but instead it has made the lives of sex workers considerably more dangerous.[162] These sex workers have now gone offline in order to escape FOSTA liability because websites have removed their safe space to vet out their clients.[163] Now, sex workers are back on the streets facing sexual abuse and physical harm.[164]

## *C. Content Virality*

The content posted online can quickly circulate and become viral because of cognitive bias[165] and evolving algorithmic

160. *See generally Id.* at 1253. (noting that Section 230 of the CDA "was never intended to provide legal protection" for prostitution or sex trafficking).

161. Karen Gullo & David Greene, *With FOSTA Already Leading to Censorship, Plaintiffs Are Seeking Reinstatement of Their Lawsuit Challenging the Law's Constitutionality*, ELEC. FRONTIER FOUNDATION (Mar. 1, 2019), www.eff.org/deeplinks/2019/02/fosta-already-leading-censorship-we-are-seeking-reinstatement-our-lawsuit [perma.cc/8752-3622].

162. Markowicz, *supra* note 102 (noting that "[s]ex workers can no longer share information or warn each other away from violent clients").

163. *See* Gullo & Greene, *supra* note 161 (discussing the consequences of FOSTA's enactment).

164. Alptraum, *supra* note 105 (examining the effects of FOSTA on sex workers in California's Bay Area).

> Johanna Breyer, interim executive director and co-founder of the Saint James Infirmary, a health clinic that supports sex workers in California's Bay Area, told [the Verge] that in the weekend following FOSTA, the infirmary's mobile van outreach saw a dramatic increase of street-based sex workers in the Mission District. Breyer estimated that there were about double or triple the usual number of workers seeking assistance. *Id.*

165. Paul Slovic et al., *The Affect Heuristic*, 177 EUR. J. OPERATIONAL RES. 1333, 1352 (2007) (describing how people let their emotions dictate their beliefs about the world).

practices.[166] The information cascade dynamic[167] and filter bubbles[168] are two phenomena that help create content virality on the internet.[169]

### *1. Informational Cascade Dynamic*

The internet and information are closely connected. Often an individual goes on the internet for information whether it is for news, to see what their friends are doing, to look something up for a school paper, or to find the right paint color for their kitchen. However, in order to not feel overwhelmed when searching the internet, a person will rely on their available knowledge and the choices that others have made.[170] A person browsing Amazon (searcher) looking to purchase a product may refer to the product's number of positive reviews as a way to make their purchase.[171] The searcher relies on other person's information and may ignore any available information they already have of the product.[172] The searcher who relied on other users then purchases the positively reviewed product, then circulates what the other reviewers have already said, and the credibility of the product continues to increase. [173]

Information cascade can be useful when trying to quickly find

---

166. *See* Kirsten Grind et al., *How Google Interferes with Its Search Algorithms and Changes Your Results*, WALL ST. J. (Nov. 15, 2019), www.wsj.com/articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753 [perma.cc/U39U-A84V] (discussing Google's advancing algorithms).

> Google made more than 3,200 changes to its algorithms in 2018, up from more than 2,400 in 2017 and from about 500 in 2010, according to Google and a person familiar with the matter. Google said 15% of queries today are for words, or combinations of words, that the company has never seen before, putting more demands on engineers to make sure the algorithms deliver useful results. *Id.*

167. *See* Wenjing Duan et al., *Informational Cascades and Software Adoption on the Internet: An Empirical Investigation*, 43 MIS QUARTERLY 23 (2008), ssrn.com/abstract=1103165 [perma.cc/9AW5-VS44] (explaining that informational cascades occur when an online user follows another user's actions or decisions without regard to his own information).

168. *Filter Bubble*, TECHNOPEDIA, www.techopedia.com/definition/28556/filter-bubble [perma.cc/PM9M-77SF] (last updated May 17, 2008) (explaining that "[a] filter bubble is the intellectual isolation that can occur when websites make use of algorithms to selectively assume the information a user would want to see, and then give information to the user according to this assumption").

169. Chesney & Citron, *supra* note 22.

170. *See* Duan, *supra* note 167, at 11 (analyzing informational cascades theory to explain online shoppers trusting other online shopper's choices).

171. *Id.* at 17.

172. *Id.* at 16.

173. *Id.* at 17-18, 33. (finding that "individuals are remarkably influenced by the information inferred from others' behavior").

a good restaurant or a reliable product, but can quickly lead to problems when the information is both negative and false.[174] A person is more likely to assign more value, importance, and weight to negative information because, for evolutionary reasons, the person must act quicker to negative events.[175] The issue is when the negative information is false, because the individual will assign as much value and trust to the false negative information as if it were true negative information.[176] M.I.T. researchers examined verified true and false news stories that spread on Twitter from 2006 to 2017.[177] They looked at about 126,000 stories that were tweeted by about 3 million people over 4.5 million times.[178] The researchers found that the false news stories, in all categories of information, traveled faster, farther, and deeper than true stories.[179] False political news was more viral than false news about terrorism, natural disasters, science, urban legends, or financial information.[180] Falsehoods were 70% more likely to be shared on Twitter and the top 1% of false stories were shared between 1,000 and 100,000 people.[181] However, true stories were rarely shared by more than 1,000 people and took six times longer for the truth to reach 1,500 people.[182] The researchers also examined whether bots spread the false information faster but found that the speed of the virality was due to people retweeting, not bots.[183] The study also suggested that people are more likely to share novel and negative information that is inspired by fear, disgust, and surprise rather than true stories that are inspired by anticipation, sadness, joy, and trust.[184]

### *2. Filter Bubbles*

The second phenomena that creates content virality are filter

174. *See* Chesney & Citron, *supra* note 22 (sharing positive information will also cascade, such as recent social movements like Black Lives Matter and the Parkland High School students).

175. *See* Felicia Pratto & Oliver P. John, *Automatic Vigilance: The Attention-Grabbing Power of Negative Social Information.,* 61 J. PERSONALITY & SOCIAL PSYCHOL. 380 (1991) (hypothesizing that people are more likely to direct their attention to "undesirable social stimuli" over "desirable social stimuli").

176. *See* Vosoughi, *supra* note 24. (suggesting that verified false news often produces stronger emotional responses than true news resulting in the faster spread of the false news).

177. *Id.*

178. *Id.*

179. *Id.*

180. *Id.*

181. *Id.*

182. *Id.*

183. *Id.*

184. *Id.*

bubbles.[185] A filter bubble is when a person shares and searches for information that confirms their preexisting beliefs.[186] There are also algorithms that put users in a filter bubble by offering news based on what the algorithm thinks the user and the user's friends like.[187] The fear is that filter bubbles will keep individuals in a like-minded community that provides inaccurate information.[188]

### *D. Section 230 and Deepfakes Crossroad*

The courts continue to rule in favor of the content provider and provide a "super immunity" that "prevents the civil liability system from incentivizing the best-positioned entities to take action against the most harmful content."[189] It is unlikely that an individual harassed by deepfakes would be able to recover damages from the content provider.[190] The court in *Doe* had concluded that "Congress did not sound an uncertain trumpet when it enacted the CDA, and it chose to grant broad protections to internet publishers."[191] The super immunity of CDA is now far more expansive than Representative Cox, one the drafters of CDA, would agree with.[192]

## IV. PROPOSAL

Section 230 of the CDA has created the internet as we know it today but also continues to shape our society. The fast flow of an enormous amount of information created within the internet allows individuals to freely discuss anything with anyone across the world or down the street. The internet affords a place for a person to learn how to direct a movie by a famous director or write a book by an accomplished author.[193] On the other hand, it also enables the

185. Filter Bubble, *supra* note 168.

186. *Id.*

187. *See* Kevin J. Delaney, *Filter Bubbles Are A Serious Problem With News, Says Bill Gates*, QUARTZ (Feb. 21, 2017), qz.com/913114/bill-gates-says-filter-bubbles-are-a-serious-problem-with-news [perma.cc/9HN6-2K3H] (discussing Bill Gates' opinion on filter bubbles and stating that education can overcome their effect). *See* Jon Keegan, *Blue Feed, Red Feed*, WALL ST. J., graphi cs.wsj.com/blue-feed-red-feed/#/president-trump [perma.cc/B3EW-FPUC] (last updated Aug. 19, 2019) (demonstrating the difference between a conservative and a liberal's Facebook feed).

188. Delaney, *supra* note 187.

189. Chesney & Citron, *supra* note 22 at 1798.

190. *Id.*

191. *Doe*, 17 F.3d at 29.

192. *See* Selyukh, *supra* note 132 (criticizing the CDA, stating that "[t]he original purpose of this law was to help clean up the Internet, not to facilitate people doing bad things on the Internet").

193. *Martin Scorsese Teaches Filmmaking,* MASTERCLASS, www.masterclass.com/classes/martin-scorsese-teaches-filmmaking [perma.cc/6FPP-T6E9] (teaching filmmaking by Martin Scorsese and storytelling by Neil Gaiman).

spread of harmful, false information and harassment of members of marginalized communities. Removing broad immunity to content providers from Section 230 will not make a safer internet. It will likely only serve as a detriment to all users, especially those most vulnerable to being silenced.[194] However, content providing websites should not be completely shielded from hosting harmful deepfake videos or negative false information.

There have been multiple solutions to the deepfake problem and amending Section 230, but each solution seems to create another problem.[195] This section discusses three potential solutions in preventing harmful deepfakes while maintaining Section 230. Congress should amend Section 230 by including a reasonable moderation test that would require content platforms to show that their content moderation practices are reasonable. Congress should also incentivize content platforms to create new technology that slows down the spread of disinformation and remove harassing deepfakes.

## *A. Three Potential Solutions to Prevent Harmful Deepfakes*

### *1. No Legal Shield for Certain Types of Behavior*

FOSTA-SESTA was an attempt at a solution to stop sex trafficking.[196] Section 230 was amended so a company's website is now liable for knowingly hosting sex trafficking content.[197] However, the amendment has made it difficult to prosecute sex traffickers and provide aid to victims.[198] The amendment criminalizes "protected speech of those who advocate for, and

194. *Joint Hearing: Fostering a Healthier Internet to Protect Consumers Before Subcomm. on Communications and Technology and the Subcommittee on Consumer Protection and Commerce of the Committee on Energy and Commerce*, 116th Cong. (Oct. 16, 2019) (providing Statement of Corynne McSherry, Ph.D., Legal Director, Electric Frontier Foundation).

195. *E.g.*, SAFE SEX Workers Study Act, H.R. 5448, 116th Cong. (2019). (requesting a "study to assess the unintended impacts on the health and safety of people engaged in transactional sex, in connection with the enactment of [FOSTA] and the loss of interactive computer services that host information related to sexual exchange, and for other purposes.").

196. *See* Markowicz, *supra* note 102 (claiming that there was "no evidence that [FOSTA-SESTA] has made any difference").

197. Allow States and Victims to Fight Online Sex Trafficking Act of 2017, 115 P.L. 164, 132 Stat. 1253, H.R. Con. Res. 1865, 115th Cong. (2018) (enacted).

198. *See* Letter from Stephen E. Boyd, Assistant Att'y Gen., U.S. Dep't of Justice, to Robert W. Goodlatte, Chairman, House Comm. on the Judiciary (Feb. 27, 2018) (noting the Justice Department's concern that the "new language would impact prosecutions by effectively creating additional elements that prosecutors must prove at trial.").

provide resources to, adult consensual sex workers."[199] Given FOSTA-SESTA's effects, it is reasonable to expect more unintended, damaging outcomes from additional amendments, particularly when attempting to draft legislation addressing something as subjective as the ways platforms should regulate fake videos and speech.

If Section 230 were to be amended again to specifically target deepfakes, there is a risk that the law would further regulate free speech and block satirical deepfakes that are socially valuable. For example, a recent deepfake of Facebook's chief executive Mark Zuckerberg showed him bragging about abusing "stolen data" from users.[200] The video was posted to challenge Facebook's policies in moderating its content after it allowed a fake video of Nancy Pelosi to remain on the site.[201] Facebook did not initially take down the Zuckerberg fake, but in early 2020, Facebook updated their policy regarding deepfakes.[202] Social media moguls like Facebook should continue to work on fighting manipulated videos intended to mislead.

### *2. No Legal Shield for Bad Actors*

The Zuckerberg deepfake may not cause any real serious harm to anyone, but what about a deepfake that does? Free speech scholar

---

199. Aaron Mackey & Elliot Harmon, *Congress Censors the Internet, But EFF Continues to Fight FOSTA: 2018 in Review*, EFF (Dec. 29, 2018), www.eff.org/deeplinks/2018/12/congress-censors-internet-eff-continues-fight-fosta-2018-review [perma.cc/KRW2-AUTR] (examining FOSTA's effects in 2018).

200. bill_posters_uk, INSTAGRAM (Jun. 7, 2019), www.instagram.com/p/ByaVigGFP2U.

201. *See* Drew Harwell, *Faked Pelosi videos, slowed to make her appear drunk, spread across social media*, WASH. POST (May 24, 2019), www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/ [perma.cc/6TG6-8CD7] (reporting on the spread of the fake Nancy Pelosi video). The faked video seems to have been removed from most social media platforms.

202. *See* Rachel Metz and Donie O'Sullivan, *A Deepfake Video of Mark Zuckerberg Presents a New Challenge for Facebook* (Jun. 11, 2019), www.cnn.com/2019/06/11/tech/zuckerberg-deepfake [perma.cc/3PMV-T5P4] (noting Facebook's decision to keep Zuckerberg's deepfake up). However, that same week, CNN requested that Facebook take the deepfake down because of the "unauthorized use of the CBSN trademark." *Id. But see* Monika Bickert, *Enforcing Against Manipulated Media* (Jan. 6, 2020), www.about.fb.com/news/2020/01/enforcing-against-manipulated-media [perma.cc/G2Q9-AQT2] (adopting new policy to remove deepfakes from Facebook). Monika Bickert, Facebook's Vice President of Global Policy Management explained that Facebook will remove content that will "likely mislead someone into thinking a subject of the video said words that they did not actually say," or if "it is the product of artificial intelligence or machine learning that merges, replaces, or superimposes content onto a video, making it appear to be authentic." *Id.*

Geoffrey Stone proposes that websites should be denied immunity when it "deliberately leave(s) up unambiguously unlawful content that clearly creates a serious harm to others."[203] The question then would be whether the content provider knew that the video was a deepfake. The deepfake may be indecipherable from a real video and it would take time to determine whether the video real or a deepfake.[204] By the time that the provider is aware that the video was manipulated, it may have spread to other websites and confusion ensues.

Removing the legal shield from bad actors may help victims of nonconsensual deepfake pornography by forcing websites to take the harmful content down. According to Amsterdam-based Deeptrace, deepfake videos on the internet almost doubled to at least 14,678 from February 2019 to September 2019.[205] They found that 96% of the total deepfake videos from the English-speaking internet was nonconsensual, deepfake pornography.[206] At the time of Deeptrace's analysis, the websites had no intention of removing the harmful content.[207] Amending Section 230 to make these websites liable for hosting nonconsensual deepfake pornography could require websites to remove the harmful content.

### *3. The "Reasonable" Moderation Solution*

Another possible solution proposed by Danielle K. Citron, Professor of Law at the Boston University School of Law and Benjamin Wittes, senior fellow in Governance Studies at The Brookings Institution, would require content platforms to show that their content moderation practices are reasonable.[208] As long as the

---

203. Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech machine and Other myths Confounding Section 230 Reform*, 3 U. OF CHI. LEGAL F.45, 70 (2020) (citing e-mail from Geoffrey Stone, Professor of Law, Univ. of Chi., to Danielle Keats Citron (Apr. 8, 2018).

204. *See* AJ Willingham, *Is That Video Real?*, CNN (Oct. 19, 2020), www.cnn.com/interactive/2020/10/us/manipulated-media-tech-fake-news-trnd/ [perma.cc/FTQ7-GBYJ] (providing tips on spotting deepfakes).

205. Henry Ajder, Giorgio Patrini, Francesco Cavalli & Laurence Cullen, *DeepTrace: The State of Deepfakes Landscape, Threat and Impacts*, DEEPTRACE (Sept. 2019), regmedia.co.uk/2019/10/08/deepfake_report.pdf (presenting data on deepfake videos). Deeptrace is an Amsterdam-based company providing deep learning and computer vision technologies for the detection and online monitoring of synthetic media. *Id.* Their mission is to protect individuals and organizations from the damaging impacts of AI-generated synthetic media. *Id.*

206. *Id.* (finding that the "top four websites dedicated to deepfake pornography received more than 134 million views on videos targeting hundreds of female celebrities worldwide.").

207. *See id.* (presenting research on the rise of deepfake pornography by separating the websites into two categories: "dedicated deepfake pornography websites and mainstream pornography websites.").

208. *Hearing on Fostering a Healthier Internet to Protect Consumers Before the H. Comm. on Energy and Commerce*, 116th Cong. (2019) (statement of

content provider has taken steps to reasonably moderate content, then their website should be immune from liability under Section 230. For example, if a harmful deepfake was posted on a website, the courts would not look at what the company did to remove the harmful deepfake. Instead, the courts would look at how that company generally moderates all their content and whether their moderating is reasonable.[209]

However, the reasonableness standard may open new issues that could dismantle Section 230's protections.[210] Electronic Frontier Foundation's (EFF) Legal Director Corynne McSherry[211] stated that a reasonableness standard would mean more litigation risk, because the courts would be trying to figure out what counts as reasonable.[212] This standard would likely rely on the machine learning algorithms and bots to handle the website's moderating. This technology is still far from perfect. In 2007, Google launched a filtering tool to help measure a conversation's "toxicity."[213] The tool was unable to differentiate between a user talking about

---

Danielle Keats Citron, Professor, B.U. Law Sch.) www.congress.gov/116/meeting/house/110075/witnesses/HHRG-116-IF16-Wstate-CitronD-20191016.pdf [perma.cc/CDH4-DGFU]. Danielle Citron and Benjamin Wittes proposed that Section 230(c)(1) is amended so that it reads:

> No provider or user of an interactive computer service that *takes reasonable steps to address known unlawful uses of its services that create serious harm to others* 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.*
>
> *Id.*

209. *See* Sophia Cope et al., *EFF Urges Congress Not to Dismantle Section 230*, EFF (Oct. 16, 2019), www.eff.org/deeplinks/2019/10/eff-urges-congress-not-dismantle-section-230 [perma.cc/7TXR-9NQD] (encouraging user empowerment and competition, not competition with respect to Section 230).

210. *Id.*

211. *Corynne McSherry*, EFF, www.eff.org/about/staff/corynne-mcsherry [perma.cc/VJ9H-BTJ6] (last visited Mar. 14, 2021) (specializing in intellectual property, open access, and free speech issues).

212. Cope, *supra* note 209. McSherry's explains that, "As a litigator, [a reasonableness standard] is terrifying. That means a lot of litigation risk, as courts try to figure out what counts as reasonable." *Id.*

213. *See* Carline Sinders, *Medium Toxicity and Tone Are Not the Same Thing: analyzing the new Google API on toxicity, PerspectiveAPI*, MEDIUM (Feb. 23, 2017), www.medium.com/@carolinesinders/toxicity-and-tone-are-not-the-same-thing-analyzing-the-new-google-api-on-toxicity-perspectiveapi-14abe4e728b3 [perma.cc/S3RQ-DLY5] (describing a machine learning tool to rate toxicity requires teaching the machine by providing negative words. sentiments, violent threats, racism, misogyny, etc.). *See also* Elliot Harmon & Jeremy Gillula, *Stop SESTA: Whose Voices Will SESTA Silence?*, EFF (Sept. 13, 2017), www.eff.org/deeplinks/2017/09/stop-sesta-whose-voices-will-sesta-silence [perma.cc/K2BM-ZPXK] (discussing the need for human moderators even with bots performing automatic filtering).

themselves or a marginalized group.[214] For example, the tool flagged "I am a Jew" as more toxic than "I don't like Jews."[215] Artificial intelligence is still in its early stages and has difficulties in detecting the subtlety of human speech.[216]

The high cost for small companies to meet the reasonable moderation standard would be another issue with amending Section 230.[217] Small companies would be unable to keep up with the Internet goliaths like Facebook, YouTube, and Google.[218] Google's Global Head of Intellectual Property Policy Katherine Oyama testified that Google employs around 10,000 people that work on content moderation.[219] Google also released a report in 2018 that they had spent over $100 millions dollars to combat piracy online.[220] It would not be reasonable for a small company to hire that many people to moderate their content. However, small companies who want to participate in the social media realm would likely have to attempt similar moderation techniques as the larger companies like Facebook and Twitter.[221] This would make it difficult for smaller companies to participate and compete with larger companies.[222] Then with no competition, there is less opportunity for innovation.

## *B. Reasonable Moderation Test Should Be Applied by the Courts*

Congress should amend Section 230 by adding a reasonable moderation test which could then be applied by the courts. This test should provide content providers broad immunity while also putting

214. Harmon & Gillula, *supra* note 213.

215. *Id.*

216. *Id.*

217. *See* Cambridge Consultants, USE OF AI IN ONLINE CONTENT MODERATION 62-64 (2019) www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf [perma.cc/GFL6-3B4X] (providing a balance of costs for content moderation).

218. *See id.* at 64-54 (discussing the skilled labor necessary using content moderation tools but shortage of individuals who are qualified).

219. *See* Cope, *supra* note 209 (reforming Section 230 would harm small Internet companies because they would be unable to meet reasonable moderation standards).

220. Google, HOW GOOGLE FIGHTS PIRACY, 13 (2018), www.blog.google/documents/25/GO806_Google_FightsPiracy_eReader_final.pdf [perma.cc/FDT9-4CC4] (reporting on Google's anti-piracy initiatives).

221. *See* Cambridge Consultants, *supra* note 217 at 64. (detailing the major benefits of moderating content such as "positive experience" for users, "positive reputation which will further attract users and improve advertiser's perception of the platform").

222. *See id.* at 65 (noting that larger companies may be able to subsidize their own access to the large amount of computational power needed for content moderation).

the burden on them to work on ways to prevent harmful deepfakes or other harmful content being posted on their site. EFF's Legal Director Corynne McSherry argues that this test will only result in more litigation.[223] However, over time, the courts may be able set out reasonable standards by examining the content that is being moderated by the social media giants, smaller companies with less resources, and third parties hired to moderate content.[224] Facebook's 2.8 billion active users will have their own reasonable standard.[225] It would be unfair for a smaller company with a small number of active users to receive the same kind of scrutiny that Facebook should receive. The standard would allow smaller companies to compete against the social media goliaths. As the smaller companies' active users grow, their moderation test should then face further scrutiny. The courts could set up different standards by setting different brackets for active user numbers. For example, if the content provider has anywhere from zero to one million active users, they will face less scrutiny under the moderation test versus a content provider that has over a million active users. Sites like Facebook with billions of active users will face the most scrutiny. However, the government should provide grants and incentivize companies who work on innovative, less costly ways for moderating content.

One possible way to moderate deepfakes may be facial recognition. Recently, a facial recognition company, Clearview AI, claimed to "scrape" billions of pictures that were uploaded by users to Facebook, YouTube, Venmo, and millions of other websites.[226] The scrape was done without the social media companies' permission and against their policies.[227] Clearview AI began to sell their facial recognition software to federal and state law enforcement officers.[228] Indiana state officers used the recognition software via an app to quickly run a suspect's image, which led to

223. Cope, *supra* note 209.

224. *See generally*, Paul M. Barrett, Deputy Director of N.Y.U. Stern Ctr. for Bus. and Hum. Rts., WHO MODERATES THE SOCIAL MEDIA GIANTS? A CALL TO END OUTSOURCING (2020), www.issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_content_moderation_report_final_version/21?fr=sZWFmYzE0OTcyNDk [perma.cc/QX22-S274] (discussing large companies like Facebook that outsource their content moderation).

225. J. Clement, *Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2020*, STATISTA (Feb. 2, 2021), www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide [perma.cc/X39N-D3MH] (providing a current estimate of Facebook's monthly active users).

226. Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES (Jan. 18, 2020), www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html?searchResultPosition=1 [perma.cc/9WN6-H293] (reporting on Clearview AI's scraping of social media).

227. *Id.*

228. *Id.*

an arrest in twenty minutes.[229]

The creation of this facial recognition database presents all sorts of privacy issues, but it may help spot deepfakes by providing a starting point for a faceprint like a fingerprint.[230] Instead of the software being used to find suspected criminals, it could be used for quickly removing deepfakes if an individual has elected to having a faceprint. Essentially, the facial recognition software could be used by companies to moderate content being published on their site. When a deepfake of a nonconsenting individual is published on a website, the individual can request that their images in the video are uploaded to the facial recognition application, or they themselves could upload their deepfake images to the application. Then the facial recognition application could be used to remove the deepfake from all the websites where the video was uploaded.

This potential technology would present numerous privacy issues and may put the burden on the user if they are tasked with moderating content that the user finds harmful.

## V. CONCLUSION

Section 230 has been a great benefit to society in jump starting the internet. The law has created an internet where people can find and talk with friends across the world, learn and share ideas, organize, and speak out against powerful individuals. However, it is time to amend Section 230 and restrain the powerful social media goliaths and websites that permit the harmful spread of disinformation. Disinformation will continue to be a problematic, and Section 230 is helpful in combatting it. Amending Section 230 by including a reasonable moderation test will force technology companies to find ways to detect deepfakes and other fake information. Deepfakes and false information are not going anywhere. It is up to the Congress and powerful content providers to find ways to fight against disinformation but also encourage the internet's growth. The spread of important, accurate information should not be hindered because disinformation is permitted in the system. It is important that marginalized communities are able to have a voice and organize against harmful actors.

229. *Id.*

230. *See generally* Diaa Salama AbdELminaam et al., *A Deep Facial Recognition System Using Computational Intelligent Algorithms*, PLOS ONE (Dec. 3, 2020), journals.plos.org/plosone/article?id=10.1371/journal.pone.0242269 [perma.cc/UB2W-HJA8] (examining facial recognition software and proposing methods "to capture the biometric measurements of a person").