DEEPFAKES ARE TAKING OVER SOCIAL MEDIA: CAN THE LAW KEEP UP?

KAVYASRI NAGUMOTU*

Abstract

Public figures are being subject to deepfakes portraying artificially created circumstances that never actually occurred. Digital impersonation is becoming increasingly realistic and convincing. Online platforms such as Facebook, Twitter, and YouTube are fueling the rapid and widespread diffusion of user-created deepfakes. Intellectual property doctrines and recent "fake news" rules are unable to handle published deepfakes. The current Section 230 of the Communications Decency Act completely shields online platforms from the liability of publishing users' deepfakes. The online platforms, controlled by a few private companies, are essentially governing the large parts of the digital world, leading to a crisis of legitimacy.

Technological and legal solutions are necessary to deter deepfakes that are primarily used to spread misinformation. As of now, the only possible ramification for public figures is to use property or tort law to claim civil liability against the individual deepfake creators. However, civil liability cannot ameliorate the harms because plaintiffs are not always able to identify the deepfake creator, and the creators can be located beyond the effective reach of the U.S. legal process. Since online platforms play a key role in enabling the distribution of deepfakes, a more effective approach would be to shift the focus and impose liability on the platforms. A discussion of First Amendment rights will remain in the background for these claims, and the courts must decide how to balance free speech rights with the

* Kavyasri Nagumotu, J.D. Candidate at University of New Hampshire School of Law 2022. B.S. University of Rochester 2018.

societal harm that deepfakes cause. While we wait for legal mechanisms to potentially fall into place, the technology of deepfakes is only going to improve, causing chaos. We need to discuss the harms of deepfakes and possible solutions to prevent the spread of misinformation now.

I. THE CREATION OF DEEPFAKES

The early developments of deepfakes can be traced back to the 1997 paper written by Christoph Bregler, Michele Covell, and Malcolm Slaney.[1] The paper laid the groundwork to develop an innovative and unique program

[1] *See* Christopher Bregler, Michele Covell & Malcolm Slaney, *Video Rewrite: Driving Visual Speech with Audio*, ACM SIGGRAPH 1 (1997).

that essentially automated what movie studios could do.[2] The proposed Video Rewrite Program could synthesize new facial animation from the audio output.[3] The particular type of media relied on not only simple editing of a video but on specific neural networks to alter audio and video.[4] The program combined prior work that interpreted faces, synthesized audio from a test, and modeled lips in 3D space.[5] This inspired further developments in the facial recognition work with researchers making drastic improvements to make realistic convincing deepfakes in the early 2000s.[6]

A new algorithm called Active Appearance Models was created and instantly gained popularity.[7] The authors of the algorithm used a statistical model to match a shape to an image which significantly improved the tracking of the facial features.[8] Such a model relied upon a generative adversarial network (GAN) to identify the patterns in images or videos and re-create a target's face as an output.[9] With rapid improvements in the field, by 2016, creating deepfakes could successfully be accomplished using consumer-grade hardware.[10] There were already developed methods for

---

[2] *Id.* at 5–6.
[3] *Id.* at 2–5.
[4] *Id.* at 2–3.
[5] *Id.* at 2–5.
[6] *See* Timothy F. Cootes, Gareth J. Edwards & Christopher J. Taylor, *Active Appearance Models*, 23 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACH. INTELL. 681 (2001).
[7] *Id.*; *see* Iain Matthews & Simon Baker, *Active Appearance Models Revisited*, 60 INT'L J. COMPUT. VISION 135 (2004).
[8] *See* Cootes et al., *supra* note 6, at 681–82.
[9] Elizabeth Caldera, *"Reject the Evidence of Your Eyes and Ears:" Deepfakes and the Law of Virtual Replicants*, 50 SETON HALL L. REV. 177, 181 (2019).
[10] Thies et. al., *Face2Face: Real-time Face Capture and Reenactment of RGB Videos*, PROC. CVPR (2016); Suwajanakorn et. al., *Synthesizing Obama: Learning Lip Sync from Audio*, 36(4) ACM TRANSACTIONS ON GRAPHICS 95:1 (2017).

synthesizing human voice available in the market, and researchers focused on manipulating the visuals.[11]

The notable Face2Face and Synthesizing Obama projects added on to prior computing techniques to update the graphical fidelity and make videos look realistic.[12] Face2Face, a project out of the University of Munich, made real-time animation by replacing the mouth area of the targeted video with an actor's mouth.[13] Synthesizing Obama, a project from a team at the University of Washington, was the former Video Rewrite Program with better animations, textures, and expressions.[14] The graphical improvements enhanced features such as wrinkles and dimples; they manipulated colors to better match the lighting and skin tone of the targeted video.[15] Algorithms were so detailed that facial expressions such as eyebrows were precisely programmed to be in sync with the moving of the mouth.[16] The combination of the developments produced a convincing model with the ability to temporally alter both audio and video.[17] Creating a nearly photo-realistic 66-second video could be achieved in about 45 minutes on NVIDIA TitanX graphics card and an Intel Core i7-5820 processor.[18] However, these developments were primarily

[11] Adobe Creative Cloud, *#VoCo. Adobe Audio Manipulator Sneak Peak with Jordan Peele*, YOUTUBE (Nov. 4, 2016), https://www.youtube.com/watch?v=I3l4XLZ59iw [https://perma.cc/AW6L-XA8T].
[12] Thies et. al., *supra* note 10; Suwajanakorn et. al., *supra* note 10.
[13] Thies et. al., *supra* note 10.
[14] Suwajanakorn et. al., *supra* note 10.
[15] *Id.*
[16] *Id.*
[17] *See* Thies et. al., *supra* note 10; *see* Suwajanakorn et. al., *supra* note 10.
[18] Suwajanakorn et. al., *supra* note 10 at 95:8.

being researched in academia by researchers and for movies by big-budget Hollywood productions.[19]

In 2017, Reddit contributed to a massive spike in the creation of deepfakes. A now-deleted subreddit named r/deepfakes had nearly 90,000 subscribed members; it primarily featured pornographic deepfakes featuring a variety of famous actors.[20] Redditors[21] were able to place celebrities' faces on existing pornographic videos using the Face2Face algorithm found on open-source libraries like the Python library Keras.[22] After users raised privacy and consent concerns, Reddit acted by updating the content policies and posing a ban on the r/deepfakes subreddit.[23] A variety of non-pornographic deepfake subreddits have since been created gaining popularity among the Reddit community.[24]

Over the years, the technology to create deepfakes has become easier to access. There are many deepfake projects on GitHub, such as TensorFlow, available to the public that provide the software for easy deepfake

---

[19] Kevin Roose, *Here Come the Fake Videos, Too,* N.Y. TIMES (Mar. 4, 2018), https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html [https://perma.cc/M552-W2KU].

[20] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked,* MOTHERBOARD (Dec. 11, 2017), https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/RT3P-EABJ].

[21] Redditors are users of the website Reddit. *Redditor,* OXFORD LEXICO**,** https://en.oxforddictionaries.com/definition/redditor [https://perma.cc/3YAF-MQ4D].

[22] Cole, *supra* note 20.

[23] *Update on site-wide rules regarding involuntary pornography and the sexualization of minors*, REDDIT (Feb. 7, 2018), https://www.reddit.com/r/announcements/comments/7vxzrb/update_on_sitewide_rules_regarding_involuntar/ [https://perma.cc/X749-BTPU].

[24] *Deepfakes that are Safe for* Work, REDDIT, https://www.reddit.com/r/SFWdeepfakes [https://perma.cc/PY6K-MH5N].

development.[25] Today, the existence of many deepfake creation applications like Zao,[26] Faceswap,[27] and AvengeThem,[28] make it convenient for anyone with a smartphone to create deepfakes. Even law school students can create deepfakes by manipulating videos of Supreme Court Justices singing.[29]

The rise in the creation of deepfakes consequentially created a need to develop tools to detect them. Sensity, a company founded in 2018 based in Amsterdam, started researching deepfakes and dubs itself "the world's first visual threat intelligence company."[30] In their 2019 report, Sensity detected 14,679 deepfakes online and, in 2020, found that the number rose to 49,081.[31] The tends indicated that the numbers of deepfakes available online were nearly doubling every six months.[32] Even though early deepfakes involved pornographic content, the recent popular deepfakes targeted people who were popular politicians and internet celebrities.[33]

---

[25] Cole, *supra* note 20.
[26] *Zao*, APP STORE, https://apps.apple.com/cn/app/zao/id1465199127 [https://perma.cc/UNA2-W3C8].
[27] *Online Deepfake Maker*, DEEPFAKES WEB, https://faceswapweb.com/ [https://perma.cc/Y5N9-8DPG].
[28] *Add your design to iconic scenes*, GIFSHIRT, https://gifshirt.com/ [https://perma.cc/8ZS5-GFQC].
[29] Victor Gray, *"Scalia when he sees a Founding Father," See Law School Memes for Edgy T14s*, FACEBOOK (Mar. 10, 2021), https://www.facebook.com/100042763871377/videos/457986038970188/ [https://perma.cc/N86T-STF4].
[30] *Our Mission*, SENSITY, https://sensity.ai/about/.
[31] Henry Ajder, *Deepfake Threat Intelligence: a statistics snapshot from June 2020*, SENSITY (Mar. 7, 2020), https://sensity.ai/deepfake-threat-intelligence-a-statistics-snapshot-from-june-2020/.
[32] *Id.*
[33] *Id.*

In September 2019, Facebook launched a Deepfake Detection Challenge (DFDC).[34] The public contest encouraged people to develop autonomous algorithmic detection systems to identify deepfake videos.[35] Participants were given a raw dataset with 38.5 days' worth of video recorded by 3,500 actors, some of which were manipulated deepfakes.[36] Over 2,000 participants submitted multiple models each with new algorithms to detect deepfakes.[37] The winning model was able to detect 82% of the deepfakes that it was exposed to.[38]

The House Intelligence Committee on Intelligence met in 2019 to have an open hearing on "the national security threats posed by AI-enabled fake content, what can be done to detect and combat it, and what role the public sector, the private sector, and society as a whole should play to counter a potentially grim, 'post-truth' future."[39] The hearing tried to examine the profound questions raised by deepfakes about national security and democratic governance.[40] The tangible result of the hearing was the U.S. Defense Advanced Research Projects Agency (DARPA) funding a media forensics project aimed at finding ways to automatically screen for deepfakes.[41]

---

[34] *Deepfake Detection Challenge Dataset*, FACEBOOK AI (June 25, 2020), https://ai.facebook.com/datasets/dfdc/ [https://perma.cc/P9NJ-EBRR].
[35] *Id.*
[36] *Deepfake Detection Challenge Results: An open initiative to advance AI*, FACEBOOK AI (June 12, 2020), https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/ [https://perma.cc/7DMM-KE2J].
[37] *Id.*
[38] *Id.*
[39] *House Intelligence Committee To Hold Open Hearing on Deepfakes and AI*, U.S. HOUSE OF REPRESENTATIVES (June 7, 2019), https://intelligence.house.gov/news/documentsingle.aspx?DocumentID=657 [https://perma.cc/Y7GU-KD3W].
[40] *Id.*
[41] *Id.*

Policymakers continue to raise concerns that deepfakes may mislead voters and sway election outcomes. Rep. Adam Schiff, the committee chair of the House Permanent Select Committee on Intelligence, noted that deepfakes allow "malicious actors to foment chaos, division or crisis," and the videos "have the capacity to disrupt entire campaigns, including that for the presidency."[42] While experts are just now starting to search for technological solutions to detect deepfakes, the technology behind the creation continues to advance at a rapid pace. Deepfakes are here to stay. The old saying of "seeing is believing" is no longer true.

II. THE RISE OF DEEPFAKES

Russian disinformation through Twitter and Facebook posts disrupted the 2016 U.S. presidential campaign. Even with a growing public awareness regarding the dangers of deepfakes, people are generally not able to discriminate between a real video and a deepfake.[43] Further, when the public is unsure whether a video is real or a deepfake, they may begin to distrust all political video footage.[44] Conveniently, supporters of Trump suggested that a video Trump shared on his Twitter conceding the 2020 election was a deepfake.[45]

The tectonic shift in the use of online platforms to spread misinformation is starting to be acknowledged by the

---

[42] Donie O'Sullivan, *Schiff sounds alarm in Congress' first hearing on deepfake videos*, CROSSROADS TODAY (June 13, 2019), https://www.crossroadstoday.com/schiff-sounds-alarm-in-congress-first-hearing-on-deepfake-videos/ [https://perma.cc/V5DA-CAL7].

[43] *See* John Ternovski, Joshua Kalla & Peter M. Aronow*, Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discernment: Evidence from Two Experiments*, OSF PREPRINTS (Jan. 14, 2021), https://osf.io/dta97/ [https://perma.cc/GD4A-9DBM].

[44] *Id.*

[45] *Id.*

federal government. In January 2021, the U.S. Department of Justice arrested far-right figure Douglass Mackey on "charges of conspiring with others in advance of the 2016 U.S. Presidential Election to use various social media platforms to disseminate misinformation designed to deprive individuals of their constitutional right to vote."[46] In the complaint, the prosecution noted that Mackey spread false messages about voting via social media.[47] Mackey utilized a coordinated campaign consisting of tweets, memes, and other forms of media.[48] He was able to trick at least 4,900 voters to cast their ballots by phone instead of going to the polling booths.[49] Mackey played a role in the 2016 presidential campaign election, which marked a significant change in how social media has been used as a news source for the public and influential in changing online interactions.

The startling effects of social media in the political landscape can be traced back to June 16, 2015, when Trump declared his run for presidency with a 45-minute television conference.[50] Trump already had a well-known television presence through his primetime programs and business connections. In just six months, he was getting more nightly news coverage than all the Democratic candidates

---

[46] *Social Media Influencer Charged with Election Interference Stemming from Voter Disinformation Campaign*, DEPARTMENT OF JUSTICE (Jan. 27, 2021), https://www.justice.gov/opa/pr/social-media-influencer-charged-election-interference-stemming-voter-disinformation-campaign [https://perma.cc/WMW6-88VK].

[47] Complaint at 4-10, United States v. Mackey, No. 21-MJ-82(RER) (E.D.N.Y 2021), https://www.justice.gov/opa/press-release/file/1360816/download [https://perma.cc/QC3L-D3LV].

[48] *Id.* at 22–23.

[49] *Id.* at 23.

[50] Jeremy Diamon, *Donald Trump jumps in: The Donald's latest White House run is officially on*, CNN POLITICS (June 17, 2015), http://edition.cnn.com/2015/06/16/politics/donald-trump-2016-announcement-elections/ [https://perma.cc/M5FG-WHM7].

combined.[51] Trump's impressive ability to generate continually controversial comments in the public and on social media aided in networks providing extensive coverage of his campaign.

Researchers generally associate the timing at which political scandals are uncovered with how the voters view each of the candidates.[52] A candidate would prefer to have a scandalous story uncovered early in the campaign to allow for the story to break down and no longer exist in the voters' minds by the time they head to the polls.[53] However, continuous coverage of controversial actions can cause "scandal fatigue," where voters become numb to the constant media questioning of candidates' actions.[54] Scandal fatigue is common around political scandals where media continuously keeps the voters informed about developments surrounding the scandal.[55] Trump likely benefited from scandal fatigue by creating a persona in the media appearing to be unwavering in the face of constant criticism.

Television networks portray themselves as employing notable journalists to broadcast unbiased news coverage, whereas social media was able to discuss without the restrictions of journalism ethics. Social media allowed for the real-time broadcasting of information, side-stepping traditional news media. The 2016 electoral campaigns focused on creating new strategies for constantly advancing

---

[51] Brian Stelter & Ken Olshansky, *How much does Donald Trump dominate TV news coverage? This much*, CNN BUSINESS (Dec. 6, 2015), http://money.cnn.com/2015/12/06/media/donald-trump-nightly-news-coverage/ [https://perma.cc/TZA3-64KX].

[52] Dona-Gene Mitchell, *Here today, gone tomorrow? Assessing how timing and repetition of scandal information affects candidate evaluations*, 35 POLITICAL PSYCHOLOGY 679, 697 (2013).

[53] *Id.* at 697-98.

[54] *See generally* JOHN B. THOMPSON, POLITICAL SCANDAL: POWER AND VISIBILITY IN THE MEDIA AGE (2013).

[55] *See generally id*.

their candidates on social media. The announcement of Hillary Clinton running for president on Twitter set a new precedent for political outreach and campaign strategies.[56] Traditional news media was simply unable to keep up with the increased media platforms, bots, and image sharing.

There was a rapid increase in the use of bots driven by unique algorithms to copy the behavior of humans on social media and post biased messages about candidates.[57] The number of bots dedicated to promoting positive content directed toward a candidate can give a misperception that the candidate has a larger number of supporters than reality suggests. In an attempt to include the ideas and expressions from online platforms, traditional media covered inaccurate online information without careful consideration of the sources. The television media saw the high number of bots posting certain content and proceeded to endorse those messages. The information that was presented to the viewers, in turn, created issues of individual interpretation. The influence of bots can have far-reaching consequences for journalists and news media looking for material to present to their viewers.[58]

After numerous claims of Russian interference in the 2016 election, President Obama directed the U.S. Senate Intelligence Committee to assess Russia's possible involvement.[59] The Intelligence Committee concluded that Russia employed over 1,000 people to create fake accounts to spread pro-Trump and anti-Clinton rhetoric during the

---

[56] Hillary Clinton (@HillaryClinton), TWITTER (Apr. 12, 2015, 2:27 PM), https://twitter.com/hillaryclinton/status/587336319321407488?lang=en [https://perma.cc/KPB9-S6KT].

[57] *See* Alessandro Bessi & Emilio Ferrara, *Social Bots Distort the 2016 U.S. Presidential Election Online Discussion*, 21 FIRST MONDAY 1, 8–11 (2016).

[58] Panagiotis T. Metaxas & Eni Mustafaraj, *Social Media and the Elections*, 338 SCI. 472, 472–73 (2012).

[59] *See* S. Rep. No. 116-XX, at 3 (2016).

election.[60] Journalists working for news media were likely swayed by the bots employed by Russia, leading to telecasts involving manipulated algorithmic online discourses. The distorted news coverage led the public to consume false information and question media credibility.

The infamous indictment issued by Robert Mueller against Russia's Internet Research Agency and affiliates detailed the Russian influence on the 2016 election.[61] Mueller's indictment outlined how a Russian campaign spent tens of millions of dollars over several years to build a broad, sophisticated system on social media to influence American opinion.[62] The Russian bots were instrumental in creating an information environment and a narrative to amplify and promote a biased story.[63] Russians are alleged to have promoted advertisements using video, visual, memetic, and textual elements to push specific narrative themes, conspiracies, and character attacks.[64] The employed algorithms and bots were designed to look like the messages were all coming from authentic American individuals and groups.[65] It was not only one-way communication; Russians also used data-driven analysis to assess how the content was being perceived to further refine the messages to make them

---

[60] Rachel Roberts, *Russia hired 1,000 people to create anti-Clinton 'fake news' in key US states during election, Trump-Russia hearings leader reveals*, THE INDEPENDENT (Mar. 30, 2017, 5:41 PM), http://www.independent.co.uk/news/world/americas/us-politics/russian-trolls-hilary-clinton-fake-news-election-democrat-mark-warner-intelligence-committee-a7657641.html [https://perma.cc/SQ2A-YWY4].

[61] *See* Indictment, U.S. v. Internet Research Agency LLC, No. 1:18-cr-00032-DLF (D. D.C. Feb. 16, 2017), https://www.justice.gov/file/1035477/download [https://perma.cc/Q7CQ-LM6F].

[62] *Id.* at 5.

[63] *See id.* at 14–23.

[64] *See id.*

[65] *See id.* at 13–16.

more effective.[66] A key goal was to infiltrate online platforms with a particular idea to make it *mainstream* and appear more widely held than it was.[67]

The use of bots was instrumental in creating and responding to discourses on Twitter, Facebook, and Reddit. With the American population engaging in the discussions on the online platforms, the Russians were able to manually craft messages to influence and alter American behavior. In one case, a Russian soldier was able to infiltrate a social media group by pretending to be a 42-year-old American housewife.[68] He engaged with other members of the group by weighing in on political debates and sending tailored messages.[69] Another common method employed was creating fake Facebook accounts to spread stories on political issues like refugee settlement to specifically target users that were susceptible to influence.[70] In another instance, a pedophilia story circulated under the hashtag #pizzagate targeting swing voters to promote anti-Clinton sentiment.[71] The search for the story was disproportionately higher in swing districts than in districts that were already likely to vote for Trump.[72]

The viral nature of social media aided the Russians in spreading the misinformation effectively. Going "viral" is defined as quickly and widely spread or popularized especially using social media.[73] The content that is pushed

---

[66] *See id.* at 15.
[67] *See id.* at 16–24.
[68] Massimo Calabresi, *Inside Russia's Social Media War on America*, TIME (May 18, 2017, 3:48 PM), https://time.com/4783932/inside-russia-social-media-war-america/ [https://perma.cc/W25B-PX4Y].
[69] *Id.*
[70] *Id.*
[71] *Id.*
[72] *Id.*
[73] *Viral*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/viral [https://perma.cc/DRG9-BU7U] (last visited Mar. 25, 2022).

on social media has the potential of getting exponentially shared. A seamless meme can receive millions of views within a couple of hours, and just about anything can go viral. The key factor in influencing the spread was the pattern of connections that participate in the social media network. When one user sees a particular type of content on social media, they have options to scroll past it, tag their friend, or send it to their friend. The user can actively participate by distributing that content to others in their social network. Even if the user does not actively share it, social media websites track what type of content the user is interested in and share similar content with others in their network.[74] The information posts of social media are transmitted over a wide range of audiences with no critical assessment.

Individuals that were interested in political engagement felt encouraged to create their own online media news sources to generate stories that they felt should receive adequate coverage. Trump Supporters felt that traditional newscasts were misleading and started to label them as "fake news."[75] An example of a creation of a new online media was Breitbart News, focusing on right-wing news.[76] Since early in the campaign, Breitbart News focused on supporting Trump's policies and was able to accumulate a large number of views comparable to other media sources.[77] This further

---

[74] *See generally* Mason A. Porter & James P. Gleeson, *Dynamic Systems on Networks*, 4 FRONTIERS IN APPLIED DYNAMICAL SYSTEMS: REVIEWS AND TUTORIALS (2010).

[75] James Carson & Michael Cogley, *Fake news: What exactly is it – and how can you spot it?*, THE TELEGRAPH (Jan. 7, 2021), https://www.telegraph.co.uk/technology/0/fake-news-exactly-donald-trump-rise/ [https://perma.cc/KVG2-R6ZZ].

[76] Yochai Benkler et. al., *Study: Breitbart-led right-wing media ecosystem altered broader media agenda*, COLUM. JOURNALISM REV. (Mar. 3, 2017) https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php [https://perma.cc/G245-FAJB].

[77] *Id.*

gave other right-wing supporters motivation to mobilize and create additional online news platforms.[78] The number of partisan right-wing sites far exceeded the number of left-wing sites, creating a large presence of right-wing online news sources with their base of followers.[79] The availability of the numerous sources of online media allowed the public to pick and choose what to read rather than relying on storylines designed by traditional news providers.

The 2016 election campaign challenged traditional forms of news coverage, with Trump perpetuating narratives that appeared outside the traditional news media. He created his anti-establishment persona that many of his supporters took an interest in as displayed by the popularity on online platforms. Trump began to be a consistent user of the term "fake news" to express his disagreement with the information that the news sources were telecasting.[80] He would take to his, now suspended,[81] Twitter account to question the value and quality of the traditional news throughout his campaign. Social media sites were able to create a fast-paced environment that allowed for all opinions to be shared with limited censorship. Whereas traditional news media could not keep by because they were motivated by a financial factor having to appeal to a vast range of social, political, and individual interests.

---

[78] *Id.*

[79] *Id.*

[80] Jessi Hempel, *Fixing Fake News Won't Fix Journalism*, WIRED (Mar. 13, 2017, 12:00 AM), https://www.wired.com/2017/03/fixing-fake-news-wont-fix-journalism/ [https://perma.cc/7RU3-8SN7].

[81] *Permanent suspension of @realDonaldTrump*, TWITTER (Jan. 8, 2021), https://blog.twitter.com/en_us/topics/company/2020/suspension.html [https://perma.cc/BDX2-UJBU].

III. THE ROLE OF ONLINE PLATFORMS

The digital platforms controlled by a few private companies are essentially governing the large parts of the digital world, leading to a crisis of legitimacy. Most of the news and information relayed to the public is through social media. A single post can quickly reach a global audience. Within the last two decades, social media has grown to be a multi-billion-dollar sector of the global economy. Social media has an expansive definition and has become a common term in our daily lexicon. It has shaped the narrative of modern political activism. The turning point of the 2016 election led to social media becoming a tool for mass manipulation, vote suppression, and propagation of false information.

Every social media interface has three particular attributes: (1) be web-based, (2) provide a means for individuals to connect and interact with other users, and (3) provide a means for users to generate and distribute content on the platform. Social media platforms are built on collecting user data and selling it to companies who use it for advertising purposes. The exchange of information enables the social media platforms to be *free* for the users who sign up and provide their personal information.

Some of the most popular social media platforms with unrivaled influence are Facebook, Twitter, and YouTube.[82] The dominance of particular social media platforms may also be viewed as genres themselves. For example, when a new social media website or app is introduced, people initially refer to it as being "like Facebook" or "like Twitter." Facebook, Twitter, YouTube, and other platforms spread the viral news across the internet

[82] Brook Auxier & Monica Anderson, *Social Media Use in 2021*, PEW RESEARCH CENTER (April 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [https://perma.cc/XZ2T-RHDW].

like wildfires by allowing users to create and share posts, pages, groups, hashtags, videos, and channels.

Numerous studies have already shown that fake news spreads faster online because of how social media has prioritized virality.[83] In one study, MIT researchers examined 126,000 rumors spread by three million people and found that false news reached more people than accurate information.[84] The phenomenon is due to the online platform's selection of trending topics is "prioritized by complex algorithms that have been coded to sort, filter, and deliver content in a manner that is designed to maximize users' engagement."[85] Through the carefully curated algorithms, the content being shared grows exponentially regardless of whether the information is true. Online platforms lack the secondary screen to check the source and validity of the information like in traditional news sources.

There has always been a tension between free speech and suppressing content that is harmful to society.[86] However, the First Amendment only limits government actors and not private entities such as Facebook, Twitter, and YouTube.[87] The leeway provided by the Constitution allows social media platforms to formulate their own policies and methods for detecting and removing misleading or false

---

[83] *See* Peter Dizikes, *Study: On Twitter, false news travels faster than true stories*, MIT NEWS (Mar. 8, 2018), https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308 [https://perma.cc/97WL-HMKB]; *see* Samantha Bradshaw & Phillip N. Howard, *Why does Junk News Spread So Quickly Across Social Media*, KNIGHT FOUNDATION (Jan. 29, 2018) https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_White-Paper_Howard_V1_ado.pdf [https://perma.cc/QM69-VARX].

[84] Dizikes, *supra* note 83.

[85] Bradshaw & Howard, *supra* note 83, at 10.

[86] *See* Evelyn Douek, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 769-70 (2021).

[87] *See* U.S. CONST. amend. I.

information.[88] The C-suite executives and policy staff of popular social media companies have a tough role in determining how platforms should be utilized in politics balancing the free speech protections.

Ahead of the 2020 election, Twitter, Facebook, and YouTube started taking steps to secure their platforms while they faced growing criticism that they are not doing enough to prevent racism, hate speech, and misinformation. Facebook announced several initiatives to "better identify new threats, close vulnerabilities and reduce the spread of viral misinformation and fake accounts."[89] These initiatives included updating the policy on user authenticity, protecting accounts of political candidates, labeling state-controlled media, and including fact-checking labels on problematic content.[90] For the most part, Facebook adopted a hands-off policy for allowing political advertisements citing their fundamental belief in free expression.[91] The permissive policy led to a large number of companies threatening to boycott paid advertising on Facebook and Instagram, which is owned by Facebook, to show support for a movement called #StopHateForProfit.[92] Even following the heavy criticism and $60 billion market value loss, Facebook responded that it would not change its policy.[93]

---

[88] *See* Douek, *supra* not 86, at 773–74.

[89] Guy Rosen et. al., *Helping to Protect the 2020 US Elections*, FACEBOOK (Oct. 21, 2019), https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/ [https://perma.cc/ML9R-TJGT].

[90] *Id.*

[91] *See* Nick Clegg, *Facebook, Elections and Political* Speech, FACEBOOK (Sept. 24, 2019), https://about.fb.com/news/2019/09/elections-and-political-speech/ [https://perma.cc/NDR3-BJTB].

[92] *Thank You to All of the Businesses that Hit Pause on Hate*, STOP HATE FOR PROFIT, https://www.stophateforprofit.org/participating-businesses (last visited Apr. 1, 2021) [https://perma.cc/X6T4-RCYM].

[93] Tyler Sonnemaker, *Mark Zuckerberg Reportedly Said Facebook is 'Not Gonna Change' in Response to a Boycott by More Than 500 Advertisers Over the Company's Hate-SpeechPolicies*, INSIDER (July 2, 2020, 12:51 AM), https://www.businessinsider.com/zuckerberg-

Twitter took a different approach by banning all political advertisements on its platform.[94] Twitter also announced new rules addressing deepfakes and other forms of synthetic and manipulated media.[95] Starting in February 2020, Twitter limited users from sharing deepfakes that were likely to cause harm and start labeling tweets with manipulated content.[96] The labeling feature examined three criteria for the posted content: (1) misleading information which are "statements or assertions that have been confirmed to be false or misleading by subject-matter experts", (2) disputed claims which are "statements or assertions in which the accuracy, truthfulness, or credibility of the claim is contested or unknown", and (3) unverified claims which contain "information that is unconfirmed at the time it is shared".[97] For the first time in May 2020, Twitter added labels on two of Trump's tweets urging Twitter users to "get the facts" before taking the tweets at face value.[98] Later in January 2021, Twitter issued a permanent suspension of

---

facebook-not-gonna-change-due-to-advertising-boycott-report-2020-7 [https://perma.cc/GYQ8-4RJJ].

[94] *Political Content*, TWITTER BUSINESS, https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html (last visited Apr. 1, 2021) [https://perma.cc/AST5-TZRJ].

[95] Yoel Roth & Ashita Achuthan, *Building Rules in Public: Our Approach to Synthetic & Manipulated Media*, TWITTER BLOG (Feb. 4, 2020), https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html [https://perma.cc/78TW-3MRV].

[96] *Id.*

[97] Yoel Roth & Nick Pickles, *Updating Our Approach to Misleading Information*, TWITTER BLOG (May 11, 2020), https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html [https://perma.cc/UV42-Y7XT].

[98] Jason Silverstein, *Twitter flags Trump Tweet With Fact-Checking Label For First Time*, CBS NEWS (May 27, 2020, 12:53 PM), https://www.cbsnews.com/news/twitter-adds-fact-check-warning-trump-tweets/ [https://perma.cc/WU96-BGS9].

@realDonaldTrump, citing a violation of its public interest framework.[99]

YouTube went a step further and created deceptive practices policies, which banned manipulated media and videos that spread birther conspiracy theories.[100] The now-infamous deepfake of Nancy Pelosi slurring words initially drew lots of attention on YouTube but was taken down per its policy.[101] The Community Guidelines on the platform have an overarching ban for manipulated media under the spam, deceptive practices, and scam policies.[102]

Twitter and Facebook soon followed and outlined a ban of manipulated media, including deepfakes.[103] However, the effectiveness of blog posts trying to address the dangers of deepfakes is unknown. These vague guidelines and laissez-faire attitudes from social media companies cannot possibly make bad actors stop the spread of misinformation. There is also little incentive, other than widespread criticism, for social media platforms to confront these issues. Banning and policing the content of well-known users such as Trump is easy, but the spread of misinformation that users otherwise view is left unregulated. Issuing countless policy changes and drafting blog posts

---

[99] *Permanent Suspension*, *supra* note 81.

[100] Leslie Miller, *How YouTube Supports Elections*, YOUTUBE OFFICIAL BLOG (Feb. 3, 2020), https://blog.youtube/news-and-events/how-youtube-supports-elections?m=1 [https://perma.cc/3U7F-KATP].

[101] Hannah Denham, *Another Fake Video of Pelosi Goes Viral on Facebook*, WASHINGTON POST, (Aug. 3, 2020), https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/ [https://perma.cc/MP3Y-3Y2X].

[102] *Community Guidelines*, YOUTUBE, https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines (last visited Apr. 1, 2021) [https://perma.cc/39MD-NDHR].

[103] Roth & Achuthan, *supra* note 95; Monika Bickert, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020), https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ [https://perma.cc/5ZE9-AD5A].

only creates an illusion of activism with minimal direct impact. There are also other closed networks like Snapchat and TikTok that continue to gain popularity and have undetermined influence on the spread of misinformation.

During summer 2020, a new startup called Clubhouse was the latest hype to join the tech world of voice-driven social media. The invite-only, iPhone-exclusive application made huge waves, gathering celebrities such as Tiffany Haddish, Virgil Abloh, and Oprah Winfrey.[104] Clubhouse provided a platform where users could join and participate in different chat rooms on a wide range of topics they choose.[105] The platform was unique in that the conversations are audio only and disappear forever when the conversation in the chat room ends.[106]

Soon after launch, Clubhouse received backlash over their lenient harassment protocols and failure to moderate rooms appropriately.[107] In January 2021, Tiffany Haddish, Jason Lee, and Chakra Bars were accused of circulating conspiracy theories about COVID-19 on the app.[108] Due to their popularity as public figures, other users on the app doxxed a physician in the chat room who pushed back on the conspiracy theories.[109] With only a voice attached to each account, it is virtually impossible to prove the identity of the person behind the voice. Users trust the name and profile

---

[104] Eni Subair*, Drake, Oprah Winfrey and Virgil Abloh Are Fans: Here's Everything You Need to Know About Clubhouse*, VOGUE (Jan. 3, 2021), https://www.vogue.com/article/everything-you-need-to-know-about-clubhouse# [https://perma.cc/GJ3F-9FBG].
[105] *Id.*
[106] *Id.*
[107] *Id.*
[108] Yohance Kyles*, Tiffany Haddish Responds To Accusations Of Bullying Black Doctor Over COVID-19 Conspiracy Theories*, ALL HIPHOP (Jan. 15, 2021), https://allhiphop.com/news/tiffany-haddish-responds-to-accusations-of-bullying-black-doctor-over-covid-19-conspiracy-theories/ [https://perma.cc/6K4Q-F9VY].
[109] *Id.*

page linked to the voice to give credibility to the voice in the chat room.

It wasn't until six months after their initial launch that the founders, Paul Davison and Rohan Seth, issued a list of guidelines to condemn hate speech and abuse on the Clubhouse app.[110] The use of moderation and anti-abuse tools seemed like an afterthought that the founders tried to cram in after backlash. Mobile applications like Clubhouse show how difficult it would be to hold an individual accountable for their actions, and precautions need to be enforced by the online platform. Implications on the mobile applications may include a person mimicking the voice of a public figure to gain popularity and influence the behavior of others on the app. In one instance entrepreneur Sriram Krishnan changed his name on Clubhouse to Tim Cook, Apple's chief executive, intending for it to be a prank.[111] But because of Tim Cook's name recognition, more than 100 people immediately joined the room Krishnan was in.[112] Hours later, someone else impersonated Elon Musk, business magnate and father to X Æ A-Xii.[113]

Seeing the hype surrounding Clubhouse, other social medial heavyweights such as Twitter, Facebook, and Spotify launched similar audio-only products on their existing platforms.[114] This left Clubhouse struggling to obtain users

---

[110] *On Community Moderation*, CLUBHOUSE BLOG (Oct. 1, 2020), https://www.joinclubhouse.com/on-community-moderation [https://perma.cc/E29T-NDWJ].

[111] Erin Griffith & Taylor Lorenz, *The Hot New Thing in Clubby Silicon Valley? An App Called Clubhouse*, N.Y. TIMES (May 19, 2020), https://www.nytimes.com/2020/05/19/technology/clubby-silicon-valley-app-clubhouse.html [https://perma.cc/9ARH-JTZA].

[112] *Id.*

[113] *Id.*

[114] Sam Shead, *Clubhouse Co-Founder Opens Up On Growing Pains: 'It's Been Quite an 18 Months*,' CNBC (Nov. 9, 2021), https://www.cnbc.com/2021/11/09/clubhouse-co-founder-opens-up-on-

who hesitated to install another application on their phone.[115] Nevertheless, Clubhouse showcases that the level of anonymity with only using one's voice also gives people with malicious intent a greater opportunity to take advantage of audio deepfakes, whose effects are yet to be uncovered.

There is a lack of infrastructure to limit the spread of misinformation among social media startups. The amplifying power of social media causes misinformation to circulate far and wide. Too many people rely on what they assume others have reliably determined and then pass the information along. Especially when a topic is interesting and presents novel information, it grabs the attention of people and lets them disregard its authenticity. This means that the current environment of social media platforms allows for deepfakes to mature and proliferate.

IV. LEGAL ANALYSIS

The most effective way to stop the demeaning effects of deepfakes may not be through restricting those that make and post the deepfakes, but rather restricting the online platforms that amplify them. No current criminal or civil law bans the creation or posting of deepfakes.[116] A flat ban on deepfakes would not be normatively appealing or constitutionally permissible.[117] After all, deepfakes do have advantages in certain contexts. Deepfakes provide for routine modifications that improve the clarity of digital content.[118] They are a cheap and accessible method of video

growing-pains-over-last-18-months.html [https://perma.cc/ZP5R-CCTR].

[115] *See id.*

[116] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1788 (2019).

[117] *See Id.*

[118] *Id.* at 1788–89.

production that can be used in films and shows.[119] Deepfakes can be used by educators in classrooms and museums.[120] For example, it is possible to manufacture videos of historical figures speaking about a specific subject, which can draw the viewer's attention.[121] Professors may use deepfakes to bring cases to life using deepfake actors.[122] Artists may also use deepfakes as their platform to express ideas.[123] They may even be used to self-express, through the creation of digital avatars, when unable to otherwise physically do so.[124]

However, there is no denying that deepfakes used to spread disinformation are a problem that plagues our society.[125] More than likely, nearly all of us, at some point, have spread misinformation through social media.[126] The level of content moderation administered by the various online platforms should be the primary inquiry to limit the spread of deepfakes.

Several problems limit the current law from holding the creators of the deepfakes accountable for the spread of misinformation. Tracing a deepfake back to its original creator can be difficult because of the fast spread of digital content through various online platforms. A careful sharer onto the web may take precautions to remain anonymous such as using various technologies like Tor.[127] These technologies can make the IP address connected to the original sharer impossible to find and trace to a particular

---

[119] *Id.* at 1769.
[120] *Id.*
[121] *Id.*
[122] *Id.* at 1769–70.
[123] *Id.* at 1770.
[124] *Id.* at 1770–71.
[125] *See id.* at 1782.
[126] *See id.* at 1766–67.
[127] *Id.* at 1792; *see also* Danielle Keats Citron, HATE CRIMES IN CYBERSPACE 142–43 (Harvard University Press 2014).

individual.[128] The creator of the deepfake may even be outside the United States beyond the reach of the jurisdiction where any legal action can be taken, making it a global concern.[129]

Banning deepfakes that are intentionally deceptive or made with malicious intent would lead to politicized enforcement.[130] The dislike of minority or unpopular viewpoints would result in prejudicial outcomes.[131] The intent of the creator of the deepfake would also be ambiguous and difficult to determine. Constitutionally, the founding fathers warned against the government picking winners and losers in the realm of ideas because it will "tend to act on behalf of the ideological powers that be."[132]

Deepfakes implicate First Amendment concerns. The First Amendment protects individuals' rights to freedom of religion, speech, press, and assembly.[133] Unless deepfakes are truly defamatory, the speech would receive First Amendment protection.[134] It would be difficult to look beyond the broad First Amendment protection to stop the individual creator from making deepfakes. Individual creators can also easily use the First Amendment to protect their work.

It may not be realistic to use copyright or tort law against a particular individual to deter or redress

---

[128] Chesney, *supra* note 116, at 1792; Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 117 (2009).
[129] Chesney, *supra* note 116, at 1792.
[130] *Id.* at 1789.
[131] *See Id.*
[132] Chesney, *supra* note 116, at 1789; Frank I. Michelman, *Conceptions of Democracy in American Constitutional Argument: The Case of Pornography Regulation*, 56 TENN. L. REV. 291, 302 (1989).
[133] U.S. CONST. amend. I.
[134] Marc Jonathan Blitz, *Deepfakes and Other Non-Testimonial Falsehoods: When Is Belief Manipulation (Not) First Amendment Speech?*, 23 Yale J. of L. & Tech. 160, 173–75 (2020).

deepfakes.[135] Some deepfakes include exploited copyrighted content which can lead to copyright infringement challenges.[136] The copyrighted content may be a televised interview with a politician or a scene from a movie. The copyright owner is the person that originally took the picture and not the person that is depicted in the picture.[137] Therefore, the victim would not be able to bring a copyright claim against the deepfake creator; only the copyright owner can.[138] The copyright owner likely has not suffered monetary damages and has little incentive to bring forth a suit. The creator that uses the footage to create the deepfake also likewise has not gained any profits from the exploitation. The court would also conduct a fact-specific inquiry of whether the deepfake was a fair use of the copyrighted content. The primary question revolves around whether the deepfake sufficiently transformed the original to earn fair use protection. There are no judicial precedents that answer the specific question. Nevertheless, the prospects for the success of a copyright claim are low.

Using tort law of defamation is not as effective for public officials and public figures since they must show clear and convincing evidence of actual malice.[139] Victims may also use false light for recklessly creating a harmful and false implication about someone in a public setting, but without knowing the original creator or the clear potential, the claim may be hard to prove.[140] The politician would also have a high burden of proving that they were harmed by the publication of the "defamation."[141]

---

[135] *See* Chesney, *supra* note 116.
[136] *See Id.* at 1793.
[137] *See Id.*
[138] *See Id.*
[139] *See Id.* at 1793–94; Gertz v. Welch, 418 U.S. 323, 342–46 (1974).
[140] *See* Chesney, *supra* note 116, at 1794–95; Citron, HATE CRIMES IN CYBERSPACE, *supra* note 127, at 132–34.
[141] Jessica Ice, *Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RES. L. REV. 417, 434–35 (2019).

The right of publicity, which is rooted in tort law, permits a monetary remedy for the misappropriation of someone's likeness for commercial gain.[142] While deepfakes do include a clear misappropriation of the victim's likeness, there is not likely a commercial gain for the creator.[143] Using privacy law is also a poor fit since public officials and public figures are of a newsworthy and public concern, further implicating successful First Amendment defenses.[144] Current options for imposing liability directly on creators of deepfakes are not effective because of the various risks and lack of clear legal recourse. The only other possibility of seeking remedy is against the owner of the platforms that enable circulation of the content.[145]

A. Section 230 of the Communications Decency Act

Content platforms play a key role in enabling the distribution of deepfakes. Given that it may be difficult to find and deter creators of deepfakes, the most efficient and effective way to mitigate harm may be to impose liability on platforms.[146] "Online platforms already have an incentive to screen content," posted because of "moral suasion, market dynamics, and political pressures."[147] The online platforms hosting the user-generated deepfakes currently do not face any liability.[148] The websites are shielded by the

[142] *See* Chesney, *supra* note 116, at 1794.
[143] *See* Chesney, *supra* note 116, at 1794; Jesse Lempel, *Combatting Deepfakes Through the Right of Publicity*, LAWFARE (Mar. 30, 2018), https://www.lawfareblog.com/combatting-deepfakes-through-right-publicity [https://perma.cc/GDP7-DQ4C].
[144] *See* Chesney, *supra* note 116, at 1794.
[145] *See id.* at 1795.
[146] *Id.*; Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CAL. L. REV. 1805, 1839–40 (2010).
[147] Chesney, *supra* note 116, at 1795.
[148] *See id.*

Communications Decency Act, which immunizes websites from being responsible for user-generated content.[149] Using classic tort claims, like defamation, the individual could sue the online platforms if they are the publisher.[150] However, Section 230(c)(1) of the Communications Decency Act provides broad immunization for online platforms.[151]

The relevant portion drafted by Congress states: "[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."[152] Therefore, the poster on the website, and not Twitter or Facebook, "will be treated as the publisher who is liable for the tort."[153] Distributor liability is very limited. Distributors, including online platforms, bookstores, and newsstands are not held liable for the content of the material that they distribute.[154] The historical reason behind the limitation was that it would be impossible for distributors to read every publication and would possibly result in excessive self-censorship.[155]

These "online platforms [currently] enjoy immunity … for user-generated content even if they deliberately encouraged the posting of that content."[156] Courts have applied Section 230 broadly to include speech-based torts,

---

[149] Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 102 (2019); 47 U.S.C. §230 (2012).
[150] Lempel, *supra* note 143.
[151] *Id.*; *see* 47 U.S.C. § 230.
[152] 47 U.S.C. § 230(c)(1).
[153] Lempel, *supra* note 143.
[154] *Immunity for Online Publishers Under the Communications Decency Act*, DIGITAL MEDIA LAW PROJECT, https://www.dmlp.org/legal-guide/immunity-online-publishers-under-communications-decency-act [https://perma.cc/B5GL-KUPF].
[155] *Id.*
[156] Chesney, *supra* note 116, at 1796; *see also* Danielle Citron & Quinta Jurecic, *Platform Justice: Content Moderation at an Inflection Point*, at 9, Hoover Institute Aegis Series (2018).

invasion of privacy, misappropriation, and fraud.[157] A key exception to broad immunity is the preservation of "liability for a violation of any 'Federal criminal statute.'"[158] "The law provide[s] . . . immunity . . . for hosting harmful content [with the exceptions] for content that violates federal criminal law, the Electronic Communications Privacy Act, and intellectual property law."[159] Specifically, Section 230(e)(2) notes that "[n]othing in this section shall be construed to limit or expand any law pertaining to intellectual property."[160] The intellectual property exception may be the opening necessary to hold "online platforms liable for . . . egregious deepfakes posted by third parties onto their [web]sites."[161]

In crafting Section 230, Congress provided a safe harbor for online platforms by "forbid[ding] civil suits against platforms based on the good-faith act of filtering to screen out offensive content [including] obscenity, harassment, violence, or otherwise."[162] However, the deepfakes with misinformation are not accounted for under the offensive content screening.

Courts have consistently held that decisions made by online platforms to publish, remove, or edit user-submitted content is immunized under Section 230.[163] The online

---

[157] *See* Doe v. MySpace, 474 F. Supp. 2d 843, 849–50 (W.D. Tex. 2007).
[158] 47 U.S.C. § 230(e)(1); Lempel, *supra* note 143.
[159] 47 U.S.C. § 230(c)(1); Chesney, *supra* note 116, at 1795.
[160] 47 U.S.C. § 230(e)(2).
[161] Lempel, *supra* note 143.
[162] Chesney, *supra* note 116, at 1796; *see also* 47 U.S.C. § 230(c)(2).
[163] *See* Johnson v. Arden, 614 F.3d 785, 792 (8th Cir. 2010) (the interactive website www.ComplaintsBoard.com immune for allegedly defamatory statements); Nemet Chevrolet Ltd. v. ConsumerAffairs.com, Inc., 591 F.3d 250, 252–53 (4th Cir. 2009) (court granted immunity from defamation and tortious interference with business expectancy claims to a website that included consumer complaints); Global Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 931–32 (D. Ariz. 2008) (website operator immune under 230 for refusing to remove post despite notification of its potentially defamatory content); Blumenthal v.

platforms, under Section 230, are not being held accountable for the amplification of deepfakes. The only way to find them liable would be using the intellectual property exception of Section 230. Intellectual property encompasses copyright and rights of publicity which would fall under the exception when applied to online platforms.

B. Copyright infringement

Several factors limit the usefulness of copyright infringement under the intellectual property exception. Identifying the owner of the underlying content is difficult.[164] It may not be financially smart to file a complaint against the infringer. A court can easily determine that "the deep fake is a 'fair use' of the copyrighted material, intended for educational, artistic, or other expressive purposes."[165] The tactic of filing lawsuits by well-funded litigants can simply be used to bully opponents with little means into bending to their will. The inquiry of whether a deepfake is transformative is very fact-specific with no case law precedence.[166] The deepfake "may be deemed 'transformative' and . . . not protected by copyright" laws.[167] Similar obstacles faced by using copyright law to stop

---

Drudge, 992 F. Supp. 44, 49–53 (D.D.C. 1998) (AOL has Section 230 immunity from liability for the content of independent contractor's news reports, despite the agreement with contractor allowing AOL to modify or remove such content); Goddard v. Google, Inc., 640 F. Supp. 2d 1193, 1201–02 (N.D. Cal. 2009) (court found that there was web-based advertisements created by allegedly fraudulent providers of services for various mobile devices).

[164] Lempel, *supra* note 143.

[165] Chesney, *supra* note 116, at 1793.

[166] *See id.; but see* David Greene, *We Don't Need New Laws for Faked Videos We Already Have Them*, EFF BLOG (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them [https://perma.cc/UX92-KKEM].

[167] Lempel, *supra* note 143.

individual creators are repeated when applying to online platforms.

C. Rights of Publicity

The right of publicity is an individual's right to exclusively make economic use of their identity independent from copyrights and trademarks.[168] A person can bring a right of publicity claim for misappropriating the commercial use of his or her identity using the appropriate state statute.[169] The person whose image is being used could use the right of publicity under the Section 230 exception to help combat deepfakes posted by third parties on the online platforms. Such a claim would have to get over three major hurdles when applied to online platforms: "(1) fitting the right of publicity into the Section 230 intellectual property exception, (2) counting deepfakes as 'commercial use' of identity for right-of-publicity claims, and (3) First Amendment protections on free speech."[170]

1. Section 230 Exception

First, the right of publicity is generally known to be an intellectual property right, which should fall within the meaning of the Section 230 exception.[171] "In *Zacchini v. Scripps-Howard Broadcasting Co.*, [the Supreme Court] described the right of publicity as 'closely analogous to the goals of patent and copyright law.'"[172] The Ninth Circuit

[168] INT'L TRADEMARK ASS'N, RIGHT OF PUBLICITY (2019) https://www.inta.org/topics/right-of-publicity/ [https://perma.cc/B83W-P9BE].

[169] The Right of publicity is a state-based right, each state may draft their own parameters for the protection of the right. A majority of states currently have distinctly recognized the right of publicity. *See id; e.g.*, Cal. Civ. Code § 3344 (Deering 2022).

[170] Lempel, *supra* note 143.

[171] *See id.*

[172] *Id.*; Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 573 (1977).

applied Section 230's intellectual property exception narrowly to mean only "federal intellectual property."[173] In *Perfect 10, Inc. v. CCBill LLC*, the court said that content on the online platforms could be viewed "in more than one state at a time;" therefore, any particular state's definition of intellectual property would be contrary to Congress's goal of protecting the internet from different state-law systems.[174] Since rights of publicity are grounded in state law, it is not written into the Section 230 immunity according to the Ninth Circuit ruling.[175] However, "[o]utside the Ninth Circuit, [the] right-of-publicity claims would likely be able to pierce Section 230 immunity."[176]

"[T]he First Circuit (in dicta) and the Southern District of New York[] have … applied the intellectual property exception to both federal and state claims."[177] Some courts have also decided that the right of publicity is an intellectual property right well within the Section 230 exception.[178]

A federal trial court in New Hampshire held that Section 230 did not bar a plaintiff's claim against an online platform for violating her right of publicity under New Hampshire law.[179] The website was curated to enable singles and *swingers* to find sexual partners by allowing users to create online profiles by providing personal information.[180] An unknown imposter created a fake profile using the plaintiff's identity and posted inappropriate

---

[173] Lempel, *supra* note 143.
[174] Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1118 (9th Cir. 2007).
[175] *See id.*
[176] Lempel, *supra* note 143.
[177] *Id.*; *see* Universal Commc'n Sys., Inc. v. Lycos, Inc., 478 F.3d 413, 421 (1st Cir. 2007); Atl. Recording Corp. v. Project Playlist, Inc., 603 F. Supp. 2d 690, 704 (S.D.N.Y. 2009).
[178] Lempel, *supra* note 143; *see* Doe v. Friendfinder Network, Inc., 540 F. Supp. 2d 288, 303–04 (D.N.H. 2008).
[179] *Doe*, 540 F. Supp. 2d at 306–07.
[180] *Id.* at 291.

photographs.[181] The plaintiff consequently sued the online platform after she found about the profile a year later.[182] The plaintiff brought several claims, including the invasion of property/intellectual property rights.[183] The defendant moved to dismiss based on Section 230.[184] The primary question was whether the plaintiff's state law right of publicity claim against the online platform was an intellectual property claim. The court answered in the affirmative, citing a federal appeals court opinion which explicitly recognizes, "there appears to be no dispute that the right of publicity is a type of intellectual property right."[185] The New Hampshire court held the state right of publicity intellectual property claim was not preempted by Section 230.[186] The case exemplifies that it is possible for state law claims, such as the right of publicity, to be exempted from the broad immunity provided under Section 230. However, circuit courts have been unable to reach a consensus regarding the treatment of misappropriation of right of publicity.[187]

2. Commercial Use of Identity

Second, the right of publicity only protects the commercial use of one's identity, most commonly in

[181] *Id.* at 292.
[182] *Id.* at 292–93.
[183] *Id.* at 293.
[184] *Id.* at 291.
[185] Almeida v. Amazon.com, Inc., 456 F.3d 1316, 1323 (11th Cir. 2006); *see also Doe*, 540 F. Supp. 2d. at 302 (citing *Almeida*, 456 F.3d at 1322).
[186] *Doe*, 540 F. Supp. 2d. at 304.
[187] *Compare Perfect 10, Inc.*, 488 F.3d at 1118 (holding that the intellectual property exception only applies to federal intellectual property laws), *with Universal Commc'n Sys., Inc.*, 478 F.3d at 421 (in dicta held that the intellectual property exception may apply to both federal and state law claims), *and* Hepp v. Facebook, et al., No. 20–2725 (3d Cir. 2021) (holding that the section 230(e) carveout can apply to federal and state laws that pertain to intellectual property).

advertisements.[188] Deepfakes that portray fake news tend to be political and do not fall into a clear commercial category.[189] Online platforms like Facebook and Twitter have business models that depend on the number of clicks and views of the posts making commercial use of the deepfakes.[190] In *Cross v. Facebook*, "a business owner sued Facebook for placing [his business] ads next to an unauthorized page critical of his business."[191] The "California state judge ruled that the right-of-publicity claim was viable under Section 230's intellectual property exception because 'Facebook's financial portfolio is based on its user base.'"[192] The decision caused alarm with commentators critiquing that the ruling was "not a proper use of the right[s] of publicity."[193] "[T]he appellate court eventually reversed [the state court because] Facebook [was] merely 'display[ing] unrelated ads from Facebook advertisers adjacent to' images of the plaintiff 'posted by third parties' [and] … did not … 'use his name or likeness in anyway.'"[194]

---

[188] White v. Samsung Elecs. Am., Inc., 971 F.2d 1395, 1398 (9th Cir. 1992).

[189] Lempel, *supra* note 143

[190] *See* Ally Mintzer, *Paying Attention: The Attention Economy*, Berkeley Econ. Rev., (Mar. 31, 2020) https://econreview.berkeley.edu/paying-attention-the-attention-economy/ [https://perma.cc/7HG6-MUNZ].

[191] Lempel, *supra* note 143; Cross v. Facebook, Inc., 222 Cal. Rptr. 3d 250, 265 (Cal. Ct. App. 2017).

[192] Lempel, *supra* note 143; Cross v. Facebook, Inc., No. CIV537384, 2016 WL 7785723, at *4 (Cal. Super. 2016).

[193] Paul Alan Levy, *California Ruling Against Facebook on Right of Publicity Blows Huge Hole in Section 230 Immunity*, CONSUMER LAW & POLICY BLOG (June 3, 2016, 12:48 PM), https://pubcit.typepad.com/clpblog/2016/06/california-ruling-against-facebook-blows-huge-hole-in-section-230-immunity.html [https://perma.cc/YB42-XH4M].

[194] Lempel, *supra* note 143; *Cross*, 222 Cal. Rptr. 3d at 266-67 (2017).

However, the implications of the ruling are only valid for posts by third parties that are not easily recognizable and exempt more publicly known individuals such as celebrities or political figures.[195] If a deepfake of a well-known figure was posted on an online platform, that video has the potential to be viewed billions of times. The user activity on the online platform is commercial and could make a right of publicity claim appropriate. The increased activity leads to more users getting on the platforms and viewing the other advertisements also posted on the website.

3. First Amendment

Third, "the most meaningful constraint on the right of publicity is . . . the First Amendment."[196] In *New York Times v. Sullivan*, the Supreme Court held that false speech has constitutional protection because a broad prohibition would lead to a substantial "chilling effect" on free speech.[197] The right of publicity is restricted from being used to specifically censor "disagreeable portrayals."[198] Courts have provided First Amendment protection to "reports of newsworthy events or matters of public interest" even when it would otherwise violate the rights of publicity.[199] In certain cases, an article or a movie about a person "may be so infected with fiction, dramatization or embellishment that it cannot be said to fulfill the purpose of the newsworthiness exception."[200] First Amendment protection is also not extended to speech about public figures when it is made with "actual malice" or with "reckless

[195] Lempel, *supra* note 143.
[196] *Id.*
[197] New York Times Co. v. Sullivan, 376 U.S. 254, 300–01 (1964).
[198] Lempel, *supra* note 143; Comedy III Productions, Inc. v. Gary Saderup, Inc., 21 P.3d 797, 807 (Cal. 2001).
[199] Lempel, *supra* note 143; Messenger ex rel. Messenger v. Gruner + Jahr Printing & Pub., 94 N.Y.2d 436, 441 (2000).
[200] *Id.* at 446.

disregard of whether it was false or not."[201] There may be deepfakes that are "so infected with fiction" that they lie beyond the newsworthy exception to the right of publicity.[202] Online platforms failing to remove deepfakes could be considered reckless disregard.

Considering First Amendment protections, the right of publicity would only be enforced against technologically deceptive impersonations that can generate commercial revenue.[203] The online platforms would only be tasked with verifying whether the content on their websites is technologically genuine or falsified by detecting whether the video or audio has been artificially manipulated to be fundamentally deceptive.[204]

Current options using the existing law to impose liability for the spread of deepfakes have mixed potential. There are limited prospects for using copyright and the right of publicity. However, the recent court interpretations may potentially leave a small opening for using the right of publicity under the intellectual property exception to hold online platforms liable. Even so, extending the control to the scope of platform liability only seems to be a partial remedy in the grand scheme of all the players involved.

V. CURRENT LEGISLATION

After intelligence agencies confirmed Russian meddling on social media during the 2016 U.S. presidential election, Americans have taken an analytical look at online platforms and government officials have shown an interest in taking various actions. Congress announced a bill in October 2017 that would require online platforms to keep

---

[201] *New York Times Co.*, 376 U.S. at 280; *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46, 52 (1988).
[202] *Messenger*, 94 N.Y.2d at 446; Lempel, *supra* note 143.
[203] Lempel, *supra* note 143.
[204] *Id.*

copies of advertisements, make them public, and record who was paying for them.[205] A month later, representatives from Facebook, Twitter, and Google testified to a Senate judiciary committee on their role in spreading disinformation during the election.[206] In mid-September 2018, two Democrats and one Republican representative sent a letter to the director of national intelligence asking the intelligence community to assess the possible national security threats posed by deepfakes.[207] The lawmakers noted that there was potential for foreign adversaries to use deepfake videos against U.S. interests.[208]

In 2018, lawmakers modified Section 230 by enacting the Allow States and Victims to Fight Online Sex Trafficking Act ("FOSTA").[209] In addressing online platforms' facilitation of sex trafficking, FOSTA added a new exception to Section 230 immunity similar to the intellectual property portion.[210] FOSTA provides an exception from Section 230 to enable victims and state

[205] S. 1989, 115th Cong. (2017), https://www.congress.gov/bill/115th-congress/senate-bill/1989 [https://perma.cc/6LCE-6FXU].
[206] Tom McCarthy, *Facebook, Google and Twitter grilled by Congress over Russian meddling – as it happened*, THE GUARDIAN (Oct. 31, 2017), https://www.theguardian.com/technology/live/2017/oct/31/facebook-google-twitter-congress-russian-election-meddling-live?page=with:block-59f8e5a5e4546a06df012051#block-59f8e5a5e4546a06df012051 [https://perma.cc/X8W4-JP4U].
[207] *Schiff, Murphy and Curbelo Request DNI Assess National Security Threats of "Deep Fakes"*, ADAM SCHIFF (Sept. 13, 2018), https://schiff.house.gov/news/press-releases/schiff-murphy-and-curbelo-request-dni-assess-national-security-threats-of-deep-fakes [https://perma.cc/67NM-5S88].
[208] *Id.*
[209] *See* Chesney, *supra* note 116, at 1798; *see also* Danielle Citron & Quinta Jurecic, FOSTA: *The New Anti-Sex-Trafficking Legislation May Not End the Internet, But It's Not a Good Law Either*, LAWFARE (Mar. 28, 2018), https://www.lawfareblog.com/fosta-new-anti-sex-trafficking-legislation-may-not-end-internet-its-not-good-law-either [https://perma.cc/U5S4-YJW9].
[210] *See* Chesney, *supra* note 116, at 1798–99.

attorney generals to sue online platforms for knowingly assisting, supporting, or facilitating sex trafficking offenses.[211]

The first U.S. federal legislation on deepfakes was signed into law in December 2019 as part of the National Defense Authorization Act (NDAA).[212] It required a comprehensive report on foreign weaponization of deepfakes, the government to notify Congress of foreign deepfake-disinformation activities and established an incentive to advance research of deepfake-detection technologies.[213]

President Trump signed an executive order in early 2020 to limit the broad legal protections afforded to social media companies.[214] The executive order states that the "growth of online platforms in recent years raises important questions about applying the ideals of the First Amendment to modern communications technology."[215] Trump's abrupt issuance is assumed to have been a consequence of Twitter issuing fact-checking labels on his tweets.[216] The order was drafted to overturn 25 years of judicial precedent by revoking Section 230 to end liability protections for social media platforms and make them responsible for the speech

---

[211] *See id.* at 1799

[212] Chipmon, Jason, Matthew Ferraro, & Stephen Preston, *First Federal Legislation on Deepfakes Signed Into Law*, JD Supra (Dec. 24, 2019), https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346/ [https://perma.cc/THQ4-HVKN].

[213] *Id.*

[214] Exec. Order No. 13925, 85 Fed. Reg. 34079 (May 28, 2020), https://www.federalregister.gov/documents/2020/06/02/2020-12030/preventing-online-censorship [https://perma.cc/X34U-5PZQ].

[215] *Id.*

[216] Bobby Allyn, *Stung By Twitter, Trump Signs Executive Order To Weaken Social Media Companies* , NPR (May 28, 2020), https://www.npr.org/2020/05/28/863932758/stung-by-twitter-trump-signs-executive-order-to-weaken-social-media-companies [https://perma.cc/82BM-GKQN].

of users that post on the sites.[217] It directs the Federal Communications Commission (FCC) to start a rulemaking process to clarify when social media companies should keep protections under the law.[218]

However, legal experts state that the order does not change existing law and will have no bearing on the federal courts.[219] The FCC is limited to regulating traditional broadcasters and has not historically been subjected to being involved in social media regulation.[220] Soon after Trump's executive order, the Justice Department announced a legislation proposal to curtail legal protections for social media platforms for the content that appears on their sites.[221] The Department's proposal sets out to update three primary objectives: (1) provide online platforms incentives to address illicit content, (2) clarify federal powers to address unlawful content, and (3) promote open discourse and greater transparency.[222]

Critics have brought up arguments regarding Section 230 on both sides of the political spectrum. Republicans argue that the social media companies should not have broad protections because they censor conservatives and take away from "a true diversity of political discourse."[223] Sen. Ben

---

[217] Maggie Haberman & Kate Conger, *Trump Prepares Order to Limiti Social Media Companies' Protections*, N.Y. TIMES (June 2, 2020), https://www.nytimes.com/2020/05/28/us/politics/trump-executive-order-social-media.html [https://perma.cc/AAS9-992N].

[218] Exec. Order No. 13925, *supra* note 214, at 34081.

[219] Allyn, *supra* note 216.

[220] *See What We Do*, FEDERAL COMMUNICATIONS COMMISSION, https://www.fcc.gov/about-fcc/what-we-do [https://perma.cc/WCA3-Q87L].

[221] *Justice Department Issues Recommendations for Section 230 Reform*, DEPARTMENT OF JUSTICE (June 17, 2020), https://www.justice.gov/opa/pr/justice-department-issues-recommendations-section-230-reform [https://perma.cc/5WT2-EJTJ].

[222] *Id.*

[223] Diasuke Wakabayashi, *Legal Shield for Social Media Is Targeted by Lawmakers*, N.Y. TIMES (May 28, 2020),

Sasse proposed rules that would make it unlawful for people to maliciously create and distribute deepfakes.[224] Meanwhile, Democrats argue that websites do not give enough effort to take down problematic content because they enjoy the Section 230 shield.[225] Rep. Yvette Clarke introduced a bill that would force creators of deepfakes to disclose that they are fabricated by including an identifier such as a watermark.[226] As of publication, no piece of proposed federal legislation has seen much traction.

The sentiments from Facebook and Twitter closely mimic their current platform policies. Facebook CEO, Mark Zuckerberg, takes a stance that private companies should not be responsible for taking down any expressions from their users.[227] Whereas Twitter co-founder and former CEO, Jack Dorsey, responds that online platforms should provide transparency to their users by marking conflicting information.[228] Throughout the years, lawmakers continue to mull over possible legislation in search for methods that platforms could apply on their own to tackle the issue of deepfakes going viral. There has yet to be a clear solution

---

https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html [https://perma.cc/5XET-ZASF].

[224] Kaveh Waddell, *Lawmakers plunge into "deepfake" war*, AXIOS (Jan. 31, 2019), https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html [https://perma.cc/24RW-YKCU].

[225] Wakabayashi, *supra* note 223.

[226] Christiano Lima, *'Nightmarish': Lawmakers brace for swarm of 2020 deepfakes*, POLITICO (June 13, 2019), https://www.politico.com/story/2019/06/13/facebook-deep-fakes-2020-1527268 [https://perma.cc/K64X-GDNM].

[227] Mike Isaac & Cecilia Kang, *While Twitter Confronts Trump, Zuckerberg Keeps Facebook Out of It*, N.Y. TIMES (May 29, 2020), https://www.nytimes.com/2020/05/29/technology/twitter-facebook-zuckerberg-trump.html [https://perma.cc/8Q4N-3N5X].

[228] @jack, TWITTER (May 27, 2020) https://twitter.com/jack/status/1265837139360485376?lang=en [**https://perma.cc/4LT5-PGQ7**].

on how to stop deepfakes aimed at spreading misinformation.

VI. POSSIBLE SOLUTIONS

So much of the regulation of deepfakes depends on technological advances for the rapid detection of deepfakes. Detection software would have to keep pace with innovations in deep-fake technology to retain efficacy.[229] As programs are written to authenticate content on online platforms, there will be a counter-effort from creators to bypass them. Growing effects of deepfakes might make grantmaking agencies like the National Science Foundation and the Defense Advanced Research Projects Agency (DARPA) continue to fund research for scalable detection systems.[230]

The large online platforms that currently host and exaggerate the effects of deepfakes should be pressured to work towards effective technological solutions.[231] Emerging market pressures may also encourage private companies to invest in finding a reliable technological solution to minimize the harms deepfakes cause.[232] Technology that is developed can then be deployed through social media platforms to reduce the harms of deepfakes. It is not likely that one program becomes effective enough for all online platforms to incorporate to screen content posted by third-party users. As seen with Facebook's public contest with identifying deepfakes, it is an extremely impossible feat to design algorithms able to effectively detect all deepfakes. Moreover, by no means is detecting deepfakes alone sufficiently to eliminate the dangers of spreading misinformation through deepfakes.

---

[229] Chesney, *supra* note 116, at 1787.
[230] *Id.* at 1788.
[231] *See id.*
[232] *Id.* at 1795.

Congress had two goals in mind when immunizing users and operators of interactive computer services from liability for content posted by third parties: to promote the free exchange of information and ideas over the Internet and to encourage voluntary monitoring for offensive material.[233] The aim is to avoid blanket censorship that would lead to a substantial "chilling effect" on free speech.[234] Justice Powell, notably in his dissent in *Zacchini*, warned of the "disturbing implications" of "media self-censorship" in the shadow of excessive right-of-publicity liability.[235]

Any viable legal solution for the next generation of fake news through the right of publicity will hinge on the capacity of online platforms to efficiently flag the fakes and the courts' ability to enforce the right of publicity in a way that protects free speech.[236] The online platforms are private companies, free to determine which speech is viewed on their platform, and have no requirements to hide or mute content. The First Amendment cannot control the speech that appears on their platforms.[237] But that does not mean that online platforms will be required to self-censor to an extreme extent, it will be up to the courts to shape what is *reasonable* self-censorship.

As a matter of public policy, online platforms should take steps to deter the spread of misinformation through deepfakes posted on their websites. Platforms should be encouraged to be transparent about their policies and be held accountable for their speech decisions.

The most effective way to remedy the split between the Ninth, First and Third Circuits is to have the legislature clarify the language of Section 230(e)(2) by including an express statutory definition of intellectual property. By

---

[233] *See* 47 U.S.C. § 230(b).
[234] Lempel, *supra* note 143.
[235] *Zacchini*, 433 U.S.,at 580 (Powell, L., dissenting).
[236] Lempel, *supra* note 143.
[237] *See supra* Part IV § C(3).

listing the right of publicity among the different types of intellectual property protected under Section 230(e)(2), courts will be bound by the strict literal interpretation of the statute, resulting in the uniform application of the law across the circuit courts.

A proposed change to the Section 230 immunity would include placing the burden on the online entity to demonstrate that it has taken *reasonable* steps to ensure that its platform is not being used for fueling deepfakes.[238] The online platforms will not be declared to be the publisher of the user-generated deepfakes they host but will be held liable for not taking effective steps to stop the spread of the misinformation propagated in the deepfake. With an amendment to Section 230(e)(2) expressly listing the right of publicity as a protected intellectual property claim, exempt from Section 230 immunity, public figures and celebrities will have a legitimate cause of action against the online platform if their image is used to perpetuate misinformation. Similar defenses used with rights of publicity can be used by online platforms such as fair use and parody.

There is no denying that it is hard to predict how rigidly the online platforms would apply their screening, it would therefore be up to the courts to shape the law. Online platforms would still be able to let users freely post their content and only be responsible for determining what deepfakes are. The determination solely hinges on online platforms determining whether the content is technologically genuine or falsified, not whether the actual claims made in the deepfakes are true or considered misinformation. In 2021, Facebook underwent a "rebrand" by changing its name to Meta, aimed at branching out, but has not yet put out any responses on misinformation

[238] *See* Chesney, *supra* note 116, at 1799.

projects.[239] It is not yet clear what Mark Zuckerberg's "metaverse" will actually consist of, but it does have adequate backing to invest in projects to accurately detect misinformation.

There are obvious technological obstacles that would have to be developed for the online platforms to flag fakes efficiently.[240] Any proposed changes that rely on a fact-based inquiry raise the question of the metes and bounds of reasonableness.[241] The inquiry should be like the Section 230(c)(2) good faith act filtering determination for offensive content. The online platforms would use good faith effort by using reasonable standards of either digital forensic tools to spot the fakes or rely on a system of digital signatures.[242]

Assuming that the online platforms have the technological capability to distinguish between genuine and fake videos, failing to remove the fakes within a reasonable time would rise to the level where the public figure can seek judicial judgment.[243] The law should compel online platforms to use technology to the best of their ability as effective authentication technology develops.[244] But maybe at the end of the day, as Elon Musk hypothesized his understanding of artificial intelligence derived from an Oxford philosopher, "the odds that we're in 'base reality' is one in billions."[245] The dynamic nature of social media

---

[239] *Introducing Meta: A Social Technology Company*, META (Oct. 28, 2021), https://about.fb.com/news/2021/10/facebook-company-is-now-meta/ [https://perma.cc/LM68-RDZL].
[240] Lempel, *supra* note 143.
[241] *See* Chesney, *supra* note 116, at 1799.
[242] Lempel, *supra* note 143.
[243] *Id.*
[244] *Id.*
[245] Recode, *Is life a video game?* at 2:30 | *Elon Musk* | *Code Conference 2016*, YOUTUBE (June 2, 2016), https://www.youtube.com/watch?v=2KK_kzrJPS8&feature=emb_title [**https://perma.cc/T9TK-YTMS**].

makes it difficult to predict the specific direction of future research, but we do know that social media is likely to proliferate and mature, contribute to new social divisions, alter how individuals organize and mobilize, and complicate the way organizations and institutions manage information.