William & Mary Bill of Rights Journal

Volume 29 (2020-2021)
Issue 2

Article 7

December 2020

Deepfakes: A New Content Category for a Digital Age

Anna Pesetski

DEEPFAKES: A NEW CONTENT CATEGORY FOR A DIGITAL AGE

Anna Pesetski*

INTRODUCTION

Technology has advanced rapidly in recent years, greatly benefitting society. One such benefit is people's ability to have quick and easy access to information through news and social media.[1] A recent concern, however, is that manipulated media, otherwise known as "deepfakes," are being released and passed off as truth.[2] These videos are crafted with technology that allows the creator to carefully change details of the video's subject to make him appear to do or say things that he never did.[3] Deepfakes are often depictions of political candidates or leaders and have the potential to influence voter choice, thereby altering the outcome of elections.[4] Deepfakes have already influenced the politics of other countries,[5] and lawmakers expressed legitimate fears about how deepfakes would affect the 2020 United States presidential election.[6]

The current unprotected categories of speech developed during a more primitive technological age.[7] Efforts have been made to combat deepfakes,[8] but they have fallen

---

* JD Candidate, William & Mary Law School Class of 2021. Thank you to my parents, Peter and Karen, for their endless support, and thank you to Paul Johnson for being my biggest encourager. I also want to thank the *William & Mary Bill of Rights Journal* editorial board for all of their hard work.

[1] *See, e.g.*, Douglas Soule, *US Lawmakers Weigh 'Deepfake' Concerns with First Amendment Rights*, GLOBE POST (June 13, 2019), https://theglobepost.com/2019/06/13/deepfakes-first-amendment/ [https://perma.cc/96YG-GL2R].

[2] *See* Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We are Outgunned,'* WASH. POST (June 12, 2019, 4:44 PM), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ [https://perma.cc/6CKV-XU2U]; *see also* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1759 (2019) (discussing the emergence of new technologies that makes deepfakes "more realistic and more difficult to debunk than they have been in the past").

[3] Carrie Mihalcik, *California Laws Seek to Crack Down on Deepfakes in Politics and Porn*, CNET, https://www.cnet.com/news/california-laws-seek-to-crack-down-on-deepfakes-in-politics-and-porn/ [https://perma.cc/78E5-QMRB] (Oct. 7, 2019, 8:32 AM).

[4] *See* Harwell, *supra* note 2.

[5] *See id.* (discussing political issues that deepfakes have caused in Gabon and Malaysia).

[6] *See* Soule, *supra* note 1.

[7] *See generally, e.g.*, Miller v. California, 413 U.S. 15 (1973); Brandenburg v. Ohio, 395 U.S. 444 (1969) (per curiam); N.Y. Times Co. v. Sullivan, 376 U.S. 254 (1964); Chaplinsky v. New Hampshire, 315 U.S. 568 (1942); *see also infra* Section III.A.

[8] *See, e.g.*, Christopher Carbone, *Google Releases 3,000 Deepfake Videos So Researchers*

short of effectively attacking the problem. It may be time for the Supreme Court to reevaluate First Amendment protections in light of the current digital age and consider the benefits of adding a new unprotected content category of speech for deepfakes.[9] The dangers deepfakes present far outweigh the concerns of the potential chilling effects from restrictions on speech.[10] Even though the Court has rejected arguments for new categories of unprotected speech in recent years,[11] deepfakes should ultimately constitute a new content category because of the dangers they pose to the election process and political systems; the "marketplace of ideas" fails to combat their falsity.[12]

This Note will argue that deepfakes can and should constitute a new content category of unprotected speech without infringing on First Amendment protections of speech.[13] Deepfakes are created with a technological process that is accessible to the average person and results in false depictions of people that even trained videographers have trouble spotting.[14] These false depictions are often of political candidates or officials, and they have the power to deceive voters and influence election outcomes.[15] This Note will first examine the evolution of deepfakes and the dangers they present for the nation's political processes.[16] Then it will analyze the justifications for First Amendment protections and discuss why deepfakes do not fit within these justifications.[17] It will next address the current categories of speech that the Supreme Court has deemed to be unprotected and explain the process for how the Court declares speech to be unprotected.[18] This Note will then argue why the federal government should be able to regulate deepfakes without offending the First Amendment through a narrowly tailored law, address possible concerns with regulation, and discuss why current solutions fail to adequately address the issue.[19] Finally, it will conclude with a proposed regulation.[20]

Can Combat Them, FOX NEWS (Sept. 25, 2019), https://www.foxnews.com/tech/google-3000-deepfake-videos-combat [https://perma.cc/3TTM-X7HK].

[9] *See* Harwell, *supra* note 2.

[10] *See* Ben Christopher, *Can California Crack Down on Deepfakes Without Violating the First Amendment?*, CAL MATTERS, https://calmatters.org/politics/2019/07/deepfake-berman-california-politics-ab730-fake-news-first-amendment/ [https://perma.cc/8NWE-KL7B] (July 8, 2019).

[11] *See generally, e.g.*, United States v. Alvarez, 567 U.S. 709 (2012) (plurality opinion); United States v. Stevens, 559 U.S. 460 (2010).

[12] This Note will touch on the fact that deepfakes have been used for fake pornography, but the primary focus will be their use in elections and political processes and what lawmakers can do to combat those deepfakes.

[13] *See infra* Part I.

[14] Herbert B. Dixon, Jr., *Deepfakes: More Frightening Than Photoshop on Steroids*, 58 JUDGES' J., Summer 2019, at 35, 35.

[15] *See* Harwell, *supra* note 2.

[16] *See infra* Part I.

[17] *See infra* Part II.

[18] *See infra* Part III.

[19] *See infra* Parts IV–V.

[20] *See infra* Part VI.

I. THE RISE OF DEEPFAKES AND THEIR IMPACT

Even though deepfakes are a relatively new phenomenon, their notoriety has quickly spread.[21] This Part will explore deepfakes generally and expound on both their current and potential dangers to society.

A. Definition of Deepfake

The term "deepfake" is relatively new to the vocabulary of American society and has been used to describe the convincingly realistic false videos that have been disseminated via social and news media.[22] Deepfakes are "video forgeries that make people appear to do or say things they didn't. They use a type of facial recognition technology to mash up identity so well you don't even question its truth."[23] Software that creates deepfakes "studies the statistical patterns in a data set, such as a set of images or videos, and then generates convincing fake videos."[24] The technology behind the creation of deepfakes is advanced, but a person with more general knowledge of videography can easily find an instructional video on YouTube explaining how to manipulate a video to create his own deepfake.[25] A recent example of manipulated media was a doctored video of House Speaker Nancy Pelosi, which depicted the political official slurring her words as if she were drunk during a speech.[26] This video could be referred to as a "shallowfake" because the type of manipulation used to alter the video, where context was removed and Pelosi's voice was simply slowed down, does not quite rise to the level of the facial recognition and mashup technology used to create deepfakes.[27] The video was discovered to be altered, but it still received more than 2.5 million views on social media.[28] Videos of this nature have the power to influence voter choice and the opinions of political officials.[29] Other recent examples of deepfakes include one of Mark Zuckerberg talking about stealing the data of Facebook users[30] and one by Jordan Peele depicting former President Barack Obama calling President Trump "a total and complete dipshit."[31] Peele's video was "[s]et in what appears to be the Oval Office" and "depicts the former president speaking fondly

---

[21] *See* Mihalcik, *supra* note 3 (noting the attention deepfakes gained after an altered video of Nancy Pelosi was disseminated on social media); *see also* Soule, *supra* note 1.

[22] *See* Dixon, *supra* note 14, at 35; Mihalcik, *supra* note 3.

[23] Mihalcik, *supra* note 3.

[24] Dixon, *supra* note 14, at 36.

[25] Christopher, *supra* note 10.

[26] Mihalcik, *supra* note 3.

[27] *Id.*; *see* Dixon, *supra* note 14, at 35.

[28] *Doctored Nancy Pelosi Video Highlights Threat of "Deepfake" Tech*, CBS NEWS, https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/ [https://perma.cc/RW2D-6ZV2] (May 26, 2019, 9:26 AM).

[29] *See* Mihalcik, *supra* note 3.

[30] Soule, *supra* note 1.

[31] Christopher, *supra* note 10.

of the militant anti-colonial villain of the 'Black Panther' comic franchise and claiming that Housing Secretary Ben Carson is brainwashed."[32] The video was created to make the public aware of deepfake technology and as a warning about its "potential misuses."[33] Former President Obama publicly commented on deepfakes and the danger of their deception.[34] The false information in these videos has the ability to influence opinions and can clearly be harmful to a video subject's reputation.

B. Deepfake Usage

Deepfakes have been used to serve various purposes by their creators.[35] One of the dominating uses is fake pornography.[36] A recent study found that 96% of over 14,000 identified deepfake videos online were pornographic, all of which depicted women, often popular celebrities.[37] Pornographic deepfakes can cause serious emotional and psychological issues and even lead to the harassment of the subject of the video.[38] Such was the case for Rana Ayyub, an Indian investigative journalist who was targeted with a fake pornographic video in retaliation for her critique of the Indian prime minister and his political party.[39] The harassment and emotional turmoil resulting from the deepfake sent her to the hospital with heart palpitations.[40]

Deepfakes have also been used to smear political candidates and officials, even to the point of interfering with elections[41] and governments at the international level.[42] Outside of the United States, an unsuccessful coup by the Gabonese military was sparked when the president's opponents claimed a video of the president, whom many in the country previously believed to be in declining health or dead, was a deepfake.[43] In Malaysia, a video of "a man's seeming confession to having sex with a local cabinet minister" has come to light as a possible deepfake.[44] Moldova recently experienced interference in its elections when a deepfake was posted on Al Jazeera's Facebook page showing, "a mayoral candidate's proposal to lease an island to the United Arab Emirates," which went viral.[45] United States lawmakers are concerned about the

---

[32] *Id.*

[33] Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U. J. L. & TECH., Fall 2018, at 51, 60.

[34] *See* Harwell, *supra* note 2.

[35] *See* Mihalcik, *supra* note 3 ("While some deepfakes are silly and fun, others are misleading and even abusive.").

[36] *Id.*

[37] *Id.*

[38] Christopher, *supra* note 10.

[39] *Id.*

[40] *Id.*

[41] *See* Hall, *supra* note 33, at 55.

[42] *See* Harwell, *supra* note 2.

[43] *Id.*

[44] *Id.*

[45] Sonya Swink & Kyle Qualls, *'Deep Fake' Videos Could Change Outcome of Electoral*

possibility of deepfakes exerting a similar influence on United States voters, particularly in the 2020 presidential election.[46]

A third usage of deepfakes is humor.[47] Parody and satire of public figures are protected uses of speech under the First Amendment,[48] and deepfakes have been created for these less sinister purposes. Jimmy Kimmel, for example, played a video of President Donald Trump on his show where the president's voice was slowed down so that he sounded drunk while giving a speech on a segment called "Drunk Donald Trump."[49] The video has a disclaimer stating that it was slowed down by the creators of the show, which allowed audiences to understand that the video was fake.[50] The use of deepfakes for humor poses little to no threat, because audiences viewing the video understand that it is not real and was only intended to be funny.[51]

C. Dangers of Deepfakes

Deepfakes present many dangers. Lawmakers are concerned that these videos "could threaten national security, the voting process—and, potentially, their reputations."[52] The most crucial and timely of these dangers is the potential influence on election outcomes.[53] A deepfake of a political candidate supporting a policy contrary to his platform, or one that he simply does not endorse, could harm the candidate's chances of winning the election.[54] House Intelligence Committee Chairman Adam Schiff has voiced his fears about deepfakes influencing the presidential election in 2020 "with the government, media, and public struggling to discern what is real and fake."[55] Schiff went on to say that part of the danger of deepfakes lies in "the ubiquity of social media and the velocity at which false information can spread."[56]

To better understand the danger of deepfakes, it is helpful to look at how false news spreads generally.[57] Massachusetts Institute of Technology (MIT) performed an in-depth Twitter study that "analyzed around 126,000 cascading news stories tweeted by 3 million users over more than 10 years."[58] The study discovered that "a

---

Races, GLOBE POST (Aug. 1, 2018), https://theglobepost.com/2018/08/01/deep-fake-videos/ [https://perma.cc/G5VW-2KVY].

[46] *See* Harwell, *supra* note 2.

[47] *See* Mihalcik, *supra* note 3.

[48] *See* Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 55–57 (1988).

[49] Jimmy Kimmel Live, *Drunk Donald Trump—Why He Got Elected*, YOUTUBE (Feb. 9, 2017), https://youtu.be/dK3LbVFgyqQ [https://perma.cc/9MXZ-LZTK].

[50] *Id.*

[51] *See, e.g.*, Christopher, *supra* note 10.

[52] Harwell, *supra* note 2.

[53] *See id.*

[54] *See* Dixon, *supra* note 14, at 37.

[55] Soule, *supra* note 1.

[56] *Id.*

[57] *See* Hall, *supra* note 33, at 55.

[58] *Id.*

fabricated story reaches 1,500 people six times more rapidly than a true story. False political stories were particularly effective in being spread, more than false stories about business, terrorism or science."[59] Given how rapidly false political stories can spread and the vast number of people they reach, it is easy to see why deepfakes are particularly dangerous. Detecting fake text-based stories is already challenging enough, and deepfakes present an even greater obstacle because viewers must question the validity of what they see instead of just what they read.[60]

Deception is a strong force that can have lasting effects on the public's opinions of political candidates or government officials.[61] Studies involving subjects that were purposely given false information about the views of a candidate and then later told that the information was false have shown that people continued to think poorly of those candidates, even with the knowledge that those views were misrepresented.[62] Voters may change their minds about which candidate they choose, or they might experience a phenomenon called "reality apathy," where they find it too difficult to discern what is true and decide to just vote along the lines of their political affiliations.[63] In essence, the right to vote becomes "nullified" because voters receive false information, which influences their voting choices.[64] Americans have stated that doctored videos and images have impeded their understanding of the facts of current events,[65] making it more difficult to form educated opinions.[66] Siwei Lyu, the director of a computer-vision lab at the State University of New York at Albany, posited that "media manipulation can have a broader psychological effect, by subtly shifting people's understandings of politicians, events and ideas."[67] People will likely question the validity of more videos they see online or in the media, even when the videos are completely true, leading to more confusion.[68]

National security is another major concern surrounding deepfakes.[69] For example, if a deepfake was made of President Trump declaring war on another country, there could be a very real national security issue if the other country perceived the threat to be legitimate.[70] Even if the video was discovered to be fake, it may be too late to avoid serious problems resulting from it.[71] Senator Marco Rubio, a member

---

[59] *Id.*

[60] *See id.* at 55–56 (noting society's vulnerability to misinformation).

[61] Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1463 (2019).

[62] *Id.* at 1463–64.

[63] Harwell, *supra* note 2.

[64] Green, *supra* note 61, at 1457–58 (discussing the potential harm to voters who base their votes on falsified positions).

[65] *See* Harwell, *supra* note 2 (discussing a Pew Research study).

[66] *See* Green, *supra* note 61, at 1458.

[67] Harwell, *supra* note 2.

[68] *See* Dixon, *supra* note 14, at 37; Harwell, *supra* note 2.

[69] Harwell, *supra* note 2; *see* CBS NEWS, *supra* note 28.

[70] *See* CBS NEWS, *supra* note 28.

[71] *See, e.g.*, *id.*; Green, *supra* note 61, at 1463.

of the Senate Intelligence Committee, expressed his concerns surrounding deepfakes and provided the example of a foreign intelligence agency producing "a deepfake of a United States soldier massacring civilians overseas."[72] A video of that nature could create fear among the American people and possibly disrupt relationships between the United States and foreign powers.[73]

A third major concern resulting from deepfakes is the possible damage to the reputation of government officials.[74] A Belgian political party recently crafted and disseminated a deepfake advertisement "featuring what appeared to be President Trump criticizing the Paris Climate Accord" in an attempt to get signatures for a climate-change petition.[75] Even though the video was "intentionally messy" to signal its inauthenticity, many people who viewed the video believed it to be real.[76] This video could have damaged the reputation of President Trump with some citizens and may have influenced voters' choices in the 2020 election. Deepfakes of this nature, especially if their falsity remains undiscovered, could lead citizens to distrust their government.[77]

II. FIRST AMENDMENT THEORIES AND DEEPFAKES

The First Amendment's protection of free speech is vital to the exchange of ideas in American society.[78] This Part will review the justifications and theories behind the First Amendment and analyze where the rationale for allowing deepfakes may fall among the different justifications.

A. Justifications for First Amendment Protections

Free speech protections under the First Amendment are at the core of our democracy.[79] "Free speech has been thought to serve three principal values: advancing knowledge and 'truth' in the 'marketplace of ideas,' facilitating representative democracy and self-government, and promoting individual autonomy, self-expression and self-fulfillment."[80] One dominant theory behind permitting deepfakes under the First Amendment is the marketplace of ideas.[81] The theory posits that false ideas should be allowed because they are needed for true ideas to emerge and that false ideas will

---

[72] Hall, *supra* note 33, at 59.

[73] *See id.* at 59–60.

[74] *See id.* at 60.

[75] *Id.*

[76] *Id.*

[77] *See id.* at 59–60.

[78] *See* Whitney v. California, 274 U.S. 357, 375–77 (1927) (Brandeis, J., concurring), *overruled in part by* Brandenburg v. Ohio, 395 U.S. 444 (1969) (discussing the Founding Fathers' belief that free speech was necessary for discussion and protection against false speech).

[79] *See id.*

[80] KATHLEEN M. SULLIVAN & NOAH FELDMAN, FIRST AMENDMENT LAW 5 (6th ed. 2016).

[81] *See* Hall, *supra* note 33, at 63–64.

be driven out by true ideas in the long run.[82] However, it can take time for true ideas to expose false ones because "[t]he problem is that the short run may be very long, that one short run follows hard upon another, and that we may become overwhelmed by the inexhaustible supply of freshly minted, often very seductive, false ideas."[83] Given the rate deepfakes can be created and distributed, they certainly have the ability to become an inexhaustible supply.[84]

B. First Amendment Theories—Why Deepfakes Cannot Coexist

The marketplace of ideas argument for free speech protections is ingrained in First Amendment jurisprudence, for "freedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth."[85] The Founding Fathers urged that conversation and counterspeech are the proper responses to false or harmful speech and that government suppression of ideas is a dangerous step toward tyranny.[86] However, the discovery of truth can become difficult when people may not be able to believe what they see: a phenomenon the Founders could not have anticipated.[87] Following the deepfake made about him by Jordan Peele, former President Barack Obama commented on the nature of deepfakes stating, "[t]he marketplace of ideas that is the basis of our democratic practice has difficulty working if we don't have some common baseline of what's true and what's not."[88]

Given the convincing nature of deepfakes, it will likely take a long time for true ideas to stamp out the false videos.[89] By that time, the damage will probably already be done and there may even be another deepfake to combat.[90] The "reliance on counterspeech is increasingly ineffectual and potentially damaging to democracy."[91] Hany

---

[82] *See* Harry H. Wellington, *On Freedom of Expression*, 88 YALE L.J. 1105, 1130 (1979).

[83] *Id.*

[84] *See* Hall, *supra* note 33, at 51–52 (discussing the advancement of deepfake technology and the increasing ability of the general public to create and disseminate fake videos).

[85] Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring), *overruled in part by* Brandenburg v. Ohio, 395 U.S. 444 (1969) (explaining the First Amendment beliefs of the Founding Fathers).

[86] *See id.*

[87] *See* Hall, *supra* note 33, at 59 (noting that the United States Defense Advanced Research Projects Agency is aware of this problem and has funded a media forensics project to aid the public in making the correct determinations about fake videos).

[88] Harwell, *supra* note 2.

[89] *See id.* (noting the difficulties of evaluating which videos are fake); Chesney & Citron, *supra* note 2, at 1777–78 (discussing how the "large-scale erosion of public faith" in empirical evidence has made it difficult for truthful information to surface in the democratic discourse).

[90] *See* Hall, *supra* note 33, at 59 ("The pervasiveness and ease of the technology could mean substantial numbers of deceptive videos in the marketplace that the government is ill-prepared to deal with.").

[91] *Id.* at 69 (quoting Philip M. Napoli, *What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMMC'NS L.J. 55, 97 (2018)).

Farid, a professor of computer science at the University of California at Berkeley, stated that "[t]he number of people working on the video-synthesis side, as opposed to the detector side, is 100 to 1."[92] In fact, "a disinformation campaign using deepfake videos probably would catch fire because of the reward structure of the modern Web, in which shocking material drives bigger audiences—and can spread further and faster than the truth."[93]

The marketplace of ideas theory may not be an adequate justification for deepfakes because true counterspeech is not enough to reveal their falsity, given their realistic appearance.[94] The theory fails because the ability to fact-check the validity of the content of a deepfake is incredibly difficult, even near impossible.[95] For example, in *United States v. Alvarez*, the lie that Alvarez told—claiming that he won the Congressional Medal of Honor—could easily be verified.[96] The list of award winners could be checked, and people could easily call Alvarez out for his false statement.[97] Deepfakes, on the other hand, are hard to prove false because of the technology and expertise needed to detect them.[98] It is much easier to prove that someone said something false, rather than prove that a convincingly realistic video is fake.

Deepfakes also fail under the First Amendment justifications of personal autonomy and self-government. Voters generally have the responsibility to thoroughly research political candidates to make an informed voting choice, which includes choosing the sources from which they glean information.[99] If a voter reads conflicting information about a candidate, it is his duty to further research the candidate's platform to discern what is true.[100] Deepfakes, however, make this task nearly impossible because there is virtually no way for a voter to confirm whether the deepfake has portrayed accurate information.[101] Deepfakes, in essence, rob voters of their personal autonomy to make educated, informed decisions concerning candidates because voters are unable to determine if the speech that they watched and heard a candidate deliver was real.[102] The First Amendment cannot serve the democratic process well if voters must make

---

[92] Harwell, *supra* note 2.

[93] *Id.* (discussing a worry of Rachel Thomas, a co-founder of the machine-learning lab Fast.ai).

[94] *See* Hall, *supra* note 33, at 52.

[95] *See* Harwell, *supra* note 2 (discussing the problems with identifying deepfake videos and how the seemingly "insurmountable" challenge "has led some researchers to instead pursue an authentication system that would fingerprint footage right as it's captured").

[96] 567 U.S. 709, 726–27, 729 (2012).

[97] *Id.* at 727, 729.

[98] *See* Harwell, *supra* note 2; Green, *supra* note 61, at 1458.

[99] *See* Green, *supra* note 61, at 1458.

[100] *See id.*

[101] *See id.* ("The voter is not only tasked with determining the authenticity of the counterfeited candidate speech (difficult even for forensic computer scientists); the voter must put all speech from all candidates to an authenticity test if counterfeit campaign speech is allowed to run rampant.").

[102] *See id.* at 1457–58.

decisions based on false information. Thus, deepfakes run afoul of the justifications for freedom of speech.

III. CURRENT UNPROTECTED CONTENT CATEGORIES

This Part will outline the current unprotected content categories of speech and analyze how deepfakes compare with those categories. Ultimately, this Note will argue that deepfakes, while bearing some similarities to defamation and fraud, are still categorically different from any of the current unprotected areas of speech.

A. Overview of Categories

The current categories of unprotected speech that have been established by the Supreme Court are obscenity,[103] defamation,[104] fraud,[105] incitement,[106] fighting words,[107] true threats,[108] speech integral to criminal conduct,[109] and child pornography.[110] These limited categories, developed through various instances of harmful speech in society, show that the Court is reluctant to restrict freedom of speech.[111] The exclusion of incitement from protected speech, for example, arose in response to a KKK rally that resulted in a violation of Ohio's Criminal Syndicalism Act, which was enacted during the First Red Scare.[112] The participants in the rally publicly advocated for violence coupled with derogatory references to minority groups.[113] Ohio's Criminal Syndicalism Act prohibited advocacy of violence for the purpose of political or industrial reform.[114] The Court determined that the statute was unconstitutional and expressed that "free speech and free press do not permit a State to forbid or proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."[115] Advocating for violence against minorities is clearly harmful speech, but the

---

[103] *See* Miller v. California, 413 U.S. 15, 24 (1973).
[104] *See* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).
[105] *See* Illinois *ex rel*. Madigan v. Telemarketing Assocs., 538 U.S. 600, 612 (2003).
[106] *See* Brandenburg v. Ohio, 395 U.S. 444, 447–48 (1969) (per curiam).
[107] *See* Chaplinsky v. New Hampshire, 315 U.S. 568, 573–74 (1942).
[108] *See* Virginia v. Black, 538 U.S. 343, 359–60 (2003).
[109] *See* United States v. Williams, 553 U.S. 285, 297–98 (2008).
[110] *See* New York v. Ferber, 458 U.S. 747, 764 (1982).
[111] *See Black*, 538 U.S. at 358–59 (discussing the Court's hesitance to restrict freedom of speech, allowing protection of "even ideas that the overwhelming majority of people might find distasteful or discomforting" and instead only restricting speech in cases where there were potential societal harms in allowing the speech).
[112] *See* Brandenburg v. Ohio, 395 U.S. 444, 444–45, 447 (1969) (per curiam).
[113] *See id.* at 446.
[114] *Id.* at 448.
[115] *Id.* at 447.

Court still stated that such speech was protected unless it rose to the level of incitement.[116] The incitement test illustrates the stringent scrutiny that speech restrictions are subject to and the heightened importance the Court places on freedom of speech.[117]

B. Process for Determining Unprotected Content Categories

The Supreme Court has rejected the notion of using an ad hoc balancing standard alone to declare a category of speech to be unprotected.[118] The government cannot prohibit a type of speech simply because it determines that the speech is valueless or that the damage of the speech outweighs its societal benefits.[119] The Court seems to take a number of factors into consideration when prohibiting speech, one such factor being a reasonable person standard.[120] When obscenity was determined to be unprotected in *Miller v. California*, the Court stated that part of the test was "whether 'the average person, applying contemporary community standards' would find that the work, taken as a whole, appeals to the prurient interest."[121] In other words, a reasonable person must find the speech in question to be offensive and not in line with the sexual interest of the community in order for it to be obscene.[122] This standard represents the fact that the government cannot be the only party to find the speech valueless.[123]

Another factor is the harm that the speech creates.[124] In all of the unprotected categories, the subject of the speech or society generally is likely to be harmed by the speech.[125] A state's prohibition of child pornography in *New York v. Ferber*, for example, was not based on categorical balancing alone; there was also the compelling interest of protecting children.[126] The arguments for any potential value of the speech were moot because the harm to children resulting from child pornography was far greater.[127] This point is also easily illustrated by the restriction of speech that

---

[116] *See id.* at 447–48.

[117] *See id.*

[118] United States v. Stevens, 559 U.S. 460, 470 (2010) ("The First Amendment's guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits.").

[119] *See id.* at 471.

[120] *See, e.g.*, Miller v. California, 413 U.S. 15, 24 (1973).

[121] *Id.* (quoting Kois v. Wisconsin, 408 U.S. 229, 230 (1972)).

[122] *See id.*

[123] *See id.* (stating that the standard looks to "the average person" rather than another entity such as the government).

[124] *See* New York v. Ferber, 458 U.S. 747, 763–64 (1982) (holding that child pornography was not protected under the First Amendment as there was a compelling interest of protecting children from harm).

[125] *See* Virginia v. Black, 538 U.S. 343, 358–59 (2003) (noting that speech restrictions under the First Amendment weigh the social value of truth against the societal interests of order and morality (citing R.A.V. v. City of St. Paul, 505 U.S. 377, 382–83 (1992))).

[126] *See* 458 U.S. at 763–64.

[127] *See id.* ("[I]t is not rare that a content-based classification of speech has been accepted

constitutes a true threat.[128] This type of speech "encompass[es] those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."[129] Clearly there is a strong interest in protecting people from violence or the fear of violence, and harm is a consideration in banning true threats.[130]

A third consideration is the narrow tailoring of a speech regulation.[131] In order for a speech regulation to be valid, it must not be overbroad or vague.[132] The phrase "sexual conduct," for example, is too broad and vague for a statute attempting to prohibit obscenity.[133] The state law would need to narrowly define what the phrase means and what specific behavior falls under it.[134] Otherwise, protected speech would be swept in under the statute.[135]

There may be other categories of unprotected speech, but the Court has not recognized a new category since child pornography in *Ferber* in 1982.[136] The Court will likely require "persuasive" evidence to consider a new category[137] in addition to the restrictions being part of a long tradition, even if thus far unrecognized.[138]

C. Evaluating Deepfakes Against Unprotected Content

Justifications for prohibiting deepfakes are similar to those for prohibiting defamation.[139] In *New York Times Co. v. Sullivan*, the *New York Times* published an advertisement about the civil rights student movement.[140] An elected commissioner

---

because it may be appropriately generalized that within the confines of the given classification, the evil to be restricted so overwhelmingly outweighs the expressive interests, if any, at stake, that no process of case-by-case adjudication is required.").

[128] *See Black*, 538 U.S. at 363 (allowing states to ban "those forms of intimidation that are most likely to inspire fear of bodily harm").

[129] *Id.* at 359.

[130] *See id.* (discussing true threats and other categories of free speech that are limited due to the potential violence they could cause).

[131] *See* Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942) (noting that the classes of unprotected speech are "well-defined and narrowly limited").

[132] *See* United States v. Stevens, 559 U.S. 460, 473 (2010) (noting that a law regulating speech may be invalidated if it is overbroad).

[133] Miller v. California, 413 U.S. 15, 23–24 (1973).

[134] *See id.* at 23–26.

[135] *See id.* at 26–27.

[136] *See generally* New York v. Ferber, 458 U.S. 747 (1982).

[137] *See Stevens*, 559 U.S. at 471 (quoting Osborne v. Ohio, 495 U.S. 103, 110 (1990)) (noting that the Court found the argument in *Ferber* that child pornography was "intrinsically related" to child abuse "persuasive" (first quoting *Ferber*, 458 U.S. at 759, 761; and then quoting *Osborne*, 495 U.S. at 110)).

[138] *See id.* at 471–72.

[139] *Compare* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964), *with* Erik Gerstner, *Face/Off: "DeepFake" Face Swaps and Privacy Laws*, 87 DEF. COUNS. J., Jan. 2020, at 1, 6.

[140] 376 U.S. at 256–57.

of Montgomery, Alabama, brought a civil libel action against the newspaper because he claimed that it had printed libelous statements, such as the incorrect number of times that Martin Luther King, Jr. was arrested.[141] It is true that some of the statements were inaccurate, but it did not amount to defamation because the statements were about a public official and not made with actual malice.[142] The Court stated that the protection of the First Amendment requires "a federal rule that prohibits a public official from recovering damages for a defamatory falsehood . . . unless he proves that the statement was made with 'actual malice'—that is, with knowledge that it was false or with reckless disregard of whether it was false or not."[143] Similar to defamation, many deepfakes are created with actual malice and intent to cause harm to the subject of the video's reputation or character.[144] Given the sinister nature of many deepfakes, particularly those that are created with the intent to influence election outcomes, the actual malice standard seems to have been met.[145]

Deepfakes should be their own category of unprotected speech separate from defamation, however, for justifications similar to those separating the category of child pornography from obscenity.[146] Just as child pornography was designated as a separate content category from obscenity because of its unique harm to children,[147] deepfakes should be in a category separate from defamation because of their unique harm to the political and democratic processes.[148] Revisiting *Sullivan*, the fact that the *New York Times* misstated some facts created little harm and the facts were easily verified.[149] The reputation of Sullivan likely did not suffer at length.[150] Deepfakes, on the other hand, are difficult to detect without advanced technology and can easily reach millions of viewers,[151] necessitating a protection greater than the actual malice standard for defamation claims concerning public officials.[152]

Justifications for prohibiting deepfakes are also similar to those for prohibiting fraud.[153] In *Illinois ex rel. Madigan v. Telemarketing Associates*, telemarketers had a

---

[141] *See id.* at 256, 258.

[142] *See id.* at 286.

[143] *Id.* at 279–80.

[144] *See* Gerstner, *supra* note 139, at 6.

[145] *See id.*; *see also* Grace Shao, *Fake Videos Could be the Next Big Problem in the 2020 Elections*, CNBC, https://www.cnbc.com/2019/10/15/deepfakes-could-be-problem-for-the-2020-election.html [https://perma.cc/KCM2-UP7C] (Jan. 17, 2020, 2:49 AM).

[146] *See* New York v. Ferber, 458 U.S. 747, 757 (1982).

[147] *See id.*

[148] *See* Shao, *supra* note 145.

[149] *See* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 260 (1964).

[150] *See id.* at 260.

[151] *See* Harwell, *supra* note 2.

[152] *See id.*

[153] *Compare* Illinois *ex rel.* Madigan v. Telemarketing Assocs., 538 U.S. 600, 623–24 (2003), *with* Soule, *supra* note 1 (The Court's determination that fraud was unprotected because of the intent to deceive is similar to the reasoning lawmakers have posited for prohibiting deepfakes; their intent is to deceive voters.).

fundraising contract with a "charitable nonprofit corporation organized to advance the welfare of Vietnam veterans; under the contracts, the fundraisers were to retain 85 percent of the proceeds of their fundraising endeavors."[154] In reality, less than fifteen cents per dollar was donated to the charity and the rest went primarily to the personal benefit of the fundraisers.[155] Although the Supreme Court had previously rejected state statutes enacted in an effort to prevent fraud that banned solicitations when fundraising costs were high,[156] it determined that fraud itself was not protected speech under the First Amendment.[157] The Court reasoned that "[f]raud actions so tailored, targeting misleading affirmative representations about how donations will be used, are plainly distinguishable" from blanket bans on soliciting funds.[158] The Court determined that "[s]tates may maintain fraud actions when fundraisers make false or misleading representations designed to deceive donors about how their donations will be used."[159]

Deepfakes of political candidates and officials most certainly contain false or misleading representations, making them similar to fraud.[160] However, deepfakes pose an even greater threat to society than misrepresentations about how charitable donations will be used because of their ability to influence voter choice and citizens' views of political officials.[161] Additionally, deepfakes are more difficult to track and can reach a larger sphere of people through social media.[162] Therefore, deepfakes should constitute their own content category separate from the category of fraud because of their unique danger.[163]

IV. DEEPFAKES: A NEW CONTENT CATEGORY

Given the dangers that deepfakes present, it may be time for the Supreme Court to prohibit them as a new content category of unprotected speech under the First Amendment.[164] Some regulations and solutions have been implemented or suggested to combat deepfakes, but these solutions are not enough to truly counteract their dangerous effects.[165]

---

[154] 538 U.S. at 605.

[155] *Id.* at 605–06.

[156] *Id.* at 617.

[157] *Id.* at 612.

[158] *Id.* at 619.

[159] *Id.* at 624.

[160] *See id.*; Harwell, *supra* note 2.

[161] *Compare* Illinois *ex rel.* Madigan, 538 U.S. at 624, *with* Mihalcik, *supra* note 3 (explaining how deepfakes may influence voters' political choices, which are fundamental to our democracy, giving them a larger reach than fraud).

[162] *See* Harwell, *supra* note 2.

[163] *See* Mihalcik, *supra* note 3; Harwell, *supra* note 2.

[164] *See* Christopher, *supra* note 10.

[165] *See* Rich Haridy, *California Bans Political Deepfake Videos Ahead of 2020 Elections*, NEW ATLAS (Oct. 7, 2019), https://newatlas.com/computers/california-bans-political-deep

A. California Statute

Responding to concerns about the effects of deepfakes on elections, California Assemblyman Marc Berman, chair of the Assembly's election committee, proposed a bill that would prohibit the dissemination of deepfakes concerning a candidate within sixty days of an election.[166] The bill goes on to explain that a candidate may seek injunctive relief if such manipulated media is released.[167] After viewing the deepfake of former President Barack Obama created by Jordan Peele, Berman stated, "I immediately realized, 'Wow, this is a technology that plays right into the hands of people who are trying to influence our elections like we saw in 2016.'"[168] The governor signed the bill on October 3, 2019, and it went into effect in 2020 with exemptions for broadcasting outlets and satire or parody.[169] Berman recognized that "technological developments make it increasingly difficult to sort fake from real news," which prompted him to find a way to neutralize manipulated media.[170]

The law seems to pass constitutional muster under strict scrutiny, the standard for content-based regulations, which requires a compelling government interest and a narrowly tailored statute.[171] California certainly has a compelling interest in protecting the integrity of its elections.[172] This statute is also narrowly tailored to fit deepfakes targeting political candidates around elections, which have the potential to influence voters, and provides limited exceptions to the prohibition.[173] The statute is a step in the right direction but only applies to California residents.[174] It is also unclear as to whether a person would violate the statute simply by sharing the video on social media, even if he did not create the deepfake.[175]

The California statute bears some similarity to the issue in *Citizens United v. FEC*.[176] In that case, Citizens United, a nonprofit group, produced a documentary

---

fake-videos-2020-elections/ [https://perma.cc/8S5Z-F46G] (discussing the apprehension surrounding the enactment of new deepfake laws).

[166] A.B. 730, 2019–2020 Reg. Sess., ch. 493 (Cal. 2019); *see also id.*, legislative counsel's digest ("This bill would, until January 1, 2023, instead prohibit a person, committee, or other entity, within 60 days of an election at which a candidate for elective office will appear on the ballot, from distributing with actual malice materially deceptive audio or visual media of the candidate with the intent to injure the candidate's reputation or to deceive a voter into voting for or against the candidate, unless the media includes a disclosure stating that the media has been manipulated.").

[167] Cal. A.B. 730.

[168] Christopher, *supra* note 10.

[169] Cal. A.B. 730.

[170] Christopher, *supra* note 10.

[171] *See* United States v. Alvarez, 567 U.S. 709, 710 (2012) (plurality opinion).

[172] *See* Christopher, *supra* note 10.

[173] *See* Cal. A.B. 730.

[174] *See id.*

[175] *See id.* (prohibiting the *distribution*, instead of the creation).

[176] *See generally* 558 U.S. 310 (2010).

criticizing then-Senator Hillary Clinton when she was a Democratic presidential nominee.[177] Section 203 of the Bipartisan Campaign Reform Act of 2002 (BCRA) specifically "prohibit[ed] corporations . . . from using their general treasury funds to make independent expenditures for speech defined as an 'electioneering communication' or for speech expressly advocating the election or defeat of a candidate."[178] An electioneering communication is "any broadcast, cable, or satellite communication" that depicts a candidate running for federal office made within thirty days of a primary election or sixty days of a general election.[179] Citizens United wanted to release the documentary to the public within thirty days of the 2008 primary election.[180] The Court determined that there was "no reasonable interpretation of *Hillary* other than as an appeal to vote against Senator Clinton,"[181] but determined that BCRA was unconstitutional in its application to Citizen United's First Amendment right to political speech.[182]

The California statute's prohibition of the dissemination of deepfakes sixty days before an election is quite similar to BCRA's prohibition of electioneering communications thirty days before a primary election and sixty days before a general election.[183] In both instances, the laws attempt to protect the integrity of the election process.[184] However, BCRA was unconstitutional in its application because it restricted political speech,[185] whereas the California statute is seeking to restrict the spread of false information.[186] Deepfakes differ from simply advocating for or against a political candidate; they seek to pass off false, and typically damaging, information as truth.[187] Deepfakes can be detrimental even beyond a thirty- or sixty-day prohibition because of the lasting negative effects of voter deception.[188] Even though it has met some backlash from a couple of groups,[189] the California statute seems to pass constitutional muster because it focuses on curbing the intentional dissemination of false information that purposefully confuses voters and does not seek to infringe on political speech.[190]

---

[177] *Id.* at 319, 325.

[178] *Id.* at 318–19 (quoting 2 U.S.C. § 441(b)).

[179] *Id.* at 321 (quoting 2 U.S.C. § 434(f)(3)(A)).

[180] *Id.*

[181] *Id.* at 326.

[182] *Id.* at 353.

[183] *Compare* 52 U.S.C.A. §§ 30118, 30104(f)(3)(A) (West), *with* A.B. 730, 2019–2020 Reg. Sess., ch. 493 (Cal. 2019). Both the federal and California laws seek to limit "targeted communications" that will reach a large number of people and thereby influence voter choice. *Id.*

[184] *See Citizens United*, 558 U.S. at 321; Cal. A.B. 730.

[185] *Citizens United*, 558 U.S. at 353.

[186] Cal. A.B. 730; Christopher, *supra* note 10.

[187] *See* Mihalcik, *supra* note 3.

[188] *See* Green, *supra* note 61, at 1457–58.

[189] *See* Haridy, *supra* note 165.

[190] *See* Christopher, *supra* note 10.

B. DEEPFAKES Accountability Act

The Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act ("DEEPFAKES Accountability Act") was introduced in Congress by Representative Yvette Clarke in June 2019.[191] This bill "requires mandatory labeling, watermarking, or audio disclosures for all 'advanced technological false personation records.'"[192] Such records are defined in the bill as "any media that falsely appears to depict speech or conduct of any person engaged in 'material activity,' created via any technical means, that a reasonable person would believe to be authentic, and that was created without the consent of the person depicted."[193] Additionally, violations of the proposed law may result in up to $150,000 in civil penalties and possible criminal penalties for "violations intended not only to harass, incite violence, interfere in an election, or perpetuate fraud, but also to 'humiliate' the person depicted."[194] In essence, the media creator must have a malicious intent.[195] Furthermore, "[t]he act also establishes a right on the part of victims of synthetic media to sue the creators and/or otherwise 'vindicate their reputations' in court."[196]

One benefit to this proposed law, similar to that of the California statute, is that it provides a legal remedy for those who are harmed by deepfakes.[197] Another benefit is that the law places "unauthorized digital recreations of people under the umbrella of unlawful impersonation statutes."[198] Analogizing deepfakes to impersonation can help guide courts as they navigate the issues surrounding deepfakes, because they are a new problem.[199] Although the proposed law has some flaws, it seems to be a step in the right direction toward deepfake regulation.[200]

C. Considerations from Ferber

In *New York v. Ferber*, the New York statute in question "prohibit[ed] persons from knowingly promoting sexual performances by children under the age of 16 by

---

[191] H.R. 3230, 116th Cong. (2019).

[192] Hayley Tsukayama, India McKinney & Jamie Williams, *Congress Should Not Rush to Regulate Deepfakes*, ELEC. FRONTIER FOUND. (June 24, 2019) (quoting H.R. 3230), https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes [https://perma.cc/9YE4-UMM8].

[193] *Id.* (quoting H.R. 3230).

[194] *Id.* (quoting H.R. 3230).

[195] *See* Devin Coldewey, *DEEPFAKES Accountability Act Would Impose Unenforceable Rules—But It's a Start*, TECHCRUNCH (June 13, 2019, 3:25 PM), https://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/ [https://perma.cc/5JRH-9ZNE].

[196] *Id.* (quoting H.R. 3230).

[197] *Id.*

[198] *Id.*

[199] *See id.*

[200] *See id.*

distributing material which depicts such performances."[201] Ferber was convicted under this statute for selling films showing underage boys masturbating.[202] The Court determined that this statute did not violate the First Amendment, reasoning that the "[s]tates are entitled to greater leeway in the regulation of pornographic depictions of children" because of the harm to the "physiological, emotional, and mental health of the child."[203] The Court even went so far as to say that the value of these materials is "exceedingly modest, if not *de minimis*."[204] The reasoning behind prohibiting deepfakes has some similarity to that of child pornography.[205] Distributing false information that appears to be true to damage the reputation of political officials or influence voters is harmful to the democratic process and may change the outcome of elections or people's view of officials.[206] Deepfakes are not just defamatory to the person that they depict, they are also harmful to society as a whole.[207] Similar to child pornography in *Ferber*, the value of disseminating intentionally false information as truth is also "exceedingly modest," and likely "*de minimis*."[208] Deepfakes also differ from buying a tabloid magazine or seeking out a fortune teller, where the person is "looking to be lied to as entertainment."[209] In those instances, the consumers understand that they are receiving false information, whereas deepfakes, in most instances, provide information that people rely on as true.[210] Thus, deepfakes present a danger worthy of prohibition.

D. Breyer Concurrence in Alvarez

The Supreme Court determined that false statements were protected speech in *United States v. Alvarez*.[211] In that case, the Court examined the Stolen Valor Act, which made it a crime to falsely claim receipt of military decorations or medals, with an enhanced penalty for claims involving the Congressional Medal of Honor.[212] Alvarez violated the Stolen Valor Act by falsely claiming that he won the Congressional Medal of Honor, and he argued that the Act violated his free speech rights under the First Amendment.[213] The Court held that the Act violated the First Amendment,

---

[201] 458 U.S. 747, 749 (1982).
[202] *Id.* at 751–52.
[203] *Id.* at 756, 758.
[204] *Id.* at 762.
[205] *See* Green, *supra* note 61, at 1466–67 (drawing similarities between the "real harm" to children resulting from child pornography and the "harm to living, identifiable humans—to candidates" from counterfeit campaign speech).
[206] *See id.* at 1457, 1463–64.
[207] *See id.* at 1457.
[208] *See Ferber*, 458 U.S. at 762.
[209] Green, *supra* note 61, at 1468.
[210] *See id.*
[211] 567 U.S. 709, 718, 730 (2012) (plurality opinion).
[212] *Id.* at 715–16.
[213] *Id.* at 715.

reasoning that false statements were a protected form of speech and the Act was an unconstitutional content-based speech restriction.[214] The government failed to show that false statements should comprise a new content category.[215] Additionally, the Act was not narrowly tailored enough, failing strict scrutiny, because it sought "to control and suppress all false statements on this one subject in almost limitless times and settings. . . . [W]ithout regard to whether the lie was made for the purpose of material gain."[216] False statements on their own are not enough to make the speech unprotected.[217] Particularly, false statements about matters of public concern should be protected to foster the marketplace of ideas.[218]

In his *Alvarez* concurrence, Justice Breyer agreed with the Court's holding but posited that intermediate scrutiny would be a more appropriate standard than strict scrutiny in this case.[219] Justice Breyer argued that the Act could be constitutional if it were narrower in its scope.[220] He recognized that false statements are deserving of some protection, as they can "prevent embarrassment, protect privacy," or even "promote a form of thought that ultimately helps realize the truth."[221] However, the government has an important objective in prohibiting false statements about military decorations because of the confusion they create about who has earned the honor, which diminishes the value of the award.[222] The Stolen Valor Act may have been constitutional if it had been narrowly tailored to "insist upon a showing that the false statement caused specific harm or at least was material, or focus its coverage on lies most likely to be harmful or on contexts where such lies are most likely to cause harm."[223]

Similar to the interest of protecting the integrity of military decorations through the Stolen Valor Act, the government would also have a legitimate interest in prohibiting the distribution of deepfakes.[224] A narrowly tailored law limiting the distribution of deepfakes—specifically depicting political figures—would likely pass constitutional muster under the First Amendment, particularly if it required a showing of material harm.[225] California has already achieved this with its new law prohibiting the dissemination of deepfakes within sixty days of an election,[226] but

---

[214] *Id.* at 724, 729–30.

[215] *Id.* at 718–22.

[216] *Id.* at 722–23.

[217] *Id.* at 721–22.

[218] *See* Marc Jonathan Blitz, *Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 66–67 (2018).

[219] *Alvarez*, 567 U.S. at 730 (Breyer, J., concurring) ("Ultimately the Court has had to determine whether the statute works speech-related harm that is out of proportion to its justifications.").

[220] *Id.* at 737.

[221] *Id.* at 733.

[222] *Id.* at 735.

[223] *Id.* at 738.

[224] *See id.*

[225] *See id.*

[226] *See* Christopher, *supra* note 10.

if Congress implemented a federal law that prohibited deepfakes of this nature entirely, it would likely be a more effective way to combat deepfakes.[227]

E. Special Virulence from R.A.V.

The Supreme Court in *R.A.V. v. City of St. Paul* reiterated that speech restrictions may only prohibit unprotected categories of speech and that content-based restrictions cannot be substantially overbroad.[228] In that case, petitioner R.A.V. burned a cross on an African American family's lawn and was charged under the Bias-Motivated Crime Ordinance of St. Paul, which prohibited the display of a "symbol, . . . which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender."[229] The Court held that the St. Paul ordinance violated the First Amendment because it was overbroad; although it appeared to be aimed at fighting words, it prohibited "speech solely on the basis of the subjects the speech addresses."[230] The Court reasoned that the government may not regulate the use of fighting words "based on hostility—or favoritism—towards the underlying message expressed."[231]

The majority expressed that there is an exception to this rule when the proscribed speech is especially harmful.[232] Essentially, the exception applies when content-based speech presents "special risks" that the government can suppress.[233] Content-based restrictions are typically prohibited to prevent the government from favoring one viewpoint over another, but "content discrimination among various instances of a class of proscribable speech often does not pose this threat."[234] Thus, "the particular virulence exception . . . cover[s] prohibitions that are not clearly associated with a particular viewpoint."[235] This principle is illustrated by the federal government's ability to criminalize threats that are specifically against the President of the United States.[236]

---

[227] *See id.*

[228] *See generally* 505 U.S. 377 (1992).

[229] *Id.* at 379–80 (quoting ST. PAUL, MINN., LEGIS. CODE § 292.02 (1990), *invalidated by id.* at 377).

[230] *Id.* at 381.

[231] *Id.* at 386.

[232] *See id.* at 388 ("When the basis for the content discrimination consists entirely of the very reason the entire class of speech at issue is proscribable, no significant danger of idea or viewpoint discrimination exists. Such a reason, having been adjudged neutral enough to support exclusion of the entire class of speech from First Amendment protection, is also neutral enough to form the basis of distinction within the class.").

[233] *See* Virginia v. Black, 538 U.S. 343, 384 (2003) (Souter, J., concurring in part and dissenting in part).

[234] *R.A.V.*, 505 U.S. at 388.

[235] *Black*, 538 U.S. at 384.

[236] *See* Watts v. United States, 394 U.S. 705, 707 (1969) (per curiam).

A statute prohibiting threats of violence against the president is a constitutional, content-based restriction because of the special importance of protecting the president.[237] It does not favor one viewpoint over the other.[238] Similarly, a statute specifically banning deepfakes of political candidates and officials would have a special importance and be a valid content-based restriction.[239] Protecting the president so that he can effectively perform his duties and ensure smooth leadership of the country is an important governmental interest that is furthered by a ban of threats of violence specifically against the president.[240] Protecting the integrity of the political process in elections and reputation of government officials is also an important governmental interest that would be furthered by a specific ban on deepfakes.[241] Such a statute would not be discriminating against a specific viewpoint and would be aimed at combatting the dangers of deepfakes.[242]

F. False Campaign Speech

Regulating deepfakes has similar reasoning to that of regulating false campaign speech generally.[243] False campaign speech has the potential to "trick voters into voting for the 'wrong' candidate or voting the 'wrong' way on a ballot measure."[244] "Wrong" in this case refers to an individual voting inconsistently with how he would normally vote because of the false campaign speech.[245] For example, a campaign lie could be that a candidate was accepting bribes.[246] This lie could lead voters who supported that candidate to vote for someone else.[247] Deepfakes certainly share this ability to lead voters to make decisions contrary to their normal voting patterns.[248]

Justifications for regulating false campaign speech include the arguments that this type of speech can manipulate the electoral process, "lower the quality of campaign discourse and debate," cause voters to become apathetic or distrustful of the voting process, and "inflict reputational and emotional injury upon the attacked individual."[249]

---

[237] *See id.*

[238] *See id.* at 708.

[239] *See R.A.V.*, 505 U.S. at 388; *Watts*, 394 U.S. at 707.

[240] *See Watts*, 394 U.S. at 707.

[241] *See* Green, *supra* note 61, at 1460–61.

[242] *See R.A.V.*, 505 U.S. at 388.

[243] *See* Richard L. Hasen, *A Constitutional Right to Lie in Campaigns and Elections?*, 74 MONT. L. REV. 53, 55–56 (2013).

[244] *Id.* at 55.

[245] *Id.* at 55–56.

[246] *Id.* at 55 (discussing McKimm v. Ohio Elections Comm'n, 729 N.E. 2d 364 (Ohio 2000), *cert. denied*, 531 U.S. 1078 (2001)).

[247] *See id.* at 55–56.

[248] *See id.*

[249] *See id.* at 63 (quoting William P. Marshall, *False Campaign Speech and the First Amendment*, 153 U. PA. L. REV. 285, 294, 296 (2004)).

These arguments have also been extended by lawmakers advocating for deepfake regulations.[250]

Protecting voter integrity is surely a compelling interest, but there are concerns surrounding the regulation of false election and campaign speech.[251] One concern rests on "the possibility that these laws will be the subject of manipulation by government authorities who want to favor one side or the other in an election."[252] Another is that "we depend upon the campaigns themselves to allow voters to separate truth from lies and decide how to vote in line with voters' preferences."[253] With fact-checking now a commonplace practice,[254] the marketplace of ideas has a chance to work as a check on false campaign speech.[255] Diligent voters can do research, examine all pertinent information, and decide for themselves what they believe to be the truth.[256] Lies are protected speech because they are still useful in the sense that they add to the debate in the exchange of ideas and allow for true ideas to reveal them as lies.[257] Additionally, false campaign speech often consists of speech made directly by candidates to voters.[258] It is easier to expose this type of lie because a voter could determine whether the statement was consistent with something the candidate said before and it would be relatively easy for the truth to expose it.[259] Deepfakes, on the other hand, while a type of false campaign speech, are different because they convey something that was never said at all.[260] Other types of false campaign speech are still true in the sense that a candidate or campaign actually made the statement.[261] The marketplace of ideas does not have the same opportunity to work because it is incredibly hard to detect and fact-check deepfakes.[262] Therefore, deepfakes require regulation to protect voter integrity.[263]

G. Other Solutions Are Not Enough

Potential solutions outside of government regulation for addressing deepfakes include, "(1) using existing laws, (2) urging additional action from social media companies, (3) developing the technology to detect deepfakes, (4) fostering the use

---

[250] *See* Christopher, *supra* note 10.
[251] *See* Hasen, *supra* note 243, at 56.
[252] *Id.*
[253] *Id.*
[254] *See id.* at 53–54.
[255] *See id.*
[256] *See* Green, *supra* note 61, at 1458.
[257] *See* Wellington, *supra* note 82, at 1130.
[258] *See* Hasen, *supra* note 243, at 58.
[259] *See* United States v. Alvarez, 567 U.S. 709, 726–27 (2012) (plurality opinion).
[260] *See* Harwell, *supra* note 2.
[261] *See Alvarez*, 567 U.S. at 726–27.
[262] *See* Harwell, *supra* note 2.
[263] *See id.*; Hasen, *supra* note 243, at 56.

of private foundations and other organizations to respond to false information, and (5) deploying digital literacy curriculum in schools."[264] These solutions are all practical approaches to deepfakes,[265] but they fall short of adequately combatting the problems that deepfakes present.

First, existing laws are not enough to stop deepfakes because they do not effectively address the dangers that deepfakes pose.[266] Deepfakes bear some similarity to defamation and fraud, but laws concerning those unprotected categories of speech are not strong enough because deepfakes present heightened dangers to elections, the reputation of government officials, and national security.[267] Existing elections laws also fall short of tackling the problems of deepfakes.[268] The Federal Election Campaign Act, for example, prohibits a candidate from fraudulently misrepresenting himself or "any committee or organization under his control."[269] While the Act prohibits misrepresentations by the candidates themselves, it does not address deepfakes made by third parties portraying candidates.[270] Deepfakes, then, cannot be prohibited under this law. Other state election codes prohibit specific instances of false political speech, which do not cover deepfakes unless they fall into one of those specific categories.[271]

Second, urging social media companies to take action is not a reliable measure against deepfakes because companies disagree on the best approach to combat deepfakes, where some will leave the videos on their websites and others will take them down.[272] Facebook, for example, has committed to the preservation of truth by flagging false news, but the social media company has stated that it will not take misrepresentations down from its website.[273] Facebook reasons that it can better preserve freedom of expression by limiting sources spreading false information by decreasing their ability to monetize and distribute content without actually removing the false news.[274] YouTube, on the other hand, quickly takes down deepfakes once they are discovered because of the "deceptive practices" policies they implemented.[275] As previously stated, deception can have a lasting negative impact on the viewers of

---

[264] Hall, *supra* note 33, at 71 (citations omitted).

[265] *See id.*

[266] *See* Harwell, *supra* note 2; Coldewey, *supra* note 195.

[267] *See* Harwell, *supra* note 2.

[268] *See, e.g.*, Green, *supra* note 61, at 1469–76.

[269] *Id.* at 1470 (quoting 52 U.S.C. § 30124).

[270] *Id.*

[271] *Id.* at 1470–71 (noting some of the "highly-specific bans" found in state election codes).

[272] Harwell, *supra* note 2 ("Officials with the Democratic and Republican parties and the nation's top presidential campaigns say they can do little in advance to prepare for the damage, and are counting on social networks and video sites to find and remove the worst fakes. But the tech companies have differing policies on takedowns, and most don't require that uploaded videos must be true.").

[273] Hall, *supra* note 33, at 72–73.

[274] *See id.* at 73.

[275] Harwell, *supra* note 2.

deepfakes,[276] and leaving deepfakes on social media websites where millions of people can view them ultimately furthers the goal of deepfakes.[277] Social media companies are private entities that are not bound by the First Amendment,[278] and they have no affirmative duty to police deepfakes.[279] Because of the ability of social media companies to filter content as they see fit, they are an unreliable solution to combat deepfakes.[280] A new content category of unprotected speech for deepfakes would be a better solution because it would prohibit the creation of the video itself, making it less likely that deepfakes would be circulated across social media and viewed by millions.[281]

Third, developing detection technologies to find deepfakes, while certainly a necessary and crucial step toward combatting deepfakes, is not enough on its own to counteract the dangers of deepfakes. Detection technology is advanced and can pinpoint minute details that reveal the falsity of a video, but skilled video editors are still able to make deepfakes seem real enough to the point where the technology cannot detect subtle errors.[282] Additionally, detection technology must be kept relatively under wraps because "making the system more widely available carries its own threat, by potentially allowing deepfake creators to examine the code and find workarounds."[283]

One recent effort to help researchers detect deepfakes was initiated by Google when the company released 3,000 deepfakes that it created.[284] Google "is hoping their release will help academic researchers and other experts develop new ways to uncover and combat the manipulated videos, potentially providing new tools for the public and for publishers to pinpoint them."[285] Another similar effort was made by Deep Video Portraits, "a company able to make and manipulate facial expressions and head poses of figures," which "released their technology as a means of informing

---

[276] *See, e.g.*, Green, *supra* note 61, at 1463 (discussing the potentially irreversible damage of counterfeit campaign speech).

[277] *See* Hall, *supra* note 33, at 55 (explaining that deepfakes have a greater chance of spreading misinformation to social media users the longer they remain on social media websites).

[278] *See* U.S. CONST. amend. I (prohibiting only actions by federal, state, and local governments).

[279] *See* Hall, *supra* note 33, at 72 (discussing the inadequacies of internet company service agreements in addressing fake news); Harwell, *supra* note 2.

[280] *See* Hall, *supra* note 33, at 72–73.

[281] *See id.* at 54–55 (discussing the role of social media in the dissemination of false information).

[282] Harwell, *supra* note 2 ("Forensic researchers have homed in on a range of subtle indicators that could serve as giveaways, such as the shape of light and shadows, the angles and blurring of facial features, or the softness and weight of clothing and hair. But in some cases, a trained video editor can go through the fake to smooth out possible errors, making it that much harder to assess.").

[283] *Id.*

[284] Carbone, *supra* note 8.

[285] *Id.*

the public of its capabilities and to hopefully deter those who might use it to trick consumers."[286] Both of these efforts by Google and Deep Video Portraits seem beneficial, but they have the potential to "backfire and create just the opposite result as hackers use it to manipulate images and audio of politicians onto anything they want."[287] For example, researchers at the State University of New York at Albany discovered that a lack of blinking was an indicator that the video was a deepfake.[288] Shortly after this discovery, one of the researchers "received an email from a deepfake creator who said they had solved the problem in their latest fakes."[289] Developing tools to detect deepfakes is important, but the tools on their own are not enough to combat deepfakes.[290] If deepfakes are detected but not removed from the internet or social media, then what benefit does detection bring? Even if deepfakes can be detected and removed from the internet, it does not deter people from continuing to create them because there are no laws regulating them.[291] Making detection technology available to the public also increases the risk of deepfake creators becoming more skilled in their craft by learning what aspects the video detection technology pinpoints.[292] Regulation of deepfakes under a new category of unprotected speech would bridge this gap and help avoid the further spread of deepfakes.

Fourth, encouraging private foundations and other organizations to address manipulated media is also inadequate. This solution includes applying industry-developed "private accreditation systems to establish industry standards, limit fraud, and ensure the quality of services or products" to identify misinformation.[293] While encouraging foundations and other organizations to contribute to deepfake detection is beneficial, they are similar to social media companies in that they are private entities with no obligation to police deepfakes, and their reach would be on a much smaller scale than government regulations.[294]

Fifth, implementing digital literacy curriculum in schools would "give children the tools they need to make smart choices online."[295] This curriculum would help students "distinguish between factual and fabricated content" and "critically evaluate the content they are consuming."[296] Deepfakes are already difficult for artificial

---

[286] Swink & Qualls, *supra* note 45.

[287] *Id.*

[288] Harwell, *supra* note 2.

[289] *Id.*

[290] *See, e.g.*, *id.*

[291] *See id.*

[292] *See, e.g.*, *id.* (discussing the email from a deepfake creator to the researchers discovering the lack of blinking in deepfake videos).

[293] Anna Gonzalez & David Schulz, *Helping Truth with Its Boots: Accreditation as an Antidote to Fake News*, 127 YALE L.J.F. 315, 317 (2017).

[294] *See id.* at 323–26 (describing the decentralized structure of journalistic organizations).

[295] David Goldberg, *Responding to "Fake News": Is There an Alternative to Law and Regulation?*, 47 SW. L. REV. 417, 428 (2018) (quoting DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT, INTERNET SAFETY STRATEGY, 2017, Cm. 48, at 26 (UK)).

[296] *Id.*

intelligence and those savvy in technology and videography to identify.[297] Teaching children about deepfakes and giving guidance for navigating the digital world are both important, but these solutions are still not enough to fix the problem or curb the negative effects of deepfakes. Identifying deepfakes is far more challenging than creating them.[298]

V. POTENTIAL ISSUES WITH REGULATION

Opponents of deepfake regulations argue that such regulations are at odds with First Amendment protections of freedom of speech and cannot pass constitutional muster.[299] Critics of California's new statute, for example, have stated that "[t]he law is overbroad, vague, and subjective" and "hinges on whether the deepfake leads to a fundamentally different impression of the candidate, which is not specific enough, and could suppress speech."[300] Essentially, opponents fear that regulations of deepfakes would sweep in protected speech in violation of the First Amendment.[301] The American Civil Liberties Union (ACLU) and the Electronic Frontier Foundation (EFF) have also opposed California's statute by writing to Governor Gavin Newsom, expressing that the "political deepfake law would not solve the problem and may only lead to more confusion."[302]

The EFF also takes issue with the DEEPFAKES Accountability Act.[303] The organization argues that it is "unclear" how provisions of the proposed law, such as mandatory labeling and watermarking, will actually address the problems of deepfakes because those creating them for nefarious purposes are typically anonymous and unlikely to adhere to the provisions of the statute.[304] Additionally, the bill presents some First Amendment concerns because criminal penalties can be imposed without a showing of harm, and it fails to identify who has the burden of proof, which could have a chilling effect on speech.[305] Other critics of the DEEPFAKES Accountability Act emphasize that the labeling and watermarks required by the Act are easy to remove.[306] Offending the First Amendment protection of freedom of speech is always a concern when creating a speech restriction.[307] However, the Court has already found valid

---

[297] *See* Harwell, *supra* note 2.

[298] *See id.*; *see also* Chesney & Citron, *supra* note 2, at 1759.

[299] *See* Will Fischer, *California's Governor Signed New Deepfake Laws for Politics and Porn, But Experts Say They Threaten Free Speech*, BUS. INSIDER (Oct. 10, 2019, 12:51 PM), https://www.businessinsider.com/california-deepfake-laws-politics-porn-free-speech-privacy-experts-2019-10 [https://perma.cc/Y5EC-4YFU].

[300] *Id.*

[301] *See id.*

[302] *Id.*

[303] *See generally* Tsukayama et al., *supra* note 192.

[304] *Id.*

[305] *Id.*

[306] *See* Coldewey, *supra* note 195.

[307] *See* Virginia v. Black, 538 U.S. 343, 358 (2003).

ways to prohibit other categories of speech without violating the First Amendment, and the same constitutional prohibition of speech is possible for deepfakes.[308]

In his article, Marc Blitz argued that "absent legally cognizable harm, fake news and fake science fall squarely in the same category as protected speech," and "[l]ike religious ideas and political opinions, they are staunchly protected against government censorship."[309] He went on to explain that the Justices in *Alvarez* "agreed that where false statements arise in public debate and concern matters where disagreement is an inevitable and desirable part of that debate . . . then the speaker of that falsity should be just as protected as she is when she speaks a truth."[310] He also warns of the danger presented by "letting government exercise coercive authority over the exchange of ideas."[311] In essence, Blitz argues that the harm of regulation outweighs any benefit that it may bring.[312] The concern of government regulation over false speech is certainly valid. The chilling of speech is always a concern when evaluating the merits of speech restrictions. However, the narrow tailoring of any ban of deepfakes would effectively limit the control that the government has over false speech, and the ban would not be viewpoint-based.[313] The problem with the Stolen Valor Act in *Alvarez* was that it was overbroad because it essentially prohibited false statements about receiving military decorations at all times, in any setting, and for any purpose.[314] Justice Breyer pointed out in his concurrence that there was an important, if not compelling, government interest in limiting this kind of false speech, and that purpose may have been achieved with a more narrowly tailored law.[315] The narrow tailoring of deepfake regulations would achieve the compelling interest of protecting the integrity of elections, voter choice, and the reputation of political figures by banning deepfakes.[316]

Other possible issues stem from existing case law.[317] The Supreme Court has determined that those who hold a public office or purposely put themselves in the public eye open themselves up to criticism by the public.[318] However, deepfakes extend far beyond comment and criticism; they create an entirely false perception

---

[308] *See id.* at 358–59 (discussing various types of unprotected speech).

[309] Blitz, *supra* note 218, at 69.

[310] *Id.* at 71.

[311] *Id.* at 83.

[312] *See id.* at 85 ("Even if fake news or false speech raises a genuine problem, government intervention may be an ineffective solution—or one that, even if it succeeds, would bring even worse distortion into public debate or personal reflection than the one it was designed to combat.").

[313] *See supra* Section III.B.

[314] *See* United States v. Alvarez, 567 U.S. 709, 722–23 (2012) (plurality opinion).

[315] *Id.* at 738–39 (Breyer, J., concurring).

[316] *See supra* Part IV.

[317] *See generally* United States v. Stevens, 559 U.S. 460 (2010); Hustler Mag., Inc. v. Falwell, 485 U.S. 46 (1988); N.Y. Times Co. v. Sullivan, 376 U.S. 254 (1964).

[318] *Hustler Mag.*, 485 U.S. at 51 (quoting *Sullivan*, 376 U.S. at 270) (noting public figures and officials are subject to "vehement, caustic, and sometimes unpleasantly sharp attacks").

of the public official or figure.[319] Unlike parody[320]—where viewers understand statements are made in jest—or valid discussions of matters of public concern,[321] deepfakes depict false realities that have the ability to thwart reputations and affect election outcomes.[322] Other arguments include the fact that the Supreme Court has been extremely hesitant to recognize new categories of unprotected speech, particularly absent persuasive evidence[323] and a tradition of such a restriction.[324] However, "misinformation has a long history in our political processes," just in a different form than deepfakes today.[325] During the presidential election of 1800, for example, Thomas Jefferson "used a questionable journalist, James Callendar, to write defamatory pieces about [John] Adams, including an inaccurate story that Adams wanted to start a war with France."[326] The "deceptive propaganda tactic" is old, but the technology is new and allows "ordinary citizens" to generate deceptively realistic information, making deepfakes far more dangerous than stories in a newspaper.[327] The harmful dangers that deepfakes present and the history of deception in political processes warrants consideration by the Supreme Court of restriction through a new content category of unprotected speech.[328]

VI. PROPOSED DEEPFAKE REGULATION

After analyzing the current categories of unprotected speech under the First Amendment and circumstances surrounding the establishment of those categories, it seems fitting to create a new category of unprotected speech for deepfakes. The prohibition should extend further than the new California statute by prohibiting deepfakes generally, instead of just in the couple of months preceding an election, and should apply to the United States as a whole.[329] The category should be narrowly tailored to deepfakes depicting political officials and candidates. Similar to the California statute, the category should have exemptions for videos that are parodies or satire or if the video includes a disclaimer stating that the video contains false information.[330] A law prohibiting deepfakes should also state that the creator

---

[319] *See* Tsukayama et al., *supra* note 192.
[320] *See Hustler Mag.*, 485 U.S. at 50.
[321] *See Sullivan*, 376 U.S. at 256–58.
[322] *See* Harwell, *supra* note 2.
[323] *See* United States v. Stevens, 559 U.S. 460, 471 (2010) (citing Osborne v. Ohio, 495 U.S. 103, 110 (1990)).
[324] *See id.* at 472.
[325] *See* Hall, *supra* note 33, at 53.
[326] *Id.* at 54.
[327] *See id.* at 53.
[328] *See id.*
[329] *Cf.* Fischer, *supra* note 299.
[330] *See id.*

of a deepfake video is the one to be held liable;[331] people that share a deepfake on social media, but who did not create the deepfake, should not be held liable for the dissemination of the video. Otherwise there could be a chilling effect on speech if people stop sharing videos of political figures on social media out of fear of unknowingly sharing a deepfake.[332] The regulation should also specify who bears the burden of proof and require a showing of harm for any criminal sanctions to solve the issues presented with current proposed regulations.[333]

Deepfakes should constitute their own category of unprotected speech because they present a unique and dangerous threat to our election system, the reputations of political officials, and national security.[334] Similar to the heightened importance of protecting children with the prohibition of child pornography, the crucial interest of protecting this country's democratic processes merits the establishment of a new category.[335] False statements have traditionally fallen under protected speech, but they may be prohibited in some instances with a narrowly tailored law when they are particularly harmful.[336] Similar to the interest in preserving the integrity of military decorations by prohibiting false claims of their receipt, there is a compelling interest in prohibiting purposefully deceptive false videos depicting political candidates and officials.[337] Additionally, this content-based regulation would pass constitutional muster because deepfakes may be singled out based on the special importance of curbing negative impacts on the political system.[338] Just as threats against the President of the United States may be specifically prohibited by the federal government, deepfakes may also be specifically banned to further the interest of the smooth functioning of our political system.[339] Regulation may not eliminate deepfakes entirely, but it could reduce the harms that they present and deter people from creating them.

CONCLUSION

In a polarized political environment, an examination of the dangers of deepfakes and their threat to the election process is exceedingly vital.[340] Deepfakes do not fall within any of the justifications for freedom of speech protections under the First

---

[331] *See* Tsukayama et al., *supra* note 192 (discussing the potential chilling effect if the unidentified burden of proof made someone other than the deepfake creator liable).

[332] *See id.* (discussing the potential chilling effect resulting from failing to identify who has the burden of proof).

[333] *See id.*

[334] *See* Harwell, *supra* note 2.

[335] *See* New York v. Ferber, 458 U.S. 747, 757–58 (1982).

[336] *See* United States v. Alvarez, 567 U.S. 709, 734 (2012) (Breyer, J., concurring).

[337] *See id.* at 737.

[338] *See* Watts v. United States, 394 U.S. 705, 707 (1969).

[339] *See id.*

[340] *See* Harwell, *supra* note 2.

Amendment.[341] The marketplace of ideas cannot function properly because citizens' inability to discern if information is true or false stifles true counterspeech.[342] Citizens now have less incentive to be involved with the political process and the debate of new ideas because there is no common starting point for truth.[343] Furthermore, deepfakes present unique issues that set them apart from current categories of unprotected speech. The potential threats to the election process, national security, and the reputation of government officials, coupled with the ability of deepfakes to rapidly spread, presents a greater danger than defamation or fraud.[344] Just as child pornography was given its own content category separate from obscenity because of the unique harm to children,[345] deepfakes should constitute their own category of unprotected speech because of the unique harm they pose to the political process. As technology continues to advance, deepfakes have the potential to become even more convincing and harder to detect.[346] A new content category of unprotected speech regulating deepfakes seems to be the best option. Other proposed solutions fail because they only address parts of the problem.[347] If nothing greater is done to attempt to neutralize their harmful effects, deepfakes will remain a threat to democracy.

---

[341] *See supra* Part II.

[342] *See* SULLIVAN & FELDMAN, *supra* note 80, at 5.

[343] *See* Chesney & Citron, *supra* note 2, at 1777–78 (discussing the public's loss of faith in "basic empirical insights" and suggesting that "[s]ometimes lies erode the factual foundation that ought to inform policy discourse").

[344] *See* Harwell, *supra* note 2.

[345] *See* New York v. Ferber, 458 U.S. 747, 764 (1982).

[346] *See* Harwell, *supra* note 2.

[347] *See supra* Section IV.G.