

2-2022

Deepfake Reckoning: Adapting Modern First Amendment Doctrine to Protect Against the Threat Posed to Democracy

Alyssa Ivancevich

Follow this and additional works at: https://repository.uchastings.edu/hastings_constitutional_law_quaterly



Part of the [Constitutional Law Commons](#)

Recommended Citation

Alyssa Ivancevich, *Deepfake Reckoning: Adapting Modern First Amendment Doctrine to Protect Against the Threat Posed to Democracy*, 49 HASTINGS CONST. L.Q. 61 (2022).

Available at: https://repository.uchastings.edu/hastings_constitutional_law_quaterly/vol49/iss1/5

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Constitutional Law Quarterly by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.

Deepfake Reckoning: Adapting Modern First Amendment Doctrine to Protect Against the Threat Posed to Democracy

BY ALYSSA IVANCEVICH*

TABLE OF CONTENTS

Introduction	61
I. Political Deepfake Problems Outstrip Current Solutions	63
A. The Danger of Manipulated Media	63
B. Present Concerns Regarding Political Speech	64
C. Leading Regulatory Proposals	66
II. Modern First Amendment Doctrine Fails to Adequately Protect Against Harm Incurred by Deepfakes	68
A. Current First Amendment Jurisprudence	68
B. The Challenges of Regulating Political Deepfakes.....	70
III. Recommendations for Adapting First Amendment Doctrine.....	72
A. The Supreme Court Should Adopt a New Categorical Exception for Political Deepfakes.....	72
C. Adopting a Proportionality Approach as an Alternative.....	75
Conclusion.....	77

Introduction

Fake news is just the beginning. Misinformation has graduated to “deepfake media”: advanced and realistic audiovisual communications specifically designed to deceive and inflict harm. Deepfake media created and distributed in the leadup to an election or those that attempt to influence the American electorate present an especially concerning threat to the integrity of our democratic election system. The issue of political deepfakes that attempt to influence voters is particularly salient now in the wake of the 2020 election as political polarization accelerates.¹

* University of California, Hastings College of the Law J.D. Candidate 2022. I offer my sincerest thanks to Professor Zachary Price for his invaluable encouragement and overall guidance

A deepfake video has the potential to confuse voters, incite violent action, influence confidence in a candidate, and manipulate voters' decisions. As artificial intelligence technology advances, manipulated media will become more difficult to detect and distinguish from unaltered media.² Additionally, manipulated media has the potential to create lasting, visceral impressions in voters' minds. However, any regulation of deepfake media must be reconciled with one of America's most cherished rights: the First Amendment freedom of expression. Although Congress has previously attempted to pass legislation that imposes penalties, categorical bans, or require disclosure of manipulated media, these efforts have largely been unsuccessful.³ Moreover, copyright law, privacy torts, and defamation actions have been largely ineffectual.⁴

Political deepfakes are an example of why First Amendment doctrine should adapt to protect against new threats. The harm to society inflicted by political deepfakes are a symptom of a larger problem: the U.S. Supreme Court has so far taken a rigid, hardline approach that will result in harm to our cherished democratic society. This Note argues that the time has finally come for the Supreme Court to evolve from their stubborn First Amendment doctrine in order to adequately protect the integrity of the electoral process that is essential for democracy.

This Note will propose that the Court should expand First Amendment doctrine to include a new, unprotected category specifically for political deepfakes and resultant harm. Part I will introduce the problem that political deepfakes pose for the stability of America's democracy and discuss current and proposed legislation. Part II will evaluate the recent trends in the Supreme Court's First Amendment jurisprudence and outline the challenges of regulation under the First Amendment. Part III will analyze scholarly proposals and recommend the establishment of a new unprotected category or at least a more reasonable judicial approach to evolving technologies. Political deepfakes present a greater threat to democracy than fake news because they are technologically superior. As technology advances, First

throughout the preparation of this Note. I am grateful to my family, and especially my parents, for their endless support. Additional thanks are owed to Professor Wendy Hill for her mentorship in legal writing and to the entire editorial staff at *Hastings Constitutional Law Quarterly*.

1. John Koetsier, *Fake Video Election? Deepfake Videos 'Grew 20X' Since 2019*, FORBES (Sept. 9, 2020), <https://www.forbes.com/sites/johnkoetsier/2020/09/09/fake-video-election-deepfake-videos-grew-20x-since-2019/?sh=3d88fc47148c> (documenting the recent increase in political deepfake media).

2. Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned'*, WASH. POST (June 12, 2019), <https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/> (recognizing the negative effect online information can have on consumer confidence).

3. H.R. 3230, 116th Cong. (2020), <https://www.congress.gov/bill/116th-congress/house-bill/3230>.

4. Shannon Reid, Comment, *The Deepfake Dilemma: Reconciling Privacy And First Amendment Protections*, 23 U. PA. J. CONST. L. 209, 220 (2021).

Amendment doctrine should adapt to protect essential governmental functions that are valued even more than the freedom of expression.

I. Political Deepfake Problems Outstrip Current Solutions

A. The Danger of Manipulated Media

Since the proliferation of “fake news” in the 2016 election, academics and scholars have grown increasingly concerned with the rise in political polarization.⁵ While polarization can be beneficial to a democracy, this extreme acceleration can be harmful if animosity scales too quickly. Political deepfakes are the latest, most extreme kind of fake news media that attempt to directly affect the political electorate.

Deepfakes are manipulated media that have been edited with deep learning artificial intelligence to replace a person’s likeness without their permission.⁶ The vast majority of deepfake media that has been detected online is pornographic in nature, but deepfake video or audio files have the potential to cause defamatory, economic, and political harm as well.⁷ The technology used to create a deepfake is readily available to anyone with a computer, specialized software system, or specialized application.⁸ As a result, deepfakes are becoming more pervasive in American society as everyone from computer engineers to cheerleader moms realize they are capable of manipulating media to their advantage.⁹

Some of the most problematic types of deepfakes are political in nature. Just in the past few years, manipulated videos of “Nancy Pelosi,” “Barack Obama,” “Donald Trump,” and Mark Zuckerberg have gone viral.¹⁰ For example, one video of “Nancy Pelosi” was altered in order to make it appear that she was slurring her words to suggest a state of intoxication.¹¹ Additionally, a public service announcement by “President Barack Obama”

5. Mike Wendling, *The (Almost) Complete History of ‘Fake News’*, BBC TRENDING (Jan. 22, 2018), <https://www.bbc.com/news/blogs-trending-42724320>.

6. Ian Sample, *What Are Deepfakes - and How Can You Spot Them?*, THE GUARDIAN (Jan. 13, 2020), <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them> (outlining common characteristics of deepfake media content).

7. Dave Johnson, *What is a Deepfake? Everything You Need to Know About the AI-powered Fake Media*, BUS. INSIDER (Jan. 22, 2021), <https://www.businessinsider.com/what-is-deepfake> (explaining the basic components of deepfake media).

8. *Id.*

9. Andrew Court, *Cheerleader Mom is Arrested for Creating ‘Deepfake’ Images and Videos Showing Her Daughter’s Rivals ‘Naked, Drinking and Smoking’ in a Bid to Have Them Kicked Off the Team*, DAILYMAIL.COM (Mar. 13, 2021), <https://www.dailymail.co.uk/news/article9359823/Cheerleader-mom-created-deepfake-images-daughters-rivals-naked-drinking-smoking.html> (reporting damage caused by a pornographic deepfake video).

10. Joseph Foley, *14 Deepfake Examples That Terrified and Amused the Internet*, CREATIVEBLOQ.COM (Feb. 10, 2021), <https://www.creativebloq.com/features/deepfake-examples> (listing examples of damaging deepfake videos).

11. *Id.*

shows his likeness edited to include gestures and audio that never existed.¹² Some of the most harmful or destabilizing media sources are political deepfakes that target politicians or those that are deployed during an election campaign. Although lawmakers at the federal and state levels have attempted to enact legislation to regulate the creation and distribution of deepfake media, these attempts have been largely unsuccessful.

The potential consequences of deepfake technology interfering with the political process are alarming. For instance, a deepfake video could falsely depict a candidate committing a criminal or morally reprehensible act in the lead up to an election. A deepfake audio file could be manipulated to mislead the electorate into believing that the candidate used a racial slur. Further, deepfake technology that facilitates manipulation of elections provides an attractive opportunity to international regimes that have an interest in American politics.

Notably, not all applications of deepfake technology are nefarious. The film industry uses deepfake technology for various artistic projects. Further, the Supreme Court has consistently held that parodies and political satire are forms of protected speech.¹³ In reality, deepfakes are most commonly used to inflict harm on individuals and spread suspicion and distrust in the society at large. Deepfakes present a particular challenge to traditional avenues of relief because it is not always possible to identify the person responsible for the creation of the manipulated media.

According to a 2019 study conducted by Pew Research Center, 68% of Americans acknowledge fake news or misinformation as “greatly impact[ing] Americans’ confidence in government institutions.”¹⁴ Furthermore, at least half of Americans believe that false information negatively affects how people feel towards each other, and 80% believe that action should be taken to regulate false information.¹⁵ At a time when political polarization is at its zenith, false information is particularly concerning because of how it can be weaponized. Recent cases predictably reflect traditional First Amendment jurisprudence: a stubborn refusal to recognize new categories of protected speech. As technology advances and poses a greater threat to our society, this hard-line approach will be hard to justify as polarization increases.

B. Present Concerns Regarding Political Speech

Addressing the problem of political deepfakes reveals the larger, consistent problems lurking in modern First Amendment jurisprudence.

12. *Id.*

13. *Hustler Magazine Inc. v. Falwell*, 485 U.S. 46, 53–54 (1988).

14. Amy Mitchell, et al., *Many Americans Say Made-Up News is a Critical Problem That Needs To Be Fixed*, PEW RSCH. CTR. (June 5, 2019), <https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>.

15. *Id.*

Technology is accelerating so quickly that legislation is unable to adequately protect against the threats manipulated media poses to the public and the electoral process. At a time when the country is more polarized and politically involved than ever, media manipulation poses a unique threat to the integrity of the electoral process and democracy itself. Political deepfakes are one of the most visceral and potentially harmful sources of false and inflammatory media that is affecting the political electorate and democracy.¹⁶ Deepfake media is just one example of an extreme, unregulated fake news source that has the potential to damage public confidence. As false media continues to evolve, it will become harder for the Court to justify excluding extremely harmful media from regulation.

Under American constitutional law, speech regulations are subject to a First Amendment analysis.¹⁷ Courts first determine the category in which the regulation falls: viewpoint, content, speaker's identity, time, place, and manner.¹⁸ The standard of review is ultimately determined by the category in which the regulation falls. Public officials such as political candidates do not have a private tort remedy for false statements without actual malice due to the Court's ruling in *New York Times Co. v. Sullivan*.¹⁹ However, scholars recognize and criticize the Court's reluctance to expand the scope of unprotected categories under the First Amendment.²⁰

In *Citizens United v. Federal Election Commission*, the Court found that the political corporate independent expenditures should not be restricted partially because they "would not interfere with governmental functions."²¹ This holding contradicted Court precedent that allowed such restrictions when speech interfered with governmental functions.²² Notably, the Court recognized that "it is inherent in the nature of the political process that voters must be free to obtain information from diverse sources in order to determine how to cast their votes."²³ The *Citizens United* Court held that "political speech does not lose First Amendment protection 'simply because its source is a corporation.'"²⁴ Further, the *Citizens United* Court recognized that "[c]orporations and other associations, like individuals, contribute to the

16. Mitchell, *supra* note 14.

17. Jamal Greene, *Foreword: Rights As Trumps?*, 132 HARV. L. REV. 28, 40 (2018).

18. See *Perry Educ. Ass'n v. Perry Local Educators' Ass'n*, 460 U.S. 37, 55 (1983); see also *Police Dep't v. Mosley*, 408 U.S. 92, 95–96 (1972); *Sorrell v. IMS Health Inc.*, 564 U.S. 552, 580 (2011); *Ward v. Rock Against Racism*, 491 U.S. 781, 802–03 (1989).

19. *New York Times Co. v. Sullivan*, 376 U.S. 254, 279–80 (1964).

20. Symposium: Hate Crime V. Hate Speech: Exploring The First Amendment, *Our Imperiled Absolutist First Amendment*, 20 U. PA. J. CONST. L. 818 (2018).

21. *Citizens United v. FEC*, 558 U.S. 310, 341 (2010).

22. *Id.* at 342.

23. *Id.* at 341.

24. *Id.* at 341 (quoting *First Nat. Bank of Boston v. Bellotti*, 435 U.S. 765, 784 (1978)).

‘discussion, debate, and the dissemination of information and ideas’ that the First Amendment seeks to foster.”²⁵

Political expression in the context of deepfake media would likely not lose First Amendment freedom of expression under the Court’s current First Amendment jurisprudence. Political speech is respected and guarded as critical to a democratic society. Although political speech itself should not lose protection, the First Amendment should adapt to protecting society against *false* and manipulated political speech that has a destabilizing and harmful impact on society. The American electorate deserves protection against false visceral media to be adequately informed and to maintain the integrity of the electoral system.

C. Leading Regulatory Proposals

At the federal level, the 2021 U.S. National Defense Authorization Act (“NDAA”) instructs the Department of Homeland Security (DHS) to compile and publish an annual report on the harms caused by deepfakes for the next five years.²⁶ The report is expected to include an analysis of the potential harm from both foreign and domestic campaign interference.²⁷ Congress amended the NDAA to specifically include reported harm caused by foreign and domestic campaign interference and abusive military deepfake videos.²⁸ Additionally, as of December 2020, the Identifying Outputs of Generative Adversarial Networks Act requires the National Science Foundation to research and track evolving deepfake technology.²⁹ This Act also requires the National Institute of Standards and Technology to establish standards related to deepfakes in the public and private sectors.³⁰

On March 11, 2021, the Senate received H.R. 1 after it was passed by the House of Representatives.³¹ Section 4421 of H.R. 1, or the For the People Act, amends Title III of the Federal Campaign Act of 1971 to include section 325 which prohibits distribution of “materially deceptive media prior to election.”³² Section 325 specifically provides that,

a person, political committee, or other entity shall not, within 60 days of an election for Federal office at which a candidate for elective office will appear on the ballot, distribute, with actual malice, materially deceptive audio or visual media of the candidate with the intent to

25. *Citizens United*, 558 U.S. at 342–43 (quoting *Bellotti*, 435 U.S. at 783).

26. Scott Briscoe, *U.S. Laws Address Deepfakes*, ASIS INT’L (Jan. 12, 2021), <https://www.asisonline.org/security-management-magazine/latest-news/today-in-security/2021/january/U-S-Laws-Address-Deepfakes/>.

27. *Id.*

28. *Id.*

29. *Id.*

30. *Id.*

31. For the People Act of 2021, H.R.1, 117th Cong. (2021).

32. *Id.*

injure the candidate's reputation or to deceive a voter into voting for or against the candidate.³³

If passed by the Senate, the For the People Act will be the first federal law to regulate the contents of deceptive video and audio media in the leadup to an election.³⁴ Because the law effectively targets media focused on candidate portrayal or political speech, it will likely trigger a First Amendment challenge. The language of H.R. 1 reflects the Court's holding in *New York Times v. Sullivan*, where the Court held that actual malice was required for a finding of defamation against public officials.³⁵ Although the statute may technically pass the standard set by this case, the realistic application of H.R. 1 would likely be ineffective in practice as discussed in Part II.

At the state level, California and Texas are recognized as leaders in deepfake regulation because of their specific laws targeting political and pornographic deepfake media.³⁶ In 2019, Texas prohibited deepfake media targeted at elections.³⁷ Additionally in 2019, Virginia enacted a law banning deepfake pornography.³⁸ Soon after, California banned the "videos, images, or audio of politicians doctored to resemble real footage within sixty days of an election."³⁹ States such as Washington, Maine, and Maryland have also proposed deepfake election legislation as well.⁴⁰ However, scholars have grown increasingly concerned that state laws that restrict political speech overstep First Amendment boundaries.⁴¹

One of the motivations behind California's deepfake legislation was the 2019 viral video of House Speaker Nancy Pelosi.⁴² The video was manipulated to make it appear as though "Speaker Pelosi" was drunk and slurring her words.⁴³ Unfortunately, social media sites are unable to

33. For the People Act of 2021, *supra* note 31.

34. *Id.*

35. *Sullivan*, 376 U.S. at 279–80.

36. Kathleen Ronayne, *California Bans 'Deepfakes' Video, Audio Close to Elections*, ASSOCIATED PRESS (Oct. 4, 2019), <https://apnews.com/article/4db02da9c1594fd1a199ee0242c39cc2>.

37. Briscoe, *supra* note 26.

38. *Id.*

39. *Id.*

40. H.B. 198, Reg. Sess. (Md. 2020); S.B. 6513, 66th Leg., Reg. Sess. (Wash. 2020), L.D. 1988, 129th Leg., Reg. Sess. (Me. 2020).

41. S.B. 751, 86th Leg., Prior Sess. (Tex. 2019–2020); Jared Schroeder, *Texas Deepfake Law Unlikely to Survive Scrutiny of the Courts*, TEXAS TRIB., (Sept. 13, 2019), <https://blog.smu.edu/opinions/2019/09/25/texas-deepfake-law-unlikely-to-survive-scrutiny-of-the-courts/>.

42. Mahin Sadiq, *Footage of the Democratic House Speaker Was Edited to Make Her Appear Drunk or Unwell, in the Latest Incident Highlighting Social Media's Struggle to Deal with Disinformation*, THE GUARDIAN (May 24, 2019), <https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video>.

43. *Id.*

effectively regulate deepfake media.⁴⁴ Facebook’s Mark Zuckerberg conceded that the platform’s algorithm did not detect the deepfake video of “Speaker Pelosi” quickly enough.⁴⁵ Even if detection technology is able to recognize deepfake media, it is difficult or sometimes impossible to identify the creator of the deepfake media.

The robust concern over manipulated media and false depictions of candidates before the 2020 election resulted in legislative action in California.⁴⁶ California State Assemblymember Marc Berman introduced Assembly Bill 730 in 2019 in order to ban the distribution of “deceptive audio or visual media of a candidate with the intent to injure the candidate’s reputation or deceive a voter into voting for or against the candidate, unless specified disclosure requirements are met.”⁴⁷ Berman emphasized that the impetus for AB 730 was the potential harm to the political electorate.⁴⁸ In a statement, Berman declared, “[d]eepfakes are a powerful and dangerous new technology that can be weaponized to sow misinformation and discord among an already hyper-partisan electorate.”⁴⁹

Berman further emphasized that deepfakes of political candidates are more harmful leading up to an election because “[d]eepfakes distort the truth, making it extremely challenging to distinguish real events and actions from fiction and fantasy. AB 730 seeks to protect voters from being tricked and influenced by manipulated videos, audio recordings, or images before an election.”⁵⁰ In the absence of effective federal legislation, states recognize the harm political deepfakes pose and are attempting to protect their residents.

II. Modern First Amendment Doctrine Fails to Adequately Protect Against Harm Incurred by Deepfakes

A. Current First Amendment Jurisprudence

The First Amendment protection of expression is unique in its scope, integrity, and strength. In the twenty-first century, the Supreme Court has developed a First Amendment doctrine that is highly protective of speech. The Court has recognized that the First Amendment “has its fullest and most

44. Kurt Wagner, *Facebook CEO: Company Was Too Slow to Respond to Pelosi Deepfake*, BLOOMBERG (June 26, 2019), <https://www.bloomberg.com/news/articles/2019-06-26/facebook-ceo-company-was-too-slow-to-respond-to-pelosi-deepfake>.

45. *Id.*

46. Kaitlin Curry, *Berman Introduces Legislation to Combat Nefarious ‘Deepfakes,’ Protect Election Integrity*, ASMDC (June 24, 2019), <https://a24.asmdc.org/press-releases/20190624-berman-introduces-legislation-combat-nefarious-deepfakes-protect-election>.

47. *Id.*

48. *Id.*

49. *Id.*

50. *Id.*

urgent application’ to speech uttered during a campaign for political office.”⁵¹ The right of citizens to inquire, hear, speak about, and use information to reach consensus is a precondition to enlightened self-government.⁵²

Traditionally under a First Amendment analysis, courts first consider whether a law is aimed at regulating conduct or speech. Additionally, courts examine whether the law establishes a time, place, or manner restriction or whether the law attempts to control speech based on content or the speaker’s viewpoint.⁵³ If the court determines that the restriction is content-based, the restriction is considered presumptively unconstitutional and subject to strict scrutiny.⁵⁴ However, the Supreme Court has allowed regulation of certain categories of content-based restrictions such as obscenity, defamation, fraud, incitement, fighting words, true threats, speech integral to criminal conduct, and child pornography.⁵⁵

Especially in the twenty-first century, the Supreme Court has stubbornly refused to recognize new categories of unprotected speech. In fact, the Court has declined to recognize a complete prohibition on false statements.⁵⁶ The Court has acknowledged that “some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation.”⁵⁷ Defamation or fraud are current legal remedies that provide potential avenues of relief to false statements. The government is allowed to regulate false speech in order to protect the public against consumer deception.⁵⁸

In *United States v. Alvarez*, the Court refused to establish a new, unprotected category for false statements.⁵⁹ Notably, however, Justice Breyer signaled a willingness to analyze exactly how well the statute’s means are tailored towards its ends.⁶⁰ Justice Breyer specifically recommended that the Court consider the “seriousness of the speech-related harm the provision will likely cause, [and] the nature and importance of the

51. *Citizens United*, 558 U.S. at 339 (quoting *Eu v. San Francisco Cnty. Democratic Cent. Comm.*, 489 U.S. 214, 223 (1989)).

52. *Id.*

53. See *Hill v. Colorado*, 530 U.S. 703, 722 (2000); see also *Reed v. Town of Gilbert, Ariz.*, 576 U.S. 155, 169 (2015).

54. *Reed*, 576 U.S. at 164.

55. See *Miller v. California*, 413 U.S. 15, 24 (1973); see also *Gertz v. Robert Welch*, 418 U.S. 323, 344 (1974); see also generally *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964); *Illinois ex rel. Madigan v. Telemarketing Assocs.*, 538 U.S. 600, 612 (2003); *Brandenburg v. Ohio*, 395 U.S. 444, 447–48 (1969); *Chaplinsky v. New Hampshire*, 315 U.S. 568, 574; *Watts v. United States*, 394 U.S. 705, 708 (1969) (per curiam); *Giboney v. Empire Storage & Ice Co.*, 336 U.S. 490, 498 (1949); *New York v. Ferber*, 458 U.S. 747, 764 (1982).

56. *United States v. Alvarez*, 567 U.S. 709, 718–19 (2012) (plurality opinion).

57. *Id.*

58. See *Illinois ex rel. Madigan*, 538 U.S. at 612.

59. *Alvarez*, 567 U.S. at 718–19.

60. *Id.* at 730 (Breyer, J., concurring).

provision's countervailing objectives."⁶¹ Furthermore, Justice Breyer suggested that a proportionality approach, discussed in Part III, would allow the Court to balance freedom of expression with important governmental interests.⁶²

Of course, it is possible that political deepfakes may not be limited to a speech analysis. The primary danger that political deepfakes pose to society is not the amount of political speech or even the identity of the speaker, but rather the act of creating and sharing deepfake media is a likely cause of increased political polarization and threatens the integrity of our democratic political process. Therefore, using a proportionality analysis, it could be possible to analyze content-neutral legislation that restricts speech or inherently expressive conduct under the four-part test set forth in *United States v. O'Brien* and intermediate scrutiny.⁶³ The challenged regulation will be upheld if it falls within "the constitutional power of the Government; if it furthers an important or substantial governmental interest; if the governmental interest is unrelated to the suppression of free expression; and if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest."⁶⁴

B. The Challenges of Regulating Political Deepfakes

Traditionally, the First Amendment freedom of expression has been fiercely guarded as one of the most important rights conferred by the Bill of Rights. The reasoning behind this hard line approach is to protect the integrity of the First Amendment.⁶⁵ As a result, legislators and the judiciary alike have been cautious to create exceptions to the First Amendment.⁶⁶ The obvious risk is that a multitude of exceptions will lead to a slippery slope that weakens the First Amendment.⁶⁷ Even though some landmark Supreme Court decisions recognize exceptions to the First Amendment freedom of expression, these exceptions are intentionally and strictly limited to specific circumstances.⁶⁸ Especially in modern times, the Supreme Court has been

61. *Id.*

62. *Id.* at 731.

63. *O'Brien*, 391 U.S. at 377.

64. *Id.*

65. John D. Moore, *The Closed and Shrinking Frontier of Unprotected Speech*, 36 WHITTIER L. REV. 1, 5 (2014) (discussing the importance of First Amendment protections).

66. Eugene Volokh, *Censorship Envy*, WASH. POST: THE VOLOKH CONSPIRACY (Jan. 16, 2015), <https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/01/16/censorship-envy/> (criticizing government restriction on freedom of expression).

67. Eugene Volokh, *The Mechanisms of the Slippery Slope*, 116 HARV. L. REV. 1026 (2003), <https://www2.law.ucla.edu/Volokh/slipperyshorter.pdf> (warning of the dangers of excessively regulating freedom of expression).

68. Symposium: An Ocean Apart? Freedom of Expression in Europe and the United States, *The Categorical Approach to Protecting Speech in American Constitutional Law*, 84 IND. L.J. 917, 932–33 (2009).

overly hostile to new categories and has declined to extend First Amendment protection beyond the limits of those few cases.⁶⁹

Does deepfake technology create new problems that are incapable of being reconciled with the Constitution? The likely answer is no. American constitutional law has adapted to confront new technological issues that the Framers could not have imagined. A robust protection of speech is important in order to protect the fundamental principle behind the First Amendment: to encourage the proliferation of discussion and debate that is fundamental and essential to democracy. Scholars recognize that deepfake videos “could hijack elections, eroding faith in . . . democracy.”⁷⁰ In *Holder v. Humanitarian Law Project*, the Court upheld a statute that prohibited knowingly giving material support, such as expert advice, to foreign terrorist organizations.⁷¹ Although the Court recognized that conduct ordinarily calls for a lower standard of scrutiny, the statute regulated on the basis of content and thus triggered a higher standard of scrutiny.⁷²

Like the statute in *Holder*, a statute controlling the creation and distribution of a politically-based deepfake video could be analyzed as regulating creators’ content rather than pure political speech.⁷³ This distinction is especially important because First Amendment analyses first inquire as to the type of speech or conduct the government is attempting to regulate.⁷⁴ Next, courts must decide what level of scrutiny applies to the regulation at issue. Generally, a content-neutral regulation would receive intermediate scrutiny and would be upheld if it advances important governmental interests not related to the suppression of free speech and it does not burden substantially more speech than necessary to further the government’s interests.⁷⁵

One example of an advancement of the Supreme Court’s jurisprudence is its ruling in *New York Times v. Sullivan* (*Sullivan*), which extended First Amendment protection to libel upon a showing of actual malice.⁷⁶ However, the Court’s holding in *Sullivan* does not help solve the deeper harm inflicted by political deepfakes. *Sullivan* is inapplicable to deepfake regulations for two reasons. First, *Sullivan* invalidated an Alabama law intended to protect

69. Symposium: What Swings the Vote? The Influence of the U.S. Legal System and the Media on Presidential Elections, *Note: The Cost of Free Speech: Combating Fake News or Upholding the First Amendment?*, 75 U. MIAMI L. REV. 572, 607–09 (2021).

70. Robert Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 888 (2019).

71. *Holder v. Humanitarian Law Project*, 561 U.S. 1, 39 (2010).

72. *Id.*

73. *Id.* at 28.

74. *Holder*, 561 U.S. at 36.

75. *O’Brien*, 391 U.S. at 377.

76. *Sullivan*, 376 U.S. at 293.

public officials from specific sources of libel.⁷⁷ Second, the creator of political deepfake content cannot always be identified, which is why state defamation laws are inadequate solutions for deepfake regulation.⁷⁸

To satisfy the standard set by *Sullivan*, a plaintiff must prove that the defendant knew or acted with reckless disregard as to the falsity of the content in order to prevail.⁷⁹ However, if the plaintiff cannot identify the creator of the deepfake content, the plaintiff would only be able to pursue an action against a third-party host such as a social media platform, which could be blocked by Section 230 immunity.⁸⁰ Additionally, it would be difficult for the plaintiff to prove that the third-party host had actual knowledge of the malicious intent. Further, it is not clear who or what qualifies as a “public figure” or “public official.”⁸¹

Moreover, plaintiffs in a potential defamation deepfake challenge must also grapple with the lasting impact of the video and the possibility of never obtaining relief. It is commonly understood today that once something is posted on the internet it is unlikely to ever be completely deleted or destroyed. Whether it be a screenshot, individual download, or private recording, any media on the internet has the potential to exist forever.

III. Recommendations for Adapting First Amendment Doctrine

A. The Supreme Court Should Adopt a New Categorical Exception for Political Deepfakes

The Supreme Court should permit the government to regulate deepfakes that are political in nature by recognizing a new unprotected category under the First Amendment. Although in recent cases the Court has declined to establish new unprotected categories, political deepfakes are distinguishable from other forms of technology because of the unique destabilizing threat they pose to democracy.

In *Brown v. Entertainment Merchants Association*, the Court considered whether a California law that restricted the sale of violent video games to minors violated the First Amendment.⁸² The *Brown* Court acknowledged its decision in *United States v. Stevens*, where just the year before, it held that new categories of unprotected speech could not be established simply because a legislature identified a certain kind of speech as harmful.⁸³ In *Stevens*, the Court struck down a statute that criminalized

77. *Id.* at 256.

78. Reid, *supra* note 4 at 218.

79. *Hustler Magazine Inc.*, 485 U.S. at 53.

80. 47 U.S.C. § 230 (1996).

81. *Hustler Magazine Inc.*, 485 U.S. at 53.

82. *Brown v. Ent. Merch. Ass'n*, 564 U.S. 786, 788–89 (2011).

83. *Id.* at 791.

the sale of depictions of animal cruelty.⁸⁴ The Court determined that the regulation of depictions of animal cruelty was a content-based restriction on speech in violation of the Free Speech Clause.⁸⁵

The *Brown* Court found that California’s research from psychologists, who found that violent videogames had a harmful effect on children, was not compelling enough to overcome strict scrutiny.⁸⁶ Notably, the Court reasoned that the studies failed to show that “violent video games *cause* minors to *act* aggressively.”⁸⁷ Therefore, the *Brown* Court declined to rely on studies that demonstrated correlation, instead of causation.⁸⁸ Further, Justice Scalia emphasized that the “new and not historically recognized categories of noncoverage were highly disfavored,” so that creating new recognized categories “would require a showing of necessity resembling the Court’s traditionally highly stringent ‘compelling interest’ standard.”⁸⁹

The regulated content in *Brown* is distinguishable from political deepfakes because videogames cannot alter elections or affect voter trust. Further, a law that would regulate the creation, sale, and distribution of political deepfakes would likely be characterized as content-based core political speech. Core political speech is one of the most protected categories under the First Amendment because it is essential for democratic debate.⁹⁰ As a result, courts apply strict scrutiny to laws that burden political speech and require the government to show that the restriction “furthers a compelling interest and is narrowly tailored to achieve that interest.”⁹¹

The purpose of the First Amendment is to protect the proliferation of viewpoints that further democracy.⁹² Although the Court has allowed some restrictions on speech, these restrictions have been justified in order to permit the government to operate democratic functions. The Court has reserved some narrow categories of operations open to regulation because of their essential democratic function. These democratically-based operations include public school education, criminal punishment, military duties, and political service.⁹³ As the Court recognized in *Buckley v. Valeo*, “it is of

84. *United States v. Stevens*, 559 U.S. 460, 468–70 (2010).

85. *Id.*

86. *Brown*, 564 U.S. at 800.

87. *Id.*

88. *Id.*

89. Frederick Schauer, *The Politics and Incentives of First Amendment Coverage*, 56 WM. & MARY L. REV. 1613, 1623 (2015).

90. Senator Marco Rubio, *Keynote Remarks at The Heritage Foundation’s Homeland Security Event on Deep Fakes* (July 19, 2018), <https://www.heritage.org/homeland-security/event/deep-fakes-looming-challenge-privacy-democracy-and-national-security>.

91. *FEC v. Wis. Right To Life, Inc.*, 551 U.S. 449, 464 (2007).

92. *See United States v. Playboy Ent. Grp., Inc.*, 529 U.S. 803, 813 (2000).

93. *See Bethel School Dist. No. 403 v. Fraser*, 478 U.S. 675, 683 (1986) (protecting the “function of public school education”); *Jones v. North Carolina Prisoners’ Lab. Union, Inc.*, 433 U.S. 119, 129 (1977) (furthering “the legitimate penological objectives of the corrections system”)

particular importance that candidates have the unfettered opportunity to make their views known so that the electorate may intelligently evaluate the candidates' personal qualities and their positions on vital public issues before choosing among them on election day."⁹⁴

As stated above, the *Citizens United* Court found that the corporate independent expenditures "would not interfere with governmental functions" and could not be restricted by the Free Speech Clause of the First Amendment, contradicting precedent that allowed restrictions when speech interfered with governmental functions.⁹⁵ Notably, the Court recognized that "it is inherent in the nature of the political process that voters must be free to obtain information from diverse sources in order to determine how to cast their votes." Further, the *Citizens United* Court recognized that "[c]orporations and other associations, like individuals, contribute to the 'discussion, debate, and the dissemination of information and ideas' that the First Amendment seeks to foster."⁹⁶ The Court has recognized political speech as "indispensable to decision making in a democracy, and this is no less true because the speech comes from a corporation rather than an individual."⁹⁷ In *First National Bank of Boston v. Bellotti*, the Court noted that the "inherent worth of the speech in terms of its capacity for informing the public does not depend upon the identity of its source, whether corporation, association, union, or individual."⁹⁸

While the Court has so far rigidly set its First Amendment jurisprudence, technology will only continue to evolve. The time has come for the Court to determine a reasonable approach to applying the First Amendment when the constitutional right of speech conflicts with democracy. The Court could justify a new unprotected category by citing what it has already recognized: maintaining the integrity of the political process. Allowing the regulation of political deepfakes would result in a more informed political electorate and a fairer electoral process. However, if the Court declines to establish a new unprotected category, the Court could instead adopt a proportionality approach that some scholars view as necessary to balance fundamental constitutional rights.

(internal quotation marks omitted)); *Parker v. Levy*, 417 U.S. 733, 759 (1974) (ensuring "the capacity of the Government to discharge its [military] responsibilities" (internal quotation marks omitted)); *Civil Serv. Comm'n v. Letter Carriers*, 413 U.S. 548, 557 (1973) ("[F]ederal service should depend upon meritorious performance rather than political service").

94. *Buckley v. Valeo*, 424 U.S. 1, 52–53 (1976).

95. *Citizens United*, 558 U.S. at 341.

96. *Citizens United*, 558 U.S. at 342–43 (quoting *Bellotti*, 435 U.S. at 783).

97. *Bellotti*, 435 U.S. at 777.

98. *Id.*

C. Adopting a Proportionality Approach as an Alternative

Scholars are recognizing that the harmful effects of fake news media have permeated society and are inflicting damage on democracy.⁹⁹ For example, Kareem Gibson analyzes the current remedies available to victims of targeted deepfake videos.¹⁰⁰ Specifically, Gibson addresses torts of defamation, false light, and intentional infliction of emotional distress as applicable causes of action available to plaintiffs.¹⁰¹ Gibson proposes that a “federal criminal statute would be the best option for plaintiffs seeking to redress the harms they have suffered.”¹⁰² Although Gibson is correct that a federal criminal statute would send a message to the states that the government is serious about confronting damaging political deepfakes, the reality of federal criminalization is unlikely to come to fruition. Congress has continuously failed to adequately regulate social media platforms.¹⁰³ Further, such a statute would trigger a tidal wave of litigation because of its potential effect and chilling of the closely-guarded freedom of expression within the First Amendment. Gibson’s prediction that intentional infliction of emotional distress actions will be the best tool to combat deepfakes fails to consider the lasting challenges that political deepfakes cause to society.

In contrast, privacy and government investigations scholar Shannon Reid advocates for the Supreme Court to adapt their First Amendment doctrine to infer stronger privacy protections that exist within the Constitution.¹⁰⁴ Reid recognizes the “threat that deepfakes pose to the integrity of public discourse itself, in addition to the reputational harm it causes targeted individuals.”¹⁰⁵ At the same time, Reid acknowledges the Court’s opinion in *Brown v. Hartlage* where the Court determined that “the State’s fear that voters might make an ill-advised choice does not provide the State with a compelling justification for limiting speech.”¹⁰⁶ Reid also emphasizes the Supreme Court’s willingness to infer a right of personal privacy in the Constitution.¹⁰⁷ Similarly, defamatory deepfakes are generally nonconsensual and “deprive individuals of their autonomous right to build their own personality and reputation.”¹⁰⁸ Reid proposes that federal courts

99. Terry Lee, *The Global Rise of “Fake News” and the Threat to Democratic Elections in the USA*, 22 PUBLIC ADMIN. AND POLICY: AN ASIA-PACIFIC J. 1, 1 (2019) (recognizing that “many fear ‘fake news’ [has] gradually become a powerful and sinister force, both in the news media environment as well as in the fair and free elections”).

100. Kareem Gibson, *Deepfakes and Involuntary Pornography: Can Our Current Legal Framework Address This Technology?*, 66 WAYNE L. REV. 259, 261 (2020).

101. *Id.* at 262.

102. *Id.* at 288.

103. *Packingham v. North Carolina*, 137 S. Ct. 1730, 1736 (2017).

104. Reid, *supra* note 4, at 228.

105. *Id.* at 228–29.

106. *Brown v. Hartlage*, 456 U.S. 45, 60 (1982).

107. Reid, *supra* note 4, at 232.

108. *Id.* at 234.

should infer a Fourteenth Amendment right to privacy in order to protect against the personal privacy and reputational harms caused by political deepfakes.¹⁰⁹ However, because it is unlikely that the Court will recognize that the Fourteenth Amendment extends to private conduct, Reid offers an alternative approach that balances constitutional freedoms.¹¹⁰ While this proportionality proposal is intriguing, it has not been widely practiced.¹¹¹ For example, in *Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Commission*, the Court refused to balance constitutional freedoms of speech and religion and instead determined the holding on a technicality.¹¹²

Also an advocate of the proportionality approach, Jamal Greene presents two frameworks for the traditional constitutional adjudication of rights in recent Supreme Court jurisprudence.¹¹³ The first framework, adopted by the Supreme Court, Greene presents constitutional rights as absolute with limited exceptions. The second framework embraced by the rest of the world characterizes rights as generally limited, but absolute in certain circumstances.¹¹⁴ Greene argues that proportionality, rather than categorical adjudication, is more helpful when confronting modern challenges such as technological advancements.¹¹⁵ Greene presents a proportionality analysis as a solution to modern constitutional conflicts, such as those related to technological advancements, because the increasing number of categorical exceptions result in incoherence and confusion.¹¹⁶ However, recent First Amendment cases show the trend of stubbornness and reflects that an adoption of a proportionality analysis is highly unlikely. Nonetheless, Greene provides the example of Justice Breyer signaling his willingness to embrace a transition to a proportionality analysis in cases such as *United States v. Alvarez* and *Craig v. Boren*.¹¹⁷

Another scholar Brittany Finnegan analyzed a hypothetical statute regulating social media and outlined the Court's awareness of evolving technologies and necessary constitutional adaptations. From there, Finnegan concluded that the government's unwillingness to regulate harmful conduct now will result in an unfortunate inability to regulate what is exchanged online in the future.¹¹⁸ Finnegan outlined the Court's awareness of evolving

109. *Id.*

110. Reid, *supra* note 4, at 235.

111. *Id.*; see also *Masterpiece Cakeshop, Ltd. v. Colo. Civil Rights Comm'n*, 138 S. Ct. 1719 (2018).

112. Reid, *supra* note 4, at 235.

113. Greene, *supra* note 17, at 30.

114. *Id.* at 30.

115. *Id.* at 35.

116. *Id.* at 120.

117. *Alvarez*, 567 U.S. at 730–31 (Breyer, J., concurring); *Craig v. Boren*, 429 U.S. 190, 211–12 (1976) (Stevens, J., concurring).

118. Brittany Finnegan, Note, *The Cost of Free Speech: Combating Fake News or Upholding the First Amendment?*, 75 U. MIAMI L. REV. 572, 618 (2021).

technologies and necessary constitutional adaptations. In *Packingham v. North Carolina*, the Court warned that they could not yet fully appreciate the “full dimensions and vast potential” of the “Cyber Age.”¹¹⁹ However, Justice Alito warned that the Court did not “hee[d] its own admonition of caution” regarding the regulation of social media.¹²⁰ Regulation of speech on social media platforms has become a polarizing issue that signals the need to adapt the First Amendment doctrine. As technology advances, Congress and the Supreme Court must decide how to reconcile fundamental constitutional freedoms.

Conclusion

The principal purpose of the First Amendment is to protect the proliferation of viewpoints that further democracy.¹²¹ However, the government has compelling interests both in furthering democratic debate and maintaining the integrity of an essential government function: the electoral process. Although the Court has allowed some restrictions on speech, these restrictions have been justified in order to permit the government to operate democratic functions.¹²² The Court has reserved some narrow categories of operations open to regulation because of their essential democratic function. The NDAA may allow the DHS to collect evidence of manipulated media affecting elections. Perhaps the Court would be willing to establish a new, unprotected category based on evidence that reveals a causal, rather than correlative, relationship between the regulation and the threat to society. Although any law that attempts to regulate political speech, including The For the People Act, would be likely subject to strict scrutiny, it is possible that the government may have enough to show its compelling interest in preventing voter misinformation that skews elections.

Technological evolution is inevitable and the First Amendment doctrine must change in order to adapt to the current times. Political deepfakes are an extreme example of altered media that is used to manipulate the minds of the political electorate and election results. The Supreme Court should adopt a new, unprotected category in order to allow government regulation of political deepfakes. Although freedom of expression is one of the most protected rights in American constitutional law, the Court must act now to

119. *Packingham*, 137 S. Ct. at 1736.

120. *Id.* at 1744 (Alito, J., dissenting).

121. *See Playboy Ent. Grp. Inc.*, 529 U.S. at 813.

122. *See Bethel School Dist. No. 403*, 478 U.S. at 683 (protecting the “function of public school education”); *Jones*, 433 U.S. at 129 (furthering “the legitimate penological objectives of the corrections system” (internal quotation marks omitted)); *Parker*, 417 U.S. at 759 (ensuring “the capacity of the Government to discharge its [military] responsibilities” (internal quotation marks omitted)); *Letter Carriers*, 413 U.S. at 557 (“[F]ederal service should depend upon meritorious performance rather than political service”).

establish an unprotected category or prioritize democratic interests to protect the foundation of American democracy.