Tailoring the First Amendment in the Age of AI and Algorithms

Jeremy Ross[♦]

Abstract: *This note discusses the new harms posed to democracy and self-governance by online AI-generated speech, altered speech, and algorithmic content moderation, before proceeding to an analysis of whether state efforts to regulate these kinds of speech can be justified under the First Amendment. It surveys various proposals throughout, including whether the government may require social media platforms to label altered and AI-generated content that appear on their platforms, whether the government may mandate certain kinds of algorithmic content moderation on platforms, and whether the government may ban certain forms of AI and altered speech altogether. Notably, this note gives substantial attention to the recent Supreme Court decision* Moody v. NetChoice *(2024). After detailing how algorithms affect the marketplace of ideas and surveying Supreme Court precedent covering the relationship between the First Amendment and technology, this note addresses three arguments for regulating altered and AI speech and algorithmic content moderation. First, the most promising avenue for regulation lies in a line of Supreme Court cases allowing First Amendment protections to be tailored according to the technology of the medium which is regulated, which in turn would lead the Supreme Court to account for the impact of algorithms on the "marketplace of ideas." Secondly and less plausibly, some of these regulations may be justified as regulations of commercial speech. Finally, the specific characteristics of AI content may permit an exception to First Amendment coverage. This note argues that an AI exception allowing more extensive regulation of AI speech is highly unlikely to emerge from existing doctrine. Throughout, this note emphasizes the application of these arguments to political speech, which receives heightened protection from the First Amendment, while maintaining that its arguments also apply to online speech more widely. It argues that while these various regulations face steep hurdles to being found constitutionally permissible, some regulations may be justified where courts conduct a searching analysis of the technological structure of the targeted communicative medium, and*

♦ J.D., University of Virginia School of Law, expected 2025. I would like to thank Professor Matthew Sanderson for his comments on this work. I am also indebted to the late, great Professor Frederick Schauer, not only for teaching his incredible course on the First Amendment, but for also providing his thoughts on an early outline of this note. I also thank the editors of the *Journal of Law & Politics* for their insightful comments which have undoubtedly improved this work.

notably where the speech in question is untethered from any direct human expressive choice.

TABLE OF CONTENTS

INTRODUCTION

"[T]he advancement & diffusion of Knowledge . . . is the only Guardian of true liberty . . . . "[1]

- James Madison

On September 17, 2024, California passed into law a ban on deceptive "deepfakes" — images, videos, or audio which are edited or generated using artificial intelligence tools — depicting politicians during election periods.[2] The law was inspired by a viral video on X (formerly known as Twitter) depicting Vice President Kamala Harris, which used artificially altered audio to deceptively produce the appearance of Harris claiming she had spent "four years under the tutelage of the ultimate deep state puppet," and calling herself "the ultimate diversity hire."[3] Elon Musk shared the video to his X account, generating over 100 million views.[4] California Governor Gavin Newsom responded to the video by stating that "[m]anipulating a voice in an 'ad' like [this] . . . should be illegal."[5] A federal court nonetheless swiftly placed a preliminary injunction on the law within several weeks of its passage.[6]

In this specific instance, the speech in question is both highly deceptive but also clearly ideological and political. As a result, regardless of the harms posed, it receives the highest level of protection that the First Amendment affords any speech.[7] Additionally, this case demonstrates that the penetration of the internet and social media into all realms of public and private life has dramatically altered both how we speak and how we are spoken to, calling into question the ability of these mediums to transmit accurate information

---

[1] Letter from James Madison to George Thomson (June 30, 1825) (on file with The James Madison Papers at The Library of Congress), https://founders.archives.gov/documents/Madison/04-03-02-0562.

[2] See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1191 (E.D. Cal. 2024); see Danielle K. Citron & Robert Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1758 (2019).

[3] *Kohls*, 752 F. Supp. 3d at 1192.

[4] Jacob Gershman, Election Deepfakes Prompt State Crackdowns—and First Amendment Concerns, Wall St. J. (Oct. 5, 2024, 5:00 AM), https://www.wsj.com/us-news/law/election-deepfakes-prompt-state-crackdownsand-first-amendment-concerns-0b992e8e.

[5] Id.

[6] *Kohls*, 752 F. Supp. 3d at 1200.

[7] Buckley v. Valeo, 424 U.S. 1, 14 (1976) ("Discussion of public issues and debate on the qualifications of candidates are integral to the operation of the system of government established by our Constitution. The First Amendment affords the broadest protection to such political expression in order 'to assure (the) unfettered interchange of ideas for the bringing about of political and social changes desired by the people.'" (quoting Roth v. United States, 354 U.S. 476, 484 (1957))).

to voters.[8] The rapidly changing nature of online political speech and the extensive protections afforded that speech by the First Amendment are already causing massive transformations in our political system.[9] In turn, academics and politicians alike have called for a fundamental rethinking of the regulation of the free flow of information[10] in the wake of express foreign efforts to influence elections,[11] the proliferation of bots and deepfakes,[12] the rise of AI,[13] the declining veracity of audio and visual media,[14] and the power of algorithms to shape our information environments and silo speakers and listeners from voices which may legitimately seek to influence them.[15]

---

[8] See Richard L. Hasen, Deep Fakes, Bots, and Siloed Justices: American Election Law in a "Post-Truth" World, 64 St. Louis Univ. L.J. 535, 537 (2020) [hereinafter Hasen, Bots]; Tim Wu, Is the First Amendment Obsolete?, 117 Mich. L. Rev. 547, 548–49 (2018).

[9] See Darrell M. West, How Disinformation Defined the 2024 Election Narrative, Brookings (Nov. 7, 2024), https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/. But see Sayash Kapoor & Arvind Narayanan, We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem, Knight First Amend. Inst. (Dec. 13, 2024), https://knightcolumbia.org/blog/we-looked-at-78-election-deepfakes-political-misinformation-is-not-an-ai-problem.

[10] See Richard L. Hasen, Cheap Speech and What It Has Done (to American Democracy), 16 First Amend. L. Rev. 200, 216 (2017) [hereinafter Hasen, Cheap Speech]; Wu, supra note 8, 549–50; see also Citron & Chesney, supra note 2, at 1754 ("The marketplace of ideas already suffers from truth decay as our networked information environment interacts in toxic ways with our cognitive biases. Deep fakes will exacerbate this problem significantly. Individuals and businesses will face novel forms of exploitation, intimidation, and personal sabotage.").

[11] See Dustin Volz & Alan Cullison, 'Putin Has Won': Mueller Report Details the Ways Russia Interfered in the 2016 Election, Wall St. J. (Apr. 19, 2019, 5:30 AM), https://www.wsj.com/articles/putin-has-won-mueller-report-details-the-ways-russia-interfered-in-the-2016-election-11555666201.

[12] One British researcher recently was easily able to use OpenAI and Midjourney to create realistic images: "[w]ithin seconds" of entering prompts into AI, including ones requesting "'a realistic photo of voter ballots in a dumpster'; 'a photo of long lines of voters waiting outside a polling station in the rain'; [and] 'a photo of Joe Biden sick in the hospital,'" AI gave him many realistic images. Mark Scott, Deepfakes, Distrust and Disinformation: Welcome to the AI Election, Politico (Apr. 16, 2024, 6:30 AM), https://www.politico.eu/article/deepfakes-distrust-disinformation-welcome-ai-election-2024/.

[13] See Global Views on A.I. and Disinformation, Ipsos, https://www.ipsos.com/sites/default/files/ct/news/documents/2023-11/Ipsos_Global_Views_on_AI_and_Disinformation_full_report.pdf (last visited Oct. 1, 2024); Lucia Mackenzie & Mark Scott, How People View AI, Disinformation and Elections — in Charts, Politico (Apr. 16, 2024, 6:30 AM), https://www.politico.eu/article/people-view-ai-disinformation-perception-elections-charts-openai-chatgpt/; see also David Klepper & Ali Swenson, AI-Generated Disinformation Poses Threat of Misleading Voters in 2024 Election, PBS (May 14, 2023, 7:52 PM), https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election (various reports detailing the rising influence of the AI in politics).

[14] See Gerritt De Vynck, The AI Deepfake Apocalypse is Here. These are the Ideas for Fighting It, Wash. Post (Apr. 5, 2024), https://www.washingtonpost.com/technology/2024/04/05/ai-deepfakes-detection/.

[15] See Citron & Chesney, supra note 2, at 1768 ("Platforms' algorithms highlight popular information, especially if it has been shared by friends, and surround us with content from relatively homogenous groups. As endorsements and shares accumulate, the chances for an algorithmic boost increase. . . . Because people tend to share information with which they agree, social media users are surrounded by information confirming their preexisting beliefs."); see also Rashi Shrivastava & Forbes Staff, AI Nudes of Celebs Like Margot Robbie And Selena Gomez Are For Sale on eBay, Forbes (Mar.

A. Roadmap

In the context of calls for a fundamental rethinking of information regulation, the main purpose of this note is to analyze various mandatory content moderation proposals under existing precedent, and to pose the question of what measures can actually be employed, within constitutional limits, to address the problems posed by AI, altered speech, algorithmic content moderation or curation, and their proliferation on social media. While this analysis applies to *all* online speech, this note emphasizes the application of its analysis to political speech, given rising concerns over the impact of technological change on democracy and governance. To assess these proposals, this note proceeds as follows: Part I outlines the "marketplace of ideas" model, which is central to First Amendment doctrine. Part II details a line of Supreme Court cases which instruct courts to evaluate regulations of particular kinds of speech in light of the specifics of the medium through which that speech is communicated. Part III details the Justices' differing approaches in *Moody v. NetChoice*, analyzing their implications for mandatory online content moderation. Part IV analyzes whether mandatory content moderation can be justified as a regulation of commercial speech, and Part V analyzes the same question by asking whether there exists the possibility that AI and altered online speech may garner fewer protections from the First Amendment. Finally, Part VI examines a proposal from Justice Barrett's *NetChoice* concurrence, which speculates that First Amendment protections may attach to a lesser degree to speech which is attenuated from any direct human expressive choice. This note terms this proposal the "human anchor test."

To the extent that AI and AI-driven disinformation is harming democracy, this note seeks to uncover any legal tools which the government (state or federal) can use to limit these harms while emphasizing that the First Amendment places profound limits on those same tools. As will quickly become apparent, these proposals and many other proposals mandating online moderation, curation,[16] labeling, and disclosure by private actors are constitutionally suspect. The core contention of this note is that the best avenue for regulating online political speech is by justifying the regulation based on the specific technological changes to online communicative ecosystems created by the rise of algorithmic curation and content-moderation. In other

---

12, 2024, 5:13 PM), https://www.forbes.com/sites/rashishrivastava/2024/03/12/ai-nudes-of-celebs-like-margot-robbie-and-selena-gomez-are-for-sale-on-ebay/?sh=3b2e335b1441 (detailing the proliferation of deceptive nude deepfake images of celebrities).

[16] "Curation" is defined as a platform or hosts' "practice of exercis[ing] 'editorial discretion in the selection and presentation' of the content it hosts." See *NetChoice*, 603 U.S. at 782 (Alito, J., concurring).

words, the Supreme Court has recognized that regulations of speech on social media should not necessarily be subject to the same analysis as newspaper speech or broadcast media speech. It follows that emphasizing the distinct technological character of contemporary social media may allow for certain forms of regulation, which would have been disfavored when applied to older media. Other avenues, such as 1) regulating online political speech and content moderation as commercial speech or 2) locating some other exception allowing this speech to be regulated, are unlikely to bear fruit. As this note will show, these alternative approaches require a strained view of existing doctrine and are likely to lead to severe over- and underinclusion problems.

Indeed, the model of internet speech as understood in the 1997 case of *Reno v. ACLU* is no longer the experience of the average internet user.[17] Our laws must adapt to our new realities, as they have in the past. Otherwise, we risk restricting lawmakers to precedents explicitly formulated in response to older technologies. In doing so, the courts are likely to hamstring lawmakers with genuine concerns that these new technologies are undermining our democracy if not our social well-being. The extent to which courts permit this adaptation will determine the viability of mandated moderation, disclosure, and labeling requirements for altered and AI speech online, as well as the ability of individuals to disseminate AI and altered content altogether.[18]

B. Strange First Amendment Bedfellows

Many commentators are well aware that state efforts to dictate content moderation are constitutionally suspect,[19] and have often focused their proposals on state-mandated labeling and disclosure requirements identifying harmful content.[20] Some proposed policies include: a requirement that

---

[17] 521 U.S. 844 (1997). *Reno* was in response to the Communications Decency Act of 1996, unanimously striking down as overbroad federal legislation which introduced serious criminal and civil penalties for those who exposed children to obscene content online.

[18] "Altered content" is broader than but can include AI content, and likewise can include anything from photoshop to simple editing that can be done quickly today on a smartphone. Altered content is any audio or visual content altered after its initial creation. This could include photoshopped images, song remixes, or advertisements where voices are distorted. It is a necessarily broad term, since it is not clear where the meaningful line is between various some sorts of altered content, for instance photoshopped beauty ads versus AI imagery which closely resembles a real photo.

[19] See Daphne Keller, Platform Transparency and the First Amendment, 4 J. Free Speech L. 1, 20 (2023); Hasen, Bots, supra note 8, at 552; Moody v. NetChoice, LLC, 603 U.S. 707, 731–33 (2024) (summarizing First Amendment precedent supporting this proposition and amounting to a rule that "the First Amendment offers protection when an entity engaging in expressive activity, including compiling and curating others' speech, is directed to accommodate messages it would prefer to exclude.").

[20] See Hasen, Bots, supra note 8, at 549.

websites and social media platforms with large numbers of users label content containing altered material as such[21] and "laws that would punish deepfakes or require their removal from websites and social media platforms."[22]

Notably, Texas and Florida have advanced complex proposals to mandate certain forms of content moderation and public disclosures related to moderation rules. These policies, which were the subject of the *NetChoice* cases, were designed to limit the ability of social media platforms to bar users over political and ideological speech. These laws generally imposed a variety of obligations on social media companies that require transparency in their content moderation activities and prohibit them from engaging in certain types of content moderation, for instance by limiting the situations in which social media companies can ban or shadow ban users and remove or limit the reach of posts about certain topics.[23] They also represent the kinds of legal mechanisms which could potentially be used to require the labeling of AI or altered content.[24] For instance, the Fourth Circuit in *Washington Post v. McManus* ruled unconstitutional a Maryland law which sought to require online platforms to publicly disclose information about political advertisers they host and publish that information on their websites.[25]

The Supreme Court dashed hopes of conservative and progressive commentators alike that it would defer to states enacting these types of regulations in its 2024 decision *Moody v. NetChoice, LLC*, which vacated and remanded opposing decisions from the Eleventh and Fifth Circuits.[26] *Moody*

---

[21] Id. at 549 ("[T]he government likely has the power under the Constitution to mandate a truth-in-labeling law requiring social media platforms and other websites with large numbers of users to deploy the best reasonably available technology to label synthetic media containing altered video and audio images as 'altered.'").

[22] Id. at 552.

[23] *NetChoice*, 603 U.S. at 719-721.

[24] NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1203 (11th Cir. 2022), *vacated, and remanded sub. nom.* Moody v. NetChoice, LLC, 603 U.S. 707 (2024); NetChoice, LLC v. Paxton, 49 F.4th 439, 444 (5th Cir. 2022), *vacated, and remanded sub nom.* Moody v. NetChoice, LLC, 603 U.S. 707 (2024).

[25] Wash. Post v. McManus, 944 F.3d 506, 511–12 (4th Cir. 2019) (the law required that "within 48 hours of an ad being purchased, platforms must display somewhere on their site the identity of the purchaser, the individuals exercising control over the purchaser, and the total amount paid for the ad. They must keep that information online for at least a year following the relevant election. Second, there is an 'inspection requirement.' Under this part, platforms must collect records concerning their political ad purchasers and retain those records for at least a year after the election so that the Maryland Board of Elections can review them upon request.").

[26] See Tim Wu, The First Amendment Is Out of Control, N.Y. Times (July 2, 2024), https://www.nytimes.com/2024/07/02/opinion/supreme-court-netchoice-free-speech.html; see also Ash Johnson, The Conservative Weaponization of Government Against Tech, ITIF (Oct. 7 2024), https://itif.org/publications/2024/10/07/the-conservative-weaponization-of-government-against-tech/ ("This conservative techlash, or backlash against large tech companies and technology more generally, arose out of real concerns."); The Federalist Soc., Fireside Chat with Hon. Andrew Ferguson and Hon. Paul B. Matey,

v. NetChoice was the first Supreme Court case to address the extent to which states may foist far-reaching content moderation requirements on social media platforms. The Court found that both circuits had failed to conduct a proper analysis of the facial First Amendment challenges to the Florida and Texas laws regulating content moderation by large online platforms.[27] Tim Wu, a high-profile progressive legal scholar and former Obama and Biden White House official, reacted to the decision by warning that "[t]he First Amendment is Out of Control."[28] Wu fretted that the *NetChoice* decision could undermine national security, privacy protections, and hamper state efforts at common and long-practiced forms of economic regulation.[29] In sum, Wu argued the decision could fundamentally undermine our democracy, which should normally permit that "the people [] have the right to react to and control [immense and concentrated] private power."[30] *NetChoice* only added to this problem, extending the degree to which "[n]early any law that has to do with the movement of information can be attacked in the name of the First Amendment."[31]

All the more surprising is that Wu's criticisms, despite his progressive credentials, align most closely, not only with many political conservatives,[32] but with the conservative wing of the Court as epitomized by Justices Alito and Thomas's *NetChoice* concurrences.[33] And yet, while Wu is correct that the *NetChoice* decision rejected extensive efforts to regulate online platforms' content moderation decisions,[34] the majority and the concurrences have left enough unaddressed that *NetChoice* does not represent a total bar on novel regulations of online speech.[35] Fleshing out the details of the majority and each of the concurrences demonstrates that *NetChoice* limits content moderation regulations while leaving the door open to less sweeping efforts for state-mandated guidance of online speech moderation. But first

---

YouTube (June 20, 2024), https://www.youtube.com/watch?v=HJovWtxAWp0, (conservative FTC Commissioner Andrew Ferguson, who is now FTC chair, offering support for the state laws in *NetChoice*).

[27] *NetChoice*, 603 U.S. 707, 708 (2024).

[28] Wu, supra note 26.

[29] Id.

[30] Id.

[31] Id.

[32] See Justin Wise & Tonya Riley, Trump FTC Pick Wants to Avoid AI Crackdown, Target 'Censorship', Bloomberg (Dec. 12, 2024, 4:45 AM), https://news.bloomberglaw.com/antitrust/trump-ftc-pick-wants-to-avoid-ai-crackdown-target-censorship.

[33] *NetChoice*, 603 U.S. at 796 (Alito, J., concurring); id. at 750 (Thomas, J., concurring).

[34] Wu, supra note 26.

[35] *NetChoice*, 603 U.S. at 795 (Alito, J., concurring); id. at 748–49 (Jackson, J., concurring); id. at 745–47 (Barrett, J., concurring).

we must understand the doctrinal bedrock on which these competing opinions stand.

I. THE MARKETPLACE OF IDEAS, ALGORITHMS, AND THE CHANGING NATURE OF ONLINE POLITICAL SPEECH

A. The Marketplace and its Critics

The most promising avenue for mandating the disclosure, labeling, takedown, or deprioritization of altered and AI content is judicial recognition of the changing nature of online communication, with particular attention to the impact of algorithms on online speech. Central to such recognition is an understanding of the "marketplace of ideas" metaphor on which First Amendment protections have long been based.[36] The marketplace of ideas was traditionally justified under the general notion that the effects of untruthful, harmful, and hateful speech could be neutralized by the persuasive force of truthful and righteous speech.[37] In turn, the marketplace naturally "resists governmental regulation."[38] As Justice Kennedy wrote in *United States v. Alvarez*: "[t]he remedy for speech that is false is speech that is true."[39] While the marketplace of ideas model had detractors even long before the rise of the internet, the evolution of online speech is arguably the single biggest challenge the marketplace model has witnessed since its promulgation a rough century ago. As written by Professors Frederick Schauer and Daniel Ho: "[A] considerable amount of existing empirical research . . . tends . . . to justify skepticism about the causal efficacy of establishing an open market-place of ideas in identifying true propositions and rejecting false ones."[40]

---

[36] See, e.g., Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting) ("[T]he best test of truth is the power of the thought to get itself accepted in the competition of the market. . . ."). The "marketplace of ideas" model refers to the theory "that speaking and writing deserve special legal, constitutional, and political protection because the unfettered exchange of ideas advances truth and knowledge." Daniel E. Ho & Frederick Schauer, Testing the Marketplace of Ideas, 90 N.Y.U. L. Rev. 1160, 1161 (2015).

[37] See, e.g., *Abrams*, 250 U.S. at 630.

[38] Wash. Post v. McManus, 944 F.3d 506, 523 (4th Cir. 2019).

[39] United States v. Alvarez, 567 U.S. 709, 727 (2012).

[40] Ho & Schauer, *supra* note 36, at 1163 (2015); see also Vincent Blasi, Toward a Theory of Prior Restraint: The Central Linkage, 66 Minn. L. Rev. 11, 73 (1981) (quoting Whitney v. California, 274 U.S. 357, 375 (Brandeis, J., concurring)) ("Brandeis's dictum that 'the fitting remedy for evil counsels is good ones' rings hollow to an age that has seen demagogues destined to perpetrate unspeakable horrors use the facilities of mass communication to acquire and retain political power."); Frederick Schauer, Facts and the First Amendment, 57 UCLA L. Rev. 897, 910–11 (2010) ("[T]he persistence of the belief that a good

Nonetheless, the Supreme Court has typically refused to treat alleged defects in the marketplace of ideas as cause for reducing the protections of the First Amendment. Despite recognizing in *Miami Herald Publishing Co. v. Tornillo*, that "[t]he First Amendment interest of the public in being informed is said to be in peril because the 'marketplace of ideas' is today a monopoly controlled by the owners of the market," the Court rejected a law requiring newspapers to allow equal access to political candidates in the case of political, editorial, or endorsement content.[41] Thus, the marketplace remains sacred in First Amendment law even when it operates in a less-than-ideal manner.[42]

Professor Eugene Volokh was ahead of the curve in anticipating online threats to the marketplace model, writing in 1995 on how the internet would usher in the coming wave of "cheap speech" that would transform mass communications.[43] As has also been argued by Toni Massaro and Helen Norton: "speedy, cheap, and abundant speech does not always translate into more ideas, nor does it always maximize listeners' choices."[44] Rather, it often allows hidden actors with technological abilities to "undermine free speech and democratic values."[45] Tim Wu has likewise argued that a new speech environment has emerged where two of the First Amendment's foundational assumptions, that information is scarce and that listeners' attention is abundant, are no longer accurate.[46] Additionally, Hillary Clinton has called

---

remedy for false speech is more speech, or that truth will prevail in the long run, may itself be an example of the resistance of false factual propositions to argument and counterexample.").

[41] Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 251 (1974). But see Moody v. NetChoice, 603 U.S. 707, 795 (2024) (Alito, J., concurring) (suggesting the court consider the "enormous power" possessed by social media platforms to shape the public square). While *Reno* did not address Section 230, that statute was enacted by Congress with the express intent of preserving "the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation." 47 U.S.C. § 230(b)(2); see also Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1099-1100 (9th Cir. 2009) (citing sections 230(a)(3) and 230(b)(2) for the proposition that free speech values underlie the immunity provision).

[42] See *NetChoice*, 603 U.S. at 710 ("[T]he government cannot get its way just by asserting an interest in better balancing the marketplace of ideas. In case after case, the Court has barred the government from forcing a private speaker to present views it wished to spurn in order to rejigger the expressive realm.").

[43] Eugene Volokh, Cheap Speech and What It Will Do, 104 Yale L.J. 1805, 1806–07 (1995).

[44] Toni M. Massaro & Helen Norton, Free Speech and Democracy: A Primer for Twenty-First Century Reformers, 54 U.C. Davis L. Rev. 1631, 1639 (2021).

[45] Id.

[46] Wu, supra note 8, at 548–49 ("The most important change in the expressive environment can be boiled down to one idea: it is no longer speech itself that is scarce, but the attention of listeners. Emerging threats to public discourse take advantage of this change. . . . More precisely, the emergent techniques of speech control depend on new punishments, like the unleashing of 'troll armies' to abuse critics, the fabrication of news, and 'flooding' tactics that distort or drown out other speech through the payment of fake commentators or the deployment of propaganda robots.").

for states and the federal government to take a stronger role in regulating online content and mandating content moderation.[47]

Justice Thomas has raised similar issues in calling for the reconsideration of current defamation precedent: "The proliferation of falsehoods is, and always has been, a serious matter. . . . [W]e should give them only the protection the First Amendment requires."[48] Justice Gorsuch, dissenting in *Berisha v. Lawson*, likewise noted that on one social network: "falsehood and rumor dominated truth by every metric, reaching more people, penetrating deeper . . . and doing so more quickly than accurate statements." As a result, "'the distribution of disinformation'—which 'costs almost nothing to generate'—has become a 'profitable' business while 'the economic model that supported reporters, fact-checking, and editorial oversight' has 'deeply erod[ed].'" [49]

These commentators of diverse ideological backgrounds point to an obvious contemporary fact of digital life: for today's average social media user, there is no shortage of raw information, but rather a scarcity of reliable methodologies to sift through the speech with which we come into contact, increasing difficulty in comprehending whether we are receiving the speech that would best inform us and whether the information we are receiving has been produced and curated with intentions traditionally seen as supported by the policy of the First Amendment (mainly truth-seeking and informed self-governance).[50] These concerns take on added significance where the ability to understand speech is essential to democratic self-governance. If AI or altered content develop to the point that the average voter cannot reliably understand which purported candidate or officeholder statements are real and which are fake, then previous warnings that we have entered a "post-truth era" will take on newfound significance.[51] Indeed, while the Harris video

---

[47] Ashleigh Fields, Hillary Clinton Pushes for Stronger Social Media Regulation, The Hill (Oct. 5, 2024, 4:11 PM), https://thehill.com/policy/4917812-clinton-social-media-safety-regulations/ (stating that "if [social media platforms] don't moderate and monitor the content we lose total control and it's not just the social and psychological effects it's real harm, it's child porn and threats of violence, things that are terribly dangerous").

[48] Berisha v. Lawson, 141 S. Ct. 2424, 2425 (2021) (Thomas, J., dissenting).

[49] Id. at 2427 (Gorsuch, J., dissenting) (citing David A. Logan, Rescuing Our Democracy by Rethinking *New York Times Co v. Sullivan*, 81 Ohio St. L.J. 759, 800, 804 (2020)).

[50] See Citron & Chesney, supra note 2, at 1768 (discussing "filter bubbles," which "can be powerful insulators against the influence of contrary information. In a study of Facebook users, researchers found that individuals reading fact-checking articles had not originally consumed the fake news at issue, and those who consumed fake news in the first place almost never read a fact-check that might debunk it."); First Nat'l Bank of Bos. v. Bellotti, 435 U.S. 765, 768–69 (1978).

[51] Hasen, Bots, supra note 8, at 537; see also Elisa Shearer, Michael Lipka, Sarah Naseer, Emily Tomasik & Mark Jurkowitz, Americans' Views of 2024 Election News, Pew Rsch. Ctr. (Oct. 10, 2024),

shared by Musk was sufficiently crude that the average user could detect it to be AI-generated,[52] we must remember that 2024 was the first national election that widely featured this content. We should expect the sophistication of altered and AI political speech to grow each election cycle.

Justice Gorsuch, Justice Thomas, and other commentators have speculated on the adoption of new policies or precedents to address the supposed scourge of mis- and disinformation.[53] While these terms have risen in prominence alongside concerns about "fake news," they are not legal terms with fixed statutory or common law meanings. Some proposed definitions state that disinformation is "deliberately deceptive whereas misinformation is false information that is created and spread regardless of an intent to harm or deceive."[54] The Supreme Court has offered no technical definition of what amounts to mis- or disinformation.[55] Nonetheless, these terms capture something real and novel: it has never been easier for the average person to spread poorly researched or intentionally false claims to a wide audience.

B. Examples of Algorithms Impacting Speech

There are countless examples of this rising difficulty listeners face in sifting true from false speech. In addition to the Russian influence campaign of 2016, countless lawmakers have called for banning TikTok,[56] arguing that the app is serving as a tool of the Chinese Communist Party to shape all manner of American cultural and political attitudes while spreading mis- and disinformation.[57] Former Republican Congressman Mike Gallagher has

---

https://www.pewresearch.org/journalism/2024/10/10/americans-views-of-2024-election-news/ (finding that roughly half of American voters struggled to discern what was true and what wasn't about election news).

[52] See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1192 (E.D. Cal. 2024).

[53] Citron & Chesney, supra note 2, at 1787–92; see also Hasen, Bots, supra note 8, at 549 (arguing why some mandatory content moderation proposals are justifiable under First Amendment doctrine).

[54] David S. Ardia & Evan Ringel, First Amendment Limits on State Laws Targeting Election Misinformation, 20 First Amend. L. Rev. 291, 370 n.375 (2022); see also Deen Freelon & Chris Wells, Disinformation as Political Communication, 37 Pol. Commc'n 145, 145 (2020) (explaining that disinformation includes "three critical criteria: 1) deception, 2) potential for harm, and 3) an intent to harm").

[55] The word "disinformation" has only been discussed once in a Supreme Court case, by Justice Gorsuch in his dissent to the denial of certiorari in *Berisha v. Lawson*. He offered no technical or exact legal refinement of what disinformation entails. 141 S. Ct. 2424, 2427 (2021) (Gorsuch, J., dissenting).

[56] Sapna Maheshwari & Amanda Holpuch, Why TikTok Is Facing a U.S. Ban, and What Could Happen Next, N.Y. Times (Jan. 17, 2025), https://www.nytimes.com/article/tiktok-ban.html; Sapna Maheshwari & David McCabe, TikTok Prompts Users to Call Congress to Fight Possible Ban, N.Y. Times (March 7, 2024), https://www.nytimes.com/2024/03/07/business/tiktok-phone-calls-congress.html.

[57] See Adonis Hoffman, TikTok is China's Trojan Horse, The Hill (Oct. 18, 2022, 4:30 PM), https://thehill.com/opinion/technology/3694346-tiktok-is-chinas-trojan-horse/; Ian Fisher, TikTok is a

accused TikTok of spreading "rampant pro-Hamas propaganda on the app" by manipulating algorithms.[58] On then-Twitter and other social media platforms, tech companies were frequently accused of "shadow banning" right-wing voices and promoting progressive "woke" speech,[59] while now Musk's X has been accused of tweaking its algorithms to promote hate speech and "extremist content."[60]

These criticisms share something in common: the locus of the concern is not only the quantity of speech through which listeners are forced to sift, but methods of promotion and delivery of that content to users. Indeed, complex algorithms often curate our environment in ways which shift the average voice we hear to better approximate our already existing views. Those algorithms typically enact certain preferences which result from decisions about what information is more or less valuable. [61] As Musk's Twitter acquisition effectively demonstrated, algorithmic curation determines a wide array of the content to which the average user is exposed, without any new effort made by the user or listener.[62] Likewise, on TikTok today, the user need not even "doomscroll—you can just sit and watch and let the platform do the rest."[63] Before the widespread adoption of this technology, users would affirmatively seek out content: on early Facebook (late 2000s to early 2010s),

---

'Massive Surveillance' Tool for China, Senators Warn as Biden Admin Weighs Proposal to Spare App from U.S. Ban, Fortune, (Nov. 20, 2022, 12:32 PM), https://fortune.com/2022/11/20/tiktok-surveillance-tool-china-senators-biden-admin-security-agreement-ban/; The Select Committee on the CCP, Intel Chiefs Warn of CCP Ties to TikTok in March 2023 at Gallagher's Request, YouTube (Oct. 1, 2024), //www.youtube.com/watch?v=WyygurBSC8U.

[58] Representative Mike Gallagher, Why Do Young Americans Support Hamas? Look at TikTok, Free Press (Nov. 1, 2023), https://www.thefp.com/p/tik-tok-young-americans-hamas-mike-gallag.

[59] NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1205 (11th Cir. 2022), *vacated, and remanded sub. nom* Moody v. NetChoice, 603 U.S. 707; Concurring Statement of Commissioner Andrew N. Ferguson, FTC v. 1661, Inc. d/b/a GOAT, F.T.C. Matter Number 2223016, at 4 (Dec. 2, 2024), https://www.ftc.gov/system/files/ftc_gov/pdf/ferguson-goat-concurrence.pdf (FTC Commissioner Ferguson arguing "Censorship, even if carried out transparently and honestly, is inimical to American democracy. The [FTC] must use the full extent of its authority to protect the free speech of all Americans. That authority includes the power to investigate collusion that may suppress competition and, in doing so, suppress free speech online. We ought to conduct such an investigation. And if our investigation reveals anti-competitive cartels that facilitate or promote censorship, we ought to bust them up.").

[60] Faiz Siddiqui & Jeremy B. Merrill, Elon Musk's Twitter Pushes Hate Speech, Extremist Content into "For You" Pages, Wash. Post (Mar. 30, 2023), https://www.washingtonpost.com/technology/2023/03/30/elon-musk-twitter-hate-speech/.

[61] Claudio Lombardi, The Illusion of a "Marketplace of Ideas" and the Right to Truth, Am. Aff., Spring 2019, https://americanaffairsjournal.org/2019/02/the-illusion-of-a-marketplace-of-ideas-and-the-right-to-truth/.

[62] See Mitchell Clark, Twitter Takes its Algorithm "Open-Source," as Elon Musk Promised, The Verge (Mar. 31, 2023, 2:52 PM), https://www.theverge.com/2023/3/31/23664849/twitter-releases-algorithm-musk-open-source.

[63] Sam Schechner et al., What TikTok Is Showing America's Youth About the War in Gaza, Wall St. J. (Dec. 22, 2023, 9:12 AM), https://www.wsj.com/tech/tiktok-israel-gaza-hamas-war-a5dfa0ee.

the user would receive updates from connections as they posted new content (or poked you), and little else. Older still, blogs, forums, and news sites required the user to seek out the desired content themselves.[64]

As algorithms came to dominate social media design, this affirmative search for content became less and sometimes totally unnecessary. For instance, in a test case where the *Wall Street Journal* created a handful of bots registered as 13-year-old users, "[w]ithin hours after signing up, TikTok began serving some accounts highly polarized content, reflecting often extreme pro-Palestinian or pro-Israel positions about the [Israel-Hamas] conflict."[65] Even when one account was set to restricted mode, which limits sensitive content, the "account's feed was almost entirely dominated by vivid images and descriptions of the conflict."[66] No affirmative choice was made to seek out mostly war-related content. YouTube in particular has been blamed for a purported rising wave of right-wing extremism, by taking users down a "rabbit hole" via algorithmic suggestions to popular conspiratorialists, such as Alex Jones.[67] YouTube's eventual and successful efforts to crack down on the algorithmic conspiratorial pipeline from 2019 onwards have produced substantial evidence indicating that, indeed, like with the acquisition of Twitter, algorithms play a prime role in shaping what users see and hear.[68]

In turn, these curated feedback loops have the demonstrated ability to take users into fundamentally distinct and siloed information spheres, which look little like a supermarket kind of marketplace featuring competing brands, but rather niche outlets offering highly specific goods to committed consumers who rarely interact with different cohorts of sellers and buyers.[69] Where scholars such as Richard Hasen have emphasized the "radical new

---

[64] Of course these websites often had front pages and search functions, but the actual obtaining of the video or the news story required an affirmative choice on the part of the user-listener; likewise, the decline of the independent blog model and the rise of major platforms has led to a consolidation of discussion and public communication onto a handful of platforms, and has reduced what the Supreme Court in *Reno* called the "astoundingly diverse content" that was present online in 1997. Reno v. ACLU, 521 U.S. 844, 846 n.30 (1997); Kyle Chayka, How the Internet Turned Us Into Content Machines, The New Yorker (June 4, 2022), https://www.newyorker.com/culture/infinite-scroll/how-the-internet-turned-us-into-content-machines (detailing changes in the structure of digital media aimed drawing in users and turning use of social media into a more passive process).

[65] Schechner et al., supra note 63.

[66] Id.

[67] See Conor Fridersdorf, YouTube Extremism and the Long Tail, The Atlantic (Mar. 12, 2018), https://www.theatlantic.com/politics/archive/2018/03/youtube-extremism-and-the-long-tail/555350/; Kaitlyn Tiffany, Very, Very Few People Are Falling Down the YouTube Rabbit Hole, The Atlantic (Aug. 30, 2023), https://www.theatlantic.com/technology/archive/2023/08/youtube-rabbit-holes-american-politics/675186/.

[68] Tiffany, supra note 67.

[69] See Citron & Chesney, supra note 2, at 1768.

opportunities for readers, viewers, and listeners to custom design what they read, see, and hear," increasingly these listeners are in fact not custom-designing their own information atmospheres, but are rather contributing to a communicative process over which they lack full control and awareness.[70]

Various commentators have mirrored these concerns about the siloing of listeners and viewers in the wake of the 2024 election, for instance in arguing that young male voters have shifted their support toward Donald Trump and the GOP as a result of mis- and disinformation supposedly found on right-wing podcasts and the "manopshere."[71] Many Democrats have reflected on the 2024 loss by speculating that they will be unable to compete with these young voters among whom traditional forms of media are far less influential unless they can break into this decentralized market of influencers, podcasters, and YouTubers.[72] Where algorithms can accelerate the movement of cohorts of listeners and voters to particular messages, and where those algorithms are designed with explicit ideological and/or disinformative purposes, government regulations may have newfound leeway where they would have been disfavored when applied to older forms of media.[73] Indeed, if it is the case that the government can play a role in targeting efforts by big tech to silence or censor conservative voices, as many conservatives in the United States including Vice President J.D. Vance[74] and Federal Trade Commission ("FTC") Chair Andrew Ferguson[75] have argued, then it follows that

---

[70] Hasen, Cheap Speech, supra note 10, at 201.

[71] See Caroline Haskins, Rogan, Musk and an Emboldened Manosphere Salute Trump's Win: 'Let that Sink In', Guardian (Nov. 7, 2024, 11:50 A.M.), https://www.theguardian.com/us-news/2024/nov/07/joe-rogan-elon-musk-heterodoxy-trump-win-reaction; Mina Haq, A Reporter's Take on the Role of Disinformation and Male Grievance in the Election, Pen Am. (Dec. 17, 2024), https://pen.org/a-reporters-take-on-the-role-of-disinformation-and-male-grievance-in-the-election/; Ja'han Jones, Right-Wingers are Spreading Misinformation to Black Voters through Podcasts and Influencers, MSNBC (June 26, 2024, 3:52 PM), https://www.msnbc.com/the-reidout/reidout-blog/black-voters-misinformation-podcasts-social-media-rcna159100.

[72] See Josh Marcus, After Trump Wins the 'Influencer Election', Why Some Democrats Want to Create their Own Joe Rogan, Indep. (Nov. 7, 2024, 9:13 PM), https://www.the-independent.com/news/world/americas/us-politics/joe-rogan-trump-kamala-harris-b2643492.html.

[73] See Moody v. NetChoice, LLC, 603 U.S. 707, 795 (2024) (Alito, J., concurring) (citing Ohio v. Am. Express Co., 585 U.S. 529 (2018)) ("Maybe we should think about the enormous power exercised by platforms like Facebook and YouTube as a result of 'network effects.' And maybe we should think about the unique ways in which social-media platforms influence public thought."); see also *NetChoice*, 603 U.S. at 732–33 ("It is critically important to have a well-functioning sphere of expression, in which citizens have access to information from many sources. That is the whole project of the First Amendment. And the government can take varied measures, like enforcing competition laws, to protect that access.").

[74] See Lulu Garcia-Navarro, The Interview: A Conversation with J.D. Vance, N.Y. Times (Oct. 12, 2024), https://www.nytimes.com/2024/10/12/magazine/jd-vance-interview.html.

[75] See Evan Swarztrauber, How Regulators Should Approach Big Tech and Corporate Power w/ Andrew Ferguson, The Dynamist (Nov. 22, 2024), https://thedynamist.simplecast.com/episodes/how-

the sophisticated if not technically indecipherable use of algorithms to achieve these goals could likewise be subject to regulatory scrutiny.

II. Supreme Court Precedent Governing Communicative Technologies

A. Debates Surrounding Speech and Technology

As Justice Bryon White wrote in the majority opinion of *Red Lion Broadcasting Co. v. FCC* (1969), the "differences in the characteristics of new media justify differences in the First Amendment standards applied to them."[76] The Court has continued to argue into the twenty-first century that the First Amendment protections afforded to speech on a given medium can be altered by the technological character of the medium in question.[77] However, the Court has yet to fully consider how it is now unnecessary to affirmatively seek out explicit or unreliable content to find it on the internet.[78]

It is clear that online speech is heard in a fundamentally different manner than when the seminal online speech case *Reno v. ACLU* was handed down in 1997. In *Reno*, the Court unanimously ruled that federal legislation mandating criminal penalties for the transmission of "obscene or indecent" material to minors was an unconstitutionally overbroad regulation of speech.[79] Indeed, in striking down federal regulations of online speech, the *Reno* Court explicitly distinguished the restrictions of broadcast television speech from

---

regulators-should-approach-big-tech-and-corporate-power-w-andrew-ferguson (arguing that were private companies to collude to deprioritize or shadow-ban certain kinds of speech that this would lead to liability under federal antitrust law).

[76] Red Lion Broad. Co. v. FCC, 395 U.S. 367, 386 (1969) (holding the First Amendment permits a federal agency to regulate the speech of broadcasters for the purpose of maintaining the public interest in equitable use of scarce broadcasting frequencies).

[77] See, e.g., Packingham v. North Carolina, 582 U.S. 98, 105 (2017); see also *NetChoice*, 603 U.S. at 793–96 (Alito, J., concurring) (discussing how technology can alter a First Amendment analysis); id. at 748–49 (Jackson, J., concurring) (doing the same); id. at 747 (Barrett, J., concurring) (doing the same).

[78] For instance, Instagram has been found to regularly recommend salacious and sexually explicit material to test accounts which were designed only to follow youth-themed content, such as young gymnasts, cheerleaders, and other teenage and preteen influencers. See Jeff Horwitz & Katherine Blunt, Instagram's Algorithm Delivers Toxic Video Mix to Adults Who Follow Children, Wall St. J. (Nov. 27, 2023, 5:30 A.M.), https://www.wsj.com/tech/meta-instagram-video-algorithm-children-adult-sexual-content-72874155; see also Todd Spangler, Instagram Algorithms Connect 'Vast' Network of Pedophiles Seeking Child Pornography, According to Researchers, Variety (June 7, 2023), https://variety.com/2023/digital/news/instagram-pedophile-network-child-pornography-researchers-1235635743/ (detailing how algorithms on Instagram have abetted those seeking child pornography); Jonathan Berr, Despite 'No Nudity Rule,' Instagram Is Chock Full Of Pornography, Forbes (Sept. 28, 2018, 2:38 P.M.), https://www.forbes.com/sites/jonathanberr/2018/09/28/despite-no-nudity-rule-instagram-is-chock-full-of-pornography/?sh=2a4ad3a33e26 (describing commonality of pornography on Instagram).

[79] 521 U.S. 844 (1997).

those that could be placed on online speech by emphasizing "the ease with which children may obtain access to broadcasts," and because televised "warnings could not adequately protect the listener from unexpected program content."[80] By contrast, *Reno* found "the risk of encountering indecent material by accident" on the internet of 1997 to be "remote because a series of affirmative steps [wa]s required to access specific material."[81]

Notably, the Fifth and Eleventh Circuits in the *NetChoice* cases engaged in extensive analysis of the technological particularities of social media platforms to reach their respective results. Indeed, the Fifth Circuit refused to extend editorial protections to social media platforms after concluding that their content moderation process hardly looks like the traditional media editorial process, and the Eleventh Circuit was keen to recognize the intense burdens of content moderation that running a social media platform entails.[82] As discussed below, the Supreme Court itself in *NetChoice* continued to analyze the permissible regulations of speech in the context of the technological specifics of the medium in question.[83] Thus, the First Amendment protections afforded to social media platforms are to some degree based in the technological structure of how those platforms moderate, curate, and promote certain kinds of speech.

This technologically-specific reasoning is a long-standing feature of First Amendment jurisprudence. In *Red Lion* (1969), Justice White wrote for the majority that the FCC could limit the granting of broadcasting licenses as well as maintain its fair reporting rules by broadcasters due to the scarce nature of radio frequencies and the resulting necessary limitations on broadcast licenses.[84] Additionally, in *FCC v. Pacifica Foundation* (1978) the Court again permitted regulations of broadcast television speech, allowing the FCC to restrict indecent material, there George Carlin's profanity-laden stand-up routine, from being played on the radio at certain hours. It did so on the basis that, were the FCC unable to limit the airing of profanities during daytime hours, they could "reach[] the ears of unsupervised children who

---

[80] Id. at 866–67.

[81] Id. at 867.

[82] See NetChoice, LLC v. Paxton, 49 F.4th 439, 488 (5th Cir. 2022) ("[E]ven if there was a different rule for disclosure requirements implicating a newspaper-like editorial process, that rule would not apply here because the Platforms have no such process."); NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1230 (11th Cir. 2022).

[83] See Moody v. NetChoice, LLC, 603 U.S. 707, 795–96 (2024) (Alito, J., concurring); id. at 748–49 (Jackson, J., concurring); id. at 746–47 (Barrett, J., concurring) (noting "the First Amendment implications of . . . [content moderation] laws might be different for [the] kind of algorithm" where "a platform's algorithm just presents automatically to each user whatever the algorithm thinks the user will like—*e.g.*, content similar to posts with which the user previously engaged").

[84] Red Lion Broad. Co. v. FCC, 395 U.S. 367, 391 (1969).

were likely to be in the audience at that hour."[85] Justice O'Connor in *Reno* extensively examined the technologies of broadcasting and the reasoning in *Red Lion* and *Pacifica*, relying on the technological specificity of those decisions, to elucidate similarly specific protections for online speech. Criticizing the overbroad design of the Communications Decency Act, Justice O'Connor noted that there was no effective way to engage in age verification of users or block users from encountering sexually explicit material without overly restricting access to non-explicit content.[86]

Regardless of their accuracy in 1997, some of these statements seem quaint when applied to the internet today. Various states have been able to successfully mandate the requirement that credit cards be linked to an individual's identity and age be used to verify access to pornographic websites.[87] Search engines today contain built-in filters that can reliably prevent the appearance of links to sexually explicit websites and imagery.[88] It is far more plausible today than in 1997 that a state or the federal government could mandate these technologies as the default mode for use of the internet, which can likewise be overcome by adults through a series of affirmative steps. Crucially, were a court to distinguish *Reno* by relying on technological change to regulate online speech more rigorously than the Court did there, this would not represent an overturning of *Reno* so much as an extension of its technologically-based reasoning that speech regulations should hew to the specific character of the medium in question.

The Supreme Court doubled down on robust protection for online speech in 2017, in *Packingham v. North Carolina*.[89] At the same time, Justice Kennedy was conscious of the inherent danger is pronouncing wide-ranging rules governing the exactitude of online speech: "[t]he Internet's forces and directions are so new, so protean, and so far reaching that courts must be conscious that what they say today may be obsolete tomorrow."[90]

---

[85] FCC v. Pacifica Found., 438 U.S. 726, 757 (1978) (Powell, J., concurring).

[86] *Reno,* 521 U.S. at 854–57.

[87] See Holly Richardson, A Simple Law Does the Unthinkable: It Sends the Porn Industry into Retreat, Deseret News (Aug. 8, 2023, 1:00 P.M.), https://www.deseret.com/utah/2023/8/8/23824414/utah-law-pornhub-industry-retreats/.

[88] See, e.g., Your SafeSearch Setting, https://www.google.com/safesearch (last visited Oct. 5, 2024).

[89] Packingham v. North Carolina, 582 U.S. 98, 99 (2017).

[90] Id. at 99, 105 ("The nature of a revolution in thought can be that, in its early stages, even its participants may be unaware of it. And when awareness comes, they still may be unable to know or foresee where its changes lead."). *Packingham* was a 5-3 decision, with Justice Gorsuch not participating and Justices Thomas, Roberts, and Alito concurring only with respect to concerns of overbreadth of the North Carolina law. The death of Justice Ginsburg and appointments of Justices Gorsuch, Kavanaugh, and Coney Barrett suggest that *Packingham*'s broad reasoning establishing an absolute right to access the internet may one day be revisited, and that more tailored restrictions on internet access, use, and speech may be upheld in the future.

Nonetheless, Kennedy did not interpret these warnings to prevent the recognition of a powerful First Amendment right to access the internet, in some shape or form, regardless of criminal background. At first glance, *Packingham* does not necessarily implicate the regulation of deceptive online speech per se: it is obvious and unremarkable that a user could be permitted access to the internet without everything they say being protected from regulation.[91]

In light of these conclusions, two important takeaways from *Packingham* remain for the purpose of the regulation of AI and altered content and content moderation. First, access to the internet is protected by the First Amendment to such a degree that if sex offenders are protected, then those spreaders of political mis- and disinformation are certainly protected as well, at least in the sense that the government cannot bar these users from accessing social media entirely.[92] At the same time, despite his expansive language, Justice Kennedy emphasized that protections governing online speech remain subject to the changing nature of the technology in question.[93] Thus, as in *Reno*, it remains a legally valid conclusion that the specific mechanics of a communicative technology undergird any analysis of the rights to speak through that technology. As technology changes, the rights to speak through or host content on that technology may be reexamined.[94]

Despite the fact that the First Amendment protections to online speech remain doctrinally as robust, if not more so, than protections offered to offline speech, the Supreme Court has not failed to notice technological changes which have altered the nature of mass communications in the twenty-first century. Dissenting in *Berisha v. Lawson* (2021), Justice Gorsuch wrote extensively to argue that the transformation of the media landscape should lead the Court to reconsider the landmark defamation case of *New York Times v. Sullivan*. Remarking on these changes, Gorsuch noted that: "some reports suggest that our new media environment also facilitates the spread of disinformation."[95] Indeed, as in *Reno*, *Packingham*, *Red Lion*, and *Pacifica*, the Court has been transparent about the nature of speech protections often being directly based on the technological foundations of the

---

[91] Id. at 108 ("Even convicted criminals—and in some instances especially convicted criminals—might receive legitimate benefits from [online] means for access to the world of ideas, in particular if they seek to reform and to pursue lawful and rewarding lives.").

[92] Citizens United v. FEC, 558 U.S. 310, 339 (2010) (citations omitted) ("The First Amendment 'has its fullest and most urgent application' to speech uttered during a campaign for political office.").

[93] *Packingham*, 582 U.S. at 98.

[94] Moody v. NetChoice, LLC, 603 U.S. 707, 745 (2024) (Barrett, J., concurring); id. at 748-49 (Jackson, J., concurring); see also Wash. Post v. McManus, 944 F.3d 506, 519 (4th Cir. 2019) ("In short, what goes for broadcasters is too much a product of their technical circumstances to serve as a template for state regulation writ large.").

[95] Berisha v. Lawson, 141 S. Ct. 2424, 2427 (2021) (Gorsuch, J., dissenting).

medium.[96] It is the changing nature of online speech which led Justice Gorsuch, as well as Justice Thomas, in *Berisha* to question the continuing relevance of the *New York Times* standard for defamation.[97] Justice Gorsuch has pondered whether defamation law designed in the 1960s can still be justified "in a new era where the old economic model that supported reporters, fact-checking, and editorial oversight is disappearing."[98] Indeed, if new technologies act "in favor of those who can disseminate the most sensational information as efficiently as possible without any particular concern for truth," then why should First Amendment protections based on a moribund economic model remain the law?[99]

B. Applying the Court's Technological Reasoning to Algorithms

Today, content is no longer affirmatively sought out in the manner it once was; rather, speakers and listeners interact in a subtle, curated, and obscure series of thousands upon thousands of interactions over months and years which deliver an individual user to a unique information atmosphere.[100] This cannot be said to be a series of purely "affirmative steps." How the Supreme Court might address these technological changes in applying the First Amendment is hard to predict, but the argument may well take the following form. First, online speech no longer arrives in front of a user through a series of affirmative steps. Rather, like broadcast television,[101] the process through which the speech comes before the user is a less-than-predictable process of sorting, performed by inscrutable algorithms which the user typically does not understand and where the designers often fail to predict the exact consequences of their design.[102] Or, worse, these algorithms are often designed to drive users toward the most shocking and/or titillating content to drive engagement.[103]

---

[96] FCC v. Pacifica Found., 438 U.S. 726, 748–49 (1978); Reno v. ACLU, 521 U.S. 844, 857 (1997); *Packingham*, 582 U.S. at 98–99, 104–05; Red Lion Broad. Co. v. FCC, 395 U.S. 367, 386 (1969).

[97] *Berisha*, 141 S. Ct. at 2427 (Gorsuch, J., dissenting); id. at 2425 (Thomas, J., dissenting).

[98] Id. at 2428 (Gorsuch, J., dissenting).

[99] Id. ("What started in 1964 with a decision to tolerate the occasional falsehood to ensure robust reporting by a comparative handful of print and broadcast outlets has evolved into an ironclad subsidy for the publication of falsehoods by means and on a scale previously unimaginable.").

[100] See also Hasen, Bots, supra note 8, at 537 ("Political campaigns . . . increasingly take place under conditions of voter mistrust and groupthink, with the potential for foreign interference and domestic political manipulation via new and increasingly sophisticated technological tools."); Citron & Chesney, supra note 2, at 1768 (discussing filter bubbles, which insulate listeners from contrary information).

[101] See *Red Lion*, 395 U.S. at 367.

[102] See Citron & Chesney, supra note 2, at 1768.

[103] See Schechner et al., supra note 63.

These processes have a record of producing harms, where speakers are often confronted with obscene, misleading, or outright false information that they themselves did not affirmatively seek out as a starting point. Furthermore, these processes are highly vulnerable to malicious and sophisticated actors who seek to manipulate public opinion.[104] As a result, social media does not resemble the 1997 internet of *Reno*, and these harms and technological designs should permit more leeway for the government to foist appropriately tailored burdens onto platforms to help users understand the nature of the speech they are hearing and seeing.[105] For the marketplace of ideas to work, consumers arguably need to be informed to some minimal degree of the nature by which the products are manufactured and with what incentives they are produced.[106] This kind of argument has real heft and is completely in keeping with the technologically fine-tuned logic at play in *Reno*, *Pacifica*, *Red Lion*, and *Packingham*.

Social media's defenders have arguments to counter this position, and concerns about passive algorithmic sorting will not serve as a blank check for state actors to regulate social media as they see fit.[107] This is especially so because much of this speech will be political and ideological in nature, the kind of speech which receives the greatest protection from the First Amendment.[108] Platforms and defenders of the maximally libertarian position on speech will contend that algorithms only function because private platforms and their users make a series of original choices about when to swipe or click on particular content, or what content to prioritize.[109] They will likewise argue that algorithms are editorial tools which are themselves vested with First Amendment protections.[110]

Yet, libertarians still must answer for technological problems which have not been present in broadcast media or newspaper cases. It is hard to argue that someone who is interested in fitness content has chosen to receive pornographic content.[111] It is hard to argue that simply because a person follows pro-gun accounts that they also want to be deluged with anti-vaccine content. Nor is it the case that simply because someone is a progressive that they

---

[104] See Massaro & Norton, supra note 44, at 1635.

[105] Reno v. ACLU, 521 U.S. 844, 866–67 (1997).

[106] See Moody v. NetChoice, LLC, 603 U.S. 707, 746–47 (2024) (Barrett, J., concurring).

[107] See Nat'l Inst. of Fam. and Life Advoc. v. Becerra, 585 U.S. 755, 766 (2018).

[108] Citizens United v. FEC, 558 U.S. 310, 339 (2010); Buckley v. Valeo, 424 U.S. 1, 48–49 (1976).

[109] E.g., *NetChoice*, 603 U.S. at 734–35 (discussing how algorithms typically function to curate speech on social media platforms); see also Eugene Volokh & Donald M. Falk, First Amendment Protection for Search Engine Results, 8 J.L. Econ. & Pol. 883, 887–88 (2012) (arguing that search engines deserve First Amendment protections as the result of editorial choices made by programmers).

[110] See *NetChoice*, 603 U.S. at 735; Volokh & Falk, supra note 109.

[111] See Berr, supra note 78.

want to receive anti-Zionist content, or that because they watched one war video they want their feed to be mostly war videos.[112] Where an algorithm leads users to content that is altered or AI-generated, the potential for mis- and disinformation and subsequent harm clearly increases. Sensible measures to alert the user to that trajectory should not be thrown out as constitutionally invalid, without a more particularized inquiry, on the basis of precedents designed for very different technologies. Yet, after *NetChoice*, it is increasingly likely that state mandates for content moderation will still face significant scrutiny.[113]

III. *NETCHOICE'S* APPROACH TO ALGORITHMIC CURATION

A. Justice Kagan's Traditional Approach to Editorial Curation

In *NetChoice*, the Supreme Court vacated both lower court rulings because neither had determined with sufficient specificity which platforms and which platform features are covered by the Florida and Texas laws, a detailed inquiry that the Justices declined to undertake themselves.[114] Nonetheless, while Justices Alito and Thomas noted in their concurrences that the majority's discussion of the First Amendment is dicta, the *NetChoice* majority offered ample guidance to lower courts in dealing with the issues raised in these cases. While the majority appeared willing to extend broad editorial rights to online platforms, the various concurrences reveal continuing divisions over the exact method by which the First Amendment should be applied to new technologies like algorithms.

There appear to be at least five votes for holding that the Eleventh Circuit more correctly applied First Amendment jurisprudence covering the editorial discretion possessed by hosts of third-party speech.[115] The Eleventh Circuit held that Florida's restrictions on content moderation unconstitutionally burdened the editorial discretion of social media platforms, preventing those platforms from acting freely on their own views "about the sorts of content and viewpoints that are valuable and appropriate for dissemination."[116]

---

[112] See Schechner et al., supra note 63.

[113] *NetChoice*, 603 U.S. at 739–45.

[114] Id. at 707–08.

[115] As Justice Kagan wrote for the majority, "the Eleventh Circuit . . . saw the First Amendment issues much as we do." Id. at 727. Justice Barrett, concurring, wrote "the Eleventh Circuit's understanding of the First Amendment's protection of editorial discretion was generally correct; the Fifth Circuit's was not," a view which Justice Jackson endorsed in full. Id. at 745 (Barrett, J., concurring); id. at 748–49 (Jackson, J., concurring).

[116] NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1209–10, 1216 (11th Cir. 2022).

However, the *NetChoice* Court continued the long debate over how to apply the First Amendment's protections to new technologies.[117] The concurrences of Justices Barrett, Jackson, and Alito (joined by Thomas and Gorsuch) all indicate a willingness to tailor regulations of algorithmic speech on social media platforms to the character of the technologies themselves.[118]

NetChoice stems from laws passed by Florida[119] and Texas[120] in 2021, both of which aimed to restrict the ability of social media platforms to deprioritize, "shadow ban," and remove altogether speech for ideological reasons, with the intent of counteracting what the law's supporters viewed as the censorship of conservative views online.[121] Justice Kagan writing for the majority was willing to largely extend the same protection to algorithmic content moderation which has long been enjoyed by newspapers and parades.[122] Relying on language from the Eleventh Circuit's opinion, Justice Kagan outlined agreement with the view that when a social media platform "removes or deprioritizes a user or post," it improperly substitutes its own "judgment" for those of the platform's owners and moderators about which content to prioritize.[123] In turn, despite their immense power and the fact that "today's social media pose dangers not" posed by "old media,"[124] the majority concluded that the social media platforms "unabashedly control the content that will appear to users, exercising authority to remove, label or demote messages they disfavor."[125] As such, the Texas and Florida laws could not withstand review because they "profoundly alter[] the platforms' choices about the views they will, and will not, convey."[126] This opinion extends a longstanding First Amendment principle to social media: mandated content moderation aiming to alter the expression of private viewpoints is disfavored.

By extending these protections to the speech hosted by social media platforms, the Court identified social media companies as precisely the kinds of

---

[117] See, e.g., Reno v. ACLU, 521 U.S. 844, 844 (1997) (rejecting federal online child safety regulations as overly burdensome of First Amendment rights); Red Lion Broad. Co. v. FCC, 395 U.S. 367, 367 (1969) (permitting regulations of speech of television broadcasters); FCC v. Pacifica Found., 438 U.S. 726, 726 (1978) (permitting regulations of obscenities on broadcast television).

[118] *NetChoice*, 603 U.S. at 794–96 (Alito, J., concurring); id. at 748–49 (Jackson, J., concurring); id. at 745–47 (Barrett, J., concurring).

[119] S. 7072, 2021 Leg., Reg. Sess. (Fla. 2021).

[120] H. 20, 87th Leg. Sess., 2nd Spec. Sess. (Tex. 2021).

[121] See *NetChoice*, 603 U.S. at 740–41 (providing some examples of the partisan motivations behind this legislation).

[122] Id. at 731–33.

[123] Id. at 722 (quoting NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1209, 1210 (11th Cir. 2022)).

[124] Id. at 733.

[125] Id. at 736.

[126] Id. at 737.

private actors that the government generally may not order to carry messages those actors wish to avoid.[127] It does not matter that the platforms are merely hosting the speech of third parties. Justice Kagan's majority and the Eleventh Circuit would allow for broad protections for all expressive choices, including the curation, assembly, editing, and prioritization of viewpoint within a collection of third-party speech,[128] the same protections which the Supreme Court granted older media like newspapers in *Miami Herald v. Tornillo*.[129]

B. A Splintered Outcome

It is not clear that there are five votes for Justice Kagan's view that algorithmic curation merits similar levels of protection as traditional editorial practices. At least five justices voiced doubts that editorial rights as understood in *Tornillo* could easily be applied to the algorithmic curation at issue in *NetChoice*.[130] In that case, the Court struck down a Florida law requiring that a newspaper give a political candidate a right to reply when it published criticism of his record.[131] Justice Alito found it hard to imagine that the law should treat "what newspaper editors did more than a half-century ago" and "what Facebook and YouTube do today" identically, given that "[n]o human being could possibly review" the monumental amount of data which these platforms process.[132] Justice Alito argued this distinction leads to the conclusion that algorithmic social media feeds cannot be said to transmit a "shared message"[133] or a '"collective point,"'[134] which would in turn render *Tornillo's* protections for editorial rights inapposite.[135]

Justices Barrett and Jackson voiced similar concerns regarding the application of older precedents to such new technology, without embracing Justice Alito's deeper skepticism that those precedents were inapplicable to the

---

[127] Id. at 736–38.

[128] Id. at 710-11 ("Like the editors, cable operators, and parade organizers this Court has previously considered, the major social-media platforms curate their feeds by combining 'multifarious voices' to create a distinctive expressive offering. Their choices about which messages are appropriate give the feed a particular expressive quality and 'constitute the exercise' of protected 'editorial control.' And the Texas law targets those expressive choices by forcing the platforms to present and promote content on their feeds that they regard as objectionable.") (internal citations omitted).

[129] 418 U.S. 241, 258 (1974).

[130] Moody v. NetChoice, LLC, 603 U.S. 707, 794–98 (Alito, J., concurring); id. at 745–47 (Barrett, J., concurring); id. at 748–49 (Jackson, J., concurring).

[131] *Tornillo*, 418 U.S. at 258.

[132] *NetChoice*, 603 U.S. at 794 (Alito, J., concurring).

[133] Id.

[134] Id. at 783 (quoting Hurley v. Irish-American Gay, Lesbian and Bisexual Grp. of Bos., Inc., 515 U.S. 557, 568 (1995)).

[135] Id. at 783–84.

regulation of algorithmically-curated online collections of third-party speech.[136] Justice Barrett generally noted that "technology may attenuate the connection between content-moderation actions (e.g., removing posts)" and the rights of human beings to decide which speech merits their attention and consideration.[137] Specifically, Justice Barrett conceded that the First Amendment protects editorial judgment where a human being uses "an algorithm to help them identify and delete [particular] content . . . even if the algorithm does most of the deleting without a person in the loop."[138] Yet, Justice Barrett elaborated her concern that "the First Amendment implications . . . might be different" when applied to an algorithm that feeds off user input and makes decisions with no guidance from a human editor.[139] Where an AI editor "relies on large language models to determine what . . . should be removed, has a human being with First Amendment rights made an inherently expressive 'choice . . . [to propound or] not to propound a particular point of view?'"[140] Justice Barrett did not indicate that this was the only possible instance where "the way platforms use this sort of technology might have constitutional significance."[141] More generally, Justice Jackson urged caution, noting that courts must analyze "how the regulated activities *actually function* before deciding if the activity in question constitutes expression and therefore comes within the First Amendment's ambit."[142]

The concerns voiced by Justices Alito, Barrett, and Jackson indicate that the technologically specific approach to First Amendment editorial rights jurisprudence is alive and well, and "the way platforms use [algorithmic] technology might have constitutional significance."[143] Five justices appear willing to restrict online editorial rights in certain instances. This mirrors the reasoning in *Reno* and *Pacifica*, that specific character by which speech is

---

[136] See id. at 745–47 (Barrett, J., concurring); id. at 748–49 (Jackson, J., concurring).

[137] Id. at 746 (Barrett, J., concurring) (quoting Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 641 (1994)).

[138] Id.at 745-46.

[139] Id. at 746; see also Larry Lessig, The First Amendment Does Not Protect Replicants, *in* Social Media, Freedom of Speech, and the Future of our Democracy 275 (Lee Bollinger & Geoffrey Stone eds., 2022) (arguing that speech that "is crafted or originated algorithmically, with the substance of that algorithm not in any meaningful sense programmed by any individual in advance" is legally distinct from the speech of the algorithm's creators). But see Toni M. Massaro, Helen Norton & Margot E. Kaminski, SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment, 101 Minn. L. Rev. 2481, 2512 (2017) (arguing that "current doctrine supports the coverage of speech regardless of its nontraditional source or form"); see also Volokh & Falk, supra note 109, at 887-8 (arguing that search engines deserve First Amendment protections).

[140] Moody v. NetChoice, LLC, 603 U.S. 707, 746 (Barrett, J., concurring) (quoting Hurley v. Irish-American Gay, Lesbian and Bisexual Grp. of Bos., Inc., 515 U.S. 557, 575 (1995)).

[141] Id.

[142] Id. at 749 (Jackson, J., concurring).

[143] Id. at 746 (Barrett, J., concurring).

produced and heard can affect its protections. Where the passive nature by which the speech in question was received in *Reno* and *Pacifica* determined those cases' outcomes, novel forms of algorithmic curation, moderation, and automated decision-making which are increasingly common on social media platforms are likely to influence future legal outcomes.

IV. MANDATING CONTENT MODERATION AS COMMERCIAL SPEECH

If the courts refuse to permit the regulation of online content moderation for reasons related to its technologically distinct character, then regulating and mandating content moderation as a form of commercial speech represents another potential avenue for doing so. This section, however, casts doubt on the idea that content moderation can be insulated from strict scrutiny if labeled as commercial speech, especially where content moderation places significant technical burdens on platforms. Some claim that, because social media companies profit from the content third parties post on their platforms, regulation of social media content can be treated as a regulation of commercial speech under current precedent.[144] For instance, Richard Hasen has conceded that any kind of direct regulation of cheap speech on social media will be subject to heightened scrutiny beyond rational basis review but has argued that disclosure requirements of altered audiovisual material could be legally justified if posts on social media are seen as commercial speech.[145] Specifically, Hasen has argued that mandating the labeling of "altered videos as 'altered' would mandate the disclosure of purely factual and uncontroversial information in a commercial context."[146] Of course, the first issue is that the speech in question comes from third parties and not the platforms, but Hasen claims that the reasons why users post on social media do "not negate the essentially commercial nature of [social media] enterprises."[147]

There is a clear problem with Hasen's analysis. The fact that social media companies profit from political speech is unremarkable. Indeed, the fact that the *Miami Herald* profited from its political advertising was of no matter to the Court—commercial regulation of speech was not even mentioned in that

---

[144] See Hasen, Bots, supra note 8, at 551–52.

[145] Id. at 545 n.46 ("Under *Zauderer* as we interpret it today, the government may compel truthful disclosure in commercial speech as long as the compelled disclosure is reasonably related to a substantial governmental interest, and involves 'purely factual and uncontroversial information' that relates to the service or product provided.").

[146] Id. at 551.

[147] Id. at 551 n.79.

decision.[148] Newspapers profit from hosting political and ideological speech just as social media companies do, but the Supreme Court and lower courts have been clear in striking down forced disclosure requirements in the press context.[149] Furthermore, the Supreme Court has been skeptical about expanding its main precedent on forced commercial disclosures, *Zauderer v. Office of Disciplinary Counsel of the Supreme Court of Ohio*, to allow for a broader definition of what amounts to commercial speech.[150] In *Zauderer*, the Supreme Court upheld regulations sanctioning an attorney for violating the state bar's advertising rules. The Court found that states may compel commercial speech so long as the information to be disclosed is "purely factual and uncontroversial," and those requirements are related to important interests of the state and not overly burdensome.[151] *Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston, Inc.* made clear that *Zauderer* does not apply outside of "purely factual and uncontroversial information," and that this right to tailor the speech one publishes is not "restricted to the press."[152] As the Court emphasized recently in *National Institute of Family and Life Advocates (NIFLA) v. Becerra* (2018), a state-mandated disclosure requirement must relate directly to the services the private target supplies.[153] There is a colorable argument that the kinds of disclosures Hasen is advocating for go directly to the nature of online media consumption.[154] Nonetheless, while the Court emphasized that incidental burdens on speech aimed at commercial regulations are permissible, they cannot overly burden the speech in question.[155]

Hasen contends that disclosure and labeling requirements do not in practice create a substantial burden: being told content is AI-generated does not reduce one's access to it, and it is not a matter of opinion.[156] Seen narrowly, this may be true.[157] The problem is that these kinds of labeling and disclosure

---

[148] Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 258 (1974).

[149] E.g., id.; Wash. Post v. McManus, 944 F.3d 506, 511–12 (4th Cir. 2019).

[150] 471 U.S. 626, 651 (1985); Milavetz, Gallop & Milavetz, P.A. v. United States, 559 U.S. 229, 255 (2010) (noting that *Zauderer* applies only to "advertisements that, by their nature," create a risk of consumer deception) (Thomas, J., concurring); Nat'l Inst. of Fam. and Life Advoc. v. Becerra, 585 U.S. 755 (2018); Eric Goldman, Zauderer and Compelled Editorial Transparency, 108 Iowa L. Rev. Online 80, 85 (2023) (summarizing the current application of the *Zauderer* standard).

[151] *Zauderer*, 471 U.S. at 651.

[152] Hurley v. Irish-American Gay, Lesbian and Bisexual Grp. of Bos., Inc., 515 U.S. 557, 573 (1995).

[153] Nat'l Inst. of Fam. and Life Advoc., 585 U.S. at 766.

[154] See Hasen, Bots, supra note 8, at 551–52.

[155] Nat'l Inst. of Fam. and Life Advoc., 585 U.S. at 766.

[156] Hasen, Bots, supra note 8, at 551–52.

[157] See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1196 (E.D. Cal. 2024) ("The safe harbor carveouts of the [California Election AI] statute attempt to implement labeling requirements, which if narrowly tailored enough, could pass constitutional muster.").

requirements foist massive burdens on platforms to seek out and moderate the relevant content.[158] Disclosure requirements are not per se barred just because they may chill underlying speech or editorial practices.[159] But disclosures must be mandated by some important state interest, and they must not be overly burdensome.[160] While AI and altered content disclosure requirements might meet the first threshold of informing users about potentially misleading content with harmful potential, the burdens could well be massive. Content moderation is extremely time-consuming and resource intensive. Even if, for instance, Facebook were to achieve a 99.9% accuracy rate in moderating the 350 million photos that are uploaded daily, this would still lead to roughly 350,000 moderation errors per day.[161] Justice Alito has likewise noted that these burdens would be substantial, as "[n]o human being could possibly review" the monumental amount of data which these platforms process regularly.[162]

Nor is there any clear precedent indicating the validity of Hasen's labeling proposals. Cases where labeling and disclosure requirements are upheld under *Zauderer* typically involve far less sweeping and intensive requirements than the disclosures at issue in the *NetChoice* cases,[163] or Hasen's proposal,[164] and often require trivial amounts of employee work time to implement.[165] Indeed, many applications of *Zauderer* by the lower courts approve short, consumer-facing labels and warnings at most.[166] This forced labeling of altered content could well force X to alter its core private mission, which has never been solely to disseminate reliable information, and thus act as a kind of censorship, nudging social media platforms to push toward certain kinds of core functions that may not be privately chosen.[167]

---

[158] E.g., id. at 1197 ("[the California Election AI statute's] size requirements for the disclosure statement in this case and many other cases would take up an entire screen, which is not reasonable because it almost certainly 'drowns out' the message a parody or satire video is trying to convey.").

[159] Herbert v. Lando, 441 U.S. 153, 154 (1979).

[160] Zauderer v. Off. of Disciplinary Couns. of Sup. Ct. of Ohio, 471 U.S. at 649; Nat'l Inst. of Fam. and Life Advoc., 585 U.S. at 769.

[161] See Keller, supra note 19, at 23.

[162] Moody v. NetChoice, LLC, 603 U.S. 707, 794 (2024) (Alito, J., concurring); see also Ruth Spence, Antonia Bifulco, Paul Bradbury, Elena Martellozo & Jeffrey DeMarco, The Psychological Impacts of Content Moderation on Content Moderators: A Qualitative Study, 17 Cyberpsychology: J. Psychosocial Rsch. on Cyberspace 4 (2023) https://doi.org/10.5817/CP2023-4-8 (detailing the psychological toll on content moderators).

[163] See Keller, supra note 19, at 55.

[164] Id. at 33.

[165] Id. at 32.

[166] Id. at 55.

[167] See Goldman, supra note 150, at 88–91.

Perhaps Instagram is happy to allow alterations and AI content because it means more interactions and engagement. Indeed, promoting interactions and engagement are core to the social media business model in general. One could argue that an influencer who has used AI-generated imagery depicting large numbers of migrants crossing the border or one generating videos of subway crime has an interest in doing so to attract more engagement, and that the existence of an altered label chills that influencers' desire to engage in social media posting, improperly burdening their speech, or affecting Instagram's privately chosen business model.[168] You could take this a step further by imagining an entirely fictitious account produced with AI and boosted by Instagram's algorithm to increase engagement.[169] This account could routinely generate stories that are technically made-up but represent exaggerated composites of political events of the day to provide general political commentary. For example, the U.S. has recorded a record number of illegal border crossings in the past several years;[170] an AI-generated video depicting this phenomenon could be technically fictitious footage without being disinformative if it is attempting to be representative of an actually occurring event. Or consider the various conflicts currently ongoing with high numbers of civilian casualties: a fictitious image or video showing a child killed by an airstrike may not as such be disinformative even if it is not literally accurate. Where a platform like Instagram hosts and actively supports these accounts as part of its business model based on maximizing interactions and engagement, regulations seeking to curtail this kind of speech are suspect. Indeed, these regulatory expectations tell "publishers what types of editorial practices regulators expect to see,"[171] and ultimately end up shaping the nature of the kind of third-party speech the user ends up receiving.

Notably, the First Amendment protects the rights of listeners to receive controversial and false speech to the same degree that it protects the rights of speakers to create this speech. As such, combined with the fact that listeners have a well-established "right to receive information and ideas, regardless of their social worth,"[172] efforts to force platforms to prioritize truth-seeking through content moderation, disclosure, and labeling seem

---

[168] Id.

[169] See John Hermann, AI Is Coming for the Influencers, N.Y. Mag (Jan. 3, 2024), https://nymag.com/intelligencer/2024/01/ai-is-coming-for-the-influencers.html.

[170] Illegal Border Crossings from Mexico Reach Highest on Record in December before January Lull, Associated Press (Jan. 26, 2024, 7:41 P.M.), https://apnews.com/article/immigration-border-crossings-mexico-biden-18ac91ef502e0c5433f74de6cc629b32.

[171] Goldman, supra note 150, at 82.

[172] Stanley v. Georgia, 394 U.S. 557, 564 (1969) (citing Winters v. New York, 333 U.S. 507, 510 (1948)).

constitutionally suspect at first glance.[173] If there is a right to receive and read obscene materials in one's home (as the Supreme Court held in *Stanley v. Georgia*), then surely that right extends to unrealistic or misleading altered or AI-generated content, and government efforts which chill the creation and diffusion of that content are suspect.[174]

Likewise, regulatory burdens matter even where the speech in question can theoretically be regulated. Consider a law requiring the labeling of all altered or AI content as such. Courts may still find that labeling and tracking such a large amount of content is not realistically feasible on platforms, and significant burdens like the proposed labeling and tracking doom a regulation even at intermediate scrutiny.[175] Lawmakers might seek to avoid this issue by limiting a labeling requirement to cover only altered or AI content which goes "viral," attaining a certain amount of engagement. This would pose less of a burden, but a labeling requirement that is targeted at more popular speech would also face over- and under-inclusion challenges. Still, perhaps a requirement that "viral" altered content be labeled is constitutional, given that the under-inclusion can be justified as essential to reduce the regulatory burden it places on platforms.[176]

Indeed, the question of whether viral content can be targeted for mandatory content moderation goes to the heart of the issues presented in *NetChoice*: this more specific regulation could not be justified to the extent that it is aimed to disfavor targeted content or reduce its reach. In other words, "the government cannot get its way just by asserting an interest in

---

[173] Nat'l Inst. of Fam. and Life Advoc. v. Becerra, 585 U.S. 755, 766 (2018) (holding that laws which compel speech and force individuals to "speak a particular message" that they would not otherwise speak dilutes the law's message and triggers strict scrutiny); see also See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1196 (E.D. Cal. 2024) (finding that California's anti-deepfake law is constitutionally disfavored because "[t]he political context is one such setting that would be especially "perilous" for the government to be an arbiter of truth in.").

[174] Id. See also Joseph Thai, The Right to Receive Foreign Speech, 71 Okla. L. Rev. 269, 283 (2018). This analysis may be altered where a law incidentally burdens speech and does not implicate a "substantial speech interest." Nonetheless, any law which directly mandates certain forms of content moderation is likely to go beyond incidentally burdening speech and would merit more serious scrutiny. See United States v. O'Brien, 391 U.S. 367, 377 (1968) (finding a regulation affecting speech justified "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"); see also Red Lion Broad. Co. v. FCC, 395 U.S. 367, 387 (1969) ("the ability of new technology to produce sounds more raucous than those of the human voice justifies restrictions on the sound level, and on the hours and places of use, of sound trucks so long as the restrictions are reasonable and applied without discrimination.").

[175] See Chesney & Citron, supra note 2, at 1787 (expressing skepticism of a technological solution for detecting deep fakes in the short term).

[176] See Hasen, Bots, supra note 8, at 552 ("[A] law banning deep fakes raises vagueness and overbreadth problems by preventing the use of synthetic media when one is reckless about distributing 'deceptive' video or audio about a candidate to a voter.").

better balancing the marketplace of ideas," and it cannot seek to "rejigger the expressive realm."[177] As such, a regulation forcing platforms to *remove* viral AI-generated disinformation, for example, would likely fail, since such a regulation targets speech on account of its popularity and truth-value. But, to the extent that mandated content moderation only seeks to require platforms to *label* such content as AI-generated, *and crucially also* to which that regulation is justified primarily as a method of reducing the burden on moderators, it may be more likely to pass constitutional muster.

Nonetheless, this kind of measure remains suspect since it seems to regulate speech on account of its reach and since, as *Buckley v. Valeo* informs us, "the concept that government may restrict the speech of some elements of our society in order to enhance the relative voice of others is wholly foreign to the First Amendment."[178] Any effort to ease the burden on platforms by limiting labeling and disclosure requirements to only viral altered and AI content would have to be justified to overcome the issue that it singles out more popular speakers for differential treatment. Maybe this singling out could be justified on the basis that, absent the great burden on platforms, the requirement could extend beyond viral posts, and that this is the least restrictive means for achieving the stated ends of the regulation—to alert viewers to altered and AI content. This is not a totally implausible answer to concerns regarding under-inclusion.

V. MANDATING CONTENT MODERATION THROUGH A NOVEL FIRST AMENDMENT EXCEPTION

If the Court declines to allow more expansive regulation of AI-generated disinformation or, more broadly, AI content as commercial speech, the only remaining route to limiting the reach of this content is by establishing an exception to First Amendment protections entirely or finding an exception which allows courts to apply intermediate rather than strict scrutiny to speech regulations. As previously noted, this would likely require the Court to base such an exception in the distinct technological character of the method of communication itself. It is without question that social media platforms are permitted a substantial amount of leeway in how they moderate content and design their platforms.[179] Nonetheless, as this section will detail,

---

[177] Moody v. NetChoice, LLC, 603 U.S. 707, 710 (2024).

[178] 424 U.S. 1, 48–49 (1976) (holding that limits on election expenditures were unconstitutional under the First Amendment because a restriction on spending for political communication necessarily reduces the quantity of expression).

[179] See *NetChoice*, 603 U.S. at 738–40.

exemptions from the protections of the First Amendment granted to speech are rare, nowhere more so than when the speech in question is political.[180] Speech cannot be regulated solely because its speaker is powerful,[181] it is pervasive in character,[182] it is created by a non-human,[183] or because it is false or deceptive.[184] Thus, in order for the government to regulate political speech and speech more generally, it must find a basis for doing so that goes beyond any single one of these factors.

Recent cases on regulating AI speech have quickly thrown cold water on prospects for its regulation, solely on the basis that the speech is AI-generated. In *Kohls v. Bonta*, California's electoral regulation bill failed strict scrutiny as a content-based regulation, which restricted speech based on its "purported truth or falsity,"[185] the exact distinction which the Supreme Court forbade in *Alvarez*.[186] *Kohls's* conclusion is a standard application of existing precedent. The Supreme Court has been sparing in doling out exceptions to speech that the First Amendment protects.[187]

As a starting matter, even if the Court were to carve out an exception allowing more regulation of AI, those exceptions would need to be viewpoint neutral.[188] That even unprotected speech cannot be discriminated against on account of viewpoint is well established.[189] This is no small matter: in practice, the label "disinformation" has been imprecisely applied. For example, the COVID-19 lab-leak theory was frequently referred to as disinformation, only to now be considered a respectable origin hypothesis of the virus.[190] As the Court wrote in *Alvarez*: "to hold that the interest in truthful discourse alone is sufficient to sustain a ban on speech … would give government a broad censorial power unprecedented in this Court's cases or in our constitutional tradition."[191] To get past this problem, the Court would have to establish a more substantial definition of mis- and disinformation

---

[180] Citizens United v. FEC, 558 U.S. 310, 339 (2010); *Buckley*, 424 U.S. at 48–49.

[181] *NetChoice*, 603 U.S. at 732–33; Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 249–50 (1974).

[182] *NetChoice*, 603 U.S. at 732–33.

[183] *Citizens United*, 558 U.S. at 392–93.

[184] United States v. Alvarez, 567 U.S. 709, 709 (2012).

[185] See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1195 (E.D. Cal. 2024) ("AB 2839 delineates acceptable and unacceptable content based on its purported truth or falsity and is an archetypal content-based regulation that our constitution considers dubious and subject to strict scrutiny.").

[186] *Alvarez*, 567 U.S. at 709.

[187] See Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 791–93 (2011).

[188] R.A.V. v. City of St. Paul, 505 U.S. 377, 387 (1992); Am. Booksellers Ass'n, Inc. v. Hudnut, 771 F.2d 323, 325 (7th Cir. 1985).

[189] *R.A.V.*, 505 U.S. at 387.

[190] See Ted Galen Carpenter, Disinformation and the Wuhan Lab Leak Thesis, CATO Inst. (Mar. 6, 2023), https://www.cato.org/commentary/disinformation-wuhan-lab-leak-thesis.

[191] *Alvarez*, 567 U.S. at 723.

were it to create such an exception where no workable, neutral, or appropriately tailored definition has been established by lower courts, the media, or the academic community.[192] This is a tall order.

This conclusion is bound to lead to what are, for many, distressing consequences. Consider the following: OpenAI is now planning the roll-out of a text to video program called Sora.[193] Imagine, for instance, a user who employs Sora to construct a series of videos and images that purport to show and prove that the government and ICE are rounding up migrants and sending them to concentration camps where they are interned and murdered. Or imagine a documentary purporting to prove that some mass school shootings were anti-gun hoaxes. These are clear examples of ideological speech.[194] Even the dissenters in *Alvarez* would find that this kind of ideological speech is a protected form of lying.[195] It is also certainly not commercial speech, and the argument for its hosting by a platform being a solely commercial action, if it is commercial at all, is tenuous.[196] As current precedent makes clear, this example cannot be regulated simply on account of being ideologically disfavored propaganda,[197] it cannot be regulated only because it is false,[198] and to the extent that it is regulated, it must be regulated using the least restrictive means possible.[199] Likewise, if private counterspeech can effectively deter the harms of this false speech, then the government may not regulate it.[200] The question is whether AI poses some harm, or poses such an insuperable obstacle to effective counterspeech, that the government has a genuine claim to regulating its production and dissemination, by way of

---

[192] Id. at 731 (Alito, J., dissenting) ("[T]here are broad areas in which any attempt by the state to penalize purportedly false speech would present a grave and unacceptable danger of suppressing truthful speech. . . . Allowing the state to proscribe false statements in these areas also opens the door for the state to use its power for political ends. Statements about history illustrate this point. If some false statements about historical events may be banned, how certain must it be that a statement is false before the ban may be upheld? And who should make that calculation?").

[193] Gerrit De Vynck, Jhaan Elker & Tyler Remmel, The Future of AI Video is Here, Super Weird Flaws and All, Wash. Post (Feb. 28, 2024, 6:00 A.M.), https://www.washingtonpost.com/technology/interactive/2024/ai-video-sora-openai-flaws/?itid=hp-top-table-main_p001_f006.

[194] See Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 802 (2011); United States v. Stevens, 559 U.S. 460, 473 (2010).

[195] *Alvarez*, 567 U.S. at 722 (where a plurality held that speech's falsity is not enough, by itself, to exclude speech from First Amendment protection).

[196] See supra Part IV.

[197] R.A.V. v. City of St. Paul, 505 U.S. 377, 387 (1992).

[198] *Alvarez*, 567 U.S. at 722.

[199] Id. at 729.

[200] Id. at 709; United States v. Stevens, 559 U.S. 460, 467; See Kohls v. Bonta, 752 F. Supp. 3d 1187, 1191 (E.D. Cal. 2024) . Likewise, the First Amendment requires that the government's chosen restriction on the speech at issue be "actually necessary" to achieve its interest. *R.A.V.*, 505 U.S. at 395.

disclosure or forced content moderation.[201] It is not clear what that harm is, other than those which have already been explicitly discounted as reasons for voiding the protections of the First Amendment.

If instead the Court were to attempt to avoid this tricky issue by limiting an exception to all AI content, irrespective of whether it represents some kind of accurate representation or portrayal of reality, this would almost certainly imperil a wide range of endeavors with artistic, satirical, and intellectual merit.[202] Regardless of concerns about AI, its far-reaching artistic and intellectual potential is widely recognized.[203] And indeed, this kind of exception would prompt threshold issues and severe under-inclusion problems: a large amount of media has already been edited, filtered, photoshopped, and airbrushed. Perhaps the majority of online content is altered in some manner.[204] Why should an AI-generated photo not garner First Amendment protection, but a heavily photoshopped picture should? What if the pictures look exactly the same? Should the AI photo be denied coverage simply on account of the method of its creation? How are the courts and social media platforms to decide when and how an image or video is sufficiently altered to merit a label or lack of coverage? Under- and over-inclusion frequently doom less rigorous regulations of speech.[205]

VI. MANDATING CONTENT MODERATION THROUGH TECHNOLOGICAL TARGETING: ALGORITHMIC MODERATION AND THE HUMAN ANCHOR TEST

While AI content moderation on its own might be constitutionally prohibited as other incursions on private editorial rights, the *NetChoice* concurrences emphasize the possibility that there may be a specific feature of algorithmic moderation that permits more government regulation.[206] One such case involves regulating the speech of "self-generating" AI or self-moderating algorithms.[207] Lawrence Lessig has outlined this prospect with his

---

[201] *Stevens*, 559 U.S. at 472 (holding that new categories of unprotected speech may not be added to the list by a legislature that concludes certain speech is too harmful to be tolerated).

[202] Milkovich v. Lorain J. Co., 497 U.S. 1, 20 (1990) (explaining that the Court has widely protected parody and satire under the First Amendment).

[203] See, e.g., Kevin Roose, An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy, N.Y. Times (Sept. 2, 2022), https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

[204] Expert: 90% of Online Content Could be Generated by AI, ODSC (Jan. 16, 2023), opendatascience.com/nina-schick-expert-90-of-online-content-could-be-generated-by-ai/.

[205] See Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 786, 802 (2011).

[206] Moody v. NetChoice, LLC, 603 U.S. 707, 745-47 (2024) (Barrett, concurring).

[207] Lawrence Lessig refers to this as "replicant speech," *i.e.*, speech that is totally self-generating and unbound from any direct human intentionality to speak (although it may be downstream from some original human expressive choice). Lessig, supra note 139, at 4.

"CLOGGER" thought experiment. Lessig theorizes a form of "replicant" speech in which "processes . . . have developed a capacity to make semantic and intentional choices, the particulars of which are not plausibly ascribed to any human or team of humans in advance of those choices."[208] Thus, one could imagine AI that can interact with real users, learn from their responses, conduct its own mass research of sources to refine its knowledge, and generate responses that are facially original in the same sense a real user would produce original speech, long after being consciously programmed to do so by a human individual. Paraphrasing Lessig: parents can produce a child, but they do not produce the business the child may start twenty-five years later.[209]

Some version of this is probably not far off, and social media platforms have already employed AI to generate new content.[210] This program could even be written to produce new fake accounts or bots on its own without additional input, to give off the appearance of mass collective speech. So, while this speech might be ideological in character, each speech act is not directly linked to any speaker, and the AI program would have the ability to speak so quickly, broadly, and rapidly that it could seriously distort the marketplace of ideas if used for dis- or misinformative purposes. Justice Barrett noted this issue when asking "[i]f the AI relies on large language models," prompting the question of whether AI's expressive choices merit First Amendment protection even if it is unclear whether a "human being with First Amendment rights made an inherently expressive 'choice'" to articulate, curate, or moderate the speech in question.[211]

Does such a "speaker" lack First Amendment rights because much of this speech is unconnected to the direct expressive choices of any human speaker? Language in *Citizens United* and elsewhere would seem to suggest the opposite. As Justice Scalia wrote, concurring in *Citizens United*, "[t]he Amendment is written in terms of 'speech,' not speakers."[212] Scalia argued this is so because "[i]ts text offers no foothold for excluding any category of speaker, from single individuals to partnerships of individuals, to unincorporated associations of individuals, to incorporated associations of individuals."[213] Likewise, as Justice Powell wrote over three decades earlier, the

---

[208] Id.

[209] Id.

[210] Id. at 5; see also Massaro, Norton & Kaminski, supra note 139, at 2481 (describing how an AI Twitter chatbot in 2016 learned to produce offensive tweets).

[211] *NetChoice*, 603 U.S. at 746 (Barrett, J., concurring) (quoting Hurley v. Irish-American Gay, Lesbian and Bisexual Grp. of Bos., Inc., 515 U.S. 557, 575 (1995)).

[212] Citizens United v. FEC, 558 U.S. 310, 392 (2010) (Scalia, J., concurring).

[213] Id. at 392-93.

question "is not whether corporations 'have' First Amendment rights," but whether the law "abridges expression that the First Amendment was meant to protect."[214] Other scholars have recognized this holding, concluding that "current doctrine supports the coverage of speech regardless of its nontraditional source or form."[215]

There are very few instances where the First Amendment permits the government to limit or prohibit political speech, and as we have seen, speech cannot solely be prohibited or limited because it is pervasive,[216] its speaker is powerful,[217] it is false,[218] or it is created by a non-human.[219] Notably, however, in *Federal Election Commission v. Bluman*, the United States District Court for the District of Columbia interpreted the protections of *Citizens United* and *Buckley* to not apply to the speech of foreign nationals, made outside or within the U.S.[220] This decision was written by then-judge, now Justice, Kavanaugh, and it was also affirmed by the Supreme Court without any written opinion.[221] The refusal to extend constitutional protections for speech to that of non-resident foreign citizens did not implicate the holdings in *Buckley* or *Citizens United* that the First Amendment protects speech and not speakers, based on the premise that the government may exclude foreigners from activities intimately related to the conduct of self-governance.[222]

This is notable for our purposes, as it represents a limit on the holdings from those cases that some speech may not be constrained (in *Bluman* non-national speech) to increase the influence of other speech (in *Bluman* the ideological speech of nationals). Similarly, Justice Barrett appears to focus on the question of AI and algorithmic speech by emphasizing the "human beings' constitutionally protected right" to speak and to curate collective speech, rather than that of users to receive speech, or that of speech itself to be heard.[223] Justice Barrett notes, for instance, that corporations "possess First Amendment rights," and are "composed of human beings" with those

---

[214] First Nat'l Bank of Bos. v. Bellotti, 435 U.S. 765, 776 (1978).

[215] Massaro, Norton & Kaminski, supra note 139, at 2512.

[216] *NetChoice*, 603 U.S. at 732.

[217] Id.; Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 249–50 (1974).

[218] United States v. Alvarez, 567 U.S. 709, 709 (2012).

[219] Citizens United v. FEC, 558 U.S. 310, 392–93 (2010).

[220] FEC v. Bluman, 800 F. Supp. 2d 281, 288–89 (D.D.C. 2011), *aff'd*, 565 U.S. 1104 (2012). Then-judge Kavanaugh sat by designation in the D.D.C., despite being a circuit judge at the time.

[221] *Bluman,* 565 U.S. 1104.

[222] *Bluman*, 800 F. Supp. 2d at 288, *aff'd*, 565 U.S. 1104 (2012).

[223] Moody v. NetChoice, LLC, 603 U.S. 707, 746 (2024) (Barrett, J., concurring); see also Stanley v. Georgia, 394 U.S. 557, 564 (1969); Thai, supra note 174, at 279 (arguing that the open nature of the First Amendment's interest in protecting speech supports the conclusion that the First Amendment protects open access to foreign speech that is made in digital spaces, such as social media platforms).

rights, but "foreign persons and corporations located abroad do not [possess such rights]."[224] In emphasizing this distinction, Barrett appears to be indicating that the First Amendment is anchored in the humanity, and the nexus to U.S. jurisdiction (either by citizenship or location), of the speaker.[225] In this argument, we see the preliminary form of a human anchor test, which would allow for greater regulation of speech which cannot trace back to some direct human expressive choice.

In so doing, Justice Barrett has provisionally adopted the view recognized by now-Justice Kavanaugh that, as Justice Stevens wrote in *Citizens United*, the First Amendment may not establish a "categorical approach to speaker identity," and that in certain cases speaker identity (or lack thereof) may be relevant in forging exceptions to First Amendment coverage.[226] By extension, the question becomes whether AI content or algorithmic moderation may be so potentially disconnected from the original speech intent of a human speaker that it may likewise be excepted from First Amendment coverage, at least in certain contexts like the political sphere or commercial advertising.[227] There is a plausible case to be made that AI bot armies, or content moderation decisions with no discernable human regulation, detract from self-governance to a greater degree than the speech of the average foreign speaker.

To the contrary, Eugene Volokh, Helen Norton, and others have argued that a wide variety of indirect human speech likely garners First Amendment protections.[228] For example, Volokh has argued that search engines may merit these protections as speakers, despite the fact that they largely if not totally direct listeners to third party content: "the speech of . . . a search engine consists almost entirely of the selected and arranged links to others' material. . . . these exercises of editorial judgment are fully protected by the First Amendment."[229] Toni Massaro, Helen Norton, and Margot Kaminski have argued that contemporary First Amendment jurisprudence "emphasizes listeners' interests in free speech outputs—rather than speakers' humanness or humanity—in ways that make it exceedingly difficult to place AI speakers beyond the First Amendment's reach."[230] Indeed, there appears to be a strong

---

[224] *NetChoice*, 603 U.S. at 746-47 (Barrett, J., concurring); see also Leslie Kendrick, Are Speech Rights for Speakers?, 103 Va. L. Rev. 1767 (2017) (surveying arguments that the First Amendment does not protect the rights of speakers, while arguing that speakers do in fact enjoy speech rights).

[225] *NetChoice*, 603 U.S. at 745-47 (Barrett, J., concurring).

[226] *Bluman*, 800 F. Supp. 2d at 289, *aff'd*, 565 U.S. 1104 (2012).

[227] Citizens United v. FEC, 558 U.S. 310, 421 (2010); id. at 424 n.51 (Stevens, J., concurring in part).

[228] Massaro, Norton & Kaminski, supra note 139, at 2512; Volokh & Falk, supra note 109, at 885.

[229] Volokh & Falk, supra note 109, at 885.

[230] Massaro, Norton & Kaminski, supra note 139, at 2483.

presumption following *Citizens United* that what we are trying to protect is the speech itself, rather than the speakers.[231]

Likewise, the fact that the replicant can potentially speak in a way that dwarfs the average unsophisticated user would not appear to offer any reprieve, based on the long-cited reasoning from *Buckley* that the "government may [not] restrict the speech of some elements of our society in order to enhance the relative voice of others."[232] Yet, perhaps there is something inherent in the interaction between algorithmic mechanisms and the incredible potential for anonymous actors to produce disinformation with AI.[233] Concerns over algorithmic sorting combined with replicant speech bring this issue to its most acute point, and even the most radically libertarian proponent of speech protections should be concerned about the dilution of the marketplace of ideas in such a scenario. These scenarios *may* lead the Court to carve out narrow exceptions that allow the government to mandate efforts to combat or label AI, bot speech, or content moderation on platforms that rely heavily on algorithmic sorting. This kind of exception would be well-grounded in the clear harms this context fosters, and the realities of this emerging, poorly understood, and highly manipulable technology.

At the same time, Justice Barrett's speculative compromise requires certain doctrinal sacrifices. It would require the Court to expressly reframe its *Citizens United* holding, which currently suggests that the First Amendment protects speech and not speakers.[234] Instead, the rule would be reformulated to approximate the holding that the First Amendment protects speech which is reasonably related to the expressive choices of a human person who is already entitled to that Amendment's coverage. To argue that the speech of an AI-editor far removed from any human expressive or editorial decision is protected by the First Amendment would make it difficult to maintain *Bluman's* holding that foreign speech is not protected as such.[235] Put another way, if foreign speakers can undermine democratic self-governance, surely unsupervised AI editors utilized by (in some cases foreign-owned) dominant media platforms have at least the potential to do so as well.[236]

Nonetheless, *Buckley*, *Citizens United*, *Alvarez*, *Stevens*, and *Brown v. Entertainment Merchants Association* lead to unwelcoming conclusions for efforts to regulate AI and altered speech. Taken together, it appears that the

---

[231] Id. at 2512 (citing *Citizens United*, 558 U.S. at 392–93 (Scalia, J., concurring)).
[232] Buckley v. Valeo, 424 U.S. 1, 48–49 (1976) (citations omitted).
[233] Moody v. NetChoice, LLC, 603 U.S. 707, 795 (2024) (Alito, J., concurring).
[234] Citizens United v. FEC, 558 U.S. 310, 349 (2010).
[235] FEC v. Bluman, 800 F. Supp. 2d 281, 289 (D.D.C. 2011), *aff'd*, 565 U.S. 1104 (2012).
[236] Lessig, supra note 139, at 2.

basis for a First Amendment exception allowing the regulation of AI speech cannot be based *solely* in the fact that it may be: 1) the speech of non-humans;[237] 2) false or misleading or have exceptional potential to be so;[238] 3) highly pervasive;[239] 4) structurally dominant in the marketplace of ideas;[240] or 5) restricting of other kinds of speech.[241] Each of these rationales, considered separately, are disfavored under current case law. For AI speech to be subject to greater regulation than the First Amendment typically permits, there must be some other element that goes beyond the specific character of this speech. That element, if it exists, is likely to be found in the very technological structure of algorithmically-moderated social media.

CONCLUSION

AI speech has more than *de minimis* social value and is not inherently evil in the way that some forms of speech are.[242] Given the U.S. Supreme Court's presumption against the creation of new exceptions to the First Amendment, and the high likelihood that any regulations or exceptions for AI and altered online speech will be over- or under-inclusive, the prospects are grim for regulating these kinds of content separately from other kinds of speech. As a result, any successful efforts to regulate AI speech or altered content will likely flow from existing precedent permitting the regulation of commercial speech under *Zauderer*, or from the distinct technological design of social media. Given that significant restrictions remain on the ability of government actors to restrict commercial speech, in terms of doctrinal soundness, the strongest arguments for the constitutionality of these regulations lie in a series of Supreme Court cases (*Red Lion*, *Pacifica*, *Reno*, and *Packingham*) which permit courts to consider the specifics of the targeted communicative technologies in their application of First Amendment principles. The *NetChoice* concurrences, which together represent the views of five justices, reflect this approach. Social media, algorithms, AI, and online speech involve fundamentally different considerations than traditional newspapers, print media, broadcast television, and even older forms of the internet. While the Supreme Court has yet to rule on the merits of a case involving

---

[237] *Citizens United*, 558 U.S. at 392–93.
[238] United States v. Alvarez, 567 U.S. 709, 709 (2012).
[239] *NetChoice*, LLC, 603 U.S. at 732–33.
[240] Id.; Mia. Herald Pub. Co. v. Tornillo, 418 U.S. 241, 249–50 (1974).
[241] *NetChoice*, 603 U.S. at 732–33; Buckley v. Valeo, 424 U.S. 1, 48–49 (1976).
[242] See New York v. Ferber, 458 U.S. 747, 757 (1982) (refusing to extend the First Amendment to child pornography).

how algorithms or AI impact the regulation of online speech, it should not shrink from these technological considerations when it inevitably does so.